В. Феллер

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИЛОЖЕНИЯ

В. Феллер

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИЛОЖЕНИЯ

1

ИЗДАТЕЛЬСТВО
«МИР»

AN INTRODUCTION TO PROBABILITY THEORY and its APPLICATIONS

WILLIAM FELLER

Eugene Higgins Professor of Mathematics
Princeton University

Volume I

second edition

New York John Wiley & Sons, Inc.
London Chapman & Hall, Limited

В. ФЕЛЛЕР

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИЛОЖЕНИЯ

Том 1

Перевод с английского
Р. Л. ДОБРУШИНА, А. А. ЮШКЕВИЧА
и С. А. МОЛЧАНОВА

Под редакцией
Е. Б. ДЫНКИНА

С предисловием
А. Н. КОЛМОГОРОВА

Издание второе, стереотипное

ИЗДАТЕЛЬСТВО «МИР»
Москва 1967

УДК 519.2

Перевод второго, переработанного автором издания (перевод первого издания выпущен Издательством иностранной литературы в 1952 г.) содержит систематическое изложение той части теории вероятностей, которая имеет дело с дискретными множествами элементарных событий (конечными и счетными). Такой выбор материала позволил автору без использования сложного аналитического аппарата ввести читателя в круг основных идей теории вероятностей и ее приложений.

Книга служит популярным введением в современную теорию вероятностей, доступным начинающим. Ее смогут читать студенты младших курсов университетов, а также инженеры и научные работники всех специальностей, желающие ознакомиться с основами теории вероятностей.

Особый интерес книга представит для биологов, для которых методы теории вероятностей являются главными математическими методами.

Редакция литературы по математическим наукам

Инд. 2-2-3

ПРЕДИСЛОВИЕ КО ВТОРОМУ РУССКОМУ ИЗДАНИЮ

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

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

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

$$F(x \mid y) = P(\xi < \mid x \mid, \ \eta = y)$$

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

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

* * *

В основу этого издания положен перевод первого издания книги (ИЛ, 1952), выполненный Р. Л. Добрушиным и А. А. Юшкевичем. Перевод дополнительного текста осуществлен С. А. Молчановым.

А. Н. Колмогоров

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Первое издание встретило доброжелательный прием, превзошедший самые светлые ожидания. Кроме неожиданно большого числа читателей, пользовавшихся книгой всерьез, нашлись, по-видимому, и такие, которые читали ее просто ради удовольствия. Больше всего ободряет то, что книга нашла друзей в самых различных кругах: от чистых математиков до чистых любителей. Нет никакой возможности поблагодарить здесь каждого из читателей, которым я обязан полезными критическими замечаниями. Эти замечания побуждали меня в течение шести лет думать об усовершенствованиях и подбирать лучшие примеры и упражнения. Я надеюсь, что это облегчит чтение книги и ее использование для преподавания.

Общий план, намеченный в предисловии к первому изданию, не изменился. Чтобы согласовать многообразные потребности читателей с различными наклонностями, интересами и степенью математической подготовки, приходилось часто отклоняться от главной линии. Поэтому изложение не всегда развивается от простого к сложному, сравнительно трудные разделы встречаются в самом начале, а простые — в гл. XV и XVIII. Малоподготовленный читатель не должен пытаться следить за многими побочными линиями, иначе ему грозит опасность за деревьями не разглядеть леса. Для облегчения ориентации и выбора возможных пропусков более систематически, чем в первом издании, используются звездочки. Главы, не помеченные звездочками, образуют неразрывное целое, в котором выделенные главы не используются.

Первоначальное введение в основные понятия теории вероятностей содержится в гл. I, V, VII и IX. Начинающий должен изучить их с возможно меньшими отступлениями. Глава II предназначена для развития у студентов вероятностной интуиции и техники, некоторое знакомство с ее содержанием желательно, но совсем не обязательно изучать главу систематически: может оказаться более полезным возвращаться к элементарным иллюстрациям по мере надобности в дальнейшем. То обстоятельство, что для целей первого ознакомления мы ограничиваемся дискретными распределениями, не должно быть серьезной помехой, так как элементарная теория непрерывных распределений требует лишь нескольких слов дополнительного объяснения.

Во вводном курсе после гл. IX можно сразу перейти к гл. XI, где рассматриваются производящие функции как пример более общих преобразований. Результаты гл. XI находят приложение в гл. XIII (рекуррентные события) и XII (цепные реакции, безгранично делимые распределения). Без производящих функций можно двигаться в одном из следующих направлений: предельные теоремы и теория случайных колебаний (гл. VIII, X, III), вероятностные процессы (гл. XVII), случайные блуждания (глава III и основная часть гл. XIV). Эти направления почти независимы одно от другого. Цепи Маркова (гл. XV) идейно зависят от рекуррентных событий, но и их можно изучать независимо, если читатель готов принять без доказательства общую эргодическую теорему.

Некоторая рационализация текста позволила включить новый материал и объединить старую главу III с главой II. Новое ударение сделано на временах ожидания, эта тема проходит теперь объединяющей нитью через всю книгу. Ударение на временах ожидания нашло свое отражение и в раннем введении этого понятия (в гл. II) и в нескольких независимых подходах к временам первого достижения при случайном блуждании.

Глава III — совершенно новая. Она иллюстрирует силу комбинаторных методов; с их помощью получается ряд важных результатов, которые обычно выводят, опираясь на развитый аналитический аппарат. Результаты, связанные со случайными колебаниями при бросании монеты, показывают, что широко распространенное представление о законе больших чисел ошибочно. Эти результаты столь неожиданны и в такой степени не согласуются с обычной интуицией, что даже искушенные люди сомневались, что монета ведет себя в самом деле так плохо, как предсказывает теория. Именно поэтому в § 7 включены данные эксперименты.

Более последовательное подчеркивание единства между рекуррентными событиями и цепями Маркова привело к ряду уточнений и упрощений, правда ценой некоторых отступлений от терминологии первого издания. Я приношу глубокие извинения за те недоразумения, которые могут из-за этого произойти.

Большое внимание было уделено тому, чтобы сделать указатель полезным. Однако он не может служить справочником, что и кем сделано в теории вероятностей; должное равновесие нарушено ссылками на все источники, которые были использованы, часто косвенным образом, при составлении примеров и задач. Я сожалею, что порой очень важные работы приходилось цитировать в контексте, не отражающем существа дела и не отвечающем их истинному значению.

Это издание готовилось в идеальных рабочих условиях, которых не нарушали повседневные обязанности. За этот покой я должен благодарить Научно-исследовательский отдел Военно-воздушных сил, Принстонский университет и гостеприимство Вольфовица. Мне, как и прежде, были весьма полезны критические замечания Дж. Л. Дуба. Рукопись книги и доказательства были внимательно просмотрены моей женой, что помогло устранить много ошибок и случайных погрешностей.

Август 1957 *В. Феллер*

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Первоначальная цель автора состояла в том, чтобы написать книгу об аналитических методах в теории вероятностей, при этом последняя должна была рассматриваться как чистая математическая дисциплина. Такой подход представляется самым естественным, а потому и самым удовлетворительным с эстетической точки зрения; при этом книга была бы обращена главным образом к чистым математикам. Однако широкая поддержка работ по теории вероятностей в Корнелльском университете, оказываемая Научно-исследовательским отделом Военно-Морского флота, привела автора к более честолюбивой, но менее благодарной попытке удовлетворить более широкие потребности.

Целью этой книги является строгое изложение теории вероятностей как самостоятельного раздела математики. В то же время в книге излагаются и опытные основания теории вероятностей. Книга рассчитана на то, чтобы познакомить читателя с разнообразными практическими приложениями теории вероятностей. Этой цели служат многочисленные задачи, расчеты и примеры, сопровождающие изложение основных понятий. При выборе задач и примеров автор обращал особое внимание на наглядность. Чтобы показать силу общих теоретических методов и сделать книгу более полезной для специалистов в разных областях, в книгу включены многие специальные вопросы (параграфы, отмеченные звездочкой). При первом чтении эти параграфы могут быть опущены без ущерба для понимания основного материала.

При написании книги была проделана серьезная методическая работа. Специалисты найдут много упрощений в существовавших ранее доказательствах ряда теорем, а также обнаружат и новые результаты. В частности, для нужд этой книги была развита теория рекуррентных событий, позволяющая упростить изложение цепей Маркова даже в случае конечного числа состояний.

Кроме разобранных в тексте примеров, в книге содержится ряд задач, большей частью с полными решениями. Некоторые из этих задач — простые упражнения, но большинство из них служат добавочным иллюстрирующим материалом или содержат различные дополнения. Примеры и задачи предназначены, в частности, для того, чтобы развить у читателя интуицию и навыки в постановке вероятностных задач. Как показывает опыт, трудные на первый взгляд задачи могут стать почти тривиальными, если их сформулировать естественным образом и включить в подходящий контекст.

При преподавании теории вероятностей существует тенденция возможно быстрее сводить вероятностные задачи к задачам чистого анализа, избегая специфических особенностей самой теории вероятностей. Такое изложение основывается на плохо определяемом понятии случайной величины, вводимом обычно вначале. В противоположность этому настоящая книга строится на понятии пространства элементарных событий, без которого случайные величины остаются искусственной выдумкой.

Чтобы иметь строгие теоретические основы, не обремененные вопросами теории меры и другими чисто аналитическими трудностями, мы ограничиваемся в этой книге *дискретными пространствами элементарных событий*. Это жесткое ограничение оказывается, однако, весьма полезным для читателей-нематематиков. Оно позволяет изложить ряд специальных вопросов, которые нелегко найти в литературе, и в то же время изложить эти вопросы вполне элементарным образом. Вместе с тем удается изложить совершенно исчерпывающим образом такие разрабатываемые главы теории вероятностей, как теория случайных блужданий и цепей Маркова. Общая теория случайных величин и их распределений, предельных теорем, диффузии и т. д. отложена до следующей книги.

Эта книга не могла бы быть написана без поддержки Научно-исследовательского отдела Военно-Морского флота. Одним из следствий этой поддержки были частые личные встречи с Дж. Л. Дубом, постоянная критика которого была для меня неоценима. Именно его мне хочется поблагодарить в первую очередь. Затем я хочу поблагодарить Д. Риордана, который внимательно прочел два варианта рукописи. Многие исправления и усовершенствования были предложены моей женой, которая читала и рукопись и корректуры.

Автор обязан также Чжун Кай-лаю, М. Донскеру и С. Гольдбергу, которые читали рукопись и исправили ряд ошибок; решения большинства задач были подготовлены С. Гольдбергом. Наконец, я благодарю К. Холленбах за терпеливую и квалифицированную работу по перепечатыванию рукописи и Е. Элиаса, В. Хоффмана, Д. Кинни за помощь при чтении корректур.

ВВЕДЕНИЕ

ПРИРОДА ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. Исходные представления

Теория вероятностей — математическая дисциплина, родственная таким дисциплинам, как, например, геометрия или теоретическая механика. В каждой дисциплине мы должны заботиться о различении трех сторон теории: а) формального логического содержания, б) интуитивных представлений, в) приложений.

Характер дисциплины в целом и ее прелесть нельзя по-настоящему оценить, не рассматривая эти три аспекта в их взаимосвязи.

а) Формальное логическое содержание. Характерной особенностью математики является то, что она занимается исключительно соотношениями между неопределяемыми вещами. Это хорошо иллюстрируется игрой в шахматы. Невозможно «определить» шахматы иначе, как сформулировав систему правил игры. Можно до некоторой степени описать условную форму фигур, но она не имеет существенного значения.

Шахматная доска и фигуры полезны, но можно обойтись без них. Важно знать, как ходят и действуют фигуры. Бессмысленно говорить об «определении» или «действительной природе» пешки или короля. Аналогично этому геометрия не беспокоится о том, чем «на самом деле» являются точки и прямые. Они остаются неопределяемыми понятиями, и аксиомы геометрии лишь устанавливают связи между ними: две точки определяют прямую и т. п. Это правила игры, и в них нет ничего таинственного. Мы изменяем аксиомы, чтобы изучить другие формы геометрии и логическая структура каждой из неевклидовых геометрий не зависит от ее отношения к действительности. Физики исследовали движение тел при действии закона притяжения, отличного от закона Ньютона, и такое изучение имеет смысл, даже если считать, что закон Ньютона выполняется в природе.

б) Интуитивные представления. Существенная разница между шахматами и геометрией состоит в том, что шахматные правила произвольны, в то время как аксиомы геометрии опираются на интуитивные представления. В самом деле, геометрическая интуиция так сильна, что кажется опережающей логическое рассуждение. Вопрос о том, до какой степени независимы логика, интуиция и физический опыт, является трудной проблемой философии, в которую нам незачем входить. Очевидно, что интуицию можно тренировать и развивать. Начинающий шахматист ходит осторожно, вспоминая отдельные правила, тогда как опытный шахматист с одного взгляда схватывает сложное положение и часто не в состоянии рационально объяснить свою интуицию. Аналогично можно развить интуитивное восприятие соотношений, скажем, в четырехмерном пространстве. Кроме того, общая интуиция человечества быстро прогрессирует. Ньютоновское понятие поля сил и действия на расстоянии или идея Максвелла об электромагнитных волнах, распространяющихся в пространстве, считались вначале «невообразимыми» и противоречащими интуиции.

Современная техника и радио в квартирах сделали эти понятия столь популярными, что они стали частью повседневного словаря. Аналогично этому современный студент не в состоянии оценить способы рассуждений, предрассудки и прочие трудности, с которыми приходилось бороться теории вероятностей при ее создании. В наше время газеты сообщают о выборочных обследованиях общественного мнения, и магия статистики охватывает все стороны жизни в такой степени, что молодые девушки следят за статистикой своих шансов выйти замуж. Поэтому каждый приобретает интуитивное представление о смысле таких утверждений, как «за это событие — три шанса из пяти». Эта интуиция является достаточной предпосылкой для первых формальных правил теории вероятностей. Она будет развиваться по мере изучения теории и ознакомления с различными более тонкими приложениями.

в) Приложения. В приложениях геометрии и механики теоретические понятия отождествляются с некоторыми физическими объектами, но способ этого отождествления гибок и меняется от случая к случаю, так что нельзя дать общих правил. Понятие твердого тела полезно и существенно в механике, и все же ни один физический объект не удовлетворяет точному определению твердого тела. Только опыт учит нас тому, какие тела можно с удовлетворительным приближением считать твердыми. Резину приводят обычно как типичный пример нетвердого тела, но при рассмотрении движения автомобиля большинство учебников обращается с колесами, включающими резиновые шины, как с твердыми телами. Это пример того, как в соответствии с удобствами и потребностями выбираются и изменяются теоретические схемы. В зависимости от наших целей мы можем пренебречь атомной теорией и рассматривать солнце как гигантский шар из непрерывного вещества, а в другом случае рассматривать его как материальную точку. Нужно всегда помнить, что математика имеет дело с абстрактными моделями и что разные модели могут описывать одно и то же действительное явление с различной степенью приближения и простоты. Способ применения математических теорий не зависит от предвзятых идей и не является предметом логики; это целеустремленная техника, меняющаяся с накоплением опыта. Конечно, любой вид деятельности человека представляет интерес для философа, и философский анализ приложений математики вполне законен. Но такой анализ находится вне области математики, физики или статистики. Философское рассмотрение основных понятий теории вероятностей должно быть отделено от математической теории и ее приложений в такой же мере, как рассмотрение наших интуитивных представлений о пространстве отделяется теперь от геометрии (хотя это и не всегда было так).

§ 2. Способ изложения

История теории вероятностей (и математики вообще) свидетельствует о стимулирующем взаимодействии теории и ее приложений: прогресс в теории открывает новое поле приложений, а каждое новое приложение создает новые теоретические проблемы и оказывает влияние на направление исследований. В настоящее время приложения теории вероятностей распространяются на многие области различной природы, и число приложений неуклонно возрастает. Только общая математическая теория является достаточно гибкой для того, чтобы дать рабочий инструмент для таких разнообразных проблем, и мы должны противостоять соблазну сохранить наши понятия, описания и термины слишком близкими к одной частной области применений. Нам нужна строгая математическая теория, развивающаяся по тому же пути, который принят в геометрии и механике.

Мы начнем с простейших опытов, таких, как бросание монеты или кости, где все утверждения имеют очевидный интуитивный смысл. Эта интуиция будет переведена на язык абстрактной модели, которая будет затем обобщена на более сложные случаи. Иллюстрирующие примеры будут объяснять эмпирические предпосылки теории и развивать интуицию читателя, но сама теория будет иметь математический характер. Мы будем пытаться объяснить «истинный смысл» вероятности не больше, чем современный физик останавливается на действительном смысле массы и энергии или геометр объясняет природу точки. Вместо этого мы будем доказывать точные теоремы и показывать, как они применяются на практике.

Первоначально теория вероятностей развивалась для описания очень ограниченного класса опытов, связанных с азартными играми. В связи с этим основной целью теории было вычисление некоторых определенных вероятностей. Мы будем следовать историческому пути и начнем с азартных игр. Мы поступим так не из-за их важности или их общего интереса, а потому, что азартные игры имеют более наглядные предпосылки, и, кроме того, для пользы читателя-нематематика мы хотим как можно дальше отложить использование специального аналитического аппарата. В соответствии с этим в нескольких первых главах мы вычислим некоторые типичные вероятности, хотя общая теория не зависит от этих частных численных значений. Ее объектом является раскрытие общих законов и зависимостей и конструирование абстрактных моделей, которые могут в некоторой степени описывать физические факты. Вероятности играют для нас ту же роль, что и массы в механике: можно обсуждать движение планетной системы, не зная индивидуальных масс планет и не рассматривая методов для их действительного измерения; можно также изучать гипотетическое движение несуществующей планетной системы. Мы будем обычно интересоваться общими законами и, реже, точными численными расчетами. Полезные вероятностные модели часто относятся к несуществующим мирам. Так, системы, используемые на автоматических телефонных станциях, основываются на простых вероятностных соображениях, и на этом основании вкладываются большие средства. Ясно, что соответствующая теория должна сравнивать различные возможные системы телефонных станций, большинство из которых никогда не будут осуществлены, так как теория докажет их невыгодность. Аналогично в страховом деле теория вероятностей используется для того, чтобы избежать некоторых нежелательных ситуаций, которые в результате никогда не будут наблюдаться. Если же желательны действительные наблюдения и численные оценки, то обычно оказывается необходимым использовать усовершенствованные методы, которые образуют главу математической статистики и лежат вне теории вероятностей в собственном смысле слова.

§ 3. «Статистическая» вероятность

Успех современной математической теории вероятностей приобретен следующей ценой: теория ограничивается лишь некоторыми сторонами общего предмета. Интуитивное понятие вероятности связано с предположениями и суждениями, подобными таким, как «Павел, вероятно, счастливый человек», «вероятно, эта книга будет неудачной», «гипотеза Ферма, вероятно, ошибочна». Вероятные мнения такого сорта интересны философам и логикам, и они являются также законным объектом математической теории [1]). Нужно, однако, установить, что мы далее будем иметь дело не с методами индуктивных заключений, а с тем, что может быть названо физической или статистической вероятностью. Грубо говоря, мы можем охарактеризовать это понятие, сказав, что наши вероятности относятся не к мнениям, а к исходам мыслимого эксперимента. Прежде чем говорить о вероятностях, мы должны условиться об идеализированной модели некоторого мыслимого эксперимента, подобного бросанию монеты, составлению выборки из кенгуру, живущих на луне, наблюдению частицы при диффузии или подсчету числа телефонных вызовов. Вначале мы должны условиться о возможных исходах нашего эксперимента (о нашем пространстве элементарных событий), и наши вероятности

---

[1]) B. O. Koopman, The axioms and algebra of intuitive probability, *Annals of Mathematics* (2), 41 (1940), 269—292; The bases of probability, *Bulletin of the American Mathematical Society*, 46 (1940), 763—774.

будут связаны с этими исходами и ни с чем другим. Это аналогично обычному образу действия в механике, когда вводится воображаемая модель, включающая две, три или семнадцать материальных точек, и эти точки лишаются их индивидуальных свойств. Аналогично если мы условимся анализировать игру с монетой, то мы говорим о последовательностях типа «герб, герб, решетка, герб, решетка, ...» и ни о чем другом. В нашей системе нет места для предположений, связанных с вероятностью того, что завтра взойдет солнце. Прежде чем говорить о такой вероятности, мы должны были бы условиться об (идеализированной) модели эксперимента, и она выглядела бы примерно так: «случайным образом выбирается один из бесконечного числа миров...» Небольшого воображения достаточно для того, чтобы построить такую модель, но она окажется и неинтересной и бессмысленной.

Астрономы говорят об измерении температуры в центре Солнца или о путешествии на Сириус. Эти операции выглядят невозможными, и все-таки размышление о них не бессодержательно. По тем же причинам мы не будем беспокоиться о том, выполним или нет наш мыслимый эксперимент; мы будем анализировать абстрактную модель. В глубине нашего сознания мы сохраняем интуитивную интерпретацию вероятностей, которая приобретает опытный смысл в некоторых приложениях. Мы представляем себе эксперимент, выполняемый очень много раз. О событии с вероятностью 0,6 можно предполагать, что в длинной серии экспериментов оно осуществится шестьдесят раз из ста. Это описание умышленно неопределенно, но оно создает образные интуитивные представления, достаточные для некоторых более элементарных приложений.

По мере развития и усложнения теории действительный смысл и интуитивные представления будут уточняться и конкретизироваться.

§ 4. Резюме

Мы будем иметь дело с теоретическими моделями, в которых вероятности входят в качестве свободных параметров, подобно массам в механике. Эти модели можно применять многими различными способами. Техника приложений и вероятностная интуиция будут развиваться вместе с теорией.

Это обычный путь, который плодотворно используется и в других математических дисциплинах. Не существует никакого другого пути, который мог бы в достаточной мере удовлетворить всем разнообразным нуждам и требованиям растущего организма, который называется теорией вероятностей и ее приложениями.

Можно сетовать на то, что интуитивное понятие вероятности недостаточно для научной теории, но это исторический факт. В примере (6.б гл. I) мы исследуем случайное размещение частиц по ячейкам. Соответствующее «естественное» распределение вероятностей казалось совершенно очевидным каждому и принималось без колебаний физиками. Однако оказалось, что физические частицы не ведут себя в соответствии со здравым человеческим смыслом, и «естественное» распределение Больцмана пришлось в одних случаях заменить на распределение Эйнштейна — Бозе, а в других — на распределение Ферми — Дирака. Не существовало никаких интуитивных доьодов в пользу того, что фотоны должны вести себя не так, как протоны, и что оба типа частиц не должны подчиняться «априорным» законам. Если обоснование этих фактов и может теперь быть найдено, то это свидетельствует только о том, что интуиция развивается вместе с теорией. Во всяком случае, даже для приложений существенны свобода и гибкость теории, и было бы пагубной ошибкой сковывать ее слишком жесткими ограничениями.

Иногда приходится слышать заявления о том, что современная теория вероятностей слишком сложна, чтобы быть полезной. Столь же воинственный крик поднимался в свое время практически мыслящими людьми против теории Максвелла. Чтобы опровергнуть эти доводы, достаточно было бы указать на многочисленные новые приложения, открытые абстрактной теорией случайных процессов или на новые достижения современной теории случайных колебаний, которые противоречат наивной интуиции и приводят к пересмотру некоторых практических рекомендаций. Однако этот спор просто бесполезен. Осуждать слишком легко. Те вещи, которые стали сегодня практически важными, еще вчера порицались как непрактические, и теории, которые станут практически важными завтра, всегда будут клеймиться как ничего не стоящая абстрактная игра практическими людьми сегодняшнего дня.

§ 5. Исторические замечания

Статистический или эмпирический подход к вероятности был развит главным образом Фишером и Мизесом. Понятие пространства элементарных событий [1]) идет от Мизеса. Это понятие сделало возможным построение строгой математической теории вероятностей на основе теории меры. Такой подход развивался постепенно в течение 20-х годов под влиянием многих авторов. На современном уровне аксиоматическое построение было дано А. Н. Колмогоровым [2]). Мы будем следовать этой линии, хотя термин «аксиоматическое построение», быть может, слишком торжествен для этой книги, рассматривающей лишь простой случай дискретных вероятностей.

---

[1]) См. книгу Мизеса Wahrscheinlichkeitsrechnung, Лейпциг и Вена, 1931, со ссылками на его предыдущие работы, датированные начиная с 1921 года. Немецкий термин — Merkmalraum (пространство меток).

[2]) А. Н. Колмогоров, Основные понятия теории вероятностей, М. — Л., ОНТИ, 1936 (в 1933 г. — на немецком языке).

ГЛАВА I

ПРОСТРАНСТВА ЭЛЕМЕНТАРНЫХ СОБЫТИЙ

§ 1. Опытные основания

Математическая теория вероятностей приобретает практическую ценность и наглядный смысл в связи с такими действительными или мыслимыми опытами и явлениями, как, например, однократное бросание монеты, бросание монеты 100 раз, бросание трех костей, сдача колоды карт, сопоставление двух колод карт, игра в рулетку, наблюдение продолжительности жизни радиоактивного атома или человека, выбор некоторой случайной группы людей и подсчет среди них числа левшей, скрещивание двух сортов растений и наблюдение фенотипов потомков, определение числа занятых линий на телефонной станции или числа телефонных вызовов, случайные шумы в электрических системах, выборочный контроль качества промышленной продукции, частота несчастных случаев, пол новорожденного, число двойных звезд на некотором участке неба, положение частицы при диффузии. Пока описание всех этих явлений довольно неопределенно, и, чтобы придать теории точный смысл, мы должны условиться о том, чтó мы понимаем под *возможными исходами рассматриваемого опыта или наблюдения*.

При бросании монеты не обязательно выпадает герб или решетка; монета может укатиться прочь или стать на ребро. Тем не менее мы условимся рассматривать герб и решетку как единственно возможные исходы бросания монеты. Это соглашение упрощает теорию и не сказывается на возможностях ее применения. Подобного рода упрощения часто неизбежны. Невозможно безошибочно измерить продолжительность существования какого-либо атома или время жизни какого-нибудь лица; однако для научных целей выгодно считать эти величины точными числами. Но при этом возникает вопрос: какие числа могут и какие не могут представлять продолжительность жизни человека? Существует ли максимальный возраст, сверх которого жизнь невозможна, или для возраста возможны любые значения? Мы, конечно, не решимся допустить, что человек может дожить до 1000 лет, и тем не менее обычная практика страхового дела не принимает даже такой границы для длительности жизни. В соответствии с формулами, на которых основаны современные таблицы смертности, доля людей, доживающих до 1000 лет, имеет величину порядка единица, деленная на $10^{10^{36}}$ — число, записывающееся 10^{27}

миллиардами знаков. Это утверждение лишено смысла с точки зрения биологии или социологии, но если его рассматривать исключительно с точки зрения статистики, то оно не противоречит опыту. В течение столетия рождается менее чем 10^{10} людей, и чтобы опровергнуть вышеуказанное утверждение статистически, потребовалось бы более чем $10^{10^{35}}$ столетий, что превышает возраст земного шара более чем в $10^{10^{34}}$ раз. Очевидно, столь исключительно малые вероятности совместимы с нашим понятием невозможности. Можно было бы подумать, что их употребление является полным абсурдом. В действительности оно совершенно безвредно и приводит к упрощению многих формул. Кроме того, если бы мы решили всерьез исключить возможность дожить до 1000 лет, то мы тут же столкнулись бы с еще большими трудностями, ибо мы должны были бы допустить существование максимального возраста. А право же, предположение, что можно дожить до x лет, но нельзя прожить x лет и 2 секунды, нисколько не привлекательнее, чем представление об отсутствии границы для продолжительности жизни.

Любая теория обязательно заключает в себе некоторое упрощение, и обычно последнее так естественно, что о нем и не говорится. Наше первое упрощение касается возможных исходов «опыта» или «наблюдения». Только эти возможные исходы являются объектами математической теории. Если мы хотим построить абстрактную модель опыта, мы должны вначале установить, что представляет собой возможный исход упрощенного (идеализированного) опыта.

Для единства терминологии результаты опытов или наблюдений будут называться *событиями*. Так, мы будем говорить о событии, заключающемся в том, что из пяти брошенных монет более трех выпали гербом вверх. Аналогично «эксперимент», состоящий в сдаче колоды карт для игры в бридж[1]), может иметь своим результатом «событие», заключающееся в том, что первый игрок получил два туза. Состав выборки («среди выбранных 85 людей — двое левшей») и результат измерения («температура 120°», «семь свободных проводов») будут также называться событиями.

Условимся различать составные (или разложимые) события и элементарные (или неразложимые) события. Например, сказать, что сумма очков, выпавших при бросании двух костей, равна шести, все равно, что сказать, что опыт привел к одному из исходов «(1.5),

[1]) *Определение бриджа и покера.* Колода карт для игры в бридж состоит из 52 карт, которые делятся на четыре равные группы по масти. Карты каждой масти различаются по значению. Имеется 13 значений (2, 3, ..., 10, валет, дама, король, туз). Четыре масти называются: пики, трефы, черви и бубны. Две последние масти — красные, две первые — черные. Карты одного значения называются однотипными. Для нас игра в бридж означает сдачу колоды карт четырем игрокам по 13 карт каждому. Игра в покер, по определению, означает выбор пяти карт из колоды.

или (2.4), или (3.3), или (4.2), или (5.1)», и это перечисление разлагает событие «сумма очков равна шести» на пять элементарных событий. Аналогично этому событие «две нечетные грани» допускает разложение «(1.1), или (1.3), или, ..., или (5.5)» на девять элементарных событий. Заметим, что если в результате нашего эксперимента мы получили (3.3), то те же два бросания привели также к событиям «сумма очков равна шести», «две нечетные грани». Эти события не являются взаимно исключающими и поэтому могут происходить одновременно. В качестве второго примера рассмотрим возраст человека. Каждое частное значение x представляет элементарное событие, тогда как утверждение о том, что данному человеку за пятьдесят, описывает составное событие «$x > 50$». Итак, каждое составное событие может быть разложено на элементарные события; другими словами, составное событие есть совокупность элементарных событий.

Если мы хотим говорить об «опытах» и «наблюдениях» научно и без каких бы то ни было неясностей, то мы должны сначала условиться, каковы элементарные (неразложимые далее) события, представляющие собой мыслимые исходы опыта или наблюдения; они определяют идеализированный опыт. По определению, *каждый неразложимый исход (идеализированного) опыта представляется одним и только одним элементарным событием.* Совокупность всех элементарных событий будем называть *пространством элементарных событий*, а сами элементарные события — *точками этого пространства.* Все события, связанные с данным (идеализированным) опытом, могут быть описаны с помощью элементарных событий.

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

§ 2. Примеры

а) **Распределение трех шаров по трем ящикам.** Таблица 1 дает все возможные исходы опыта, состоящего в распределении трех шаров по трем ящикам.

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

Таблица 1

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

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

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

Мы воспользуемся этим примером, чтобы проиллюстрировать тот важный факт, что природа пространства элементарных событий не играет роли для нашей теории. Для нас это пространство (вместе с заданным на нем распределением вероятностей) определяется идеализированным экспериментом. Мы пользуемся образным языком шаров и ящиков, но то же самое пространство допускает большое число практически важных интерпретаций. Чтобы пояснить нашу точку зрения, а также для удобства дальнейших ссылок, мы приведем ряд схем, внешне весьма различных, но по существу эквивалентных абстрактной схеме размещения r шаров по n ящикам, в том смысле, что соответствующие исходы отличаются лишь словесным выражением. Вероятностные распределения при этом не одинаковы, но этот вопрос будет обсуждаться ниже.

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

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

(б.3) При стрельбе по n мишеням пули соответствуют шарам, мишени — ящикам.

(б.4) *Выбор.* Пусть группа из r человек классифицируется, скажем, по возрасту или по профессии. Классы играют роль ящиков, люди — шаров.

(б.5) *Облучение в биологии.* Когда сетчатка глаза подвергается воздействию света, кванты света выступают в роли шаров, а ячейки сетчатки соответствуют ящикам. Аналогично при изучении генетического влияния излучения хромосомы соответствуют ящикам, α-частицы — шарам.

(б.6) При экспериментах с космическими лучами частицы, попадающие в счетчики Гейгера, играют роль шаров, а сами счетчики — ящиков.

(б.7) Лифт отправляется с r пассажирами и останавливается на n этажах. Распределение пассажиров по группам соответственно этажу, на котором они выйдут, соответствует размещению r шаров по n ящикам.

(б.8) *Кости.* Возможному исходу эксперимента, состоящего в бросании r костей, соответствует распределение r шаров по $n=6$ ячейкам. Если бросают монеты, то имеет место та же ситуация, только $n=2$.

(б.9) *Случайные цифры.* Каждой последовательности r случайных цифр отвечает размещение r шаров (мест) по $n=10$ ящикам с номерами 0,1, ..., 9.

(б.10) *Распределение r человек по признаку пола.* В этом случае имеется $n=2$ ящика и r шаров.

(б.11) *Коллекция купонов.* Различные типы купонов соответствуют ящикам, все собранные купоны — шарам.

(б.12) *Распределение тузов между игроками при игре в бридж.* Четыре игрока соответствуют четырем ящикам, четыре туза — шарам.

(б.13) *Распределение генов.* Каждый потомок некоторой особи (человека, растения или животного) наследует определенные гены. Если некоторый ген может находиться в одной из n форм $A_1, \ldots, A_n$, то живые организмы можно классифицировать по генотипам. Потомки особи соответствуют шарам, генотипы $A_1, \ldots, A_n$ — ящикам.

(б.14) *Химические реакции.* Предположим, что молекулярные цепочки некоторого полимера реагируют с кислородом. Каждая цепочка может прореагировать с 0, 1, 2, ... молекулами кислорода. Реагирующие молекулы кислорода играют роль шаров, а цепочки полимера — роль ящиков, в которых размещаются шары.

(б.15) *Теория фотоэмульсий.* Фотографическая пластинка покрыта слоем светочувствительных зерен, каждое зерно реагирует, если в него ударяется определенное число r квантов. Для теории важно знать число прореагировавших частиц (т. е. частиц, в которые попало r квантов). В этом случае мы получаем типичную задачу о размещении, где зерна эмульсии соответствуют ящикам, а кванты света —

шарам. (В действительности ситуация более сложна, так как пластинка обычно покрыта зернами различной чувствительности.)

(б.16) *Опечатки.* Возможные распределения r опечаток в книге, состоящей из n страниц, соответствуют размещениям r шаров по n ящикам, если только r меньше, чем число букв на странице.

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

Различимы ли шары на самом деле, для нашей теории не существенно. Если даже это и так, мы можем условиться считать их неразличимыми. Тузы в бридже [пример (б.12)] или люди в лифте [пример (б.7)], конечно, различны, и тем не менее часто предпочтительнее считать их неразличимыми. Кости в примере (б.8) можно окрасить и сделать их тем самым различимыми, но при обсуждении конкретной задачи мы можем воспользоваться и схемой с различимыми шарами, и схемой с неразличимыми шарами, исходя целиком из наших целей и удобства. Природа задачи может подсказать выбор, но в любом случае нашу теорию можно развивать только после того, как соответствующая модель выбрана, т. е. после того, как определено пространство элементарных событий.

В приведенном выше примере мы рассматривали неразличимые шары, но табл. 2 еще связана с первым, вторым, третьим ящиком, и их порядок существен. Мы можем пойти еще дальше и предположить, что даже ящики неразличимы (например, ящики можно выбирать случайно, вне зависимости от их содержимого). Если и шары и ящики неразличимы, возможно только три размещения, а именно

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

Таблица 2

1.	$\{***\mid - \mid - \}$	6.	$\{ * \mid ** \mid - \}$
2.	$\{ - \mid *** \mid - \}$	7.	$\{ * \mid - \mid ** \}$
3.	$\{ - \mid - \mid *** \}$	8.	$\{ - \mid ** \mid * \}$
4.	$\{ ** \mid * \mid - \}$	9.	$\{ - \mid * \mid ** \}$
5.	$\{ ** \mid - \mid * \}$	10.	$\{ * \mid * \mid * \}$

г) **Выбор.** Предположим, что с целью оценить число курящих выбрана группа в 100 человек. При этом единственное интересующее нас свойство данной выборки — число курящих x; оно может быть равно любому целому числу между 0 и 100. В данном случае следует рассмотреть пространство элементарных событий, состоящее из 101 «точки» $x = 0, 1, \ldots, 100$. Исход каждого отдельного выбора (или наблюдения) полностью описывается заданием соответствующей точки x. Примером составного события является исход «большинство людей данной выборки — курящие». Это означает, что опыт привел к одному из 50 событий $x = 51, 52, \ldots, 100$ (безразлично, к какому именно). Аналогично любое свойство этой выборки можно описать, перечислив соответствующие случаи или элементарные события. Для единства терминологии мы предпочитаем говорить о событиях, а не о свойствах выборки. Выражаясь математически, событие является просто множеством соответствующих точек пространства элементарных событий.

д) **Выбор (продолжение).** Допустим теперь, что выбранные нами 100 человек классифицируются не только на курящих и некурящих, но и по полу. Выбранная группа может быть теперь охарактеризована четверкой чисел $M_к$, $Ж_к$, $M_н$, $Ж_н$, означающих по порядку число курящих мужчин, число курящих женщин, число некурящих мужчин, число некурящих женщин. В качестве пространства элементарных событий следует рассмотреть множество всех четверок целых чисел, заключенных между 0 и 100 и дающих в сумме 100. Таких четверок имеется 176 851; они и образуют пространство элементарных событий (см. § 5 гл. II). Утверждение, что в выборке «среди мужчин доля курящих больше, чем среди женщин», означает, что для нашей группы дробь $M_к/M_н$ больше дроби $Ж_к/Ж_н$. Точка (73, 2, 8, 17) обладает этим свойством, точка (0, 1, 50, 49) — не обладает. В принципе наше событие может быть описано путем перечисления всех четверок чисел, обладающих заданным свойством.

е) **Бросание монеты.** Если монета бросается 3 раза, то пространство элементарных событий состоит из восьми точек, которые можно условно обозначить: *ГГГ*, *ГГР*, *ГРГ*, *РГГ*, *ГРР*, *РГР*, *РРГ*, *РРР* (*Г* означает выпадение герба, *Р* — выпадение решетки). Событие A «выпало не менее двух гербов» можно описать множеством первых четырех точек. Событие B «выпала ровно одна решетка» означает или *ГГР*, или *ГРГ*, или *РГГ*; мы говорим, что указанное событие B содержит эти три точки.

ж) **Возраст супругов.** Страховые компании интересуются распределением возрастов супружеских пар. Пусть x — возраст мужа, y — возраст жены. Каждое наблюдение дает пару чисел (x, y). За пространство элементарных событий следует принять первый квадрант плоскости x, y, так что каждая точка $x > 0$, $y > 0$ будет элементарным событием. Событие A «мужу больше 40 лет» представляется всеми точками, лежащими справа от прямой $x = 40$; событие B «муж старше жены» представляется областью, заключенной между осью x и прямой $y = x$, т. е. совокупностью всех точек, для которых $x > y$; событие C «жене больше 40 лет» представляется частью первого квадранта, расположенной над прямой $y = 40$. Для геометрического представления совместного распределения возрастов двух супружеских пар нам потребовалось бы четырехмерное пространство.

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

§ 3. Пространство элементарных событий. События

Из предыдущего должно быть ясно, что мы никогда не будем говорить о вероятностях вне связи с каким-нибудь пространством элементарных событий (или физически — вне связи с некоторым мыслимым опытом). *Мы отправляемся от понятия пространства элементарных событий и его точек; впредь они будут рассматриваться как данные. Они являются первоначальными и неопределяемыми понятиями теории*, так же как понятия «точка» и «прямая» остаются неопределяемыми при аксиоматическом построении евклидовой геометрии. Природа элементарных событий не интересует нашу теорию. Пространство элементарных событий служит моделью идеализированного опыта в том смысле, что, по определению, *любой мыслимый исход опыта полностью описывается одним и только одним элементарным событием.* О каком-либо событии A имеет смысл говорить лишь в том случае, если для любого исхода опыта известно, произошло или не произошло событие A. Совокупность точек, представляющих все те исходы, при которых событие A происходит, полностью описывает это событие. Обратно, произвольное заданное множество A, содержащее одну или более точек нашего пространства, можно рассматривать как событие; оно происходит или не происходит в зависимости от того, принадлежит или не принадлежит множеству A точка, представляющая исход опыта. Поэтому мы определяем слово *событие* как термин, означающий *некоторое множество элементарных событий.* Мы будем говорить, что *событие A состоит из определенных точек*, а именно точек, представляющих те исходы идеализированного опыта, при которых происходит событие A.

Пример. В пространстве элементарных событий примера 2, а рассмотрим событие U, состоящее из точек с номерами 1, 7, 13. Это определение является самым прямым и формальным, но U можно описать и многими другими способами, Например, U можно определить как событие, состоящее в одновременном выполнении следующих трех условий: (1) второй ящик пуст, (2) шар a находится в первом ящике, (3) шар b не появляется после c. Каждое из этих трех условий в свою очередь определяет событие. Событие U_1, описанное только условием (1), состоит из точек 1, 3, 7—9, 13—15.

Событие U_2, определяемое (2), состоит из точек 1, 4, 5, 7, 8, 10, 13, 22, 23 и, наконец, событие U_3, определяемое (3), состоит из точек 1—4, 6, 7, 9—11, 13, 14, 16, 18—20, 22, 24, 25. Событие U можно теперь описать как одновременное осуществление всех трех событий U_1, U_2, U_3.

Хотя термины «событие» и «элементарное событие» имеют и наглядный смысл, эти понятия эквивалентны точечным множествам и точкам, как их понимают во всех разделах математики.

Предыдущий пример и пример 2, а показывают, что в терминах двух или более заданных событий можно определить новые события. Опираясь на подобные элементарные примеры, мы введем систему обозначений формальной алгебры событий (т. е. алгебры точечных множеств).

§ 4. Отношения между событиями

Теперь мы предположим, что задано произвольное, но фиксированное пространство элементарных событий $\mathfrak{S}$.

Определение 1. *Мы будем пользоваться записью $A = 0$ для выражения того, что событие A не содержит элементар-*

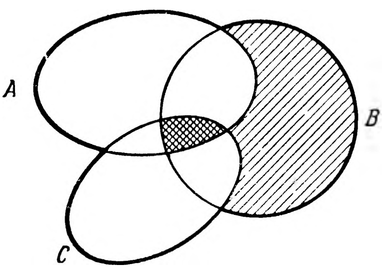


Рис. 1. Область, ограниченная внешним контуром, является объединением $A \cup B \cup C$. «Треугольная» (заштрихованная) область является пересечением ABC. «Лунообразная» (заштрихованная) область является пересечением B и события, противоположного $A \cup C$.

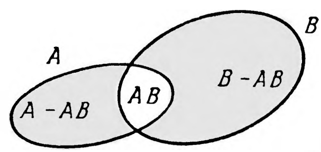


Рис. 2. Пересечение и разности событий.

ных событий (событие A невозможно). Нуль следует понимать в символическом смысле, а не как число.

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

Определение 2. *Событие, состоящее из всех точек, не содержащихся в событии A, называется событием,* *противоположным событию A (или отрицанием A), и обозначается $\bar{A}$.*

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

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

которое должно читаться: «события A и B несовместны». Событие $A\bar{B}$ означает, что произошло и A и $\bar{B}$, т. е. что событие A произошло, а событие B не произошло. Аналогично $\bar{A}\bar{B}$ означает, что ни A, ни B не произошло. Событие $A \cup B$ означает, что по крайней мере одно из событий A или B произошло; оно содержит все точки, за исключением точек, не принадлежащих ни A, ни B.

С точки зрения теории вероятностей мы можем охарактеризовать событие AB как одновременное осуществление событий A и B. По стандартной математической терминологии событие AB называется (логическим) пересечением событий A и B. Подобным же образом событие $A \cup B$ называется объединением событий A и B[1]). Введенные нами понятия переносятся и на случай многих событий A, B, C, D, ... Именно, можно определить два новых события $ABCD\ldots$ и $A \cup B \cup C \cup D \cup \ldots$, состоящих соответственно в одновременном осуществлении всех событий A, B, C, D, ... и в осуществлении по крайней мере одного из событий A, B, C, D, ...

Определение 3. *Любой совокупности событий A, B, C, ... мы сопоставляем два новых события следующим образом. Множество, состоящее из элементарных событий (точек), принадлежащих одновременно всем заданным событиям, будет обозначаться $ABC\ldots$ и называться пересечением (или одновременным осуществлением) событий A, B, C, ... Множество, состоящее из элементарных событий, каждое из которых принадлежит по крайней мере одному из заданных событий, будет обозначаться $A \cup B \cup C\ldots$ и называться объединением или осуществлением по крайней мере одного из заданных событий. События A, B, C, ... попарно несовместны (взаимно исключают друг друга), если никакие два из них не имеют общей точки, т. е. если $AB = 0$, $AC = 0$, ..., $BC = 0$, ...*

---

[1]) В русской литературе по теории вероятностей событие AB часто называют произведением, а событие $A \cup B$ — суммой событий A и B. Вместо $A \cup B$ пользуются также обозначением $A + B$, а вместо AB — обозначением $A \cap B$. — *Прим. ред.*

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

Определение 4. *Если каждая точка события A содержится в событии B, то мы будем писать $A \subset B$ и говорить, что событие A влечет за собой событие B. С другой стороны, мы будем также писать $B \supset A$ и говорить, что событие B является следствием события A. В этом случае мы будем также писать $B - A$ вместо $B\bar{A}$ для обозначения того, что событие B произошло, а событие A не произошло.*

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

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

Примеры. а) Если события A и B несовместны, то наступление события A влечет ненаступление события B, и наоборот. Следовательно, $AB = 0$ означает то же самое, что $A \subset \bar{B}$ и $B \subset \bar{A}$.

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

в) В примере 2, ж событие AB означает, что мужу больше 40 лет и что он старше своей жены, тогда как $A\bar{B}$ означает, что ему свыше 40 лет, но он не старше своей жены. Событие AB поэтому изображается областью, заключенной между осью x и прямыми $x = 40$ и $y = x$. Событие $A\bar{B}$ изображается областью между прямыми $x = 40$ и $y = x$ с включенной второй границей. Событие AC означает, что обоим супругам больше 40 лет. Событие $A \cup C$ означает, что по крайней мере одному из них больше 40 лет, тогда как $A \cup B$ означает, что либо мужу свыше 40 лет, либо, если это не так, он по крайней мере старше своей жены (иными словами, возраст мужа превосходит меньшую из двух величин: 40 лет и возраст жены).

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

д) *Бридж* (см. примечание на стр. 18). Пусть A, B, C, D означают события, состоящие соответственно в том, что при сдаче колоды карт для игры в бридж первый, второй, третий и четвертый игрок получил по крайней мере одного туза. Ясно, что хотя бы один из игроков имеет туза, так что по крайней мере одно из четырех событий должно иметь место. Следовательно, $A \cup B \cup C \cup D = \mathfrak{S}$ есть все пространство элементарных событий. Событие $ABCD$ происходит тогда и только тогда, когда каждый игрок имеет туза. Событие «четвертый игрок получил всех четырех тузов» означает, что ни одно из событий A, B, C не произошло; это означает, что одновременно осуществились события $\bar{A}$, $\bar{B}$ и $\bar{C}$, т. е. произошло событие $\bar{A}\bar{B}\bar{C}$.

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

§ 5. Дискретные пространства элементарных событий

Простейшими пространствами элементарных событий являются те, которые содержат конечное число n точек. Если n достаточно мало (как в случае бросания нескольких монет), то пространство легко себе представить. Пространство элементарных событий при сдаче карт для игры в бридж гораздо сложнее. Тем не менее можно представить себе, что каждое элементарное событие изображается какой-то фишкой, и затем рассматривать множество этих фишек как изображение пространства элементарных событий. Событие A (например, «первый игрок получил двух тузов») изображается некоторой группой фишек; событие $\bar{A}$ изображается остальными фишками. Остается сделать лишь один шаг для представления о ящике с бесконечным числом фишек или о пространстве элементарных событий, состоящем из бесконечной последовательности точек E_1, E_2, E_3, ...

Примеры. а) Условимся бросать монету до тех пор, пока не выпадет герб. Тогда элементарными событиями будут: $E_1 = Г$, $E_2 = РГ$, $E_3 = РРГ$, $E_4 = РРРГ$ и т. д. Мы можем допускать или не допускать возможности того, что герб никогда не выпадет. Если это допускается, то соответствующая возможность должна представляться специальной точкой E_0.

а) Три игрока a, b и c проводят шахматный турнир по следующей системе. В первом туре играют a и b, а игрок c свободен. Проигравший заменяется игроком c, и во втором туре играют победитель и c, а игрок, потерпевший поражение в первом туре, свободен. Турнир продолжается таким образом до тех пор, пока один из игроков не выиграет двух партий подряд, в последнем случае он объявляется победителем турнира. Для простоты мы исключаем возможность ничьей в отдельной партии. Возможные исходы турнира описываются следующей схемой:

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

Наряду с этим вполне мыслимо, что ни один из игроков не выиграет двух партий подряд, т. е. что турнир не окончится. Этому случаю отвечает одна из двух последовательностей:

$$acbacbacbacb\ \ldots,\quad bcabcabcabca. \qquad (**)$$

Пространство элементарных событий, соответствующее нашему идеализированному «опыту», определяется формулами $(*)$ и $(**)$ и бесконечно. Однако ясно, что точки этого пространства можно расположить в простую последовательность. Для этого достаточно, например, на первые два места поставить точки $(**)$, а на последующие — точки $(*)$ в таком порядке: aa, bb, acc, bcc, ... (Продолжение этого примера см. в задачах 5 и 6, примере 2, а гл. V и задаче 5 гл. XV.)

Определение. *Пространство элементарных событий называется дискретным, если оно состоит лишь из конечного числа точек или из такого бесконечного числа точек, что они могут быть расположены в простую последовательность* $E_1, E_2, \ldots$

Не всякое пространство элементарных событий дискретно. Существует теорема (принадлежащая Г. Кантору) о том, что пространство элементарных событий, состоящее из всех положительных чисел, не дискретно. Здесь мы сталкиваемся с обстоятельством, известным в механике, где обычно сначала рассматривают дискретные материальные точки, имеющие каждая конечную массу. Этой концепции противопоставляется непрерывное распределение массы, когда каждая отдельная точка имеет массу, равную нулю. В первом случае масса системы получается попросту сложением масс отдельных точек; во втором случае массы вычисляются интегрированием плотности. Совершенно аналогично вероятности событий в дискретном пространстве элементарных событий получаются простым сложением, тогда как в других пространствах необходимо интегрирование. Кроме используемого технического аппарата, эти два случая ничем существенным не отличаются. Желая изложить собственно вероятностные рассуждения, не обремененные техническими трудностями, мы сначала займемся лишь дискретными пространствами элементарных событий. Мы увидим, что даже этот частный случай приводит ко многим интересным и важным результатам.

В этой книге мы будем рассматривать только дискретные пространства элементарных событий.

§ 6. Вероятности в дискретных пространствах элементарных событий

Вероятности различных событий — числа той же природы, что и расстояния в геометрии или массы в механике. Абстрактная теория предполагает, что они даны, и не нуждается ни в каких предположениях об их действительном численном значении или о способе их измерения на практике. Целый ряд наиболее важных приложений носит качественный характер и не зависит от численных значений вероятностей событий; общие же выводы теории находят себе многочисленные применения, совершенно так же, как теоремы геометрии служат основой и физических теорий и технических приложений. В тех сравнительно редких случаях, когда требуется знать численное значение вероятностей событий, вычислительные приемы варьируются так же широко, как меняются методы определения расстояний. Когда плотник, землемер, летчик и астроном измеряют расстояния, то в их действиях мало общего. В нашем круге вопросов мы будем, например, рассматривать коэффициент диффузии, определяемый с помощью понятий теории вероятностей. Чтобы найти численное значение этого коэффициента, требуются физические рассмотрения, связывающие явление диффузии с другими теориями; прямое же измерение невозможно. Таблицы продолжительности жизни, наоборот, составляются на основании наблюдений. В наиболее важных приложениях определение вероятностей событий или сравнение результатов теории с данными наблюдений требуют применения довольно сложных статистических методов, основанных в свою очередь на тонкой вероятностной теории. Другими словами, хотя наглядный смысл вероятностей событий и ясен, но лишь по мере развития теории мы увидим, как следует применять это понятие.

Бросая «правильную» монету, мы не колеблемся связать вероятность $1/2$ с выпадением герба или решетки. Это приводит к выводу, что при n бросаниях монеты все 2^n возможных случаев равновероятны. С теоретической точки зрения, это — *допущение*. Часто говорят, что такое допущение — логически неизбежное и единственно возможное. Однако некоторые статистики отрицают это и исходят из противоположных допущений. Говорят также, что вероятности $1/2$ следуют из опыта. В действительности, когда для исследования бросания монеты использовались утонченные статистические методы, неизменно обнаруживалось, что выпадение герба и решетки *не* одинаково вероятно. И все-таки мы придерживаемся нашей модели «идеальной» монеты, хотя на самом деле правильных монет не существует. Мы сохраняем эту модель не столько ради ее логической простоты, сколько в основном из-за ее полезности в приложениях: во многих приложениях эта модель достаточно хорошо описывает действительность. Еще важнее тот эмпирический факт, что отклонения от нашей схемы всегда связаны с такими явлениями, как, например, эксцентрическое положение центра тяжести. Таким образом, наша идеализированная модель может быть крайне полезной, даже несмотря на то, что она не соответствует вполне точно ни одной реальной монете. Например, в современном статистическом контроле качества продукции идеализированные вероятностные модели используются для выявления причин явных отклонений от этих моделей с целью устранения уже на ранней стадии производства нарастающих нарушений правильной работы машин и регулярного течения производственного процесса.

Подобные замечания относятся и к другим случаям. Число возможных распределений колоды карт для игры в бридж равно почти 10^{30}. Обычно мы соглашаемся считать эти распределения карт равновероятными. Для проверки этого допущения потребовалось бы более 10^{30} опытов, т. е. если каждая игра продолжается 1 секунду, миллион миллиардов лет непрерывной игры всех живущих на Земле людей. Тем не менее следствия из нашего предположения могут быть проверены опытным путем, например наблюдением за частотой появления нескольких тузов у одного игрока. Оказывается, что идеализированная модель описывает явление достаточно хорошо, если только карты тщательно тасуются. Еще важнее то, что в случаях, когда идеализированная схема оказывается неприменимой, она позволяет обнаружить причины расхождений, например, неправильную тасовку карт. Эти примеры имеют ограниченное значение, но они указывают на пользу принимаемых моделей. Более интересные случаи мы сможем приводить лишь по мере развития теории.

Примеры. а) *Различимые шары.* В примере 2,а кажется естественным предположение о том, что все элементарные события *равновероятны*, т. е. *каждое из них имеет вероятность* $^1/_{27}$. Отправляясь от этого *определения*, мы можем получить различные следствия. Будет или не будет наша модель разумным приближением к действительности, зависит от типа явлений, к изучению которых она применяется. В некоторых приложениях гипотезу о равновероятности исходов выдвигают исходя из физических рассмотрений, в других она служит для упрощения рассуждений и получения общих закономерностей, хотя совершенно очевидно, что это лишь самое первое, грубое приближение. [Сказанное хорошо иллюстрируется примерами 2, б, 1, дни рождения; 2, б, 7, лифт; 2, б, 11, коллекция купонов.]

б) *Неразличимые шары; статистика Бозе — Эйнштейна.* Обратимся теперь к примеру 2, в, связанному с размещением трех неразличимых шаров по трем ящикам. Можно предположить, что невозможность различать шары не отражается на сущности физического эксперимента; остается по-прежнему 27 исходов, хотя только

10 из них действительно различимы. Эти рассуждения показывают, что точкам таблицы 2 разумно приписать следующие вероятности:

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

Следует отметить, что для большинства приложений, перечисленных в примерах 2, б, эти аргументы вполне удовлетворительны и выбор вероятностей разумен. Исторически эта точка зрения долгое время господствовала, не вызывая сомнений, и в статистической механике служила основанием для введения статистики Максвелла — Больцмана для размещения r шаров по n ящикам. Тем больше было общее удивление, когда Бозе и Эйнштейн показали, что определенные типы частиц подчиняются статистике Бозе — Эйнштейна (подробнее см. в § 5 гл. II). В рассматриваемом случае при $r = n = 3$ модель Бозе — Эйнштейна сопоставляет каждому из десяти элементарных событий вероятность $^1/_{10}$.

Этот пример показывает, что в одном и том же пространстве элементарных событий можно ввести различные вероятности, и еще раз подчеркивает сложный характер связи теории с практикой. В частности, он учит нас не слишком полагаться на априорные аргументы и быть готовыми принять новую и непредвиденную схему.

в) *Бросания монеты.* Частотная интерпретация постулата о равных вероятностях требует знания результатов действительных экспериментов. Любая реальная монета не является абсолютно симметричной, но можно осуществить физический эксперимент, который значительно ближе подходит к идеальной модели бросания монеты, чем этого можно достичь с реальной монетой. Для того чтобы дать представление о колебаниях, которые можно ожидать, мы приводим запись результатов этого опыта, соответствующую 10 000 бросаний монеты [1]).

Таблица 3 содержит числа появлений «герба» в сериях из ста экспериментов. Каждая серия отвечает последовательности ста бросаний правильной монеты. Полное число появлений герба равно 4979. Поглядев на приведенную таблицу; читатель, по всей вероятности, испытает чувство некоторой неудовлетворенности: «Ну и что же?» Истина состоит в том, что судить о степени согласия таких эмпирических данных с нашей абстрактной моделью можно лишь с помощью более развитой теории. (Мы вернемся к этому кругу вопросов в § 7 гл. III.)

---

[1]) На самом деле в этой таблице приведены частоты появления четных цифр в одном из разделов A Million Random Digits with 100 000 Normal Deviates, RAND Corporation, Free Press, Glenese, Illinois, 1955.

Таблица 3

Число испытаний	Число гербов										Общее число гербов
0— 1000	54	46	53	55	46	54	41	48	51	53	501
— 2000	48	46	40	53	49	49	48	54	53	45	485
— 3000	43	52	58	51	51	50	52	50	53	49	509
— 4000	58	60	54	55	50	48	47	57	52	55	536
— 5000	48	51	51	49	44	52	50	46	53	41	485
— 6000	49	50	45	52	52	48	47	47	47	51	488
— 7000	45	47	41	51	49	59	50	55	53	50	500
— 8000	53	52	46	52	44	51	48	51	46	54	497
— 9000	45	47	46	52	47	48	59	57	45	48	494
—10000	47	41	51	48	59	51	52	55	39	41	484

§ 7. Основные определения. Основные допущения

Пусть дано дискретное пространство элементарных событий $\mathfrak{S}$ с точками E_1, E_2, ... Мы допускаем, что с каждой точкой E_j связано число, называемое вероятностью элементарного события E_j и обозначаемое $P\{E_j\}$. Эти числа неотрицательны и таковы, что

$$P\{E_1\}+P\{E_2\}+\ldots=1. \qquad (7.1)$$

Заметим, что мы не исключаем для некоторой точки возможность иметь вероятность, равную нулю. Это допущение может показаться искусственным, но оно необходимо во избежание осложнений. В случае дискретного пространства элементарных событий вероятность нуль интерпретируется на практике как невозможность, и элементарное событие, имеющее вероятность, равную нулю, может быть безнаказанно изъято из пространства элементарных событий. Однако часто численные значения вероятностей заранее неизвестны, и нужны сложные рассмотрения, чтобы решить, имеет ли некоторая точка положительную вероятность.

Определение. *Вероятность $P\{A\}$ события A есть сумма вероятностей элементарных событий, составляющих событие A.*

Основное уравнение (7.1) устанавливает, что вероятность всего пространства элементарных событий равна единице: $P\{\mathfrak{S}\}=1$. Далее из (7.1) следует, что для любого события A

$$0 \leqslant P\{A\} \leqslant 1. \qquad (7.2)$$

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

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

Если E — точка, содержащаяся одновременно и в A_1 и в A_2, то $P\{E\}$ входит 2 раза в правую и один раз в левую часть неравенства. Поэтому правая часть превосходит левую на величину $P\{A_1A_2\}$, и нами доказана простая, но важная теорема.

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

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

Если $A_1A_2 = 0$, т. е. если событие A_1 и A_2 несовместны, то (7.4) *превращается в*

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

Пример. Монета бросается 2 раза. В качестве пространства элементарных событий возьмем четыре точки *ГГ*, *ГР*, *РГ*, *РР* и каждой из них сопоставим вероятность $^1/_4$. Пусть A_1 и A_2 означают соответственно события «при первом бросании выпал герб» и «при втором бросании выпал герб». Тогда событие A_1 состоит из точек *ГГ*, *ГР*, событие A_2 состоит из точек *РГ*, *ГГ*. Далее, событие $A = A_1 \cup A_2$ содержит три точки *ГГ*, *ГР*, *РГ*, тогда как событие A_1A_2 состоит из единственной точки *ГГ*. Таким образом, $P\{A_1 \cup A_2\} = {^1/_2} + {^1/_2} - {^1/_4} = {^3/_4}$.

Вероятность $P\{A_1 \cup A_2 \cup \ldots \cup A_n\}$ осуществления по крайней мере одного из n событий может быть вычислена по формуле, аналогичной формуле (7.4); это будет сделано в § 1 гл. IV. Здесь мы лишь заметим, что неравенство (7.3), очевидно, выполняется и в общем случае. Таким образом, *для произвольных событий A_1, A_2 ...* справедливо неравенство

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

В частном случае, когда события A_1, A_2, ... попарно несовместны,

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

Иногда неравенство (7.6) называется неравенством Буля.

Мы исследуем сначала простой частный случай, когда пространство элементарных событий состоит из конечного числа N точек, обладающих каждая вероятностью $1/N$. В этом случае вероятность какого-нибудь события A равна числу точек, входящих в A, деленному на N. В старой литературе элементарные события назывались «случаями», а точки, входящие в A, — «благоприятными случаями» (благоприятными для события A). Если все элементарные события равновероятны, то вероятность события A равна отношению числа благоприятных случаев к числу всех возможных случаев. К сожалению, этим утверждением много злоупотребляли для получения «определения» вероятности. Часто утверждают, что в *любом* конечном пространстве вероятности всех элементарных событий равны между собой. Это не так. При единственном бросании неправильной монеты пространство элементарных событий по-прежнему содержит всего лишь две точки (герб и решетку), но эти элементарные события могут иметь произвольные вероятности p и q, лишь бы было $p+q=1$. Новорожденный — либо мальчик, либо девочка, но в приложениях нам приходится допускать, что эти две возможности не равновероятны. Еще один контрпример дает пример 6, б. Применения пространств с равновероятными элементарными событиями почти полностью ограничены азартными играми и комбинаторикой.

§ 8. Задачи

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

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

3. Рассмотрим все 24 возможные перестановки символов 1, 2, 3, 4 и свяжем с каждой вероятность $^1/_{24}$. Обозначив через A_i событие, состоящее в том, что символ i оказался на i-м месте ($i = 1, 2, 3, 4$), проверить формулу (7.4).

4. Монета бросается до тех пор, пока 2 раза подряд она не выпадет одной и той же стороной. Каждому возможному исходу, требующему n бросаний, припишем вероятность $1/2^n$. Описать пространство элементарных событий. Найти вероятности следующих событий: а) опыт окончится до шестого бросания; б) потребуется *четное* число бросаний.

5. В пространстве элементарных событий примера 5, б припишем каждой точке из (*), содержащей ровно k букв, вероятность $1/2^k$. (Иными словами, aa и bb имеют вероятность $^1/_4$, acb имеет вероятность $^1/_8$ и т. д.) а) Показать, что сумма вероятностей всех точек из (*) равна единице (в силу этого каждая из двух точек (**) имеет вероятность нуль). б) Установить, что с вероятностью $^5/_{14}$ победителем будет первый игрок. Вероятность того, что победит игрок b, та же самая, а аналогичная вероятность для c равна $^2/_7$. в) Вероятность того, что победитель не определится до k-го тура, равна $1/2^{k-1}$. Проверить.

6. Усовершенствовать пример 5, б, допустив возможность ничьей в отдельной партии. Описать соответствующее пространство элементарных событий. Как ввести вероятности?

7. В задаче 3 показать, что $A_1A_2A_3 \subset A_4$ и $A_1A_2\bar{A}_3 \subset \bar{A}_4$.

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

9. Бросаются две кости. Пусть A — событие, состоящее в том, что сумма очков нечетная; B — событие, заключающееся в том, что хотя бы на одной из костей выпала единица. Описать события AB, $A \cup B$, $A\bar{B}$. Найти их вероятности при условии, что все 36 элементарных событий равновероятны.

10. Выяснить смысл следующих событий в примере 2, ж: а) ABC, б) $A - AB$, в) $A\bar{B}C$.

11. Проверить, что в примере 2, ж $A\bar{C} \subset B$.

12. Бридж (см. примечание на стр. 18). Пусть N_k, где $k = 1, 2, 3, 4$ — событие, состоящее в том, что первый игрок получил по крайней мере k тузов. Пусть S_k, E_k, W_k — аналогичные события для второго, третьего и четвертого игроков. Что можно сказать о числе находящихся у четвертого игрока тузов в случае событий: а) $\overline{W}_1$; б) N_2S_2; в) $\overline{N}_1\overline{S}_1\overline{E}_1$; г) $W_2 - W_3$; д) $N_1S_1E_1W_1$; е) N_3W_1; ж) $(N_2 \cup S_2)E_2$?

13. Проверить, что в предыдущей задаче

а) $S_3 \subset S_2$;
б) $S_3W_2 = 0$;
в) $N_2S_1E_1W_1 = 0$
г) $N_2S_2 \subset \overline{W}_1$;
д) $(N_2 \cup S_2)W_3 = 0$;
е) $W_4 = \overline{N}_1\overline{S}_1\overline{E}_1$.

14. Проверить следующие соотношения [1]):

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

15. Найти простые выражения для событий а) $(A \cup B)(A \cup \bar{B})$; б) $(A \cup B)(\bar{A} \cup B)(A \cup \bar{B})$; в) $(A \cup B)(B \cup C)$.

16. Установить, какие из следующих соотношений правильны:

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

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

а) произошло только A;
б) произошли A и B, но C не произошло;

---

[1]) Заметим, что $(\overline{A \cup B})$ означает событие, противоположное событию $A \cup B$; это событие не совпадает с $\bar{A} \cup \bar{B}$. Аналогично $(\overline{AB})$ не то же самое, что $\bar{A}\bar{B}$.

в) все три события произошли;
г) произошло по крайней мере одно из этих событий;
д) произошло по крайней мере два события;
е) произошло одно и только одно событие;
ж) произошло два и только два события;
з) ни одно событие не произошло;
и) произошло не больше двух событий.

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

19. Используя результат задачи 18, проверить, что

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

[Эта формула является специальным частным случаем формулы (15) гл. IV.]

ГЛАВА II

ЭЛЕМЕНТЫ КОМБИНАТОРНОГО АНАЛИЗА

Цель этой главы — получить ряд важных формул и развить соответствующие вероятностные основания. Подготовленный читатель может перейти прямо к гл. V, являющейся непосредственным теоретическим продолжением гл. I.

При рассмотрении простых азартных игр, процесса случайного выбора, задач о размещении и порядке и т. п. мы имеем обычно дело с конечными пространствами элементарных событий, в которых всем точкам отнесена одна и та же вероятность. Чтобы вычислить вероятность произвольного события A, достаточно поделить число точек, содержащихся в A (число «благоприятствующих случаев»), на общее число точек (число «возможных случаев»). Это облегчается систематическим применением нескольких правил, к выводу которых мы теперь приступаем. В дальнейшем мы будем пользоваться этими правилами, вместо того чтобы искать специальные вычислительные методы отдельно для каждой конкретной задачи [1]).

§ 1. Предварительные сведения

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

Доказательство. Составим из этих пар прямоугольную таблицу, содержащую m строк и n столбцов, так, чтобы пара $(a_j\ b_k)$ стояла на пересечении j-й строки и k-го столбца. Каждая из пар встретится в этой таблице один и только один раз, и утверждение становится очевидным.

Примеры. а) *Карты для игры в бридж* (см. примечание на стр. 18 § 1 гл. I). За множества элементов примем соответственно

---

[1]) Читателю, желающему более обстоятельно ознакомиться с комбинаторикой, можно рекомендовать классический курс W. A. Whitworth, Choice and chance, fifth edition, London, 1901. Переиздан G. E. Stechert, New York, 1942. Близкий по содержанию учебник того же автора, DCC exercises, переизданный в Нью-Йорке (New York, 1945), содержит около 700 задач с полными решениями. [См. также М. Холл, Комбинаторный анализ, ИЛ, М., 1963. — *Прим. ред.*]

четыре масти и 13 возможных значений карты. Каждая карта определяется мастью и значением. Существует $4 \cdot 13 = 52$ таких комбинаций масти и значения. Это — число карт в колоде.

б) Многоярусные электрические лампы с несколькими способами включения содержат три обычные лампочки и светящую арматуру, которая может работать в трех положениях или может быть выключена. Каждая из этих возможностей комбинируется с 0, 1, 2, 3 включенными лампочками: Следовательно, всего имеется $4 \cdot 4 = 16$ возможных комбинаций, из которых одна (0, 0) означает, что лампа не светит. Остается 15 способов включения лампы.

Комбинации элементов по одному из каждой группы. *Дано r групп элементов: n_1 элементов первой группы $a_1, \ldots, a_{n_1}$; n_2 элементов второй группы $b_1, \ldots, b_{n_2}$; ...; n_r элементов r-й группы $x_1, \ldots, x_{n_r}$; можно образовать ровно $n_1 \cdot n_2 \cdot \ldots \cdot n_r$ различных комбинаций $\left(a_{j_1}, b_{j_2}, \ldots, x_{j_r}\right)$, содержащих по одному элементу из каждой группы.*

Доказательство. Если $r = 2$, то утверждение сводится к ранее доказанному. Если $r = 3$, то возьмем пары (a_i, b_j) за элементы новой группы. Мы имеем $n_1 n_2$ таких пар и n_3 элементов c_k. Каждую тройку элементов (a_i, b_j, c_k) можно рассматривать как пару, составленную из (a_i, b_j) и элемента c_k; следовательно, число таких троек равно $n_1 n_2 n_3$. Проведением математической индукции утверждение доказывается для любого r.

Может быть, всего проще и полезней описать последнюю теорему следующим образом. Для того чтобы образовать последовательность r элементов $\left(a_{j_1}, b_{j_2}, \ldots, x_{j_r}\right)$, мы должны выбрать одно a, одно b и т. д. Надо произвести всего r выборов и при каждом из них имеется последовательно $n_1, n_2, \ldots, n_r$ возможностей. Утверждается, что этот процесс может привести к $n_1 n_2 \ldots n_r$ различным результатам.

Примеры. в) *Сложная классификация.* Предположим, что люди классифицируются по полу, семейному положению (состоит или не состоит в браке) и профессии. Различные категории играют здесь роль элементов. Если рассматривать 17 возможных профессий, то всего мы получим $2 \cdot 2 \cdot 17 = 68$ различных классов.

г) В агрономическом опыте испытываются три различных фактора (например, применение удобрений, орошение, температура). Пусть каждый из этих факторов характеризуется одним из r_1, r_2, r_3 возможных уровней (или концентраций) соответственно; тогда всего будет $r_1 r_2 r_3$ комбинаций или различных способов обработки.

д) «*Размещение шаров по ящикам*» состоит в выборе ящика для каждого шара. Если имеется r шаров, то для каждого ящика возможно r независимых исходов эксперимента, поэтому r шаров можно разместить по n ящикам n^r различными способами. Пример (2, б) гл. I показывает, что самые разнообразные мыслимые эксперименты с абстрактной точки зрения эквивалентны размещению шаров по ящикам. Например, если грани игральной кости рассматривать как «ящики», предыдущее утверждение позволяет заключить, что при r-кратном бросании кости возможно 6^r различных исходов, из которых 5^r удовлетворяют условию «ни разу не выпало очко». Значит, в предположении, что все исходы опыта равновероятны, событие «при r бросаниях кости ни разу не выпало очко» имеет вероятность $\left(\frac{5}{6}\right)^r$. Интуитивно можно рассчитывать, что при шести бросаниях «очко выпадает наверняка», но вероятность этого события равна только $1-\left(\frac{5}{6}\right)^6$, или менее $\frac{2}{3}$. [См. пример 3, б.]

§ 2. Выборки

Рассмотрим множество или, как мы будем говорить, *генеральную совокупность* из n элементов $a_1, a_2, \ldots, a_n$. Произвольное упорядоченное множество $(a_{j_1}, \ldots, a_{j_r})$ из r элементов, входящих в генеральную совокупность, будем называть *выборкой объема* r из этой генеральной совокупности. Наглядно мы можем просто представлять себе, что элементы выбираются один за другим. Возможны два способа выбора. Первый — *выбор с возвращением*; в этом случае выбор производится каждый раз из всей генеральной совокупности, и один и тот же элемент может быть выбран более одного раза. В выборке могут встречаться повторения. Второй — *выбор без возвращения*. При таком способе выбора однажды выбранный элемент удаляется из генеральной совокупности, так что выборка не содержит повторяющихся элементов. Очевидно, что в этом случае объем выборки r не может превысить объема генеральной совокупности n.

При выборе с возвращением каждый из r членов выборки может быть выбран n различными способами. Поэтому, применив предшествующую теорему при $n_1 = n_2 = \ldots = n$, получаем, что число различных возможных выборок равно n^r. В случае выборки без возвращения первый член выборки может быть выбран n способами, второй $n-1$, третий $n-2$ и т. д. Воспользовавшись той же теоремой, получим, что общее число выборок равно $n(n-1)\ldots(n-r+1)$. Так как подобное произведение встречается довольно часто, то оказывается удобным ввести следующее обозначение[1]):

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

---

[1]) Обозначение $(n)_r$ не является общепринятым, но мы будем постоянно использовать его в этой книге, даже если n не целое.

Ясно, что $(n)_r = 0$ при $r > n$. Нами получена, таким образом, следующая

Теорема. *Число различных выборок объема r из генеральной совокупности, содержащей n элементов, равно n^r, если производится выбор с возвращением, и $(n)_r$, если производится выбор без возвращения.*

Особо отметим частный случай, когда $r = n$. Если объем выборки n совпадает с объемом генеральной совокупности, то множество выборок без возвращения совпадает с множеством способов упорядочения (перестановок) этих элементов. Поэтому n элементов $a_1, a_2, \ldots, a_n$ можно упорядочить $(n)_n = n \cdot (n-1) \ldots 2 \cdot 1$ различными способами. Вместо $(n)_n$ мы будем пользоваться обычным обозначением $n!$. Итак, мы вывели из теоремы следующее

Следствие. *Число различных перестановок из n элементов равно*

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

Муж и жена Смит образуют выборку объема два из генеральной совокупности всех людей; но в то же время они образуют выборку объема единица из генеральной совокупности супружеских пар. Этот пример показывает, что объем выборки определен только по отношению к данной генеральной совокупности. Бросание монеты r раз является одним из способов получения выборки объема r из генеральной совокупности двух букв Γ и P (Γ — выпадение герба, P — выпадение решетки). Одновременно та же упорядоченная совокупность r букв Γ и P является элементарным событием в пространстве, соответствующем бросанию монеты r раз.

Выбор r элементов из генеральной совокупности объема n является экспериментом, возможными результатами которого являются выборки объема r. Общее число их равно n^r или $(n)_r$, в зависимости от того, как производится выбор: с возвращением или без возвращения. В обоих случаях наш опыт описывается пространством элементарных событий, точками которого являются всевозможные выборки объема r.

До сих пор мы интересовались вероятностями, соответствующими нашим выборкам. В дальнейшем мы будем обычно приписывать равные вероятности всем выборкам и говорить при этом о случайных выборках. Термин «случайный», вообще говоря, не имеет строгого математического определения, но в тех случаях, когда он прилагается к выборкам или выбору, ему придается вполне определенный смысл, именно: *в выражении «случайная выборка фиксированного объема r» прилагательное «случайная» означает, что все выборки имеют одинаковую вероятность* (равную n^{-r} в случае выбора с возвращением и $1/(n)_r$ в случае выбора без возвращения, где через n обозначен объем генеральной совокупности, из которой производится выбор). Если n велико, а r сравнительно мало, то отношение $(n)_r/n^r$ близко к единице. Это говорит о том, что при большой генеральной совокупности и относительно малой выборке различие между двумя способами выбора незначительно (см. задачи 1 и 2 § 11 гл. II и задачу 35 гл. VI).

Мы ввели терминологию, взятую из практики, но не говорили еще о применимости модели случайного выбора к действительным опытам. Бросание монеты или кости и другие аналогичные действия могут быть истолкованы как опыты, практически дающие случайный выбор с возвращением, и наши вероятности численно близки к частотам, наблюдаемым в длинной серии опытов, несмотря даже на то, что абсолютно симметричной монеты или кости не существует. Типичным примером случайного выбора без возвращения может служить вытаскивание карт из перетасованной колоды (причем предполагается, что колода тасуется гораздо тщательнее, чем это делается обычно). При выборе из совокупностей, состоящих из людей, статистики встречаются со значительными и часто неожиданными трудностями, и опыт показывает, что очень трудно получить даже грубое приближение к случайной выборке.

Упражнение. При выборе без возвращения вероятность того, что выборка объема r содержит заданный элемент генеральной совокупности, равна

$$1-\frac{(n-1)_r}{(n)_r}=1-\frac{n-r}{n}=\frac{r}{n}.$$

При выборе с возвращением соответствующая вероятность равна

$$1-\left\{\frac{n-1}{n}\right\}^r.$$

§ 3. Примеры

Из генеральной совокупности n элементов $a_1, a_2, \ldots, a_n$ извлекается с возвращением выборка объема r. Рассмотрим событие A, состоящее в том, что все элементы выборки различны, иными словами, что наша выборка могла бы быть получена при выборе без возвращения. Последняя теорема показывает, что различных выборок с возвращением существует n^r, из них $(n)_r$ обладают указанным свойством. Приписывая всем точкам пространства элементарных событий равные вероятности, получаем, что вероятность отсутствия одинаковых элементов в выборке (вероятность события A) равна

$$p=\frac{(n)_r}{n^r}=\frac{n(n-1)\ldots(n-r+1)}{n^r}. \qquad (3.1)$$

Интересные свойства этой формулы вскрываются следующими конкретными ее интерпретациями.

а) **Случайные числа.** Допустим, что генеральная совокупность состоит из десяти цифр: 0, 1, 2, ..., 9. Любая последовательность из пяти цифр представляет выборку объема $r=5$. Свяжем с каждой такой последовательностью вероятность 10^{-5}. Согласно (3.1), вероятность того, что пять последовательных случайных цифр различны, равна $p=(10)_5\,10^{-5}=0{,}3024$.

Интуиция подсказывает нам, что в математических таблицах, содержащих величины, вычисленные с большим числом десятичных знаков, последние пять знаков должны обладать многими свойствами случайных чисел. (В обычных логарифмических и других таблицах с малым числом знаков разность между соседними величинами близка к постоянной, и поэтому последние знаки изменяются закономерно). Для опыта были выбраны 16-значные таблицы [1]) и было сосчитано, для скольких чисел пять последних знаков различны. В первых 12 сериях по 100 чисел частоты наступления этого события оказались следующими:

$$\begin{array}{cccccc} 0{,}30 & 0{,}27 & 0{,}30 & 0{,}34 & 0{,}26 & 0{,}32 \\ 0{,}37 & 0{,}36 & 0{,}26 & 0{,}31 & 0{,}36 & 0{,}32 \end{array}$$

Эти частоты колеблются около 0,3024, и методы математической статистики показывают, что амплитуда этих колебаний не превышает ожидаемой. Средняя частота оказывается равной 0,3142, т. е. довольно близкой к вероятности 0,3024 (см. гл. VII, 3, в).

Рассмотрим теперь число $e=2{,}71828\ldots$ Из первых 800 десятичных знаков [2]) этого числа образуем 160 групп по пяти знаков в каждой; эти группы объединим в 16 серий по 10 групп в каждой серии. В этих сериях числа групп, в которых все пять знаков различны, окажутся следующими:

$$\begin{array}{cccccccc} 3, & 1, & 3, & 4, & 4, & 1, & 4, & 4, \\ 4, & 2, & 3, & 1, & 5, & 4, & 6, & 3. \end{array}$$

Как видно, частоты колеблются вокруг 0,3024, и теория подтверждает, что размах этих колебаний не больше, чем нужно было ожидать. Средняя частота появления группы в пять различных цифр в 160 группах равна $52/160=0{,}325$, т. е. близка к $p=0{,}3024$.

б) *Если n шаров случайно размещаются по n ящикам, то вероятность того, что каждый ящик будет занят, равна* $\dfrac{n!}{n^n}$.

Эта вероятность неожиданно мала: при $n=7$ она равна лишь $0{,}00612\ldots$ Отсюда следует, например, что если в городе каждую

---

[1]) Tables of Probability Functions, v. I. Nat. Bureau of Standards, 1941.
[2]) *Intermédiaire des recherches mathématiques*, 2 (1946), 112.

неделю происходят семь автомобильных катастроф, то (в предположении, что все возможные распределения равновероятны) практически каждая неделя будет содержать дни, когда происходят две или более катастрофы. Равномерное распределение числа катастроф встречается в среднем лишь для одной недели из 165. Этот пример иллюстрирует неожиданное свойство чистой случайности. (Все возможные исходы эксперимента, состоящего в размещении 7 шаров по 7 ящикам, приведены в табл. 1 § 5. С вероятностью 0,87 две или более ячейки окажутся пустыми.) При $n=6$ вероятность $n!\,n^{-n}$ равна 0,01543... . Это показывает, как мала вероятность того, что при шести бросаниях правильной кости появятся все грани. [Вероятность того, что при шести бросаниях ни разу не выпадет данная грань, близка к $^1/_3$; см. пример 1, д.]

в) **Лифт.** Лифт отправляется с $r=7$ пассажирами и останавливается на $n=10$ этажах. Чему равна вероятность p того, что никакие два пассажира не выйдут на одном и том же этаже? Для того чтобы точно поставить этот вопрос, мы предположим, что все возможные способы распределения пассажиров по интересующим их этажам имеют одинаковые вероятности (что является, конечно, только грубым приближением). Всего существует 10^7 таких способов, и из них при $(10)_7$ способах на любом этаже выходит не более чем один пассажир. Следовательно, $p=10^{-7}\cdot(10)_7=(10\cdot9\cdot8\cdot7\cdot6\cdot5\cdot4)\cdot10^{-7}=0{,}06048$. Если событие произошло, то его осуществление можно считать чем-то удивительным и ставить 1000 к 1 против повторения (ср. с ответом к задаче 10.43).

г) **Дни рождения.** Дни рождения r людей образуют выборку объема r из генеральной совокупности всех дней в году. Не все годы имеют одинаковую длину, и известно, что рождаемость в течение года не вполне постоянна. Однако в первом приближении можно считать, что случайный выбор людей приводит к случайному выбору дней рождения. Кроме того, мы игнорируем для простоты существование високосных годов и рассмотрим случайную выборку r дней рождения из года в 365 дней.

Если принять эти соглашения, то применимо уравнение (3.1), и вероятность p того, что все r дней рождения различны, равна[1])

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

Численные результаты, к которым приводит эта формула, опять идут вразрез с обычными представлениями. Так, для $r=23$ человек

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

$p < 1/2$, т. е. с вероятностью, превосходящей $1/2$, среди 23 человек найдутся по крайней мере двое, имеющие общий день рождения.

Формула (3.2) выглядит сложной, но для p нетрудно получить хорошую числовую оценку. Если r мало, можно пренебречь произведениями, содержащими более чем по одному сомножителю, и написать с довольно грубым приближением [1])

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

При $r = 10$ точное значение p равно 0,883, в то время как формула (3.3) дает приближенное значение 0,877.

При больших r можно получить значительно лучшее приближение, используя логарифмы. Если x мало, то $\log(1-x) \approx -x$, поэтому из (3.2) вытекает

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

При $r = 30$ эта формула дает для p значение 0,3037, в то время как точное значение равно $p = 0{,}294$. При $r \leqslant 40$ абсолютная ошибка формулы (3.4) меньше, чем 0,08. [Продолжение см. в § 7. См. также ответ к задаче (10.44).]

§ 4. Соединения

Как и раньше, мы будем пользоваться термином совокупность объема n для обозначения множества из n элементов безотносительно к их порядку. Две совокупности считаются различными, если в одной из них содержится элемент, не содержащийся в другой. Выбирая r элементов из заданной совокупности объема n, мы образуем новую совокупность (или выборку) объема r. Сколькими способами можно это сделать? Элементы каждой из таких выборок могут быть упорядочены $r!$ способами, что приводит к $r!$ различным выборкам без повторения. Обратно, каждая из выборок объема r без повторения содержит r различных элементов и поэтому определяет новую совокупность. Мы знаем, что существует $(n)_r$. выборок указанного типа Если x — число совокупностей объема r, которые можно образовать из n элементов, то, очевидно, число упорядоченных выборок объема r равно $xr!$ Отсюда следует, что $x = \frac{(n)_r}{r!}$. Эти числа называются биномиальными коэффициентами. Их принято обозначать

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

Итак, доказана

[1]) Знак $\approx$ означает, что равенство только приближенное.

Теорема 1. *Из совокупности n элементов можно образовать $\binom{n}{r}$ различных выборок объема $r \leqslant n$.*

Другими словами, из n элементов можно выделить группу в r элементов $\binom{n}{r}$ различными способами. Выбор r элементов из заданной совокупности объема n эквивалентен выбору $n-r$ оставшихся элементов. Поэтому ясно, что при любом $r \leqslant n$

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

Для того чтобы получить прямое доказательство формулы (4.2), воспользуемся новой формой записи биномиальных коэффициентов (4.1):

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

[К этому выражению нетрудно прийти, умножая числитель и знаменатель в (4.1) на $(n-r)!$.] При $0 < r < n$ формула (4.2) является тривиальным следствием формулы (4.3). Заметим, что $\binom{n}{0}$ не определено, в то время как $\binom{n}{n} = 1$. Для того чтобы уравнение (4.2) имело место при любом целом $0 \leqslant r \leqslant n$, следует положить

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

и

$$(n)_0 = 1.$$

Примеры. а) *Бридж и покер* (см. примечание на стр. 18). Так как порядок карт на руках несуществен, то по последней теореме первый игрок может получить карты для игры в бридж $\binom{52}{13} = 635\,013\,559\,600$ различными способами, а для игры в покер — $\binom{52}{5} = 2\,598\,960$ различными способами. Вычислим вероятность x того, что при игре в покер первый игрок получит карты пяти различных значений. Эти значения можно выбрать $\binom{13}{5}$ способами. Кроме того, мы свободны в выборе любой из четырех мастей для каждой карты. Поэтому $x = 4^5 \cdot \binom{13}{5} : \binom{52}{5}$, или приблизительно 0,5071. Соответствующая вероятность в бридже равна $4^{13} : \binom{52}{13}$, или приблизительно 0,0001057.

б) в США каждый из 48 штатов представлен двумя сенаторами. Мы исследуем события, состоящие в том, что в некотором комитете из 48 сенаторов, выбранных случайно, (1) представлен данный штат, (2) представлены все штаты.

В первом случае удобнее вычислить вероятность q дополнительного события, а именно что данный штат не представлен в комитете. Имеется 96 сенаторов, из них 94 — не из данного штата. Значит,

$$q = \binom{94}{48} : \binom{96}{48} = \frac{48 \cdot 47}{96 \cdot 95} = 0{,}24737 \ldots .$$

Далее теорема § 2 показывает, что комитет, включающий по одному сенатору от каждого штата, может быть составлен 2^{48} различными способами. Следовательно, вероятность того, что в комитете будут представлены все штаты, равна $p = 2^{48} : \binom{96}{48}$. Используя формулу Стирлинга (см. § 9), можно показать, что $p \approx (3\pi)^{1/2} 2^{-46} \approx$ $\approx 4 \ldots 10^{-14}$.

в) *Задача о размещении.* Рассмотрим еще раз случайное распределение r шаров по n ящикам (это означает, другими словами, что каждое из n^r возможных размещений имеет вероятность n^{-r}). Чтобы найти вероятность p_k того, что заданный ящик содержит ровно k шаров ($k = 0, 1, 2, \ldots, r$), заметим, что k шаров среди r можно выбрать $\binom{r}{k}$ различными способами. Оставшиеся $(r - k)$ шаров можно разместить по $(n - 1)$ ящикам $(n - 1)^{r-k}$ способами. Отсюда следует, что

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

Это частный случай так называемого биномиального распределения, которое будет подробно исследовано в гл. VI. Численное значение можно найти в табл. 3 гл. IV.

г) *Последовательности с двумя типами элементов.* Рассмотрим набор из $n = a + b$ элементов, из которых a элементов одного типа и b элементов — другого. Для удобства обозначим эти элементы через $\alpha_1, \alpha_2, \ldots, \alpha_a, \beta_1, \ldots, \beta_b$. Полный набор из n элементов можно упорядочить $n!$ различными способами. Однако если мы условимся считать элементы каждого типа неразличимыми (т. е. пренебрегать индексами), то некоторые последовательности станут неразличимыми. Нетрудно заметить, что теперь последовательность полностью определяется выбором a мест, на которых располагаются элементы типа α, и этот выбор может быть сделан $\binom{a+b}{a} = \binom{a+b}{b}$ различными способами. Таким образом, из набора a неразличимых элементов типа α и b неразличимых элементов типа β можно составить $\binom{a+b}{a} = \binom{a+b}{b}$ различных последовательностей. (Например, набор

$\alpha\alpha\alpha\beta\beta$ может быть упорядочен 10 различными способами.) Любая перестановка элементов типа α (или β) приводит к той же последовательности, так что $a!\,b!$ последовательностей неразличимы. Отсюда вытекает, что если мы припишем каждой из $(a+b)!$ перестановок одну и ту же вероятность $1/(a+b)!$, то все различимые расположения равновероятны, а именно каждое из них имеет вероятность $\frac{a!\,b!}{(a+b)!}$. Итак, если мы говорим о равновероятных расположениях, то этот термин применим как к различным расположениям, так и к множеству всех перестановок. (Рассмотренная ситуация заметно отличается от случая размещения шаров по ящикам; см. § 5.)

Теорема 2. *Пусть $r_1, r_2, \ldots, r_k$ — целые неотрицательные числа, причем*

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

Тогда число способов, посредством которых n элементов могут быть разделены на k групп, из которых первая содержит r_1 элементов, вторая r_2 элементов и т. д., равно

$$\frac{n!}{r_1!\,r_2!\ldots r_n!} \tag{4.7}$$

[Числа (4.7) называются полиномиальными коэффициентами.]

[Заметим, что порядок групп существен в том смысле, что разбиения $(r_1=2,\ r_2=3)$ и $(r_1=3,\ r_2=2)$ считаются различными; однако порядок элементов внутри группы не учитывается. Заметим также, что формула $0!=1$ позволяет избежать осложнений, связанных с тем, что некоторые r_i равны нулю.]

Доказательство. Применяя несколько раз формулу (4.3), нетрудно привести выражение (4.7) к виду

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

С другой стороны, разделение можно производить следующим образом: сначала выберем первую группу, состоящую из r_1 элементов; среди оставшихся $n-r_1$ элементов выберем вторую группу объема r_2 и т. д. После образования $(k-1)$-й группы остается $n-r_1-r_2-\ldots-r_{k-1}=r_k$ элементов, которые и образуют последнюю группу. Мы заключаем отсюда, что число возможных способов разделения на группы определяется формулой (4.8). Теорема доказана.

Примеры. д) При игре в бридж между четырьмя игроками распределяют $n=52$ карт по 13 карт каждому $(r_1=r_2=r_3=r_4=13)$. Следовательно, *число возможных положений на карточном столе равно*

$$\frac{52!}{(13!)^4}=(5{,}3645\ldots)\cdot 10^{28}. \tag{4.9}$$

Найдем вероятность того, что каждый игрок получит по одному тузу. Четыре туза могут быть распределены в этом случае между игроками $4!=24$ различными способами. Оставшиеся 48 карт могут быть распределены $(48!)(12!)^{-4}$ способами. Следовательно, искомая вероятность равна

$$24\cdot\frac{48!\,13^4}{52!}=0{,}105\ldots.$$

е) Одновременное бросание 12 игральных костей может иметь 6^{12} различных исходов. Всем им мы приписываем одинаковую вероятность. Событие, состоящее в том, что каждое из значений $1, 2, \ldots, 6$ выпадет дважды, может быть осуществлено числом способов, равным числу способов разделения 12 костей на 6 групп по 2 кости в каждой. Следовательно, вероятность этого события равна $12!/(2^6\cdot 6^{12})=0{,}003438\ldots$

(В теореме 2 допускалось, что r_i могут быть равны нулю, так что на самом деле n элементов разделяются на k или менее групп. Случай $r_i>0$ или разделение ровно на k групп рассматривается в задаче 11.7.)

§ 5*). Приложение к задачам о размещении

Примеры § 2 гл. I свидетельствуют о том, что модель случайного размещения r шаров по n ячейкам может быть применена к анализу самых разнообразных явлений. Перейдем теперь к рассмотрению этой модели, предполагая, конечно, что каждое из n^r возможных размещений имеет вероятность n^{-r}. Самой важной характеристикой отдельного размещения служит набор чисел $r_1, r_2, \ldots, r_n$, где r_i есть число шаров в i-й ячейке. В этом случае, очевидно,

$$r_1+r_2+\ldots+r_n=r,\qquad r_i\geqslant 0. \tag{5.1}$$

Условимся считать шары неразличимыми. Введенный только что набор чисел r_i теперь полностью определяет размещение, и два размещения различимы тогда и только тогда, когда соответствующие упорядоченные наборы чисел $(r_1, \ldots, r_n)$ не совпадают. В первую очередь будет доказана

Лемма. *Число различимых размещений шаров* [*т. е. число различных решений уравнения* (5.1)] *определяется формулой*[1])

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

*) Результаты этого параграфа полезны и поучительны, но в дальнейшем непосредственно использоваться не будут

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

Из них $\binom{r-1}{n-1}$ *обладают тем свойством, что ни один из ящиков не пуст.*

Доказательство. Будем представлять ящики как промежутки между $n+1$ черточками, а шары условимся обозначать звездочками. Так, символ $|***|*||||****|$ означает, что $r=8$ шаров размещены по $n=6$ ящикам, причем эти ящики содержат последовательно 3, 1, 0, 0, 0, 4 шаров. Такие символы всегда начинаются и кончаются черточками, но оставшиеся $n-1$ черточку и r звездочек можно разместить в произвольном порядке. Отсюда вытекает, что число различимых размещений равно числу способов выбора r мест среди $n+r-1$ мест. Условие отсутствия пустых ящиков приводит к следующему ограничению: каждая черточка должна быть заключена между двумя звездочками. Всего имеется $r-1$ промежутков между звездочками, и в любые $n-1$ из них можно поместить черточки. Это можно сделать $\binom{r-1}{n-1}$ способами. Лемма доказана.

Примеры. а) Число различимых исходов при одновременном бросании r одинаковых игральных костей равно $\binom{r+5}{5}$.

б) *Частные производные.* Частные производные порядка r от аналитической функции $f(x_1, x_2, \ldots, x_n)$ не зависят от порядка дифференцирования, а зависят лишь от того, сколько раз мы дифференцируем по каждому переменному. Каждое переменное играет роль ящика, и поэтому число различных частных производных порядка r равно $\binom{n+r-1}{r}$. Функция трех переменных имеет 15 производных четвертого порядка и 21 производную пятого порядка.

Размещение r шаров по n ящикам есть один из способов разбиения генеральной совокупности, состоящей из r шаров, на части. По теореме 2 § 4 число размещений, при которых ящики содержат последовательно $r_1, r_2, \ldots, r_n$ шаров, равно $\frac{r!}{r_1!r_2!\ldots r_k!}$. В этой формуле еще важен порядок чисел $r_1, r_2, \ldots, r_n$ или порядок ящиков, но часто этот порядок несуществен. Следующий пример приведен для того, чтобы проиллюстрировать простой и стандартный метод, применимый к решению целого ряда элементарных комбинаторных задач.

Пример. в) *Распределение* $r=7$ *шаров по* $n=7$ *ящикам* (под ящиками здесь можно понимать дни недели, а шары интерпретировать как телефонные звонки, письма, автомобильные катастрофы и т. д.). Для большей определенности рассмотрим частный случай, когда ящики содержат 2, 2, 1, 1, 1, 0, 0 шаров, причем порядок этих чисел роли не играет. Отправляясь от заданной семерки чисел можно разбить группу из 7 ящиков на три подгруппы следующим образом. Первая подгруппа состоит из двух ящиков, содержащих по два шара, вторая — из трех ящиков, содержащих по одному шару, и, наконец, последняя образована двумя пустыми ящиками. Такое разбиение на три подгруппы можно осуществить $\frac{7!}{2!\,3!\,2!}$ способами. Каждому такому разбиению соответствует $\frac{7!}{2!2!1!1!1!0!0!}=\frac{7}{2!\,2!}$ различных размещений 7 шаров по 7 ящикам. Поэтому полное число размещений (без учета порядка чисел 2, 2, 1, 1, 1, 0, 0) равно

$$\frac{7!}{2!\,3!\,2!}\cdot\frac{7!}{2!\,2!}. \tag{5.3}$$

Заметим, что этот результат был получен двойным применением формулы (4.7), а именно сначала к ящикам, затем к шарам. Тот же результат можно было бы получить и другими способами, но настоящий способ является стандартным и может быть с успехом применен к решению целого ряда разнообразных комбинаторных задач (см. задачи 43—45 § 10). В табл. 1 приведены результаты, аналогичные (5.3), и вероятности возможных размещений неразличимых шаров для случая $n=r=7$.

Таблица 1

Случайное размещение 7 шаров по 7 ящикам

Число шаров в ящиках	Число размещений равно $7!\times7!$, деленному на	Вероятность (число размещений, деленное на 7^7)
1, 1, 1, 1, 1, 1, 1	$7!\times1!$	0,006120
2, 1, 1, 1, 1, 1, 0	$5!\times2!$	0,128518
2, 2, 1, 1, 1, 0, 0	$2!\,3!\,2!\times2!\,2!$	0,321295
2, 2, 2, 1, 0, 0, 0	$3!\,3!\times2!\,2!\,2!$	0,107098
3, 1, 1, 1, 1, 0, 0	$4!\,2!\times3!$	0,107098
3, 2, 1, 1, 0, 0, 0	$2!\,3!\times3!\,2!$	0,214197
3, 2, 2, 0, 0, 0, 0	$2!\,4!\times3!\,2!\,2!$	0,026775
3, 3, 1, 0, 0, 0, 0	$2!\,4!\times3!\,3!$	0,017850
4, 1, 1, 1, 0, 0, 0	$3!\,3!\times4!$	0,035699
4, 2, 1, 0, 0, 0, 0	$4!\times4!\,2!$	0,026775
4, 3, 0, 0, 0, 0, 0	$5!\times4!\,3!$	0,001785
5, 1, 1, 0, 0, 0, 0	$2!\,4!\times5!$	0,005355
5, 2, 0, 0, 0, 0, 0	$5!\times5!\,2!$	0,001071
6, 1, 0, 0, 0, 0, 0	$5!\times6!$	0,000357
7, 0, 0, 0, 0, 0, 0	$6!\times7!$	0,000008

Замечание о статистиках Бозе — Эйнштейна и Ферми — Дирака. До сих пор мы предполагали, что каждое из n^r возможных размещений имеет вероятность n^{-r}. Интересно, что в ряде случаев экспериментальные данные заставляют физиков отказаться от этой гипотезы и задавать начальные вероятности по-другому.

Рассмотрим механическую систему, состоящую из r неразличимых частиц. В статистической механике обычно подразделяют фазовое пространство на большое число n маленьких областей или ячеек, так что каждая из частиц попадает в одну из ячеек. Таким образом, состояние всей системы описывается как распределение r частиц по n ячейкам. С первого взгляда кажется, что (по крайней мере при подходящем выборе ячеек) все n^r распределений будут иметь равные вероятности. Если это верно, то физики говорят о статистике Максвелла — Больцмана (термин «статистика» использован здесь в смысле, специфическом для физики). Попытки доказать, что физические частицы подчиняются статистике Максвелла — Больцмана, были весьма многочисленны, но современная теория с уверенностью утверждает, что *эта статистика неприменима ни к каким известным частицам*; ни в одном случае все n^r распределений не оказываются даже приблизительно равновероятными. Поэтому были созданы две другие вероятностные модели, каждая из которых удовлетворительно описывает поведение некоторого класса частиц. Оправдание этих моделей основано на успешности их применения. Ни одна из них не может претендовать на универсальность, и возможно, что когда-нибудь будет введена третья модель для других видов частиц.

Вспомним, что мы имеем дело с *неразличимыми* частицами. Всего у нас r частиц и n ячеек. *В статистике Бозе — Эйнштейна рассматриваются только различимые распределения, и каждому из них приписывается вероятность*

$$\binom{n+r-1}{r}^{-1}. \tag{5.4}$$

В статистической механике показывается, что это предположение верно для фотонов, атомных ядер и атомов, содержащих четное число элементарных частиц[1]). Для описания других частиц необходим третий способ введения вероятностей. Под *статистикой Ферми — Дирака* мы понимаем статистику, основанную на следующих гипотезах: 1) *в одной ячейке не могут находиться две или больше частиц* и 2) *все различимые размещения, удовлетворяющие первому условию, имеют равные вероятности.* Первая гипотеза требует, в частности, чтобы было $r \leqslant n$. Размещение полностью описывается указанием, в каких из n ячеек содержатся частицы, и так как частиц у нас r, то эти ячейки могут быть выбраны $\binom{n}{r}$ способами. Следовательно, *в случае статистики Ферми — Дирака всего имеется $\binom{n}{r}$ возможных размещений, каждое из которых имеет вероятность $\binom{n}{r}^{-1}$.* Эта модель применима к электронам, протонам и нейтронам. Мы имеем здесь поучительный пример невозможности выбора удовлетворительной вероятностной модели без помощи эксперимента. Действительно, никакое абстрактное рассуждение не может объяснить, почему фотоны и электроны не

---

[1]) Ср. H. Margenau, G. M. Murphy, The mathematics of physics and chemistry, New York, 1943, chapter 12.

подчиняются одному и тому же вероятностному закону. (Существенные различия между статистиками Максвелла — Больцмана и Бозе — Эйнштейна обсуждаются в задачах 14 — 19 § 11.)

Итак, *вероятность того, что ячейки с номерами* 1, 2, ..., n *содержат* $r_1, r_2, \ldots, r_n$ *частиц соответственно, равна*

$$\frac{r!}{r_1!\, r_2! \ldots r_n!} n^{-r} \tag{5.5}$$

в случае статистики Максвелла—Больцмана; она определяется формулой (5.4) *в случае статистики Бозе — Энштейна и равна* $\binom{n}{r}^{-1}$ *в случае статистики Ферми — Дирака, если только все* r_j *равны или нулю или единице* Заметим что термин «статистика Максвелла — Больцмана», принятый в физике, означает просто случайное размещение частиц по ячейкам.

Примеры. а) Пусть $n=5$, $r=3$. Распределение $(*|-|*|*|-)$ имеет вероятности $\frac{6}{125}$, $\frac{1}{35}$ и $\frac{1}{10}$ при использовании соответственно статистик Максвелла — Больцмана, Бозе — Эйнштейна и Ферми — Дирака. См. также пример (6.б) гл. I.

б) *Опечатки.* В книге содержится n символов (букв), из которых r неверны. Распределение опечаток соответствует размещению r шаров по n ящикам, причем ни один ящик не может содержать более одного шара. Естественно поэтому предположить, что для опечаток приближенно имеет место статистика Ферми — Дирака. (См. задачу 10.38.)

5а. Серии. Будем называть *серией* каждую максимальную подпоследовательность элементов одинакового типа в любой упорядоченной последовательности элементов двух типов. Например, последовательность *AAABAABBBA* открывается *A*-серией длины 3; за ней следуют серии длин 1, 2, 3, 1 соответственно. *A*-серии и *B*-серии всегда чередуются, так что общее чисел серий равно единице плюс число перемен типа элементов в данной последовательности.

Примеры применений. Теория серий применяется в статистике многими способами, но ее главные применения связаны с критериями случайности или с критериями однородности.

а) При проверке случайности задача состоит в том, чтобы решить, является ли данное наблюдение случайным или следует искать определенные причины. Рассмотрим простой пример. Пусть наблюдение [1]) дало следующее распределение пустых (*П*) и занятых (*З*) мест в ресторане: *ПЗППЗП ППЗПППЗПЗП*. Заметим, что ни одна пара смежных мест не оказалась занятой. Случайно ли это? Из пяти занятых и одиннадцати пустых мест можно образовать не более чем 11 серий; это число серий и было наблюдено. Позднее будет показано, что если все размещения равновероятны, то вероятность наличия 11 серий равна 0,0578 .... Малость этой вероятности подтверждает до некоторой степени подозрение, что такое распределение занятых мест не случайно. Это подозрение не может быть, однако, доказано статистическими методами, и путем дальнейшего наблюдения должны быть собраны добавочные сведения. Если ресторан часто посещается целыми семьями, то обнаружится тенденция занимать соседние места, а

---

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

это повлечет за собой относительную малость числа серий. Аналогично рассмотрение числа серий, образуемых мальчиками и девочками, сидящими в классной комнате, может обнаружить перемешивание, большее или меньшее случайного. Неправдоподобные расположения дают ключ к пониманию вызывающих их причин; *избыток серий указывает на сильное перемешивание, недостаток — на сильную группировку*. Конечно, такие заключения не могут быть совершенно безошибочными. Даже при полной случайности могут возникнуть неправдоподобные ситуации, и исследователь будет искать причины таких ситуаций. Однако это происходит редко, и на практике при подходящем выборе критерия для проверки случайности мы будем ошибаться только, например, один раз из 100, зато 99 раз из 100 будем находить соответствующие причины.

Примером статистических приложений теории серий может служить использование этой теории для контроля качества промышленной продукции. Изготовленные шайбы могут несколько меняться по своей толщине. Длинные серии толстых шайб указывают на возможные неполадки в ходе производственного процесса и необходимость устранения причин этих неполадок; таким образом, предупреждается возникновение аварий и достигается бóльшая однородность изготовляемой продукции.

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

б) На основании теории серий можно построить критерий сравнения. Допустим, что на двух группах пациентов применяются два лекарства, или нас интересует сравнение эффективности двух методов (в медицине, сельском хозяйстве или промышленности). На практике мы располагаем двумя наборами наблюдений, скажем, $\alpha_1, \alpha_2, \ldots, \alpha_a$ и $\beta_1, \beta_2, \ldots, \beta_b$, соответствующих этим методам или отражающих некоторую характеристику (такую, как вес) элементов двух групп. Допустим, что совокупности чисел α_i и β_i упорядочены по величине: $\alpha_1 \leqslant \alpha_2 \leqslant \ldots \leqslant \alpha_a$ и $\beta_1 \leqslant \beta_2 \leqslant \ldots \leqslant \beta_b$. Объединим две последовательности в одну, которую также упорядочим по величине. Крайний случай состоит в том, что все α_i предшествуют всем β_j, и это можно расценивать как указание на то, что соответствующие методы или группы элементов заметно различаются. С другой стороны, если два метода идентичны, то члены последовательностей α_i и β_j будут расположены более или менее в случайном порядке. Вальд и Вольфовиц [2]) показали, что теория серий может быть с успехом применена для обнаружения малых систематических ошибок. (См. пример § 1 гл. III, хотя он и основан на другом методе.)

Многие задачи, связанные с сериями, можно решить совершенно элементарными средствами. Пусть даны a неразличимых элементов α_i и b неразличимых элементов β_j, из примера (4.2) мы знаем, что из них можно образовать $\binom{a+b}{a}$ различимых последовательностей. Если имеется n_1, α-серий, то число β-серий может равняться $n_1 \pm 1$ или n_1. Разбиение a элементов α_i на группы, соответствующие n_1 сериям, эквивалентно размещению a шаров по n_1 ячейкам, при котором нет пустых ячеек. По последней лемме

---

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

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

это можно сделать $\binom{a-1}{n_1-1}$ различными способами. Отсюда следует, например, что имеется $\binom{a-1}{n_1-1}\binom{b-1}{n_1}$ последовательностей с n_1 α-сериями и n_1+1 β-сериями (продолжение в задачах 20—25 § 11).

в) В физике теория серий используется при изучении процессов соединения. В теории одномерных цепочек, которую предложил Айзинг, энергия цепочек зависит от числа перемен типа элементов в цепочке, т. е. от числа серий.

§ 6. Гипергеометрическое распределение

Многие задачи из комбинаторики могут быть сведены к следующей схеме. В генеральной совокупности, состоящей из n шаров, имеются n_1 красных и $n_2=n-n_1$ черных шаров. Из этой совокупности выбирается группа в r шаров (без возвращения и без учета порядка шаров внутри группы). Нужно найти вероятность q_k того, что группа, выбранная таким образом, содержит ровно k красных шаров. Здесь k является целым неотрицательным числом, меньшим n_1 и r.

В группе, содержащей k красных и $r-k$ черных шаров, красные шары могут быть выбраны $\binom{n_1}{k}$ различными способами, а черные — $\binom{n-n_1}{r-k}$ способами. Так как любой способ выбора k красных шаров может комбинироваться с любым способом выбора $r-k$ черных шаров, получаем

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

Система вероятностей q_k, определенных таким образом, называется *гипергеометрическим распределением*[1]). Используя формулу (4.3), можно переписать (6.1) в виде

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

Замечание. Вероятности q_k определены только при k, не превосходящем r и n_1. Однако из определения (4.1) вытекает, что

[1]) Это наименование объясняется тем, что производящая функция (см. гл. XI) для $\{q_k\}$ выражается через гипергеометрические функции.

$\binom{a}{b}=0$ при $b>a$. Следовательно, определения (6. 1) и (6. 2) дают $q_k=0$ при $k>n_1$ или $k>r$. Таким образом, определения (6. 1) и (6. 2) могут быть использованы для всех $k\geqslant 0$, если предположить при этом, что соотношение $q_k=0$ понимается как невозможность получить k красных шаров.

Примеры. а) При *выборочном контроле качества промышленной продукции* проверяется партия, состоящая из n изделий. Бракованные изделия, входящие в эту партию, играют роль «красных шаров». Число их n_1 неизвестно. Из этой партии берется случайная выборка объема r, и затем определяется число бракованных изделий, содержащихся в этой выборке. Формула (6. 1) позволяет сделать некоторые заключения относительно наиболее правдоподобного значения n_1; это типичная задача статистической оценки, выходящая, однако, за рамки этой книги[1]).

б) В примере 4, б генеральная совокупность состоит из $n=96$ сенаторов, из которых $n_1=2$ представляют заданный штат (играют роль «красных шаров»). Случайным образом образуется группа из $r=48$ сенаторов. В эту группу могут войти 0, 1 или 2 сенатора из заданного штата. Соответствующие вероятности находим по формуле (6. 2), используя (4. 4):

$$q_0=q_2=\frac{48\cdot 47}{96\cdot 95}=0{,}24737\ldots,\quad q_1=\frac{48}{95}=0{,}50527.$$

Значение q_0 другим способом было уже получено в примере 4, б.

в) *Оценка генеральной совокупности по выборке*[2]). Из озера вылавливается 1000 рыб. Каждая из пойманных рыб метится красным пятнышком, а затем выпускается обратно в озеро. При следующем улове среди 1000 выловленных рыб оказались мечеными 100 рыб. Какие выводы можно сделать отсюда относительно числа рыб, живущих в озере? Это типичная задача *статистической оценки*. Мы зашли бы слишком далеко, если бы стали изучать различные

---

[1]) Различные методы статистического контроля качества промышленной продукции рассматриваются в работе А. Н. Колмогорова «Несмещенные оценки» (*Изв. АН СССР*, серия математич., 14 (1950), 303—326) и «Статистический приемочный контроль при допустимом числе дефектных изделий, равном нулю» (Ленинград, 1951, изд. Всесоюзного общества по распространению политических и научных знаний). — *Прим. ред.*

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

методы, используемые современной статистикой, но сейчас мы покажем, как гипергеометрическое распределение дает ключ к решению этой задачи. Предположим, конечно, что результат двух уловов можно рассматривать как случайную выборку из генеральной совокупности всех рыб в озере (на практике это означает, что уловы не должны производиться в одном и том же месте или через короткое время один после другого). Предположим также, что число рыб в озере не изменилось за время между двумя уловами.

Несколько обобщим задачу, рассмотрев произвольный объем выборки. Пусть

n — (неизвестное) число рыб в озере;

n_1 — число рыб, пойманных при первом улове (оно играет роль числа красных шаров);

r — число рыб, пойманных при втором улове;

k — число меченых рыб, пойманных при втором улове;

$q_k(n)$ — вероятность того, что второй улов содержит ровно k меченых рыб.

При такой формулировке сразу ясно, что $q_k(n)$ находится по формуле (6.1). На практике наблюдаются n_1, r и k, а n неизвестно. Заметим, однако, что n — фиксированное число, не зависящее от случая, и поэтому бессмысленно говорить о вероятности того, что, скажем, n больше 6000. Мы знаем, что было поймано n_1+r-k различных рыб, и, следовательно, $n \geqslant n_1+r-k$. Это все, что известно с полной достоверностью. В нашем примере $n_1=r=1000$ и $k=100$, и, вообще говоря, не исключено, что в озере обитает ровно 1900 рыб. Однако, отправляясь от этой гипотезы, мы придем к выводу, что случилось событие фантастически-малой вероятности. Действительно, вероятность того, что выборка объема 1000 из генеральной совокупности объема 1900 будет содержать 100 красных шаров, если общее число черных шаров 1000, равна

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

Формула Стирлинга (§ 9) показывает, что это число порядка 10^{-430}, и, следовательно, мы должны отбросить нашу гипотезу, как абсурдную. Аналогичное рассуждение заставляет нас откинуть гипотезу о том, что n очень велико, скажем равно миллиону. Все это приводит нас к отысканию такого n, при котором $q_k(n)$ является наибольшим, так как при этом значении n наблюденный результат имеет наибольшую вероятность. Для каждого частного набора наблюдений n_1, r и k значение n, при котором $q_k(n)$ максимально, обозначается через $\hat{n}$ и называется ***оценкой максимального правдоподобия***. Это понятие было введено Р. А. Фишером. Для нахождения $\hat{n}$ рассмотрим отношение

$$\frac{q_k(n)}{q_k(n-1)} = \frac{(n-n_1)(n-r)}{(n-n_1-r+k)\,n}. \tag{6.3}$$

Несложное вычисление показывает, что это отношение больше единицы при $nk < n_1 r$ и меньше единицы при $nk > n_1 r$. Это значит, что при возрастании n числа $q_k(n)$ сначала возрастают, а затем убывают, и максимум достигается, когда n есть максимальное целое число, не большее чем $n_1 r/k$, так что $\hat{n} \approx \frac{n_1 r}{k}$.

В нашем примере оценкой максимального правдоподобия для числа рыб является число $\hat{n} = 10\,000$.

Истинное значение n может быть, однако, больше или меньше $\hat{n}$, и естественно поставить задачу о построении по опытным данным интервала, относительно которого можно было бы с достаточной уверенностью заявить, что этот интервал содержит внутри себя истинное значение n. С этой целью рассмотрим гипотезу, что n меньше 8500. Подставим в формулу (6.1) $n = 8500$, $n_1 = r = 1000$ и подсчитаем вероятность того, что второй улов содержит 100 или меньше меченых рыб. Эта вероятность есть $x = q_0 + q_1 + \ldots + q_{100}$. Прямое вычисление x громоздко, но, используя нормальное распределение (гл. VII), легко убедиться, что $x \approx 0{,}04$. Аналогично если $n = 12\,000$, вероятность того, что второй улов содержит 100 или больше меченых рыб, равна примерно 0,03. Эти цифры оправдывают гипотезу, что истинное число рыб n заключено между 8500 и 12 000. Существуют и другие способы вывести это заключение и другие методы оценки, но мы не хотим вдаваться сейчас в детали.

Из определения вероятностей q_k вытекает, что $q_0 + q_1 + \ldots = 1$. Из формулы (6.2) следует поэтому, что при любых целых и положительных n, n_1 и r

$$\binom{r}{0}\binom{n-r}{n_1} + \binom{r}{1}\binom{n-r}{n_1-1} + \binom{r}{2}\binom{n-r}{n_1-2} + \ldots$$

$$\ldots + \binom{r}{n_1}\binom{n-r}{0} = \binom{n}{n_1}. \tag{6.4}$$

Это тождество оказывается полезным довольно часто. Мы доказали его только для положительных целых n и r, но оно верно и для любых положительных или отрицательных чисел n и r. (Тождество бессмысленно, если n_1 не является целым и положительным. Указания к двум способам вывода этого тождества даны в § 12, задачи 8 и 9.)

Гипергеометрическое распределение легко может быть обобщено и на тот случай, когда первоначальная генеральная совокупность объема n содержит более двух классов элементов. Пусть, например, генеральная совокупность состоит из трех классов с объемами n_1, n_2 и $n - n_1 - n_2$ соответственно. Вероятность того, что выборка объема r содержит k_1 элементов первого класса, k_2 элементов второго класса и $r - k_1 - k_2$ элементов третьего класса, находится аналогично (6.1) и оказывается равной

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

Здесь предполагается, конечно, что $k_1 \leqslant n_1$, $k_2 \leqslant n_2$ и $r - k_1 - k_2 \leqslant n - n_1 - n_2$.

Пример. г) *Бридж.* Генеральная совокупность из 52 карт делится на четыре класса по 13 карт в каждом. Вероятность того, что случайная выборка, состоящая из 13 карт, будет содержать пять пик, четыре червы, три бубны и одну трефу, равна

$$\frac{\binom{13}{5}\binom{13}{4}\binom{13}{3}\binom{13}{1}}{\binom{52}{13}}.$$

§ 7. Примеры, связанные с временем ожидания

Содержание этого параграфа стоит несколько в стороне от основной темы настоящей главы. В ней исследуются пространства элементарных событий нового типа, к которым можно прийти простой модификацией задачи о размещении. Рассмотрим еще раз мысленный эксперимент, состоящий в случайном размещении шаров по n ящикам. Теперь, однако, мы не фиксируем заранее число r шаров, но продолжаем размещать шары один за другим до тех пор, пока не получится определенная ситуация. Мы рассмотрим две такие возможные ситуации: (I) Шары случайно размещаются по ящикам до тех пор, пока какой-нибудь шар впервые не попадет в уже занятый ящик. После этого процесс прекращается. (II) Мы фиксируем ящик [скажем, ящик с номером (I)] и продолжаем процесс случайного размещения шаров до тех пор, пока этот ящик остается пустым. Процесс прекращается в тот момент, когда впервые какой-нибудь шар попадает в выбранный нами ящик.

Приведем несколько интерпретаций этих моделей, которые облегчат понимание задачи.

Примеры. а) *Дни рождения.* В примере (3.2) о днях рождения люди соответствуют шарам, а $n = 365$ дней года — ящикам. Модель (I) теперь сводится к следующему: мы случайно выбираем людей одного за другим до тех пор, пока не найдутся двое с общим днем рождения. Из скольких человек состоит выборка? Модели (II) соответствует продолжение процесса до тех пор, пока не найдется человек с тем же днем рождения, что и у меня.

б) *Задача о ключах*. Человек хочет открыть дверь своей комнаты. У него есть n ключей, из которых только один подходит к двери. По причинам, о которых можно только догадываться, он подбирает ключ случайно так, что после любой попытки каждый ключ с вероятностью n^{-1} может быть выбран для следующего опыта и все возможные исходы, отвечающие фиксированному числу опытов, равновероятны. Найти вероятность того, что человек добьется успеха на r-м шагу. Это специальный случай модели (II). Интересно сравнить эти случайные поиски ключа с более систематическим подходом. [См. задачу (10, 11), а также задачу 5, гл. V.]

в) В предыдущем примере мы можем заменить выбор ключей выбором из произвольной совокупности предметов, скажем, из коллекции купонов. Опять можно поставить вопросы о том, когда впервые будут выбраны два одинаковых предмета и когда впервые будет выбран заранее фиксированный предмет.

2) *Бросание монет и костей*. В примере 5, а гл. I монету подбрасывают до тех пор, пока не выпадет орел. Это специальный случай модели (II) при $n=2$. Если бросать правильную кость до выпадения в первый раз очка, то опять придем к модели (II) при $n=6$. (Времена ожидания для некоторых других экспериментов рассмотрены в задачах 21, 22 и 36 § 10 и 12 § 11.)

Мы начнем с модели (I), которая в некотором отношении более проста. Удобно ввести символы вида $(j_1, j_2, \ldots, j_r)$ для обозначения того, что первый, второй, ..., r-й шары помещены в ящики с номерами $j_1, j_2, \ldots, j_r$ соответственно и что процесс окончился на r-м шагу. Очевидно, что j_i — целое число, заключенное между 1 и n. При этом числа $j_1, j_2, \ldots, j_{r-1}$ попарно различны, а j_r равно одному из них. Каждый конечный набор, обладающий такими свойствами, представляет элементарное событие. Ясно, что r может принимать только значения $2, 3, \ldots, n+1$, так как после первого испытания заведомо не будет ящика, содержащего два шара, а после $n+1$ испытания такой ящик заведомо найдется. Связь этой задачи с уже обсуждавшейся моделью случайного размещения фиксированного числа шаров по n ящикам наводит на мысль связать с каждым элементарным событием $(j_1, j_2, \ldots, j_r)$ вероятность n^{-r}. Сейчас мы покажем, что это соглашение допустимо (т. е. что сумма вероятностей, соответствующих всем элементарным событиям, равна 1) и что оно приводит к разумным результатам.

При фиксированном r множество всех элементарных событий $(j_1, j_2, \ldots, j_r)$ образует событие, состоящее в том, что процесс окончится на r-м шагу. Согласно (3. 1), числа $j_1, j_2, \ldots, j_{r-1}$ можно выбрать $(n)_{r-1}$ различными способами, j_r может принимать любое из $r-1$ значений $j_1, j_2, \ldots, j_{r-1}$. Поэтому *вероятность того, что процесс окончится на r-м шаге, равна*

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

при $r>2$, $q_1=0$, $q_2=\frac{1}{n}$. *Вероятность того, что процесс будет продолжаться после r-го испытания, равна* $p_r=1-(q_1+q_2+\ldots+q_r)$, *и поэтому* $p_1=1$, *а при* $r>1$

$$p_r = \frac{(n)_r}{n^r} = \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{r-1}{n}\right), \tag{7.2}$$

что легко проверить по индукции. В частности, $p_{n+1}=0$ и $q_1+\ldots+q_{n+1}=1$, что и требовалось доказать. Кроме того, при $n=365$ формула (7.2) сводится к формуле (3.2), и вообще наша новая модель дает те же количественные результаты, что и модель с фиксированным числом шаров.

Модель (II) отличается от модели (I) тем, что ей соответствует пространство элементарных событий, состоящее из бесконечного (счетного) числа точек. Последовательности $(j_1, j_2, \ldots, j_r)$ подчинены тому условию, что числа $j_1, j_2, \ldots, j_{r-1}$ отличны от фиксированного числа $a \leqslant n$, но $j_r = a$. Не существует никаких априорных доводов в пользу того, что процесс должен рано или поздно прекратиться. При фиксированном r мы опять свяжем с каждым элементарным событием $(j_1, j_2, \ldots, j_r)$ вероятность n^{-r}. Любое из чисел $j_1, j_2, \ldots, j_{r-1}$ можно выбрать $n-1$ способами, а j_r уже определено. Отсюда легко получить для вероятности q_r^* того, что процесс закончится на r-м шаге, следующее выражение:

$$q_r^* = \left(\frac{n-1}{n}\right)^{r-1}\frac{1}{n}, \quad r=1, 2 \ldots . \tag{7.3}$$

Суммируя эту геометрическую прогрессию, находим $q_1+q_2+\ldots=1$. Так как вероятности q_r^* дают в сумме единицу, то нет необходимости вводить элементарные события, которые бы отвечали возможности того, что ни один из шаров не попадет в ящик a (т. е. процесс никогда не окончится). Для вероятности

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

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

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

Этот результат можно было предвидеть.

Медианы распределений $\{p_r\}$ и $\{p_r^*\}$ определяются как такие числа r, для которых p_r и p_r^* ближе всего к $\frac{1}{2}$; это означает, что процесс приблизительно с одной и той же вероятностью закончится до r-го испытания или после него. [Распределение, фигурирующее в примере (3. г) («дни рождения») имеет медиану $r=23$.] Для вычисления медианы $\{p_r\}$ перейдем к логарифмам, как это мы делали в (3. 4). Если r мало сравнительно с n, то легко показать, что — $-\log p_r$ близко к $\frac{r^2}{2n}$. Отсюда вытекает, что медиана распределения $\{p_r\}$ близка к $(n \cdot 2 \cdot \log 2)^{1/2}$, или приблизительно к $6/5\, n^{1/2}$. Интересно отметить, что медиана возрастает лишь как корень квадратный из объема совокупности ящиков. Наоборот, *медиана* $\{p_r^*\}$ *близка к* $n \cdot \log 2 = 0{,}7\, n$ и возрастает линейно с ростом n. Вероятность того, что в модели (II) процесс не закончится до n-го шага, равна $\left(1-\frac{1}{n}\right)^n$, или приблизительно $e^{-1}=0{,}36788\ldots$

§ 8. Биномиальные коэффициенты

Биномиальные коэффициенты $\binom{n}{r}$ были определены только для целых положительных n и r, но теперь удобно расширить это определение. Так как r обозначает число сомножителей, оно должно быть целым. Однако число

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

определено для всех действительных x при единственном предположении, что r — целое положительное число. При $r=0$ мы положим $(x)_0=1$. Тогда формула

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

определяет биномиальные коэффициенты при любых x и любых целых положительных r. При $r=0$ полагаем [так же, как в (4.4)] $\binom{x}{0}=1$ и $0!=1$. *Для отрицательных целых r положим*

$$\binom{x}{r}=0 \qquad (r<0). \tag{8.3}$$

Мы не будем пользоваться символом $\binom{x}{r}$, если r не является целым числом.

Нетрудно проверить, что при таком определении имеем, например,

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

В дальнейшем мы будем пользоваться тремя важными свойствами биномиальных коэффициентов. Первое: *для любого целого положительного* n

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

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

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

Эти соотношения легко вывести из определения биномиальных коэффициентов. Третье соотношение предполагается известным из элементарных учебников математического анализа: *при любом* a *и любом* t, *удовлетворяющем условию* $-1 < t < 1$, *имеет место формула*

$$(1+t)^a = 1 + \binom{a}{1} t + \binom{a}{2} t^2 + \binom{a}{3} t^3 + \dots \tag{8.7}$$

Если a — целое положительное число, то все члены правой части равенства (8.7), содержащие степени t выше t^a, автоматически исчезают и формула оказывается верной при всех t (бином Ньютона). Если a не есть целое положительное число, то правая часть равенства (8.7) представляет собой *бесконечный ряд*.

Используя соотношение (8.4), легко убедиться в том, что при $a = -1$ разложение (8.7) сводится к геометрической прогрессии

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + t^4 - + \dots \tag{8.8}$$

Это равенство верно при $-1 < t < 1$. Интегрируя соотношение (8.8), мы получим другую формулу, которая также будет использована в дальнейшем, а именно *ряд Тейлора для натурального логарифма*

$$\log(1+t) = t - \frac{1}{2}t^2 + \frac{1}{3}t^3 - \frac{1}{4}t^4 + \dots \tag{8.9}$$

Это разложение также верно при $-1 < t < 1$. Часто пользуются двумя другими формами записи соотношения (8.9). Заменяя t на $-t$, получим при $-1 < t < 1$

$$\log\frac{1}{1-t} = t + \frac{1}{2}t^2 + \frac{1}{3}t^3 + \frac{1}{4}t^4 + \dots \tag{8.10}$$

Складывая две последние формулы, приходим к разложению

$$\frac{1}{2}\log\frac{1+t}{1-t} = t + \frac{1}{3}t^3 + \frac{1}{5}t^5 + \dots, \tag{8.11}$$

которое имеет место при $-1<t<1$. Правая часть формулы (8.11) больше, чем t, но меньше, чем $t+t^2+t^3+\ldots=\frac{t}{1-t}$. Следовательно, верно следующее двойное неравенство:

$$e^{-\frac{t}{1-t}}<1-t<e^{-t}, \qquad 0<t<1. \tag{8.12}$$

Большое количество полезных соотношений и тождеств будет выведено из (8.7) в § 12. Здесь мы только отметим, что при любом целом положительном n, положив $t=1$, получим

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{n}=2^n. \tag{8.13}$$

Между прочим, эта формула допускает простую комбинаторную интерпретацию. Левая часть представляет число способов, которыми совокупность n элементов можно разделить на две группы, без учета порядка элементов в каждой группе, если объем первой группы может иметь любое значение $k=0, 1, 2, \ldots, n$. С другой стороны, такое деление можно осуществить непосредственно, указав для каждого элемента, принадлежит ли он к первой или ко второй группе. (Аналогичные рассуждения показывают, что сумма полиномиальных коэффициентов (4.7) равна k^n.)

§ 9. Формула Стирлинга

Одним из важных инструментов аналитической теории вероятностей является следующая классическая формула, известная как

Формула Стирлинга:

$$n!\sim(2\pi)^{1/2}n^{n+\frac{1}{2}}e^{-n}. \tag{9.1}$$

Здесь символ $\sim$ означает, что отношение двух выражений, соединенных этим символом, стремится к единице при неограниченном возрастании n.

Эта формула неоценима во многих теоретических рассуждениях и может быть также применена для получения приближенных числовых результатов. Хотя разность правой и левой части соотношения (9.1) неограниченно возрастает, действительный интерес представляет лишь относительная ошибка, которая быстро убывает. Формула Стирлинга дает очень точное приближение к $n!$ даже при малых n. Действительно, из формулы (9.1) для $1!$, $2!$, $5!$ получаем следующие приближенные значения 0,9221, 1,919 и 118,019. Относительные ошибки равны 8%, 4% и 2% соответственно. Для $10!=3628800$ приближенное значение равно 3598600 с ошибкой 0,8%. Для $100!$ относительная ошибка равна лишь 0,08%.

Доказательство формулы Стирлинга. Рассмотрим выражение

$$a_n = \log 2 + \log 3 + \ldots + \log(n-1) + \frac{1}{2}\log n, \tag{9.2}$$

которое отличается от $\log n!$ только множителем $^1/_2$ при последнем члене в правой части. Мы покажем, что a_n представляет площади

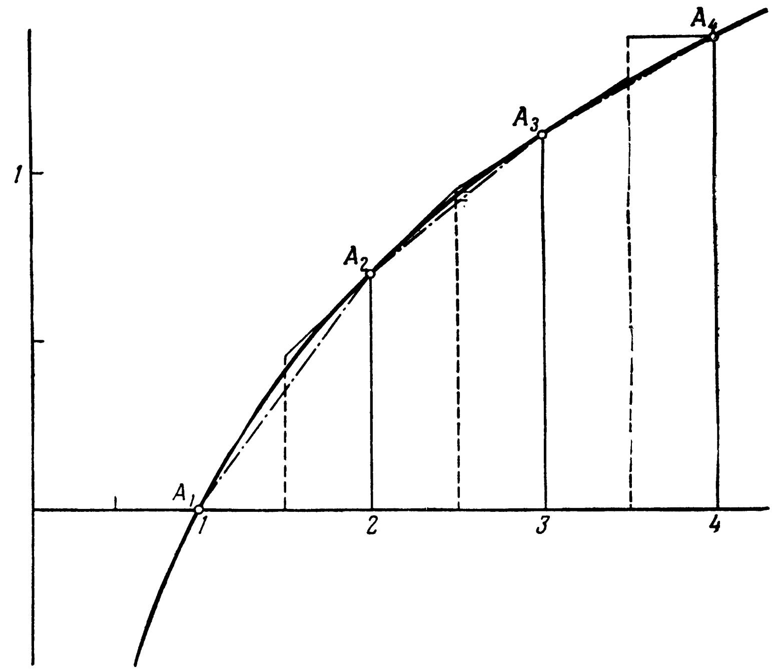


Рис. 1. Иллюстрация вывода формулы Стирлинга и приближение сумм интегралами ($y = \log x$).

двух различных многоугольников, откуда получим для $\log n!$ верхнюю и нижнюю оценки. Рис. 1 иллюстрирует положение для частного случая $n = 4$. Записав a_n в виде

$$a_n = {}^1/_2\,\{\log 1 + \log 2\} + \frac{1}{2}\{\log 2 + \log 3\} + \ldots +$$
$$+ {}^1/_2\,\{\log(n-1) + \log n\}, \tag{9.3}$$

замечаем, что a_n равно площади многоугольника, вершинами которого являются точки $A_1, A_2, \ldots, A_n$ кривой $y = \log x$ с абсциссами $x = 1, 2, \ldots, n$ и точка $(n, 0)$ оси x. Этот многоугольник лежит под кривой, поэтому его площадь меньше, чем площадь области, ограниченной кривой $y = \log x$, осью x и прямой $x = n$.

С другой стороны, $\log k$ равен площади трапеции с основанием $k - {}^1/_2 < x < k + {}^1/_2$, ограниченной сверху касательной к кривой в точке $A_k = (k, \log k)$. Отсюда следует, что $\log (n-1)!$ больше, чем площадь области, ограниченной кривой $y = \log x$, осью x и вертикальными прямыми $x = {}^3/_2$ и $x = n - {}^1/_2$. Кроме того, совершенно очевидно, что $\frac{1}{2} \log n$ превосходит площадь криволинейной трапеции с основанием $n - \frac{1}{2} < x < n$, и, значит, a_n превосходит площадь под кривой между вертикалями $x = \frac{3}{2}$ и $x = n$. Другими словами, мы показали, что

$$\int_{\frac{3}{2}}^{n} \log x \cdot dx < a_n < \int_{1}^{n} \log x \cdot dx. \tag{9.4}$$

Неопределенный интеграл от $\log x$ равен $x \log x - x$, и поэтому соотношение (9.4) сводится к двойному неравенству:

$$\left(n + \frac{1}{2}\right) \log n - n + \frac{3}{2}\left(1 - \log \frac{3}{2}\right) < \log n! < \left(n + \frac{1}{2}\right) \log n - n + 1. \tag{9.5}$$

Положим, для краткости,

$$\delta_n = \log n! - \left(n + \frac{1}{2}\right) \log n + n. \tag{9.6}$$

Тогда $1 - \delta_n$ представляет разность между правой частью неравенства (9.5) и $\log n!$, т. е. $1 - \delta_n$ равно площади между кривой $y = \log x$ и ломаной линией $A_1 A_2 A_3 \ldots A_n$. Отсюда следует, что δ_n *монотонно убывает*. Но неравенство (9.5) показывает, что $\frac{3}{2}\left(1 - \log \frac{3}{2}\right) < \delta_n < 1$. В силу известной теоремы анализа последовательность $\{\delta_n\}$ стремится к пределу, заключенному между $\frac{3}{2}\left(1 - \log \frac{3}{2}\right)$ и 1. Обозначив этот предел через $\log c$, получим

$$\delta_n \to \log c, \quad \text{где} \quad 2{,}45 < c < 2{,}72. \tag{9.7}$$

В логарифмических обозначениях формула Стирлинга сводится к формуле (9.7), если только $c = (2\pi)^{1/2}$ (или приблизительно 2,507). Число π можно определить разными способами, и для наших целей всего проще и естественнее положить $\pi = \frac{c^2}{2}$. При таком определении формула (9.7) эквивалентна формуле Стирлинга, но зато возникает вопрос о том, совпадает ли введенная постоянная π с общепринятой постоянной π, фигурирующей в других соотношениях. Этот факт будет получен попутно в гл. VII, после чего доказательство формулы Стирлинга можно будет считать завершенным.

Уточнение формулы Стирлинга. Формулу (9.1) можно усилить, добавив еще несколько членов разложения. Хотя уточненная формула Стирлинга нигде использоваться не будет, мы наметим доказательство двойного неравенства[1]):

$$(2\pi)^{\frac{1}{2}}\, n^{n+\frac{1}{2}}\, e^{-n+\frac{1}{12n+1}} < n! < (2\pi)^{\frac{1}{2}}\, n^{n+\frac{1}{2}}\, e^{-n+\frac{1}{(12n)}}. \tag{9.8}$$

Для доказательства (9.8) заметим, что

$$\delta_{n+1}-\delta_n=\left(n+\frac{1}{2}\right)\log\frac{n+1}{n}-1=\frac{1}{3(2n+1)^2}+\frac{1}{5(2n+1)^5}+\ldots. \tag{9.9}$$

[Это разложение следует из формулы (8.11), если в ней положить $t=\frac{1}{(2n+1)}$.] Мы увеличим все члены в правой части равенства (9.9), заменив коэффициенты $\frac{1}{5}, \frac{1}{7}, \frac{1}{9}, \ldots$ на $\frac{1}{3}$. После этой замены получается геометрическая прогрессия со знаменателем $\frac{1}{(2n+1)^2}$, и поэтому

$$\delta_{n+1}-\delta_n<\frac{1}{3[(2n+1)^2-1]}=\frac{1}{12n}-\frac{1}{12(n+1)}. \tag{9.10}$$

Таким образом, последовательность $\left\{\delta_n+\frac{1}{12n}\right\}$ монотонно убывает. Предел этой последовательности определяется формулой Стирлинга. Отсюда вытекает правая часть двойного неравенства (9.8). Левая часть этого неравенства получается аналогично, если заметить, что из (9.9) следует соотношение

$$\delta_{n+1}-\delta_n>\frac{1}{3(2n+1)^2}>\frac{1}{12n+1}-\frac{1}{12(n+1)+1}. \tag{9.11}$$

Формулы (9.8) дают замечательное приближение для $n!$. Даже при $n=1$ верхняя и нижняя оценки, полученные по этим формулам, равны 0,9958 ... и 1,0023. Верхняя оценка несколько лучше [см. (12.28)]. При $n=2$ она дает значение 2,0007, при $n=5$ получим 120,01, и при $n=10$ первые пять цифр верны.

§ 10. Примеры и упражнения[2])

Замечание. Во всех случаях предполагается, что все комбинации имеют равные вероятности.

1. Сколько можно образовать различных инициалов, если каждый человек имеет: а) ровно два, б) не больше двух, в) не больше трех имен.

---

[1]) H. Robbins, A remark on Stirling's formula, *Amer. Math. Monthly*, 62 (1955), 26—29.

[2]) § 11 и § 12 содержат задачи другого характера и различные дополнения к тексту.

2. Сколькими способами могут быть поставлены на шахматной доске две ладьи различного цвета так, чтобы каждая могла взять другую?

3. Буквы азбуки Морзе образуются как последовательности точек и тире. Сколько букв можно составить не более чем из 10 символов?

4. Кости для игры в домино метятся двумя числами. Кости симметричны, и поэтому порядок чисел не существен. Сколько различных костей можно образовать, используя числа $1, 2, \ldots, n$?

5. Числа $1, 2, \ldots, n$ расставлены случайным образом. Найти вероятность того, что числа а) 1 и 2, б) 1, 2 и 3 расположены рядом и притом в порядке возрастания.

6. а) Найти вероятность того, что среди трех выбранных наугад цифр встретятся 2, 1, 0 повторений. б) Решить ту же задачу для четырех выбранных наугад цифр.

7. Найти вероятность p_r того, что среди r выбранных наугад цифр не будет двух равных. Вычислить значение p_{10}, используя формулу Стирлинга.

8. Найти вероятность того, что в k выбранных наугад цифр а) не входит 0, б) не входит 1, в) не входит ни 0, ни 1, г) не входит или 0, или 1. Пусть через A и B обозначены события, описанные в „а" и „б". Выразить через A и B остальные события.

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

10. Двенадцать мест для стоянки автомобилей расположены в один ряд. Некто заметил, что на стоянке имеется восемь автомобилей, причем четыре пустых места следуют одно за другим (образуют серию). Следует ли считать такое расположение неожиданным (указывающим на отсутствие случайности).

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

12. Каждая из n палок разламывается на две части — длинную и короткую. Затем $2n$ полученных обломков объединяются в n пар, каждая из которых образует новую «палку». Найти вероятность а) того, что все обломки объединены в первоначальном порядке, б) того, что все длинные палки соединены с короткими [1]).

13. Проверка статистической гипотезы. Один профессор Корнельского университета двенадцать раз штрафовался за незаконную ночную стоянку машины, причем все двенадцать раз это происходило или во вторники или в четверги. Найти вероятность этого события. (Была бы оправдана аренда гаража только во вторники и в четверги?)

14. Продолжение. Из двенадцати штрафов ни один не пришелся на воскресенье. Согласуется ли это с гипотезой о случайности?

15. Ящик содержит 90 годных и 10 бракованных деталей. Найти вероятность того, что среди 10 вынутых из ящика деталей нет бракованных.

16. Из совокупности пяти символов a, b, c, d, e производится выборка с возвращением объема 25. Найти вероятность того, что в этой выборке будет содержаться по пяти символов каждого вида (ср. результаты с таблицей случайных чисел, отождествив цифры 0, 1 с a, цифры 2, 3 с b и т. д.)

---

[1]) Если клетки подвергаются воздействию вредной радиации, некоторые хромосомы расщепляются и играют роль «обломков». «Длинными» объявляются те «обломки», которые содержат так называемые центромы. Объединение двух «длинных» или двух «коротких» частей приводит к гибели клетки. См. D. G. Catcheside, The effect of X-ray dosage upon the frequency of induced structural changes in the chromosomes of Drosophila Melanogaster, *Journal of Genetics*, **36** (1938), 307—320.

17. n человек, в том числе A и B, располагаются в ряд в случайном порядке. Найти вероятность того, что между A и B будет стоять ровно r человек. Показать, что если n человек располагаются не в ряд, а в круг, то эта вероятность не зависит от r и, следовательно, равна $1/(n-1)$.

18. Найти вероятность того, что два бросания трех игральных костей дадут один и тот же результат, если кости а) отличимы друг от друга, б) не отличимы.

19. Показать, что более вероятно при одновременном бросании четырех костей получить хотя бы одну единицу, чем при 24 бросаниях двух костей получить хотя бы один раз две единицы. (Ответ известен как парадокс де-Мере. Игрок Шевалье де-Мере считал эти вероятности равными и обвинял математиков в своих проигрышах.)

20. Производится выборка объема r из генеральной совокупности, состоящей из n элементов. Найти вероятность того, что ни один из данных N элементов не будет входить в выборку, если выборка производится а) без возвращения б) с возвращением. Сравнить численные результаты для разных способов выбора, если (I) $n=100$, $N=r=3$ и (II) $n=100$, $r=N=10$.

21. Распространение слухов. В городе с населением в $(n+1)$ человек некто узнает новость. Он передает ее первому встречному, тот еще одному и т. д. На каждом шагу впервые узнавший новость может сообщить ее любому из n человек с одинаковыми вероятностями. Найти вероятность того, что в продолжение r единиц времени а) новость не возвратится к человеку, который узнал ее первым, б) новость не будет никем повторена. Решить ту же задачу в предположении, что на каждом шагу новость сообщается группе из N случайно выбранных людей. (Первая часть задачи является частным случаем этой более общей постановки при $N=1$.)

22. Цепь писем. Рассматривается генеральная совокупность, состоящая из $(n+1)$ человек. Человек, которого условимся называть «прародителем», пишет два письма случайно выбранным адресатам, которые образуют «первое поколение». Те в свою очередь делают то же самое, в результате чего образуется «второе поколение». Вообще каждый из людей, входящих в «r-е поколение», посылает два письма случайно выбранным адресатам. Найти вероятность того, что «прародитель» не входит ни в одно из «поколений» с номерами $1, 2, \ldots, r$. Найти медиану распределения, предполагая n достаточно большим.

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

24. Чему равна вероятность того, что а) все дни рождения 12 человек придутся на разные месяцы года (при условии, что вероятности попадания дня рождения на каждый из месяцев остаются равными для всех месяцев, б) дни рождения 6 человек придутся в точности на два месяца.

25. Найти вероятность того, что для данных 30 человек среди 12 месяцев года 6 месяцев содержат по два и 6 месяцев по три дня рождения.

26. В чулане находится n пар ботинок. Из них случайно выбирается $2r$ ботинок $(2r < n)$. Найти вероятность того, что а) среди выбранных ботинок отсутствуют парные, б) имеется одна комплектная пара, в) имеется две комплектные пары.

27. На автомобильной стоянке вновь прибывшая машина занимает место в ряду, состоящем из N автомобилей (не в том и не в другом конце этого ряда). Когда хозяин машины возвратился, он заметил, что r из N мест еще заняты. Какова вероятность того, что оба соседних места пусты?

28. Группа из $2N$ мальчиков и $2N$ девочек делится на две равные части. Найти вероятность того, что в каждой части число мальчиков и девочек одинаково. Вычислить эту вероятность, используя формулу Стирлинга.

29. Доказать, что вероятность игроку A получить ровно k тузов при сдаче колоды карт для игры в бридж равна вероятности того, что в 13 картах, случайно выбранных из колоды, будет k тузов (интуитивно это понятно, но заметим, что нахождение этих двух вероятностей приводит к различным опытам: во втором случае выбирается 13 карт, а в первом — распределяются все 52 карты).

30. В условиях предыдущей задачи доказать, что вероятность того, что игрок A получит ровно m, а игрок B ровно n пик, равна вероятности того, что в двух наборах, по 13 карт каждый, содержится соответственно m и n пик.

31. Чему равна вероятность того, что полученные A и B карты содержат вместе ровно k пик, где $k = 0, 1, 2, 3, 4$?

32. Пусть a, b, c, d — четыре неотрицательных целых числа, таких, что $a+b+c+d=13$. Найти вероятность $p(a, b, c, d)$ того, что при игре в бридж первый, второй, третий и четвертый игроки получат a, b, c и d пик соответственно. Определить схему размещения красных и черных шаров по ящикам, которая содержит сформулированную задачу в качестве частного случая.

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

34. Пусть a, b, c, d — целые числа, такие, что $a+b+c+d=13$. Найти вероятность $q(a, b, c, d)$ того, что полученные при игре в бридж карты содержат a пик, b червей, c бубен и d треф, и показать, что задача не сводится ни к одному из случайных размещений 13 шаров по 4 ящикам.

35. Распределение тузов в наборе из r карт. Найти вероятности $p_0(r)$, $p_1(r)$, ..., $p_4(r)$ того, что среди r карт, случайно выбранных из колоды карт для игры в бридж, встретится 0, 1, ..., 4 туза соответственно. Проверить, что $p_0(r) = p_4(52-r)$.

36. Продолжение. Время ожидания. Из полной колоды последовательно выбираются карты. Найти вероятности $f_1(r)$, ..., $f_4(r)$ того, что первый, ..., четвертый туз появятся при r-м испытании. Попытаться сначала угадать медианы времен ожидания первого, ..., четвертого появлений, а затем вычислить их.

37. Найти вероятность того, что каждый из двух наборов карт содержит k тузов, если объем каждого из наборов равен r и выбор производится а) из одной колоды, б) из двух колод карт для игры в бридж. Показать, что при $r=13$ вероятность в задаче (а) совпадает с вероятностью того, что каждый из двух заданных игроков в бридж получит k тузов.

38. Опечатки. На каждой странице книги напечатано N символов, возможны опечатки. Книга содержит $n=500$ страниц и $r=50$ опечаток. Показать, что а) вероятность того, что 1, 2, ..., n-я страницы содержат соответственно r_1, r_2, ..., r_n опечаток, равна

$$\binom{N}{r_1}\binom{N}{r_2}\cdots\binom{N}{r_n} : \binom{nN}{r};$$

б) при больших N эта вероятность приблизительно совпадает с вероятностью (5.5). Заключить отсюда, что распределение r опечаток по n страницам приблизительно соответствует случайному размещению r шаров по n ящикам. (Замечание. Это свойство можно сформулировать как общую предельную теорему для статистики Ферми — Дирака. См. § 5.)

Замечание. Последующие задачи имеют отношение к материалу § 5.

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

40. Чему равно число различимых результатов совместного бросания r_1 игральных костей и r_2 монет?

41. Сколькими различимыми способами могут быть расположены r_1 белых, r_2 черных и r_3 красных шаров?

В задачах 4—6 через n обозначается число ячеек, через r — число объектов; предполагается, что все объекты различимы и все n^r распределений имеют равные вероятности.

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

43. Лифт. В примере 3, в) лифт отправляется с семью пассажирами и останавливается на 10 этажах. Различные размещения пассажиров по этажам могут быть записаны символами, подобными символу (3, 2, 2), который означает, что 3 пассажира направляются на один этаж, 2 — на другой и 2 — на третий. Найти вероятности всех 15 возможных размещений пассажиров по этажам, начиная от (7) и кончая (1, 1, 1, 1, 1, 1, 1).

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

45. Найти вероятность того, что в покере получатся следующие наборы карт:

а) десятка, валет, дама, король, туз одной и той же масти;
б) четыре карты одинакового значения;
в) одна пара и одна тройка карт с одинаковыми значениями;
г) пять последовательных карт независимо от масти;
д) три карты одинакового значения плюс две дополнительные карты;
е) две пары карт одинакового значения плюс одна другая карта;
ж) две карты одинакового значения плюс три различные карты.

§ 11. Задачи и дополнения теоретического характера

1. Имеется np красных и nq черных шаров $(p+q=1)$. Из генеральной совокупности всех шаров производится выборка с возвращением. Показать, что вероятность того, что в выборку входит k красных шаров, равна

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

2. Предельная теорема для гипергеометрического распределения. Если n велико, а $n_1/n = p$, то вероятности q_k, даваемые формулами (6.1) и (6.2), приближаются к выражению (11.1). Точнее:

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

Сравнение результатов этой и предшествующей задач показывает, что *при больших генеральных совокупностях практически нет разницы между выборкой с возвращением и без возвращения.*

3. Из генеральной совокупности, состоящей из n элементов, выбирается группа в r элементов (без возвращения). Доказать, что вероятность u_r того, что каждый из N данных элементов содержится в выборке, равна

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

[Соответствующая формула для выбора с возвращением дается в (11.10) и не может быть получена прямыми рассуждениями. Другой вид формулы (11.3) см. в задаче 9 гл. IV.]

4. Предельная теорема. Если $n \to \infty$ и $r \to \infty$ так что $\frac{r}{n} \to p$, то $u_r \to p^N$ (см. задачу 13).

Замечание. Задачи 5—13 связаны с классической задачей о размещении (статистикой Максвелла — Больцмана). Другими словами, r предметов размещаются по n ячейкам и каждое из n^r возможных распределений имеет вероятность n^{-r} [1]).

5. Вероятность p_k того, что заданная ячейка содержит ровно k предметов, определяется биномиальным распределением (4.5). Наиболее вероятным является целое число ν, такое, что $(r-n+1)/n < \nu \leqslant (r+1)/n$. (Другими словами, утверждается, что $p_0 < p_1 < \ldots < p_{\nu-1} \leqslant p_\nu > p_{\nu+1} > \ldots > p_r$; см. задачу 15.)

6. Предельная теорема. Если $n \to \infty$ и $r \to \infty$ таким образом, что среднее число $\lambda = r/n$ шаров на ячейку остается постоянным, то

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

Это — распределение Пуассона, которое подробно исследуется в гл. VI см. задачу 15.

7. Пусть $A(r, n)$ — число способов расположения r предметов по n ячейкам, при которых нет пустых ячеек. Показать с помощью комбинаторных рассуждений, что

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

Вывести отсюда, что

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

Указание. Воспользоваться индукцией. Предполагая формулу (11.6) верной, выразить $A(r-k, n)$ в (11.5) через выражение (11.6). Изменить порядок суммирования и, воспользовавшись биномом Ньютона, представить $A(r, n+1)$ в виде разности двух простых сумм. Заменить во второй сумме $\nu+1$ новым индексом суммирования и применить (8.6).

Замечание. *Формула* (11.6) *представляет теоретическое решение поставленной задачи, но, конечно, было бы неблагодарной задачей использовать ее для вычисления, скажем, вероятности x того, что в деревне с населением $n = 1900$ человек каждый день в году является чьим-то днем рождения. В* § 2 *гл.* IV *мы выведем формулу* (11.6) *другим методом и получим для нее простое приближение* (*эта формула показывает, например, что приблизительно $x = 0{,}135$*).

[1]) Задачи 5—19 играют некоторую роль в квантовой статистике, теории фотопластинок, счетчиков Гейгера — Мюллера и т. п. В силу этого некоторые формулы многократно обсуждались и заново открывались в физической литературе, зачастую без выявления их классического и совершенно элементарного характера. Вероятно, все эти задачи имеются (хотя и в измененном виде) в книге Витворта, цитированной в самом начале главы.

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

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

9. Показать, не опираясь на предыдущие результаты, что *вероятность*

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

того, что ровно m ячеек окажутся пустыми, удовлетворяет уравнению

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

10. Используя результаты задач 7 и 8, прямым вычислением показать, что равенство (11.8) справедливо. Показать, что этот метод предусматривает *новый вывод (индукцией по r)* формулы (11.6).

11. Из формулы (11.6) и задачи 8 вывести, что *вероятность того, что m или более ячеек окажутся пустыми, равна*

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

(При $m \geqslant n$ это выражение сводится к нулю, что интуитивно очевидно.)

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

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

Вывести отсюда, что

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

[Воспользоваться формулой бинома Ньютона. При $N = n$ имеем $u(r, n) = n^{-r} A(r, n)$. Заметим, что формула (11.11) аналогична формуле (11.3) для выбора с возвращением [1]). Другой вывод см. в задаче 8 гл. IV.]

---

[1]) Заметим, что $u(r, n)$ можно интерпретировать как вероятность того, что при выборе без возвращения время ожидания момента, когда N-й элемент присоединится к выборке, меньше r. Результат можно применить к случайным цифрам; в этом случае $u(r, 10) - u(r-1, 10)$ — вероятность того, что потребуется последовательность из r элементов для того, чтобы получить все 10 цифр. Этот результат можно использовать для проверки случайности. Гринвуд [R. E. Greenwood, Coupon collector's test for random digits, *Mathematical Tables and other Aids to Computation*, 9 (1955), 1—5] табулировал это распределение и сравнил его с результатами, которые дают соответствующие времена ожидания для первых 2035 десятичных знаков числа π и первых 2486 десятичных знаков числа e. Медиана времени ожидания полного набора цифр равна 27. Вероятность того, что это время ожидания больше, чем 50, превосходит 0,05 и больше, чем 75, равна приблизительно 0,0037.

13. Предельная теорема. При предельном переходе, описанном в задаче 4, получим

$$u(r, n) \to (1 - e^{-p})^N.$$

Замечание. *Задачи* 14—19 *имеют тот же смысл, что и выше, но теперь мы предполагаем, что предметы неразличимы и что все различимые расположения равновероятны (статистика Бозе—Эйнштейна).*

14. Доказать, что вероятность наличия в данной ячейке ровно k предметов равна

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

15. Показать, что при $n > 2$ нуль является самым вероятным числом предметов, попавших в данную ячейку, или, более точно, $q_0 > q_1 > q_2 > \ldots$ (ср. с задачей 5).

16. Предельная теорема. *Пусть n и r неограниченно возрастают, причем среднее число r/n частиц, приходящихся на одну ячейку, стремится к λ. Тогда*

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

[Правая часть известна под наименованием *геометрического распределения.*]

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

$$\rho_m = \binom{n}{m}\binom{r-1}{n-m-1} : \binom{n+r-1}{r}. \tag{11.14}$$

18. Вероятность того, что в заданной группе из r ячеек окажется в целом j объектов, равна

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

19. Предельная теорема. *При предельном переходе, описанном в § 4, получим*

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

(Правая часть представляет частичный случай отрицательного биномиального распределения, которое будет введено в гл. VI.)

Теоремы о сериях. *В задачах* 20—25 *мы рассматриваем расположение r_1 букв α* и *r_2 букв β, предполагая, что все расположения равновероятны [см. пример* (4.2)]. *Эта группа задач связана с* § 5а.

20. Вероятность того, что расположение содержит ровно k серий, равна

$$P_{2\nu} = 2\binom{r_1-1}{\nu-1}\binom{r_2-1}{\nu-1} : \binom{r_1+r_2}{r_1} \tag{11.17}$$

при четном $k = 2\nu$ и

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

при нечетном $k = 2\nu + 1$.

21. Продолжение. Вывести, что наиболее вероятным числом серий является целое число k, удовлетворяющее неравенству $\frac{2r_1r_2}{r_1+r_2} < k < \frac{2r_1r_2}{r_1+r_2}+3$. (Указание. Рассмотреть отношения

$$P_{2\nu+2}:P_{2\nu} \text{ и } P_{2\nu+1}:P_{2\nu-1}.)$$

22. Вероятность того, что расположение начинается с α-серии длины $\nu \geqslant 0$, равна $(r_1)_\nu r_2:(r_1+r_2)_{\nu+1}$. (Указание. ν букв α и одна буква β, которая следует за ними, фиксированы.) Каков смысл этой теоремы при $\nu=0$?

23. Вероятность наличия ровно k α-серий равна

$$\pi_k = \binom{r_1-1}{k-1}\binom{r_2+1}{k}:\binom{r_1+r_2}{r_1}. \tag{11.19}$$

Указание. Этот результат является простым следствием второй части леммы § 5. Кроме того, уравнение (11.19) можно получить из (11.17) и (11.18), но такой способ довольно затруднителен.

24. Вероятность того, что n-й букве α предшествует в точности m букв β, равна

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

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

$$\frac{k!}{k_1!\,k_2!\,\ldots\,k_\nu!}\binom{r_2+1}{k}:\binom{r_1+r_2}{r_1}. \tag{11.21}$$

§ 12. Задачи и тождества, связанные с биномиальными коэффициентами

1. Для целого $n \geqslant 2$

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

(Указание. Использовать формулу бинома Ньютона.)

2. Доказать, что при любых целых положительных n, k

$$\binom{n}{0}\binom{n}{k}-\binom{n}{1}\binom{n-1}{k-1}+\binom{n}{2}\binom{n-2}{k-2}\pm\ldots\pm\binom{n}{k}\binom{n-k}{0}=0. \tag{12.2}$$

Это соотношение является частным случаем более общего тождества [1])

$$\sum \binom{n}{\nu}\binom{n-\nu}{k-\nu} t^{\nu} = \binom{n}{k}(1+t)^k. \tag{12.3}$$

3. Для любого $a > 0$

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

Доказать это непосредственно, а также с помощью рассмотрения геометрической прогрессии $\sum x^k = (1-x)^{-1}$.

4. Доказать, что

$$\binom{2n}{n} 2^{-2n} = (-1)^n \binom{-1/2}{n}. \tag{12.5}$$

5. При неотрицательных целых n и r и любом действительном a

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

[Указание. Воспользоваться уравнением (8.6). Особенно часто используется частный случай этой формулы при $n = a$.]

6. При произвольном a и целом $n \geqslant 0$

$$\sum_{\nu=0}^{n} (-1)^{\nu} \binom{a}{\nu} = (-1)^n \binom{a-1}{n}. \tag{12.7}$$

[Указание. Воспользоваться формулой (8.6).]

7. При целых положительных r и k

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

а) Установить это, используя уравнение (8.6).

б) Показать, что (12.8) является частным случаем (12.7).

в) Показать по индукции, что, используя тождество (12.8), можно дать новое доказательство первой части леммы § 5.

8. В § 6 мы отмечали уже, что сумма членов гипергеометрического распределения равна единице, т. е. для любых положительных целых чисел a, b и n

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

Доказать это по индукции. (Указание. Доказать сначала, что формула (12.9) верна при $a = 1$ и всех b.)

---

[1]) Читателю следует вспомнить соглашение (8.5): если ν пробегает все целые числа, то только конечное число членов суммы (12.3) отлично от нуля.

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

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

доказать, что формула (12.9) верна при произвольных a и b (и целом n).

10. Воспользовавшись (12.9), доказать, что

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

11. Воспользовавшись (12.10), доказать, что

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

12. Доказать, что для целых $0 < a < b$

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

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

$$(1-t)^a (1-t)^{-b-2} = (1-t)^{a-b-2}.$$

13. Вывести из соотношения (12.9) тождества

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

и

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

которые имеют место, если k, n и r — целые положительные числа. [Указание. Использовать (12.4).]

14. Используя соотношение (12.9), доказать, что

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

[Указание. Дважды применить уравнение (12.4).] Рассмотреть важные частные случаи, соответствующие $b = 1,2$.

15. Обращаясь к задачам § 11, отметим, что формулы (11.12), (11.14), (11.15) и (11.16) определяют вероятности, поэтому сумма соответствующих величин равна единице. Показать, что отсюда следуют соотношения (12.8), (12.9), (12.16) и формула бинома Ньютона.

16. Из определения величин $A(r, n)$ в задаче 7 § 11 следует, что $A(r, n) = 0$ при $r < n$ и $A(n, n) = n!$. Другими словами,

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

а) Доказать (12.17) непосредственно редукцией от n к $n-1$. б) Дать другое доказательство (12.17), исследуя r-ю производную функции $(1-e^t)^n$ в точке $t=0$. в) Обобщить (12.17), отправляясь от (11.11), а не от (11.6).

17. Доказать по индукции, что при любом целом $r \geqslant 0$ и $0 \leqslant N \leqslant n$

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

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

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

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

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

19. Показать, что при любом целом положительном m

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

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

20. Воспользовавшись формулой Стирлинга (7.7), доказать, что

$$\binom{2n}{n} \sim (\pi n)^{-(1/2)} 2^{2n}. \tag{12.21}$$

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

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

22. Гамма-функция определяется следующей формулой:

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

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

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

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

[Постоянная C равна $(2\pi)^{1/2}/\Gamma(a/r)$.]

24. Воспользовавшись результатом предшествующей задачи, показать, что

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

25. Получить следующий вид формулы Стирлинга:

$$n! \sim (2\pi)^{1/2} (n + 1/2)^{n+1/2} e^{-(n+1/2)}. \tag{12.26}$$

26. Продолжение. С помощью использованного в тексте метода показать, что

$$(2\pi)^{\frac{1}{2}}\left(n+\frac{1}{2}\right)^{n+\frac{1}{2}} e^{-\left(n+\frac{1}{2}\right)-1/24\left(n+\frac{1}{2}\right)} < n! < (2\pi)^{\frac{1}{2}}\left(n+\frac{1}{2}\right)^{n+\frac{1}{2}} e^{-\left(n+\frac{1}{2}\right)}. \tag{12.27}$$

27. Уточняя формулу Стирлинга, доказать, что

$$n! \sim (2\pi)^{\frac{1}{2}} n^{n+\frac{1}{2}} \exp\left\{-n + \frac{1}{12n} - \frac{1}{360n^3} + \ldots\right\}. \tag{12.28}$$

ГЛАВА III*)

КОЛЕБАНИЯ ПРИ ИГРЕ С БРОСАНИЕМ МОНЕТЫ И СЛУЧАЙНЫЕ БЛУЖДАНИЯ

Эта глава преследует две цели. Во-первых, будет показано, что чрезвычайно простыми методами можно получить далеко идущие и важные результаты. Во-вторых, в ней мы впервые встретимся с теоретическими заключениями, которые не только неожиданны, но, казалось бы, противоречат логике и здравому смыслу. Мы увидим, что общепринятое мнение о характере случайных колебаний лишено оснований и что выводы из закона больших чисел часто истолковываются неправильно[1]).

Эти вопросы рассмотрены здесь только потому, что они вполне элементарны. Основная линия книги будет продолжена в гл. V. Книга в целом не зависит от настоящей главы. Некоторые формулы впоследствии появятся в разделах, посвященных исследованию времен первого достижения и возвращения, однако они будут выведены заново аналитическими методами. Сравнение методов должно оказаться интересным и поучительным. *Настоящая глава может быть прочитана, по усмотрению читателя, или независимо, или параллельно остальной части книги.* Главу можно читать в *два приема*, основной текст напечатан обычным шрифтом. Мелким шрифтом изложен дополнительный материал (относящийся главным образом к временам первого достижения и возвращения). Его можно пропустить при первом чтении. В § 7 приведены некоторые опытные данные.

---

*) Эта глава может быть опущена или прочитана одновременно с последующими главами. Близкие вопросы будут рассматриваться в гл. X (законы больших чисел), XI (времена первого достижения), XIII (рекуррентные события), IV (случайные блуждания), но материал главы нигде в последующем непосредственно использоваться не будет.

[1]) Хотя мы формально имеем дело лишь с бросаниями монеты, основные заключения широко приложимы. Действительно, Спарре Андерсен сделал замечательное открытие, что многие факты теории колебаний сумм независимых случайных величин имеют чисто комбинаторную природу и являются общими для широкого класса таких величин. Так, в частности, справедливы два закона арксинуса. См. *Mathematica Scandinavica*, 1 (1953), 263—285 и 2 (1954), 195—223.

§ 1. Основные понятия

Удивительно много результатов относительно случайных колебаний выводится из следующей не привлекающей внимания леммы, отмеченной Бертраном в 1887 г. Эту лемму можно связать с рассматривающимися в комбинаторном анализе[1]) *задачами о баллотировке* (ballot problems). Предположим, что на выборах кандидат P *собрал* p *голосов, а кандидат* Q — q *голосов, где* $p > q$. *Вероятность, что в течение всего времени* P *был впереди* Q *равна* $(p-q)/(p+q)$.

На математическом языке мы здесь имеем дело с расстановками $x = p + q$ символов $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_x$, из которых p равны $+1$, а остальные -1 (голоса, поданные соответственно за P и Q). Частная сумма $s_k = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_k$ есть разность числа голосов, поданных за P и Q после того, как подан k-й голос. Ясно, что $s_x = p - q$ и

$$s_i - s_{i-1} = \varepsilon_i = \pm 1, \quad s_0 = 0 \quad (i = 1, 2, \ldots, x). \tag{1.1}$$

Обратно, каждый набор $\{s_1, s_2, \ldots, s_x\}$ целых чисел, удовлетворяющих (1.1), представляет возможный исход голосования. Мы будем пользоваться геометрической терминологией и изображать такой набор ломаной линией, i-й отрезок которой имеет наклон ε_i и i-я вершина — ординату s_i. Такие линии будем называть путями.

Определение. *Пусть* $x > 0$ *и* y — *целые числа. Путем* $\{s_1, s_2, \ldots, s_x\}$ *из начала координат в* (x, y) *называется ломаная с вершинами, которые имеют абсциссы* $0, 1, 2, \ldots, x$ *и ординаты* $s_0, s_1, \ldots, s_x$, *удовлетворяющие условиям* (1.1) *и такие, что* $s_x = y$.

В рассматриваемом случае, если p среди символов ε_i положительны, а q отрицательны, то

$$x = p + q, \qquad y = p - q. \tag{1.2}$$

Произвольная точка (x, y) может быть соединена путем с началом координат тогда и только тогда, когда x и y имеют вид (1.2). В этом

[1]) Историю вопроса и литературу см. A. Dvoretsky and Motzkin, A problem of arrangements, *Duke Mathematical Journal*, 14 (1947), 305—313. Как отмечают эти авторы, большинство формально различных доказательств опирается в действительности на принцип отражения (лемма 1 § 2). Однако без геометрической интерпретации этот принцип теряет свою простоту и выглядит курьезным трюком Дворецкий и Моцкин дают новое простое и изящное доказательство. Они обобщают баллотировочную задачу, потребовав, чтобы на каждом шагу P имел по крайней мере α-ю долю голосов, поданных за Q. Эта работа была продолжена M. T. L. Bizley, Derivation of a new formula for the number of minimal lattice paths, etc., *The Journal of the Institute of Actuaries*, 80, 1, 354 (1954), 55—62.

случае p положительных ε_i можно разместить на $x=p+q$ возможных местах

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

различными способами. Удобно *доопределить* $N_{x,y}=0$, *если* x *и* y *не имеют форму* (1.2). Тогда *существует ровно* $N_{x,y}$ *различных путей из начала координат в точку* (x, y). Лемма Бертрана устанавливает, что при $y>0$ существует $(y/x)N_{x,y}$ путей, удовлетворяющих условиям $s_1>0$, $s_2>0, \ldots, s_{x-1}>0$, $s_x=y$. Это будет доказано в § 2.

Пример. Рис. 1 изображает путь в точку $N_1=(5, 1)$. Имеется 10 таких путей, из которых два удовлетворяют условию $s_i>0$ —

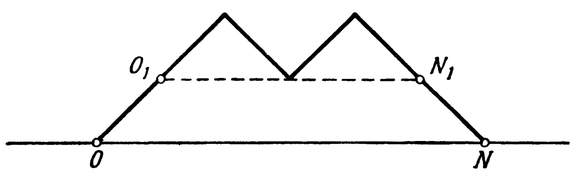


Рис. 1. Иллюстрация положительных путей и доказательства теоремы 2 § 2.

— $\{1, 2, 1, 2, 1\}$ и $\{1, 2, 3, 2, 1\}$. На рисунке — первый из них. Мы можем получить очень интересные следствия из леммы Бертрана, если отбросим предположение, что конечная точка (x, y) пути фиксирована заранее. Тогда существует 2^n различных путей из начала в точку (n, y) с произвольной ординатой y. Как разъясняется в § 3, эти 2^n путей можно использовать для представления 2^n возможных исходов n последовательных бросаний правильной монеты. В классическом изложении вводится воображаемый игрок Петр, который при каждом испытании выигрывает или теряет один рубль. Числа $\{s_1, s_2, \ldots, s_n\}$ *представляют тогда последовательные чистые выигрыши Петра*, т. е. разности между полным числом гербов и решеток.

Если $s_n=0$, то чистый выигрыш после n-го испытания равен нулю и имеет место ничья. Частота ничьих столь мала, что они не отражаются на общей картине, но постоянное упоминание о них нежелательно. Условимся поэтому говорить, что *после n-го испытания Петр находится в выигрыше, если или* $s_n>0$, *или* $s_n=0$, *но* $s_{n-1}>0$. Фраза «Петр находится в выигрыше после n испытаний» означает геометрически, что n-й отрезок пути лежит выше оси x.

Баллотировочная задача относится к путям, расположенным выше оси x, т. е. к играм, при которых один из игроков всегда находится в выигрыше. Следующим шагом является исследование вопроса о том, сколь долго один из игроков находится в выигрыше. В этой связи мы придем к выводам, которые разрушат наши интуитивные представления. В самом деле, обычно полагают, что при достаточно длинной серии бросаний правильной монеты Петр будет в выигрыше около половины всего времени. Это, однако, совершенно неверно. *Если монету бросать* 20 000 *раз, то приблизительно в* 88 *раз более вероятно, что Петр будет находиться в выигрыше после каждого испытания, чем что он будет находиться в выигрыше после* 10 000 *испытаний.* Вообще, лидерство в игре меняется более редко, чем подсказывает интуиция. Как бы длинны ни были серии бросаний, *самое вероятное число смен лидерства есть* 0, точно так же одна смена лидерства более вероятна чем две, две чем три и т. д. Короче говоря, если бы результаты достаточно длинных игр с бросанием монеты описывал современный адвокат или психолог, он бы классифицировал большинство монет как неправильные. Если несколько монет подбросить n раз каждую, то удивительно большая доля этих монет будет вести себя таким образом, что один из играющих окажется почти все время в выигрыше и в относительно немногих случаях лидерство будет меняться и колебаться так, как это обычно ожидается от правильной монеты.

Это образец заключения, полученного из закона арксинуса (см. § 5 и примеры § 7). С. Андерсен показал, что этот закон широко применим в ситуациях, сходных с описанной. Большинство случайных процессов в физике, экономике и образовании подчинены этому закону, и наши выводы должны послужить предупреждением тем, кто склонен подчеркивать различие между общей тенденцией и отклонениями от средних норм.

На эту ситуацию можно взглянуть и с несколько иной точки зрения. Здравый смысл подсказывает, что если не принимать во внимание случайных колебаний, то, бросая монету через равные промежутки времени за два дня игры, мы получим в два раза больше ничьих, чем за один день. Другими словами, мы ожидаем интуитивно, что число ничьих приблизительно пропорционально продолжительности игры. Пародоксально, что это не так. *Число ничьих возрастает как корень квадратный из времени.* При 10 000 бросаний медиана числа ничьих равна 67, но при 1 000 000 бросаний она возрастает только до 674, типичная «*длина волны*» увеличивается с 150 до 1500. *С увеличением числа испытаний средняя «длина волны» растет* (§ 6 и § 8). Формулы, на которых основаны эти заключения, играют важную роль при изучении времен первого достижения и первого возвращения в общей теории случайных блужданий и теории диффузии.

Теорема 3 § 2 стоит несколько в стороне от основной темы главы и в дальнейшем нигде не используется. Она представляет вариант баллотировочной задачи, соответствующей тому случаю, когда оба кандидата собрали одинаковое число n голосов. Тогда P находится впереди после четного числа $2k$ туров, а Q — после оставшихся $2n-2k$ туров. Мы опять ошибемся, если скажем, что каждый из кандидатов лидирует приблизительно половину времени (т. е. $2k$ близко к n). В действительности, в этом случае все $n+1$ *возможных распределений времени лидерства* (*а именно*: $2n:0$, $2n-2:2$, $2n-4:4, \ldots, 2:2n-2$, $0:2n$) *имеют одну и ту же вероятность* $(n+1)^{-1}$. Этот результат заметно отличается от того, который был описан выше. Там конечный результат не был фиксирован заранее и крайние распределения времени лидерства $2n:0$ и $0:2n$ имели наибольшую вероятность.

Д. Ходжес[1]) отмечал, что эту теорему можно использовать в статистике для получения *порядковых критериев*. Для иллюстрации этой точки зрения приведем пример.

Пример. Допустим, что некоторая величина (скажем, высота растения) измерена для каждого из n объектов, подвергнутых определенному воздействию, а также для каждого из n контрольных объектов. Для определенности предположим, что результаты наблюдений в каждой группе расположены в порядке убывания: $a_1 > a_2 > \ldots > a_n$ и $b_1 > b_2 > \ldots > b_n$. Объединим эти две последовательности и напишем $2n$ чисел $a_1, \ldots, b_n$ в порядке убывания. Результирующую последовательность n символов a и n символов b можно рассматривать как отчет о выборах, на которых каждый из кандидатов получил по n голосов. Если воздействие очень эффективно, то все a должны предшествовать всем b; воздействие, лишенное любого эффекта, приведет к случайному расположению символов. Условимся говорить, что символы a лидируют ровно $2k$ раз, если k различных a предшествуют b с тем же номером, т. е. если неравенство $a_i > b_i$ выполняется для k индексов. Если предположить случайность, то вероятность этого события равна $1/(n+1)$, а значит вероятность того, что a лидирует не менее $2k$ раз, равна $(n-k+1)/(n+1)$. Классический пример такого рода аргументов (использованных качественно, без знания теоретических вероятностей) дан Галтоном, который в 1876 г. применил эти рассуждения к данным, сообщенным ему Чарльзом Дарвином. В его примере $2n$ было равно 30, и символы a лидировали 26 раз. Галтон заключил, что воздействие было эффективным. Однако даже при отсутствии всякого эффекта символы a лидировали бы не менее 26 раз в трех из шестнадцати подобных экспериментов. Это показывает, что качественный анализ может быть ценным дополнением к нашей довольно шаткой интуиции. (Близкий критерий, основанный на теории серий см. в гл. II, § 5а.)

§ 2. Задачи о расположении

Пусть $A=(a, \alpha)$ и $B(b, \beta)$ — точки с целочисленными координатами, расположенные в первом квадранте $b > a \geqslant 0$, $\alpha > 0$, $\beta > 0$. *Точку* $A'=(a, -\alpha)$ *назовем отражением точки* A *относительно* x (см. рис. 2). *Путь из точки* A *в точку* B определяется так же, как и в § 1, с тем отличием, что роль начала выполняет точка A.

Лемма[2]). (*Принцип отражения*). *Число путей из* A *в* B, *которые касаются или пересекают ось* x, *равно числу всех путей из* A' *в* B.

---

[1]) Hodges J. L., Galton's rank-order test, *Biometrika*, 42 (1955), 261—262.

[2]) Вероятностная литература приписывает этот метод Андрэ [D. André (1887)]. В тексте он сводится к лемме о случайных блужданиях. Классические разностные уравнения для случайных блужданий (гл. IV) тесно связаны с дифференциальными уравнениями, и принцип отражения (даже в более сильной форме) хорошо известен в этой теории как *метод отражения Кельвина*.

Доказательство. Рассмотрим путь $\{s_a = \alpha, s_{a+1}, \ldots, s_b = \beta\}$ из точки A в точку B, одна или несколько вершин которого лежат на оси x. Пусть t — абсцисса *первой* такой вершины (см. рис. 2), т. е. t выбрано так, что $s_a > 0, \ldots, s_{t-1} > 0, s_t = 0$. Тогда $\{-s_a, -s_{a+1}, \ldots, -s_{t-1}, s_t = 0, s_{t+1}, s_{t+2}, \ldots, s_b\}$ — путь из точки A' в точку B, пересекающий ось x впервые в точке $T = (t, 0)$. Участки AT и $A'T$ этих путей получаются один из другого отражением относительно оси x, так что устанавливается взаимно однознач-

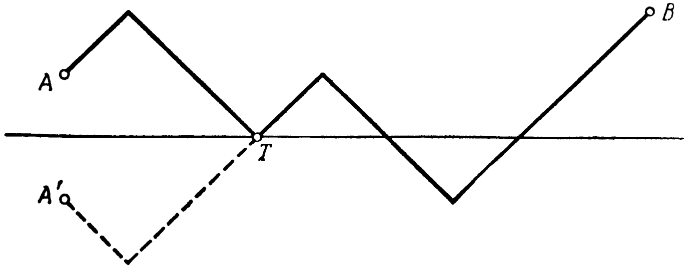


Рис. 2. Иллюстрация принципа отражения.

ное соответствие между всеми путями, соединяющими A' и B, и путями из A в B, имеющими вершину на оси x. Лемма доказана.

Теорема 1. *Теорема о баллотировке. Пусть $x > 0$, $y > 0$, число путей $\{s_1, s_2, \ldots, s_x = y\}$ из числа координат в точку (x, y), таких, что $s_1 > 0, s_2 > 0, \ldots, s_x > 0$ равно $(y/x)N_{x,y}$.*

Доказательство. Так как $s_1 = \pm 1$, то для каждого из допустимых путей $s_1 = 1$. Поэтому общее число допустимых путей равно числу путей из точки $(1, 1)$ в точку (x, y), которые не касаются оси x — *п* и не пересекают эту ось. Согласно предшествующей лемме это число равно

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

$$= \frac{p-q}{p+q}\binom{p+q}{p} = \frac{y}{x} N_{x,\, y}. \qquad (2.1)$$

Принцип двойственности. Почти всякую теорему о путях можно переформулировать так, чтобы получить формально отличную двойственную теорему. Рассмотрим путь $\{s_1, \ldots, s_n\}$ и путь, полученный из него изменением порядка ε_i на обратный, т. е. путь $\{s_1^*, \ldots, s_x^*\}$, где $s_1^* = \varepsilon_x$, $s_2^* = \varepsilon_x + \varepsilon_{x-1}$, $s_3^* = \varepsilon_x + \varepsilon_{x-1} + \varepsilon_{x-2}, \ldots$ или

$$s_0^* = 0,\ s_1^* = s_x - s_{x-1}, \ldots,\ s_i^* = s_x - s_{x-i}, \ldots,\ s_x^* = s_x. \qquad (2.2)$$

Эти два пути конгруэнтны и получаются один из другого поворотом на 180°, они имеют общие концевые точки. *Каждой теореме о путях соответствует теорема об обращенных путях* (2.2).

Например, теорема 1 дает нам число обращенных путей $\{s_1^*, \ldots, s_x^*\}$, соединяющих начало координат с точкой (x, y) и удовлетворяющих условиям $s_i^* > 0$ при $i = 1, 2, \ldots, x$. Но это означает, с другой стороны, что $s_x > s_{x-i}$ при $i = 1, 2, \ldots, x - 1$. Поэтому справедлива следующая теорема, двойственная теореме 1.

Теорема 1*. *Число путей $\{s_1, s_2, \ldots, s_x\}$ из $(0, 0)$ в (x, y), таких, что $s_1 < s_x$, $s_2 < s_x, \ldots, s_{x-1} < s_x$ (где $s_x = y > 0$), равно $(y/x) N_{x, y}$.*

На языке геометрии это означает, что теорема 1 относится к путям, левая концевая точка которых самая низкая, а теорема 1* — к путям, у ко-

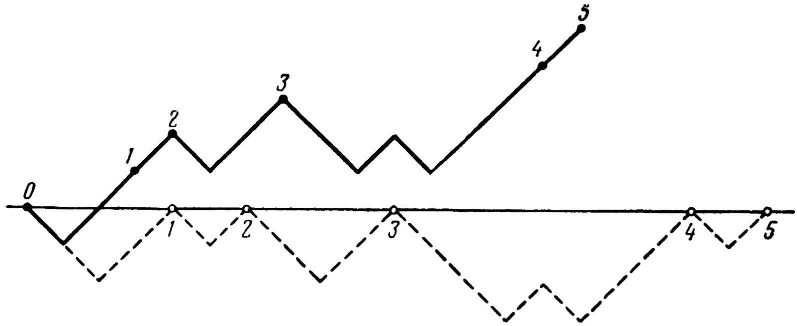


Рис. 3. Иллюстрация первого достижения и возвращения в начало.

торых правая концевая точка самая высокая. (См. рис. 3.) Теорема 1* используется при исследовании времени первого достижения в случайном блуждании.

Теперь приступим к изучению путей, соединяющих начало координат с точкой $N = (2n, 0)$, расположенной на оси x (вершины с нечетной абсциссой не могут быть расположены на оси x).

Положим для краткости

$$L_{2n} = \frac{1}{n+1} \binom{2n}{n}. \tag{2.3}$$

Теорема 2. *Среди $\binom{2n}{n}$ путей, соединяющих начало с точкой $2n$, по оси x, существует*

а) *ровно L_{2n-2} путей, таких, что*

$$s_1 > 0, \ s_2 > 0, \ \ldots, \ s_{2n-1} > 0 \quad (s_{2n} = 0); \tag{2.4}$$

б) *ровно L_{2n} путей, таких, что*

$$s_1 \geqslant 0, \ s_2 \geqslant 0, \ \ldots, \ s_{2n-1} \geqslant 0 \quad (s_{2n} = 0). \tag{2.5}$$

(Другими словами, существует столько же путей из начала координат в точку $2n$, все внутренние вершины которых лежат выше оси x, сколько существует путей в точку $2n - 2$, у которых нет вершин ниже оси x.)

Доказательство. (См. рис. 1) Каждый путь, удовлетворяющий условию (2.4), проходит через точку $N_1=(2n-1,\ 1)$, а по теореме 1 число путей из начала координат в точку N_1, таких, что $s_1>0,\ \ldots,\ s_{2n-2}>0$, равно

$$\frac{1}{2n-1}\binom{2n-1}{n-1}=\frac{1}{n}\binom{2n-2}{n-1}=L_{2n-2}. \tag{2.6}$$

Это доказывает (а). Пусть путь опять удовлетворяет условию (2.4). Выбрасывая первый и последний отрезки, получим путь, соединяющий точки $O_1=(1,\ 1)$ и $N_1=(2n-1,\ 1)$, причем ни одна из его вершин не лежит ниже линии $y=1$. Перенося начало координат в точку O_1, убеждаемся, что путь из O_1 в N_1 [эта точка имеет теперь координаты $(2n-2,\ 0)$] удовлетворяет условиям (2.5). Итак, установлено взаимно однозначное соответствие между такими путями и всеми путями, удовлетворяющими условию (2.4). Теорема доказана.

В § 1 было отмечено, что следующая теорема стоит несколько в стороне и в дальнейшем не будет использована.

Теорема 3[1]. *Пусть $L_{2k,\,2n}$ — число путей из начала координат в точку $2n$, таких, что $2k$ сторон лежит выше оси x, а остальные $2n-2k$ — ниже $(k=0,\ 1,\ 2,\ldots,\ n)$. Тогда $L_{2k,\,2n}=L_{2n}$ независимо от k.*

Доказательство. При $n=1$ утверждение тривиально, и мы предположим по индукции, что $L_{2k,\,2\nu}=L_{2\nu}$ при $\nu=1,\ 2,\ldots,\ n-1$ и $0\leqslant k\leqslant\nu$. Мы хотим сосчитать число путей $\{s_1,\ s_2,\ldots,\ s_{2n}=0\}$, у которых ровно $2k$ сторон расположены выше оси x. Сначала предположим, что $1\leqslant k\leqslant n-1$. Такие пути пересекают ось x, и мы обозначим через $2r$ абсциссу точки первого пересечения Рассмотрим теперь два класса путей.

Путь первого класса положителен между 0 и $2r$, а его часть, расположенная между $2r$ и $2n$, содержит ровно $2k-2r$ сторон, лежащих выше оси x. Здесь $k\geqslant r$. По теореме 2 (а) существует ровно L_{2r-2} путей $\{s_1,\ \ldots,\ s_{2r-1},\ s_{2r}=0\}$, для которых $s_1>0,\ldots,\ s_{2r-1}>0$, а по предположению индукции имеется $L_{2k-2r,\,2n-2r}=L_{2n-2r}$ путей, соединяющих точки $(2r,\ 0)$ и $(2n,\ 0)$, и таких, что ровно $2k-2r$ сторон расположены выше оси x-ов. Поэтому существует $L_{2r-2}L_{2n-2r}$ путей первого класса.

Путь второго класса отрицателен между 0 и $2r$, а его часть, расположенная между $2r$ и $2n$, имеет ровно $2k$ сторон, расположенных выше оси x. Аналогичными рассуждениями устанавливаем, что существует $L_{2r-2}L_{2n-2r}$ путей этого класса, но теперь $n-r\geqslant k$.

Отсюда следует, что при $k=1,\ldots,\ n-1$

$$L_{2k,\,2n}=\sum_{r=1}^{k}L_{2r-2}L_{2n-2r}+\sum_{r=1}^{n-k}L_{2r-2}L_{2n-2r}. \tag{2.7}$$

---

[1] Впервые эта теорема была доказана с помощью запутанных аналитических рассуждений в работе K. L. Chung and W. Feller, Fluctuations in coin tossing, *Proceedings National Academy of Sciences USA*, **35** (1949), 605—608 (см. также первое издание настоящей книги, гл. XII, задача 4). Изящное комбинаторное доказательство было дано Ходжесом (см. примечание 1 на стр. 84).

Заменяя индекс суммирования $\rho = n - r + 1$ и имея в виду, что во второй сумме $L_{2r-2}L_{2n-2r} = L_{2\rho-2}L_{2n-2\rho}$, получаем

$$L_{2k,\,2n} = \sum_{\rho=1}^{n} L_{2\rho-2}L_{2n-2\rho}, \tag{2.8}$$

а это выражение от k *не зависит.*

Пути, у которых все $2n$ сторон расположены выше оси x, рассматривались в теореме 2 (б), и в силу этой теоремы $L_{2n,\,2n} = L_{2n}$. Из соображений симметрии получаем также $L_{0,\,2n} = L_{2n}$. Полное число путей, соединяющих начало координат с точкой $(2n, 0)$, равно $(n+1)\,L_{2n}$, отсюда сразу следует $L_{2k,\,2n} = L_{2n}$ при $k = 0, 1, 2, \ldots, n$.

В качестве *следствия* мы получаем тождество

$$L_{2n} = \sum_{\rho=1}^{n} L_{2\rho-2}L_{2n-2\rho}. \tag{2.9}$$

Прямое аналитическое доказательство соотношения (2.9) приведено в § 8 (а).

§ 3. Случайное блуждание и игра с бросанием монеты

Пусть производится N бросаний правильной монеты. Положим $\varepsilon_k = +1$, если k-е испытание привело к выпадению герба, и $\varepsilon_k = -1$ в противном случае. Тогда $s_k = \varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_k$ есть разность между числом гербов и числом решеток после k-го испытания. В классических терминах игры в орлянку s_k — «чистый выигрыш Петра после k игр». Каждый возможный исход N бросаний правильной монеты можно интерпретировать как некоторый путь из начала координат, и обратно, каждому такому пути соответствует возможный исход N испытаний.

Это наводит на мысль *выбрать в качестве пространства элементарных событий* 2^N *путей* $\{s_1, \ldots, s_N\}$, *начинающихся в начале координат, и связать с каждым из них вероятность* 2^{-N}.

Такое событие, как «первые два испытания привели к выпадению герба» нужно интерпретировать как множество всех путей $\{s_1, s_2, \ldots$ $\ldots, s_N\}$, для которых $s_1 = 1$, $s_2 = 2$. Таких путей имеется 2^{N-2}, и, значит, вероятность соответствующего события равна 2^{-2}, что можно было ожидать заранее. Вообще если $k < N$, то существует ровно 2^{N-k} различных путей $\{s_1, s_2, \ldots, s_N\}$, таких, что первые k чисел $\{s_1, s_2, \ldots, s_k\}$ фиксированы. Отсюда следует, что *событие, зависящее от исхода первых k испытаний, имеет вероятность, которая не зависит от N.* Практически это означает, что число N не играет никакой роли, если оно достаточно велико. Лучше всего было бы рассматривать каждую конечную последовательность испытаний как начало бесконечной серии, однако это привело бы нас к несчетному пространству элементарных событий. Поэтому мы будем рассматривать конечные последовательности, где N больше, чем число опытов, фигурирующих в формулах. Вспоминать об этом N в дальнейшем нам почти не придется.

Полезно иметь в дополнение к употребляемому нами геометрическому языку еще и другой способ выражения. Мы предположим, что бросания производятся через равные промежутки времени, так что n-й опыт происходит в момент времени n. Петр может отмечать свой суммарный выигрыш после каждого испытания с помощью фишки, которую мы условимся называть частицей. Частица движется по вертикальной оси, выходя из нуля. В момент времени 1, 2, ... она движется на один шаг вверх, если монета выпадает гербом, и на один шаг вниз — в противном случае. Мы будем говорить, что *частица совершает симметрическое случайное блуждание* (физики рассматривают это блуждание как простейшую модель одномерной диффузии; см. гл. XIV).

В момент времени n частица расположена в точке s_n вертикальной оси. *Путь* $\{s_1, s_2, \ldots, s_N\}$ *есть пространственно-временной график случайного блуждания*, причем ось x играет роль оси времени.

Учитывая все сказанное, мы будем пользоваться следующей терминологией.

Терминология. *Будем говорить, что в момент времени n имеет место*:

Возвращение в начало координат, если $s_n = 0$.

Первое возвращение в начало координат, если

$$s_1 \neq 0,\ s_2 \neq 0,\ \ldots,\ s_{n-1} \neq 0,\ s_n = 0. \tag{3.1}$$

Первое достижение точки $r > 0$, если

$$s_1 < r,\ s_2 < r,\ \ldots,\ s_{n-1} < r,\ s_n = r. \tag{3.2}$$

Второе, третье и т. п. возвращения в начало координат и первое достижение точки $r < 0$ определяются аналогично. Заметим, что возвращение в начало координат может произойти только в *четные* моменты времени, поэтому соответствующие формулы будут часто относиться только к четным моментам времени. На игровом языке возвращение в начало координат означает, что после некоторого испытания *число гербов стало равно числу решеток*. (На рис. 3 представлены два пути, иллюстрирующие первое достижение и возвращение в начало координат. Особенностью второго пути является то, что он полностью расположен в отрицательной полуплоскости.)

§ 4. Новая формулировка комбинаторных теорем

В этой главе будут использоваться следующие обозначения:

$$u_{2n}=\binom{2n}{n}2^{-2n}, \qquad n=0,\ 1,\ 2,\ \ldots \tag{4.1}$$

и

$$f_0=0,\ f_{2n}=\frac{1}{2n}u_{2n-2}, \qquad n=1,\ 2,\ \ldots \tag{4.2}$$

Нетрудно проверить, что

$$f_{2n}=u_{2n-2}-u_{2n}, \qquad n=1,\ 2,\ \ldots . \tag{4.3}$$

Теорема 1. *Для любого* $n\geqslant 1$

$$u_{2n}=P\{s_{2n}=0\}, \tag{4.4}$$

$$u_{2n}=P\{s_1\neq 0,\ s_2\neq 0,\ \ldots,\ s_{2n}\neq 0\}, \tag{4.5}$$

$$u_{2n}=P\{s_1\geqslant 0,\ s_2\geqslant 0,\ \ldots,\ s_{2n}\geqslant 0\}, \tag{4.6}$$

или словами: следующие три события имеют одну и ту же вероятность u_{2n}: (а) *в момент* $2n$ *имеет место возвращение в начало координат,* (б) *до момента* $2n$ (*включительно*) *частица ни разу не возвратилась в начало координат,* (в) *путь неотрицателен в интервале времени между* 0 *и* $2n$.

Далее,

$$f_{2n}=P\{s_1\neq 0,\ \ldots,\ s_{2n-1}\neq 0,\ s_{2n}=0\}, \tag{4.7}$$

$$f_{2n}=P\{s_1\geqslant 0,\ \ldots,\ s_{2n-2}\geqslant 0,\ s_{2n-1}<0\}, \tag{4.8}$$

т. е. два события: (а) *первое возвращение в начало координат произошло в момент* $2n$ и (б) *частица впервые достигла точку* -1 *в момент* $2n-1$, *имеют одну и ту же вероятность* f_{2n}.

Доказательство. Как было установлено в § 3, достаточно исследовать вероятностное пространство, состоящее из путей фиксированной длины $2n$. Согласно (1.3) существуют $\binom{2n}{n}$ путей, соединяющих начало координат с точкой $(2n,\ 0)$, и это доказывает (4.4). По теореме 2 (а) § 2 существует L_{2n-2} путей, соединяющих точки $(0,\ 0)$ и $(2n,\ 0)$, таких, что $s_1>0,\ \ldots,\ s_{2n-1}>0$. Поэтому существует в два раза больше путей удовлетворяющих условиям (4.7), и соответствующая вероятность равна $2L_{2n-2}2^{-2n}=f_{2n}$. Теорема 2 (б) § 2 совершенно аналогично доказывает (4.8).

Вероятность того, что в течение рассматриваемого промежутка времени частица ни разу не возвратится в начало координат, равна единице минус вероятность того, что первое возвращение в начало имеет место не позже момента $2n$. Используя (4.7), эту разность можно переписать следующим образом

$$1-f_2-f_4-\ldots-f_{2n}=$$
$$=1-(1-u_2)-(u_2-u_4)-\ldots-(u_{2n-2}-u_{2n})=u_{2n}, \quad (4.9)$$

что доказывает (4.5). Аналогично правая часть в (4.6) равна единице минус вероятность того, что частица впервые достигнет точку -1 не позже момента $2n$, и используя (4.8), эту разность можно опять представить в виде (4.9). Это завершает доказательство теоремы.

Следствие. *При* $n \geqslant 1$

$$u_{2n}=\sum_{r=1}^{n} f_{2r}u_{2n-2r}. \quad (4.10)$$

Доказательство. Если частица возвращается в начало координат в момент $2n$, то первое возвращение должно произойти в некоторый момент $2r \leqslant 2n$. Мы только что установили, что число путей, соединяющих точки (0, 0) и ($2n$, 0), для которых в момент $2r$ имеет место первое возвращение в начало координат, равно $2^{2r}f_{2r}2^{2n-2r}u_{2n-2r}$. Суммируя по r, приходим к уравнению (4.10). (Прямое аналитическое доказательство можно найти в § 8 (а). В гл. XIII, § 3, мы увидим, что соотношение (4.10) является специальным частным случаем основного уравнения теории рекуррентных событий.)

Теорема 1* § 2 относится к путям, впервые достигающим точку y в момент x. Сумма $x+y$ должна быть четной, и поэтому удобно положить $x=2n-y$. Тогда теорему 1* можно переформулировать следующим образом.

Теорема 2. *Вероятность того, что точка* $y>0$ *впервые достигается в момент* $2n-y$, *равна*

$$f_{2n}^{(y)}=\frac{y}{2n-y}\binom{2n-y}{n}2^{-2n+y}, \quad n \geqslant y > 0. \quad (4.11)$$

Удивительна простота, с которой принцип двойственности позволяет получить эту важную теорему как прямое следствие теоремы о баллотировке. Прямое аналитическое доказательство соотношения (4.11) труднее и требует применения искусственных приемов.

В принципе, вероятности $f_{2n}^{(y)}$ можно было бы вычислять индукцией по y. Путь, для которого в момент времени $2n-y-1$ имеет место первое достижение точки $y+1$, можно разбить на две части (см. рис. 3 для $y=4$). Первый участок есть путь из начала координат до точки первого достижения y, которое происходит в некоторый момент $2\nu-y<2n-y-1$. За этим участком следует второй участок длины $2n-2\nu-1$, который весь, за исключением последнего отрезка, расположен ниже уровня y. Другими словами, если левый конец этого участка принять за начало отсчета, то правый конец будет точкой первого достижения 1. По определению, существует $2^{2\nu-y}f_{2\nu}^{(y)}$ возможных путей первого типа и $2^{2n-2\nu-1}f_{2n-2\nu}^{(1)}$ — второго, и комбинация любых двух из них дает путь, для которого первое достижение точки $y+1$ происходит в момент времени $2n-y-1$. Поэтому

$$f_{2n}^{(y+1)}=\sum_{\nu=y}^{n-1} f_{2\nu}^{(y)} f_{2n-2\nu}^{(1)}, \quad n \geqslant y+1. \tag{4.12}$$

Формула (4.8) показывает, что вероятность достигнуть точку -1 (а значит, и $+1$) впервые в момент времени $2n-1$ равна f_{2n}, т. е.

$$f_{2n}^{(1)}=f_{2n}, \quad n \geqslant 1. \tag{4.13}$$

Соотношения (4.12) и (4.13) определяют рекуррентно все $f_{2n}^{(y)}$, но совсем не просто проверить, что из формулы (4.11) следует (4.12), и наоборот.

Формулы (4.12) — (4.13) позволяют сделать новые заключения. Из (4.13) видно, что $f_{2n}^{(1)}$ есть вероятность впервые возвратиться в начало координат в момент $2n$. Забывая на время о предыдущей теореме, определим $f_{2r}^{(y)}$ как вероятность того, что в момент $2n$ имеет место y-е возвращение в начало координат. Воспользуемся идеей предыдущего доказательства. Разбивая путь, для которого в момент времени $2n$ имеет место $y+1$-е возвращение в начало координат, на две части: первый участок — до точки y-го возвращения в начало координат и второй — между точками y-го и $y+1$-го возвращения, легко установить, что соотношение (4.12) выполнено.

Так как это соотношение однозначно определяет все $f_{2n}^{(y)}$, мы приходим к следующей теореме.

Теорема 3. *Вероятность того, что в момент времени $2n$ имеет место y-е возвращение в начало координат, определяется формулой* (4.11).

Приведем еще одно, *чисто геометрическое доказательство* этой теоремы. Рассмотрим путь, для которого в момент времени $2n-y$ впрвые достигается точка y (на рис. 3 представлен случай $y=5$, $2n-y=15$). Построим новый путь, вставив в предыдущий путь y новые отрезки с наклоном -1 так, чтобы их левые концы совпали соответственно с началом координат и с точками первого достижения состояний $1, 2, \ldots$ $\ldots, y-1$. Этот путь, скажем, $\{\sigma_1, \sigma_2, \ldots, \sigma_{2n}\}$, состоит из $2n$ отрезков. Ясно, что $\sigma_1 \leqslant 0, \ldots, \sigma_{2n-1} \leqslant 0$, $\sigma_{2n}=0$ и ровно $y-1$ внутренних вершин лежат на оси x. Обратно, любой путь $\{\sigma_1, \sigma_2, \ldots, \sigma_{2n}\}$, удовлетворяющий этим условиям, можно получить с помощью описанной конструкции из пути, впервые достигающего точку y в момент $2n-y$. Если $f_{2n}^{(y)}$ определено, как в теореме 2, то существует ровно $2^{2n-y} f_{2n}^{(y)}$ путей $\{\sigma_1, \sigma_2, \ldots$ $\ldots, \sigma_{2n}\}$, таких, что $\sigma_i \leqslant 0$, $\sigma_{2n}=0$ и $y-1$ внутренних вершин лежат на оси x. Каждый такой путь можно разбить на y частей так, чтобы концы каждого участка лежали на оси x. Изменяя знак σ_i на одном или нескольких таких участках, можно получить 2^y различных путей, у которых $y-1$ внутренних вершин лежат на оси x и $s_{2n}=0$. Поэтому общее число таких путей равно $2^{2n} f_{2n}^{(y)}$, что и доказывает теорему.

§ 5. Первый закон арксинуса

Мы будем говорить, что *частица проводит время от $k-1$ до k на положительной стороне, если k-й отрезок ее пути лежит выше оси x*, т. е. если по крайней мере одна из вершин

s_{k-1} и s_k положительна (в этом случае другая вершина положительна или 0). В терминах игры в «орлянку» это означает, что как после $(k-1)$-го, так и после k-го испытаний суммарный выигрыш Петра неотрицателен.

Парадоксальные свойства путей, о которых мы упоминали в § 1, будут получены из следующего предложения:

Теорема 1[1]). *Пусть* $p_{2k,\,2n}$ — *вероятность того, что в интервале времени от* 0 *до* $2n$ *частица проводит* $2k$ *единиц времени на положительной стороне и* $2n-2k$ — *на отрицательной стороне. Тогда*

$$p_{2k,\,2n}=u_{2k}u_{2n-2k}. \tag{5.1}$$

(Заметим, что полное время, проведенное на положительной стороне, обязательно четно.)

Доказательство. Вероятность того, что частица остается на положительной стороне в течение всего времени от 0 до $2n$, дается формулой (4.6), и мы видим, что $p_{2n,\,2n}=u_{2n}$ в согласии с (5.1). Из соображений симметрии следует также, что $p_{0,\,2n}=u_{2n}$, поэтому остается проверить (5.1) только при $1\leqslant k\leqslant n-1$. С этой целью мы повторим рассуждения, которые привели нас к (2.7). Частица, которая проводит время $2k>0$ на положительной стороне и $2n-2k>0$ — на отрицательной стороне обязательно пройдет через нуль. Пусть $2r$ — момент первого возвращения в нуль. Тогда путь принадлежит к одному из следующих двух классов.

Первый класс характеризуется тем, что частица до момента $2r$ остается на положительной стороне, а в течение промежутка времени от $2r$ до $2n$ проводит на положительной стороне ровно $2k-2r>0$ единиц времени. Существует $2^{2r}f_{2r}$ путей длины $2r$, впервые возвращающихся в начало отсчета в момент $2r$, из них половина лежит в верхней полуплоскости. Кроме того, по определению, существует $2_{2n-2r}p_{2k-2r,\,2n-2r}$ путей длины $2n-2r$, выходящих из точки $(2r,\,0)$ и имеющих $2k-2r$ сторон выше оси x. Поэтому полное число путей первого класса равно

$$\frac{1}{2}\cdot 2^{2r}f_{2r}2^{2n-2r}p_{2k-2r,\,2n-2r}=2^{2n-1}f_{2r}p_{2k-2r,\,2n-2r}.$$

Путь принадлежит ко второму классу, если частица до момента $2r$ находится на отрицательной стороне, а в интервале времени от $2r$ до $2n$ проводит $2k$ единиц времени на положительной стороне.

---

[1]) Впервые эта теорема была доказана с помощью запутанных аналитических рассуждений в работе K. L. Chung and W. Feller (см. примечание на стр. 87 и первое издание настоящей книги, гл. XII, § 5 и § 6). Теорема была навеяна работой E. Sparre Andersen (см. примечание на стр. 80).

Здесь $2k \leqslant 2n-2r$ и рассуждения, аналогичные только что проведенным, показывают, что число путей этого класса равно $2^{2n-1} f_{2r} p_{2k,\,2n-2r}$.

Отсюда следует, что при $1 \leqslant k \leqslant n-1$

$$p_{2k,\,2n} = \frac{1}{2}\sum_{r=1}^{k} f_{2r} p_{2k-2r,\,2n-2r} + \frac{1}{2}\sum_{r=1}^{n-k} f_{2r} p_{2k,\,2n-2r}. \tag{5.2}$$

Предположим теперь по индукции, что соотношение $p_{2k,\,2\nu} = u_{2k} u_{2\nu-2k}$ имеет место для $\nu = 1, 2, \ldots, n-1$ (для $\nu = 1$ оно тривиально). Тогда формулу (5.2) можно привести к виду

$$p_{2k,\,2n} = \frac{1}{2} u_{2n-2k} \sum_{r=1}^{k} f_{2r} u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r}. \tag{5.3}$$

В силу тождества (4.10) первая сумма равна u_{2k}, а вторая — u_{2n-2k}, и поэтому выполняется соотношение (5.1).

Интуиция подсказывает нам, что *доля времени k/n, которое частица проводит на верхней стороне*, должна быть близка к $\frac{1}{2}$. Однако верно обратное: *наименее вероятно, что величина отношения k/n близка к $\frac{1}{2}$ и, наоборот, крайние значения $k/n = 0$ и $k/n = 1$ имеют наибольшую вероятность*. Это свойство можно проверить, опираясь на соотношение (5.1).

Таблица 1 иллюстрирует парадокс. В терминах игры в «орлянку» она устанавливает удивительный факт: при $2n = 20$ бросаниях правильной монеты *с вероятностью* 0,3524 *менее удачливый игрок никогда не будет впереди*. В большинстве случаев (с вероятностью 0,5379) суммарный выигрыш менее счастливого игрока будет положительным не более одного раза. Для сравнения заметим, что вероятность события $\frac{k}{n} = 10:10$ равна всего 0,0606.

Таблица 1

Распределение лидерства при 20 бросаниях монеты

	$k=0$ $k=20$	$k=2$ $k=18$	$k=4$ $k=16$	$k=6$ $k=14$	$k=8$ $k=12$	$k=10$
$p_{k,\,20} =$	0,1762	0,0927	0,0736	0,0655	0,0617	0,0606
$P_{k,\,20} =$	0,3524	0,5379	0,6851	0,8160	0,9394	1

$p_{k,\,20} = u_k u_{20}$ есть вероятность того, что k отрезков пути лежат выше оси, т. е. «Петр был впереди после k из 20 испытаний».

$P_{k,\,20}$ есть вероятность того, что один из игроков лидировал не менее чем после k испытаний, а другой — не более чем после $20-k$ испытаний.

Хотя формула (5.1) и является точной, но она не очень удобна и предпочитают пользоваться простым асимптотическим выражением. С помощью формулы Стирлинга легко получить, что $u_{2n}(\pi n)^{1/2} \to 1$ при $n \to \infty$. (Это составляет содержание задачи (12.20) гл. II.) Отсюда вытекает, что

$$p_{2k,2n} \sim \frac{1}{\pi k^{1/2}(n-k)^{1/2}}, \tag{5.4}$$

где отношение обеих частей стремится к единице, если $k \to \infty$ и $n-k \to \infty$. Вероятность того, что доля времени k/n, проведенного на положительной стороне, лежит между $\frac{1}{2}$ и $\alpha\left(\frac{1}{2} < \alpha < 1\right)$, равна

$$\sum_{\frac{1}{2}n<k<\alpha n} p_{2k,2n} \sim \frac{1}{\pi n} \sum_{\frac{1}{2}n<k<\alpha n} \left\{\frac{k}{n}\left(1-\frac{k}{n}\right)\right\}^{-\frac{1}{2}}. \tag{5.5}$$

Справа стоит сумма Римана, аппроксимирующая интеграл

$$\pi^{-1}\int_{1/2}^{\alpha} \frac{dx}{\{x(1-x)\}^{1/2}} = 2\pi^{-1}\arcsin\alpha^{1/2} - \frac{1}{2}. \tag{5.6}$$

Из соображений симметрии вероятность того, что $k/n \leqslant \frac{1}{2}$, стремится к $\frac{1}{2}$ при $n \to \infty$. Прибавляя эту вероятность к обеим частям соотношения (5.5), приходим к следующей теореме.

Теорема 2[1]. (Первый закон арксинуса.) *При фиксированном α $(0<\alpha<1)$ и $n \to \infty$ вероятность того, что доля k/n времени, проведенного частицей на положительной стороне, меньше α, стремится к*

$$\pi^{-1}\int_{0}^{\alpha} \frac{dx}{\{x(1-x)\}^{1/2}} = 2\pi^{-1}\arcsin\alpha^{1/2}. \tag{5.7}$$

На практике формула (5.7) обеспечивает превосходное приближение уже при таких значениях n, как $n=20$. Подинтегральное

---

[1]) P. Lévy, Sur certains processus stochastiques homogènes, *Compositio Mathematica*, 7 (1939), 283—339, открыл закон арксинуса для некоторых непрерывных диффузионных процессов и указал на связь с игрой в орлянку. Общий закон арксинуса для числа положительных частных сумм в последовательности независимых случайных величин был доказан P. Erdös and M. Kac, On the number of positive sums of independent random variables, *Bulletin of the American Mathematical Society*, 53 (1947), 1011—1020. E. Sparre Andersen открыл комбинаторную природу закона арксинуса и возможность его приложения к общим классам случайных величин.

выражение в (5.7) графически представляет собой U-образную кривую, которая уходит в бесконечность в точках 0 и 1. Это убедительнейшим образом показывает, что доля времени, проведенного частицей на положительной стороне, значительно чаще оказывается близка к 0 и 1, чем к «нормальному» или «ожидаемому» значению $^1/_2$. Рис. 4 иллюстрирует сказанное.

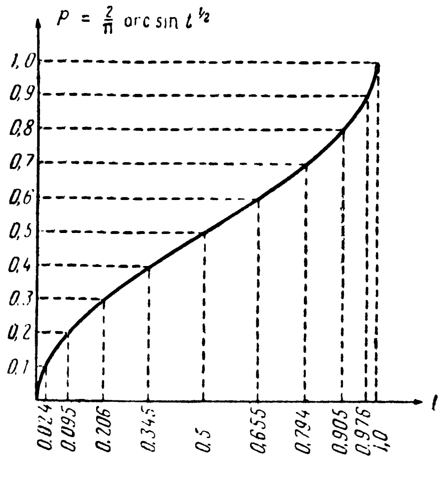


Рис. 4. Закон арксинуса.

Таблица 2

Иллюстрация к закону арксинуса

p	t_p
0,9	153,95 дня
0,8	126,10 дня
0,7	99,65 дня
0,6	75,23 дня
0,5	53,45 дня
0,4	34,85 дня
0,3	19,89 дня
0,2	8,93 дня
0,1	2,24 дня
0,05	13,5 часа
0,02	2,16 часа
0,01	32,4 мин.

Монету подбрасывают через каждую секунду на протяжении 365 дней. Пусть Z означает часть времени, в течение которого менее удачливый игрок был впереди. Тогда t_p есть такое число, что вероятность события $Z < t_p$ равна приблизительно p.

С вероятностью 0,20 *частица проводит не менее* 97,6% *времени на одной стороне. В одном из десяти случаев частица проводит на одной стороне не менее* 99,4% *всего времени.* Табл. 2 содержит данные другого рода.

В таблице приведены вероятности p того, что менее счастливый игрок будет находиться в выигрыше не более чем t_p дней в году. Взяв, к примеру, любимый статистиками уровень значимости $p = 0,05$, мы видим, что *в одном из двадцати случаев более удачливый игрок будет в выигрыше не менее 364 дней и 10 часов в году.* Мало кто поверит, что бросания правильной монеты могут привести к такой последовательности исходов, когда один из игроков все время находится в выигрыше, однако и это вполне реально.

В следующем параграфе мы обсудим другой аспект этого явления, а в § 7 приведем экспериментальные данные, иллюстрирующие теорию.

§ 6. Число возвращений в начало координат

Объяснение закона арксинуса состоит в том, что проходит слишком много времени, прежде чем частица возвращается в начало координат. На языке геометрии, пути пересекают ось x очень редко.

Интуиция подсказывает нам, что если Петр и Павел производят бросания правильной монеты достаточно долгое время $2n$, то *число ничьих* (моментов времени, когда суммарный выигрыш один и тот же у обоих игроков) должно быть *приблизительно пропорционально* $2n$. *Но это не так.* В действительности число ничьих возрастает по вероятности лишь как $(2n)^{1/2}$, т. е. с увеличением продолжительности игры относительное число ничьих быстро убывает, а «волны» возрастают по длине. Для анализа этого вопроса надо исследовать число возвращений в начало координат. Следует уяснить, что число моментов времени, когда частица переходит с положительной стороны на отрицательную, или наоборот, равно приближенно *половине* числа возвращений.

Теорема 1. *Пусть $z_{2n}^{(r)}$ — вероятность того, что за время $2n$ частица ровно r раз возвратится в начало координат. Тогда*

$$z_{2n}^{(r)} = \frac{1}{2^{2n-r}}\binom{2n-r}{n}, \qquad n \geqslant 1. \tag{6.1}$$

В частности, $z_{2n}^{(0)} = z_{2n}^{(1)} = u_{2n}$ *и*

$$z_{2n}^{(0)} = z_{2n}^{(1)} > z_{2n}^{(2)} > z_{2n}^{(3)} \ldots . \tag{6.2}$$

Другими словами, формула (6.2) устанавливает, что *независимо от продолжительности $2n$ игры более вероятно или отсутствие возвращений, или ровно одно возвращение, чем любое другое их число.*

Доказательство. Согласно формулам (4.4) и (4.5) существует столько же путей длины 2ν, возвращающихся в момент 2ν в начало координат, сколько существует путей длины 2ν, вообще не возвращающихся в нуль. Рассмотрим теперь путь длины $2n$, для которого r-е и последнее возвращение в начало происходит в некоторый момент $2n-2\nu < 2n$. Участок длины 2ν, начинающийся в точке последнего возвращения, можно выбрать столькими способами, сколькими можно выбрать путь постоянного знака, соединяющий точки $(2n-2\nu, 0)$ и $(2n, 0)$. Другими словами, *вероятность того, что до момента $2n$ произойдет ровно r возвращений, равна вероятности того, что в момент $2n$ произойдет возвращение и что ему будет предшествовать по меньшей мере r возвращений.* По теореме 3 § 4 это означает, что

$$z_{2n}^{(r)} = f_{2n}^{(r)} + f_{2n}^{(r+1)} + f_{2n}^{(r+2)} + \ldots , \tag{6.3}$$

где $f_{2n}^{(y)}$ определяется формулой (4.11). Легко проверить, что

$$f_{2n}^{(y)} = \frac{1}{2^{2n-y}}\binom{2n-y}{n} - \frac{1}{2^{2n-y-1}}\binom{2n-y-1}{n}, \tag{6.4}$$

и суммируя эти соотношения для $y = r, r+1, \ldots$, мы приходим к формуле (6.1). Неравенство (6.2) тривиально вытекает из (6.1), теорема доказана.

Снова желательно заменить точную формулу (6.1) более простым приближенным выражением. С этой целью перепишем (6.1) следующим образом:

$$z_{2n}^{(r)} = u_{2n}\frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)}{\left(1-\frac{1}{2n}\right)\left(1-\frac{2}{2n}\right)\ldots\left(1-\frac{k-1}{2n}\right)}. \tag{6.5}$$

При доказательстве закона арксинуса мы отмечали, что $u_{2n}(\pi n)^{1/2} \to 1$, если $n \to \infty$. По формуле Тейлора для $\log(1+\alpha)$, (8.10) гл. II, $\log(1-\nu/n) \sim -\nu/n$ с ошибкой порядка $(\nu/n)^2$. Отсюда следует, что с ошибкой порядка r^3/n^2 имеет место следующее приближенное равенство:

$$\log\left\{z_{2n}^{(r)}\pi^{1/2}n^{1/2}\right\} \approx -\frac{1}{2n}\sum_{\nu=1}^{r-1}\nu \approx -\frac{r^2}{4n}, \tag{6.6}$$

или

$$z_{2n}^{(r)} \approx \pi^{-1/2}n^{-1/2}e^{-r^2/4n}. \tag{6.7}$$

Вероятность того, что произойдет не более k возвращений, а именно $z_{2n}^{(0)} + z_{2n}^{(1)} + \ldots + z_{2n}^{(k-1)}$, можно оценить, используя формулу (6.7). Полученная сумма является суммой Римана для интеграла от $\pi^{-1}e^{-x^2/4}$, распространенного от 0 до k/n, относительная ошибка при этом будет величиной порядка k^3/n^2. Окончательно приходим к следующей теореме.

Теорема 2. *При фиксированном $\alpha > 0$ вероятность того, что в течение интервала времени от 0 до $2n$ частица возвратится в начало координат не более $\alpha(2n)^{1/2}$ раз, стремится при $n \to \infty$ к следующей величине*[1]:

$$f(\alpha) = \left(\frac{2}{\pi}\right)^{1/2}\int_0^{\alpha} e^{-\frac{1}{2}s^2}\,ds. \tag{6.8}$$

В частности, вероятность того, что произойдет не более 0,6745 $(2n)^{1/2}$ *возвращений близка к* $1/2$.

В гл. VIII, § 2, читатель может найти таблицу нормальной функции распределения $\Phi(\alpha) = \frac{1}{2}\{1 + f(\alpha)\}$; из нее величину $f(\alpha)$ можно получать по формуле $f(\alpha) = 2\left\{\Phi\left(\alpha - \frac{1}{2}\right)\right\}$ при $\alpha > 0$.

Пусть монета бросается 10 000 раз. С вероятностью $\frac{1}{2}$ произойдет не более 68 возвращений в точку $x = 0$, из которых только половина соответствует действительному изменению лидерства. Другими словами, с вероятностью $\frac{1}{2}$ средняя длина «волны» между последовательными изменениями лидерства равна приблизительно 300. При 1 000 000 испытаний медиана числа возвращений возрастает лишь в 10 раз, и поэтому средняя длина «волны» будет близка к 3000. Чем продолжительней серия бросаний, тем реже возвращения в нуль и длиннее волны.

Вероятность того, что при 10 000 *бросаний правильной монеты лидерство не будет меняться*, равна приблизительно 0,0085, и с той же вероятностью при 1 000 000 бросаний число перемен лидерства будет не более 10.

---

[1]) Читателей, знакомых с центральной предельной теоремой, предупреждаем, что число возвращений имеет распределение, отличное от нормального. В (6.8) появляется *отраженное* нормальное распределение со средним значением $(2/\pi)^{\frac{1}{2}}$ и дисперсией $1 - 2/\pi$.

§ 7. Экспериментальные данные

Рисунок 5 представляет собой результат эксперимента, отвечающего 10 000 бросаниям правильной монеты, соответствующие числовые данные приведены в примере 6, в гл. I. Верхняя часть графика соответствует первым 550 бросаниям, на двух других рисунках изображены результаты всех 10 000 испытаний в меньшем масштабе вдоль оси x. Масштаб вдоль оси y один и тот же у обоих графиков.

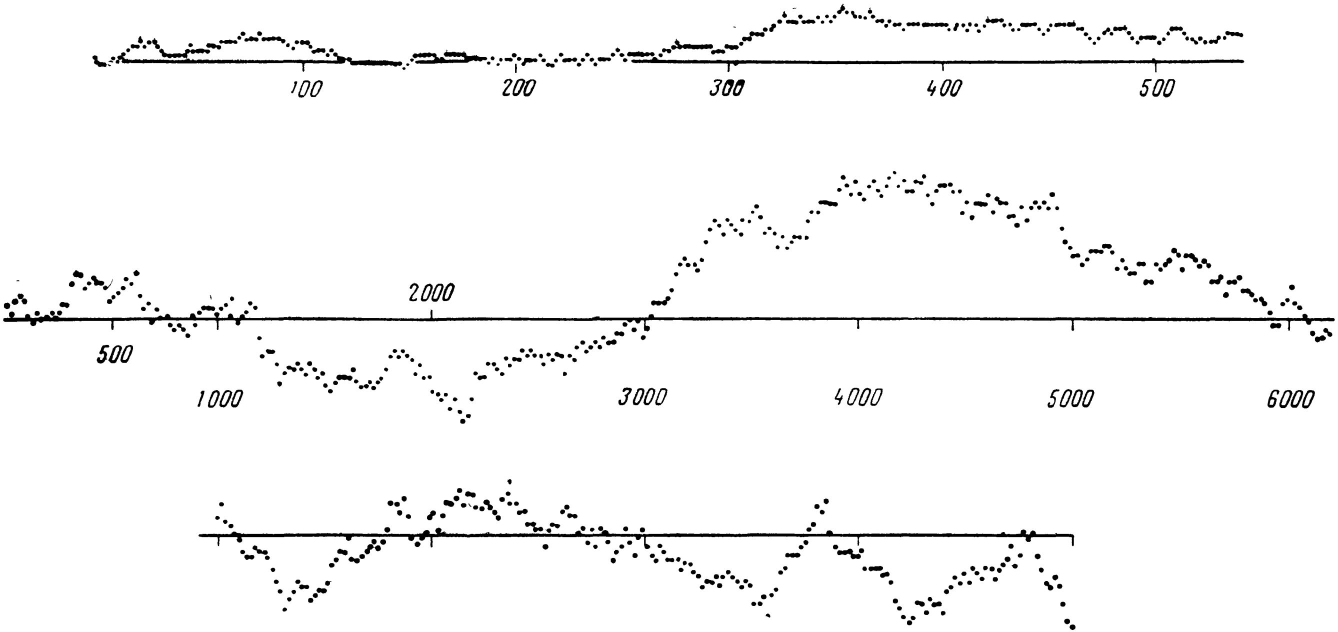
Взглянув на график, многие бывают удивлены длинами волн между последовательными пересечениями оси x (т. е. последовательными изменениями лидерства). Однако график представляет собой сравнительно умеренный случай. Он был выбран как самый умеренный из трех графиков, имевшихся в нашем распоряжении. Читателю рекомендуется взглянуть на тот же график в обратном направлении, т. е. принять его конец за начальную точку. [Аналитически обращенный путь определяется формулами (2.2).] Теоретически последовательности испытаний, отвечающие графику и его обращению, эквивалентны, и каждая из них представляет случайное блуждание. Обращенное случайное блуждание в нашем случае имеет следующие характеристики. Отправляясь из начала координат

«частица» остается на

отрицательной стороне	*положительной стороне*
первые 7804 *шага*	*следующие* 8 *шагов*
следующие 2 *шага*	*следующие* 54 *шага*
следующие 30 *шагов*	*следующие* 2 *шага*
следующие 48 *шагов*	*следующие* 6 *шагов*
следующие 2 046 *шагов*	
в целом 9930 *шагов*	*в целом* 70 *шагов*
доля времени: 0,9930	*доля времени* 0,007

Это *кажется* абсурдным, но тем не менее вероятность того, что при 10 000 бросаний одна из сторон находится в выигрыше более чем при 9930 испытаниях, а другая — менее чем при 70, несколько больше 0,1. Другими словами, в среднем *более чем один эксперимент из десяти приведет к результатам, которые «хуже» только что описанных*. Заметим для сравнения, что вероятность получить «лучшее» соотношение времени лидерства, чем на рис. 5, меньше (около 0,072).

Путь, изображенный на рис. 5, содержит 142 возвращения в начало, из которых 78 представляют действительное изменение лидерства. Обращенная серия, описанная выше, содержит 14 возвращений, из которых 8 отвечают изменению лидерства. Опрос специалистов показал, что даже опытные статистики считают 142


100
200
300
400
500
2000
500
1000
3000
4000
5000
6000

положения равновесия при 10 000 бросаний неожиданно малым числом, а 14 вообще выходящим за всякие рамки возможного. В действительности вероятность более чем 140 возвращений в начало равна приблизительно 0,157, в то время как вероятность не более 14 возвращений составляет примерно 0,115. Итак, вопреки интуиции, получение лишь 14 положений равновесия вовсе не неожиданно, с точки зрения числа изменений лидерства обращенная серия является парной с начальной серией рис. 5.

§ 8. Различные дополнения

а) **Аналитическое доказательство некоторых тождеств.** Нетрудно проверить, что

$$u_{2n}=(-1)^n\binom{-\frac{1}{2}}{n}, \quad f_{2n}=(-1)^{n+1}\binom{\frac{1}{2}}{n}. \tag{8.1}$$

Основное тождество (4.10) можно теперь рассматривать как частный случай уравнения (12.9) гл. II при $a=\frac{1}{2}$, $b=-\frac{1}{2}$. Та же самая формула показывает наряду с этим, что $\sum\limits_{r=0}^{n} u_{2r}u_{2n-2r}=1$.

Формулу (2.8) можно переписать в терминах f_{2k}, после чего она сводится к частному случаю соотношения (12.9) гл. II при $a=b=\frac{1}{2}$. Кроме того, формулу (2.8) можно вывести из (4.10), используя очевидное тождество $nr^{-1}(n-r)^{-1}=r^{-1}+(n-r)^{-1}$.

б) **Положение максимумов: второй закон арксинуса.** Условимся говорить, что *путь* $\{s_1, s_2, \ldots, s_x\}$ *имеет первый максимум в точке* k, *если*

$$s_k>0, \quad s_k>s_1, \ \ldots, \ s_k>s_{k-1}, \quad s_k\geqslant s_{k+1}, \ \ldots, \ s_k\geqslant s_x. \tag{8.2}$$

В частности, первый максимум находится в точке $x=0$, если $s_j\leqslant 0$ при $1\leqslant j\leqslant x$. Согласно формуле (4.6), вероятность того, что путь длины $x=2n$ имеет своим первым максимумом точку $x=0$, равна u_{2n}. Аналогичное предположение имеет место и для пути длиной $x=2n-1$.

Событие «первый максимум на последнем месте» означает то же самое, что и система неравенств $s_j<s_x$ при $j=0,1, \ldots, x-1$. Для обращенного пути (2.2) в этом случае получим $s_1^*>0$, $s_2^*>0, \ldots, s_x^*>0$, а вероятность этого определяется формулой (4.5). Она равна $\frac{1}{2}u_{2n}$ при $x=2n$ и также при $x=2n+1$.

Путь длины $2n$ с первым максимумом в k состоит из двух участков: начальный участок имеет первый максимум в последней точке $x=k$, а второй участок имеет первый максимум в начальной точке. Обратно, комбинируя два пути, обладающие такими свойствами, всегда можно получить путь с первым максимумом на k-м месте. Итак, доказана

Теорема. *Вероятность того, что путь длины $2n$ имеет первый максимум в точке $x=\nu$, равна*

$$\frac{1}{2}u_{2k}u_{2n-2k}, \quad \begin{array}{ll} \text{если} & \nu=2k \ (k=1, 2, \ldots, n), \\ \text{или} & \nu=2k+1 \ (k=0, 1, \ldots, n-1), \end{array} \tag{8.3}$$

и u_0, если $\nu=0$.

Замечательно, что *вероятность получить первый максимум в точках $2k$ или $2k+1$ равна вероятности $p_{2k,2n}$* [см. формулу (5.1)] *того, что частица проводит в верхней полуплоскости $2k$ из $2n$ единиц времени.* Отсюда вытекает, что применима приближенная формула закона арксинуса и, значит, существует *сильная тенденция к расположению максимумов вблизи начальной или конечной точек пути.*

Неожиданная близость между вероятностями распределения времени лидерства и положения первого максимума не является особенностью игры с бросанием монеты. Аналогичная теорема была доказана Андерсеном для широкого класса случайных величин, и комбинаторные основания его доказательства аналогичны аргументам, которыми мы только что пользовались.

в) **Предельная теорема для времени первого достижения и возвращения в начало координат**[1]. Оценки, которыми мы пользовались в § 6, можно использовать для того, чтобы показать, что вероятность $f_{2n}^{(y)}$ [см. (4.11)] удовлетворяет предельному соотношению

$$f_{2n}^{(y)} \sim \left(\frac{2}{\pi}\right)^{1/2} \frac{y}{(2n-y)^{3/2}} e^{-\frac{1/2 y^2}{2n-y}}, \tag{8.4}$$

где знак $\sim$ означает, что отношение двух величин, соединенных этим знаком, стремится к единице при $n\to\infty$. Методы, которые применялись при выводе предельных теорем § 5 и 6, позволяют теперь прийти к следующему заключению. *Вероятность того, что y-е возвращение в начало координат (или первое достижение y) имеет место до момента ty^2, стремится при возрастании y к $1-f(t^{-1/2})$, где $f(\alpha)$ определяется формулой* (6.8).

---

[1]) Для подготовленного читателя заметим, что формула $1-f(t^{-1/2})$ представляет так называемое положительное устойчивое распределение с параметром $\alpha=\frac{1}{2}$.

Отсюда следует, что *с вероятностью, близкой к* $^1/_2$ *y-е возвращение в нуль произойдет после испытания с номером* $(2.21\ldots)\,y^2$, так что *среднее время между последовательными возвращениями возрастает приблизительно линейно относительно y*. Заключение такого рода совершенно неожиданно для физиков, которые обычно рассматривают среднее значение y «измерений некоторой величины» как приближение к «истинному» значению. В нашем случае тщательный анализ показывает, что с большой вероятностью одно из y измерений будет иметь тот же порядок, что и величина всей суммы, именно y^2.

ГЛАВА IV[1])

КОМБИНАЦИИ СОБЫТИЙ

В этой главе будет идти речь о событиях A, определяемых при посредстве некоторых других событий $A_1, A_2, \ldots, A_N$. Так, при игре в бридж событие A «по крайней мере один из игроков имеет полную масть» является соединением четырех событий A_k «k-й игрок имеет полную масть» ($k=1, 2, 3, 4$). Одновременно могут произойти одно, два или более событий A_k; ввиду такого их взаимного наложения вероятность события A не равна сумме четырех вероятностей $P\{A_k\}$. Мы покажем, как зная множество событий $A_1, \ldots, A_N$, вычислить вероятности появления 0, 1, 2, 3, ... из них.

Материал, изложенный в этой главе, и разнообразные приложения содержатся в монографии М. Фреше[2]), к которой мы и отсылаем читателя за дальнейшими сведениями.

§ 1. Объединение событий

Если A_1 и A_2 — два события, то $A = A_1 \cup A_2$ обозначает событие, состоящее в появлении либо события A_1, либо события A_2, либо обоих этих событий. По формуле (7.4) гл. I имеем

$$P(A) = P(A_1) + P(A_2) - P(A_1 A_2). \tag{1.1}$$

Обобщим эту формулу на случай N событий $A_1, A_2, \ldots, A_N$, т. е. вычислим вероятность того, что произойдет по крайней мере одно из событий A_k. Символически это событие записывается в виде $A = A_1 \cup A_2 \cup \ldots \cup A_N$. Для нашей цели недостаточно знать вероятности отдельных событий A_k; мы должны быть также полностью осведомлены о том, как эти события могут перекрываться. Это значит, что для любой пары (i, j), любой тройки (i, j, k) и т. д. мы должны знать вероятность одновременного появления событий A_i

---

[1]) Материал этой главы в дальнейшем непосредственно использоваться не будет. Только первая теорема представляет принципиальный интерес.

[2]) M. Fréchet, Les probabilités associées à un système d'événements compatibles et dépendants. *Actualités scientifiques et industrielles*, № 859, 1940, № 942, 1943, Paris.

и A_j, событий A_i, A_j и A_k и т. д. Для удобства будем обозначать эти вероятности буквами p с соответствующими индексами. Итак,

$$p_i = P\{A_i\}; \quad p_{i,j} = P\{A_iA_j\}; \quad p_{i,j,k} = P\{A_iA_jA_k\}, \ldots . \tag{1.2}$$

Порядок индексов несуществен, но ради единообразия мы всегда будем *писать индексы в возрастающем порядке*; так, мы пишем $p_{3,7,11}$, а не $p_{7,3,11}$. Два индекса никогда не совпадают. Для суммы всех вероятностей p, имеющих r индексов, мы введем обозначение S_r, т. е. мы положим

$$S_1 = \sum p_i; \quad S_2 = \sum p_{i,j}; \quad S_3 = \sum p_{i,j,k}; \quad \ldots . \tag{1.3}$$

Здесь $i < j < k < \ldots \leqslant N$, так что каждая комбинация индексов появляется в суммах один и только один раз; поэтому S_r содержит $\binom{N}{r}$ членов. Последняя сумма S_N сводится к единственному члену $p_{1,2,3,\ldots,N}$, являющемуся вероятностью одновременного появления всех N событий. При $N=2$ имеются только две суммы S_1 и S_2, и формула (1.1) может быть записана в виде

$$P(A) = S_1 - S_2. \tag{1.4}$$

Обобщение на случай произвольного числа N событий дает следующая

Теорема. *Вероятность P_1 появления по крайней мере одного из событий $A_1, A_2, \ldots, A_N$ равна*

$$P_1 = S_1 - S_2 + S_3 - S_4 + - \ldots \pm S_N. \tag{1.5}$$

Доказательство. Докажем формулу (1.5) так называемым методом включения и исключения. Для вычисления P_1 нужно сложить вероятности всех точек пространства элементарных событий, содержащихся по крайней мере в одном из событий A_i; при этом каждую точку нужно брать по одному разу. Будем действовать систематически: возьмем сначала точки, входящие только в одно из событий A_i, затем точки, входящие ровно в два события из числа A_i, и т. д., наконец, точки, входящие во все A_i (если такие точки найдутся). Пусть E — какое-то элементарное событие, содержащееся в n из N событий A_i. Не нарушая общности, можно занумеровать события так, чтобы точка E содержалась в событиях $A_1, A_2, \ldots, A_n$ и не содержалась в событиях $A_{n+1}, A_{n+2}, \ldots, A_N$. Тогда $P\{E\}$ будет слагаемым в тех вероятностях p_i, p_{ij}, p_{ijk}, индексы которых меняются от 1 до n. Поэтому $P\{E\}$ войдет n раз в состав S_1, $\binom{n}{2}$ раз в состав S_2 и т. д. В итоге, выражая правую часть формулы (1.5) через вероятности элементарных событий, мы убедимся, что вероятность $P\{E\}$ входит в P_1 с коэффициентом

$$n - \binom{n}{2} + \binom{n}{3} \pm \ldots \pm \binom{n}{n}. \tag{1.6}$$

Для доказательства теоремы остается показать, что это число равно 1. Но в этом можно убедиться, сравнив формулу (1.6) с разложением бинома $(1-1)^n$ [см. формулу (8.7) гл. II]. Это разложение начинается с 1, затем в нем следуют члены (1.6) с обратными знаками. Поэтому для любого $n \geqslant 1$ выражение (1.6) равно 1, что и доказывает теорему.

Примеры. а) Обозначим через A_i событие, состоящее в том, что при игре в бридж «игрок с номером i имеет полную масть». Тогда $p_i = 4\Big/\binom{52}{13}$; событие, заключающееся в появлении полной масти у двух игроков с номерами i и j, может произойти $4 \cdot 3$ способами и имеет вероятность $p_{i,j} = 12\Big/\binom{52}{13}\binom{39}{13}$; аналогично $p_{i,j,k} = 24\Big/\binom{52}{13}\binom{39}{13}\binom{26}{13}$. Наконец, $p_{1,2,3,4} = p_{1,2,3}$, так как если три игрока имеют по полной масти, то и четвертый игрок имеет полную масть. Поэтому вероятность того, что какой-то игрок имеет полную масть, равна $P_1 = 4p_1 - 6p_{1,2} + 4p_{1,2,3} - p_{1,2,3,4}$. Используя формулу Стирлинга, получаем, что приближенно $P_1 = \frac{1}{4}10^{-10}$. В данном случае P_1 очень близко к сумме вероятностей событий A_i; но это скорее исключение, чем правило.

б) *Совпадения.* Следующая задача, имеющая много вариантов и получающая неожиданное решение, впервые сформулирована Монмортом (1708), а затем обобщалась Лапласом и многими другими авторами.

Имеются две одинаковые колоды карт, состоящие каждая из N различных карт. Карты каждой колоды располагаются в случайном порядке, после чего сравниваются расположения карт в колодах. Если какая-нибудь карта занимает одно и то же положение в обеих колодах, то мы говорим, что имеет место *совпадение*. Совпадения могут происходить на любом из N мест и на нескольких местах одновременно.

Такой схемой можно описать много шуточных задач. Например, две колоды можно представить как группу из N писем и их конвертов, причем своенравный секретарь устанавливает случайное соответствие. Или можно вообразить, что в гардеробе шляпы посетителей оказались перепутанными и распределены между ними случайно. Совпадение происходит в том случае, когда кто-либо получает свою собственную шляпу. Поучительно попытаться угадать, каким образом вероятность совпадения зависит от N: как относится вероятность получения хотя бы кем-нибудь из гостей своей шляпы при наличии восьми гостей к соответствующей вероятности в собрании из 10 000 человек? Кажется странным, что эта вероятность практически не зависит от N и близка к 2/3. (Более серьезные задачи, описываемые изложенной схемой, см. в задачах 10 и 11.)

Вероятности получить ровно 0, 1, 2, 3, ... совпадений будут вычислены в § 4. Здесь мы вычислим только вероятность хотя бы одного совпадения. Для удобства перенумеруем карты числами 1, 2, ..., N в том порядке, в каком они расположены в одной из колод. Вероятность любой перестановки карт второй колоды будем считать равной $1/N!$. Пусть A_k — событие, состоящее в том, что совпадение произошло на k-м месте. Это означает, что карта с номером k находится на k-м месте, тогда как оставшиеся $N-1$ карт могут быть расположены произвольно. Очевидно, $p_k=(N-1)!/N!=1/N$. Аналогично для любой комбинации i, j мы имеем $p_{i,j}=(N-2)!/N!=1/N(N-1)$ и т. д. Сумма S_r содержит $\binom{N}{r}$ членов, каждый из которых равен $(N-r)!/N!$. Отсюда $S_r=1/r!$, и по (1.5) получаем искомую вероятность в виде

$$P_1=1-\frac{1}{2!}+\frac{1}{3!}\pm\ldots\pm\frac{1}{N!}. \tag{1.7}$$

Заметим, что $1-P_1$ представляет собой сумму первых $N+1$ членов разложения

$$e^{-1}=1-1+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}\pm\ldots. \tag{1.8}$$

Поэтому с хорошим приближением получаем

$$P_1\approx 1-e^{-1}=0{,}63212\ldots. \tag{1.9}$$

Степень точности приближения видна из следующей таблицы истинных значений P_1:

$N=$	3	4	5	6	7
$P_1=$	0,66667	0,62500	0,63333	0,63196	0,63214

§ 2. Приложение к классической задаче о размещении

Возвратимся теперь к задаче о случайном размещении r шаров по n ящикам, считая, что каждое размещение имеет вероятность n^{-r}. Найдем вероятность $p_m(r, n)$ того, что m ящиков останутся пустыми[1]).

Пусть A_k — событие, состоящее в том, что ящик с номером k пуст ($k=1, 2, \ldots, n$). При событии A_k все r шаров расположены в остальных $n-1$ ящиках; это может быть сделано $(n-1)^r$ различными способами. Аналогично существуют $(n-2)^r$ размещений,

---

[1]) Эта вероятность была вычислена совсем другими способами в задаче (11.8) гл. II. См. также пример § 3.

оставляющих пустыми два фиксированных ящика, и т. д. Соответственно

$$p_i=\left(1-\frac{1}{n}\right)^r, \quad p_{ij}=\left(1-\frac{2}{n}\right)^r, \quad p_{ijk}=\left(1-\frac{3}{n}\right)^r, \ \ldots, \tag{2.1}$$

откуда для любого $\nu \leqslant n$

$$S_\nu=\binom{n}{\nu}\left(1-\frac{\nu}{n}\right)^r. \tag{2.2}$$

Вероятность того, что по крайней мере один ящик пуст, можно определить по формуле (1.5), и поэтому *вероятность того, что все ящики заняты*, равна $1-S_1+S_2 \pm \ldots$ или

$$p_0(r, n)=\sum_{\nu=0}^{n}(-1)^\nu\binom{n}{\nu}\left(1-\frac{\nu}{n}\right)^r. \tag{2.3}$$

Рассмотрим теперь размещение, при котором m ящиков оказались пустыми. Эти m ящиков могут быть выбраны $\binom{n}{m}$ способами. r шаров распределяются по оставшимся $n-m$ ящикам так, что каждый из этих ящиков занят; число таких распределений равно $(n-m)^r p_0(r, n-m)$. Деля на n^r, мы находим *вероятность того, что при случайном размещении m ящиков останутся пустыми*:

$$p_m(r, n)=\binom{n}{m}\left(1-\frac{m}{n}\right)^r p_0(r, n-m)=$$

$$=\binom{n}{m}\sum_{\nu=0}^{n-m}(-1)^\nu\binom{n-m}{\nu}\left(1-\frac{m+\nu}{n}\right)^r. \tag{2.4}$$

Мы уже пользовались схемой из r случайных цифр для иллюстрации случайного распределения r предметов по $n=10$ ящикам. Пустым ящикам в этом случае соответствуют отсутствующие цифры: если m ящиков пусты, то в данной последовательности встречается $10-m$ различных цифр. Численную иллюстрацию мы приводим в табл. 1.

Ясно, что непосредственное вычисление вероятностей (2.4) возможно лишь в случае относительно малых n и r. С другой стороны, задача о размещении представляет особый интерес именно при больших n. Каковы шансы найти свободный ящик, если 10 000 шаров размещаются в 1000 ящиков? Каковы шансы найти день в году, не являющийся днем рождения ни для одного лица из группы в 2000 человек? К счастью, на подобные вопросы можно ответить с помощью замечательно простого приближенного выражения для $P_{[m]}$, ошибка которого стремится к нулю при $n \to \infty$. Как это приближение, так и приводящие к нему рассуждения типичны для многих *предельных теорем* теории вероятностей.

Таблица 1

Вероятности $p_m(r, 10)$, полученные по формуле (2.4)

m	$r=10$	$r=18$
0	0,000363	0,134673
1	0,016330	0,385289
2	0,136080	0,342987
3	0,355622	0,119425
4	0,345144	0,016736
5	0,128596	0,000876
6	0,017189	0,000014
7	0,000672	0,000000
8	0,000005	0,000000
9	0,000000	0,000000

$p_m(r, 10)$ — вероятность того, что ровно m цифр 0,1, ..., 9 *не* появятся в последовательности из r случайных цифр.

Итак, наша цель — найти предельный вид формулы для $p_m(r, n)$ при $n \to \infty$ и $r \to \infty$. Соотношение между r и n, вообще говоря, произвольно, но дробь r/n представляет собой среднее число предметов на один ящик. Если оно чрезмерно велико, то мы не можем ожидать наличия свободных ящиков; в этом случае $p_0(r, n)$ близко к единице, а все $p_m(r, n)$ с $m \geqslant 1$ малы. С другой стороны, если дробь r/n стремится к нулю, то практически все ящики должны быть пустыми, и в этом случае при любом фиксированном m должно быть $p_m(r, n) \to 0$. Поэтому действительно интересен лишь промежуточный случай.

Сначала выведем приближенное выражение для S_ν. Так как $(n-\nu)^\nu < (n)_\nu < n^\nu$, то, очевидно, имеем

$$n^\nu \left(1 - \frac{\nu}{n}\right)^{\nu+r} < \nu! S_\nu < n^\nu \left(1 - \frac{\nu}{n}\right)^r. \tag{2.5}$$

Используя двойное неравенство (8.12) гл. II при $t = \nu/n$, мы получим

$$\{ne^{-(\nu+r)/(n-\nu)}\}^\nu < \nu! S_\nu < \{ne^{-r/n}\}^\nu. \tag{2.6}$$

Положим теперь для краткости

$$ne^{-r/n} = \lambda \tag{2.7}$$

и предположим, что n и r возрастают таким образом, что $\lambda = ne^{-r/n}$ остается ограниченным. Тогда для любого фиксированного ν

отношение крайних членов в (2.6) стремится к единице, и мы заключаем, что

$$S_\nu < \frac{\lambda^\nu}{\nu!} \text{ и } \frac{1}{\nu!}\lambda^\nu - S_\nu \to 0. \tag{2.8}$$

Отсюда вытекает, что

$$p_0(r, n) - \sum_{\nu=0}^{\infty} (-1)^\nu \frac{\lambda^\nu}{\nu!} \to 0 \tag{2.9}$$

или $p_0(r, n) - e^{-\lambda} \to 0$. Кроме того, множитель при $p_0(r, n-m)$ в (2.4) можно переписать просто как S_m, поэтому при любом фиксированном m

$$p_m(r, n) - e^{-\lambda}\frac{\lambda^m}{m!} \to 0. \tag{2.10}$$

Тем самым доказана следующая

Теорема[1]). *Если n и r возрастают таким образом, что $\lambda = ne^{-r/n}$ остается ограниченным, то при любом фиксированном m имеет место соотношение* (2.10).

Аппроксимирующее выражение

$$p(m; \lambda) = e^{-\lambda}\frac{\lambda^m}{m!} \tag{2 11}$$

определяет так называемое *распределение Пуассона*, которое имеет большое значение и описывает разнообразные явления. Оно будет подробно исследовано в гл. VI.

Практически $p(m, \lambda)$ можно использовать как приближенное выражение для $P_{[m]}$ лишь при большом n. При небольших значениях n нужна оценка ошибки. Мы, однако, не будем этим заниматься.

Примеры. а) В табл. 2 приведены приближенные значения вероятностей того, что m ящиков окажутся пустыми при общем числе ящиков, равном 1000, и числе предметов, меняющемся от 5000 до 9000. Для $r = 5000$ медиана[2]) числа пустых ящиков равна шести: шесть или более пустых ящиков примерно так же вероятны, как шесть или менее пустых ящиков. Даже при 9000 предметов

[1]) Заимствовано из книги: R. Mises, Über Aufteilungs- und Besetzungswahrscheinlichkeiten, *Revue de la Faculté des Sciences de l' Université d' Istanbul*, N. S., 4 (1939), 145—163. Доказательство изменено.

[2]) Если какая-то величина может принимать ряд значений $x_1 \leqslant x_2 \leqslant \leqslant x_3 \leqslant \ldots \leqslant x_n$ с вероятностями $p_1, p_2, \ldots, p_n$, то медианой этого ряда называется такое значение x_k, что

$$\sum_{i=1}^{k} p_i \geqslant {}^1/_2 \quad \text{и} \quad \sum_{i=k}^{n} p_i \geqslant {}^1/_2. \text{ — Прим. ред.}$$

Таблица 2

Пуассоновское приближение (2.11) для вероятностей (2.4) найти ровно m пустых ящиков, если r предметов случайно распределены по $n = 1000$ ящикам

r	λ	$p(m; \lambda)$											
		$m=0$	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	$m=11$
5000	6,74	0,0012	0,0080	0,0269	0,0604	0,1017	0,1371	0,1540	0,1482	0,1249	0,0935	0,0630	0,0386
5500	4,09	0,0167	0,0685	0,1400	0,1909	0,1951	0,1596	0,1088	0,0636	0,0325	0,0148	0,0060	0,0023
6000	2,48	0,0838	0,2077	0,2575	0,2128	0,1320	0,0655	0,0271	0,0096	0,0030	0,0008	0,0002	
6500	1,50	0,2231	0,3347	0,2510	0,1255	0,0471	0,0141	0,0035	0,0008	0,0001			
7000	0,91	0,4027	0,3661	0,1666	0,0506	0,0115	0,0021	0,0003					
7500	0,55	0,5777	0,3163	0,0873	0,0162	0,0023	0,0003						
8000	0,34	0,7126	0,2406	0,0414	0,0049	0,0004							
8500	0,20	0,8187	0,1637	0,0164	0,0011	0,0001							
9000	0,12	0,8869	0,1064	0,0064	0,0003								

в 1000 ящиках мы имеем один шанс из 9 за то, что найдется пустой ящик.

б) В статистике дней рождения $n=365$ и r равно числу людей. Для $r=1900$ мы приближенно находим $\lambda=2$. *Вероятности* $P_{[m]}$ *найти* m *дней в году, не являющихся днями рождения людей, проживающих в поселке с* 1900 *жителями, приближенно таковы*:

$$P_{[0]}=0{,}135;\ P_{[1]}=0{,}271;\ P_{[2]}=0{,}271;\ P_{[3]}=0{,}180;$$
$$P_{[4]}=0{,}090;\ P_{[5]}=0{,}036;\ P_{[6]}=0{,}012;\ P_{[7]}=0{,}003.$$

Вероятность найти m ящиков, содержащих каждый ровно k предметов, может быть выведена тем же путем. Как показал Мизес, эта вероятность также аппроксимируется выражением Пуассона (2.11), только на этот раз λ должно быть определено по формуле

$$\lambda=ne^{-r/n}\left(\frac{r}{n}\right)^k\Big/k! \tag{2.12}$$

§ 3. Осуществление m из N событий

Теорема § 1 может быть усилена следующим образом.

Теорема. *Для любого целого числа* m, *удовлетворяющего условию* $1\leqslant m\leqslant N$, *вероятность* $P_{[m]}$ *одновременного осуществления* m *из* N *событий* $A_1, \ldots, A_N$ *дается формулой*

$$P_{[m]}=S_m-\binom{m+1}{m}S_{m+1}+\binom{m+2}{m}S_{m+2}\pm\ldots\pm\binom{N}{m}S_N. \tag{3.1}$$

Замечание. Согласно (1.5), вероятность $P_{[0]}$ того, что ни одно из событий A_j не осуществится, равна

$$P_{[0]}=1-P_1=1-S_1+S_2\pm\ldots\mp S_N. \tag{3.2}$$

Это показывает, что формула (3.1) дает правильный результат и при $m=0$, если только положить $S_0=1$.

Доказательство. Будем рассуждать так же, как при доказательстве формулы (1.5). Пусть E — произвольное элементарное событие; предположим, что оно содержится в n из N событий A_j. Тогда $P\{E\}$ входит в состав $P_{[m]}$ только при $n=m$. Чтобы выяснить, сколько раз $P\{E\}$ входит в правую часть (3.1), заметим, что $P\{E\}$ входит в суммы $S_1, S_2, \ldots, S_n$ и не входит в суммы $S_{n+1}, \ldots, S_N$. Отсюда следует, что $P\{E\}$ не входит в правую часть (3.1) при $n<m$. При $n=m$ вероятность $P\{E\}$ появляется в одном и только одном члене суммы S_m. Для завершения доказательства теоремы остается показать, что при $n>m$ члены $P\{E\}$ в суммах $S_m, S_{m+1}, \ldots, S_n$ в правой части (3.1) взаимно уничтожаются. Для этого заметим, что из n событий, содержащих E, мы можем образовать $\binom{n}{k}$ групп по k событий; поэтому $P\{E\}$ входит в S_k с коэффициентом $\binom{n}{k}$. Отсюда вытекает, что при $n > m$ вероятность $P\{E\}$ входит в правую часть (3.1) с коэффициентом

$$\binom{n}{m} - \binom{m+1}{m}\binom{n}{m+1} + \binom{m+2}{m}\binom{n}{m+2} \pm \ldots . \tag{3.3}$$

Однако $\binom{m+\nu}{m}\binom{n}{m+\nu} = \binom{n}{m}\binom{n-m}{\nu}$, и поэтому выражение (3.3) сводится к

$$\binom{n}{m}\left\{\binom{n-m}{0} - \binom{n-m}{1} \mp \ldots \mp \binom{n-m}{n-m}\right\}. \tag{3.4}$$

В скобках мы имеем разложение бинома $(1-1)^{n-m}$, так что коэффициент (3.3), как и утверждалось, равен нулю.

Пример. Читателю рекомендуется проверить, что подстановка формулы (2.2) в (3.1) приводит непосредственно к формуле (2.4).

§ 4. Приложения к задачам о совпадениях и к задаче угадывания

В примере 1, б мы рассматривали соответствие двух колод карт и убедились, что $S_k = 1/k!$ Подставляя это значение в формулу (3.1), приходим к следующему результату.

При установлении случайного соответствия между двумя одинаковыми колодами из N различных карт вероятности $P_{[m]}$ ровно m совпадений даются формулами

$$\begin{aligned}
P_{[0]} &= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} \mp \ldots \mp \frac{1}{(N-2)!} \pm \frac{1}{(N-1)!} \mp \frac{1}{N!}. \\
P_{[1]} &= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} \mp \ldots \mp \frac{1}{(N-2)!} \pm \frac{1}{(N-1)!}; \\
P_{[2]} &= \frac{1}{2!}\left\{1 - 1 + \frac{1}{2!} - \frac{1}{3!} \mp \ldots \pm \frac{1}{(N-3)!} \mp \frac{1}{(N-2)!}\right\}; \\
P_{[3]} &= \frac{1}{3!}\left\{1 - 1 + \frac{1}{2!} - \frac{1}{3!} \mp \ldots \pm \frac{1}{(N-3)!}\right\}; \qquad (4.1)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
P_{[N-2]} &= \frac{1}{(N-2)!}\left\{1 - 1 + \frac{1}{2!}\right\}; \\
P_{[N-1]} &= \frac{1}{(N-1)!}\{1 - 1\} = 0; \\
P_{[N]} &= \frac{1}{N!}.
\end{aligned}$$

Последнее соотношение очевидно. Равенство нулю $P_{[N-1]}$ выражает невозможность получить $N-1$ совпадений, не имея одинакового порядка всех N карт.

Скобки в правых частях (4.1) содержат сумму первых членов разложения e^{-1}. *Следовательно, при больших N приближенно получаем*

$$P_{[m]} \approx \frac{1}{m!} e^{-1}. \tag{4.2}$$

В табл. 3 в столбцах, озаглавленных $P_{[m]}$, приведены точные значения $P_{[m]}$ для $N = 3, 4, 5, 6, 10$. Последний столбец дает предельные значения:

$$p_m = \frac{1}{m!} e^{-1}. \tag{4.3}$$

Следует отметить, что p_m является хорошим приближением к $P_{[m]}$ даже при не слишком больших значениях N.

Для чисел p_m, определенных по (4.3), имеем

$$\sum p_k = e^{-1}\left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots\right) = e^{-1}e = 1.$$

Вследствие этого p_k могут рассматриваться как вероятности.

Таблица 3

Вероятности правильных угадываний в колоде из N различных карт

	$N=3$		$N=4$		$N=5$		$N=6$		$N=10$		p_m
	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	$P_{[m]}$	b_m	
0	0,333	0,296	0,375	0,316	0,367	0,328	0,368	0,335	0,36788	0,34868	0,367879
1	0,500	0,444	0,333	0,422	0,375	0,410	0,367	0,402	0,36788	0,38742	0,367879
2	...	0,222	0,250	0,211	0,167	0,205	0,187	0,201	0,18394	0,19371	0,183940
3	0,167	0,037	...	0,047	0,083	0,051	0,056	0,053	0,06131	0,05740	0,061313
4			0,042	0,004	...	0,006	0,021	0,008	0,01534	0,01116	0,015328
5					0,008	0,000	...	0,001	0,00306	0,00149	0,003066
6							0,001	0,000	0,00052	0,00014	0,000511
7									0,00007	0,00001	0,000073
8									0,00001	...	0,000009
9									...	...	0,000001
10									...	...	0,000000

$P_{[m]}$ полученные по формуле (4.1), b_m — по формуле (4.4). В последнем столбце даны предельные вероятности (4.3) (распределение Пуассона).

Заметим, что (4.3) представляет *распределение Пуассона* (2.11) при частном значении параметра $\lambda = 1$.

Формулы (4.1) полезны при оценке *способности угадывать*. При дегустации вин, в некоторых физических опытах и т. п. от угадывающего требуется назвать неизвестный порядок N предметов, скажем, карт. Любое проявление проницательности с его стороны немедленно сказывается в отклонениях от случайности. Для оценки степени этой проницательности мы должны уметь оценивать вероятность случайной удачи. Угадывающий может придерживаться разных систем, среди которых мы отметим три крайние возможности: 1) Он называет все время одну и ту же карту. При этой системе он может быть уверен, что у него будет одна и только одна верная догадка; случайные колебания исключены. 2) Угадывающий называет каждую карту по одному разу; если считать все возможные порядки равновероятными, то применимы формулы (4.1). 3) Третья возможность заключается в том, что N догадок делаются совершенно независимо друг от друга. Здесь имеется N^N возможных комбинаций. Несомненно, что каждое лицо склонно называть некоторые комбинации чаще, чем другие; но в первом приближении мы можем принять все N^N комбинаций равновероятными. Поскольку m верных и $N-m$ неверных догадок могут произойти $\binom{N}{m}(N-1)^{N-m}$ различными способами, вероятность получить m правильных догадок равна

$$b_m=\binom{N}{m}\frac{(N-1)^{N-m}}{N^N}. \tag{4.4}$$

[Это частный случай биномиального распределения. Он уже был получен нами в примере 4, в гл. II.]

В табл. 6 сопоставлены вероятности удач в случаях, когда угадывание производится по способам (2) и (3). Чтобы оценить достоинства этих двух методов, необходимо знание теории средних значений и случайных колебаний. Оказывается, что среднее число верных случайных догадок одно и то же для всех способов угадывания; случайные колебания при угадывании по способу (2) несколько больше, чем при способе (3). Табл. 1 показывает, что практически разница не очень велика.

§ 5. Различные дополнения

1. Осуществление по крайней мере m событий. В обозначениях § 3 *вероятность одновременного осуществления m или более из N событий $A_1, \ldots, A_N$ равна*

$$P_m=P_{[m]}+P_{[m+1]}+\ \ldots\ +P_{[N]}. \tag{5.1}$$

Чтобы найти формулу, выражающую P_m через S_k, проще всего действовать методом индукции, начиная с формулы (1.5) и используя рекуррентное соотношение $P_{m+1}=P_m-P_{[m]}$. При $m\geqslant 1$ получаем

$$P_m=S_m-\binom{m}{m-1}S_{m+1}+\binom{m+1}{m-1}S_{m+2}-\binom{m+2}{m-1}S_{m+3}\pm\ldots$$
$$\ldots\pm\binom{N-1}{m-1}S_N. \tag{5.2}$$

Формулу (5.2) можно вывести также и непосредственно, пользуясь рассуждениями, приведшими нас к (3.1).

2. Дальнейшие тождества. *Коэффициенты* S_ν *могут быть выражены через* $P_{[k]}$ *или* P_k *следующим образом*:

$$S_\nu = \sum_{k=\nu}^{N} \binom{k}{\nu} P_{[k]} \tag{5.3}$$

и

$$S_\nu = \sum_{k=\nu}^{N} \binom{k-1}{\nu-1} P_k. \tag{5.4}$$

Идея доказательства. При данных значениях величин $P_{[m]}$ равенства (3.1) можно рассматривать как линейные уравнения с неизвестными S_ν. Нужно доказать, что (5.3) является единственным решением этих уравнений. Если подставить (5.3) в выражение (3.1) для $P_{[m]}$, то в правой части коэффициент при $P_{[k]}$ $(m \leqslant k \leqslant N)$ окажется равным

$$\sum_{\nu=m}^{k} (-1)^{\nu-m} \binom{\nu}{m}\binom{k}{\nu} = \binom{k}{m} \sum_{\nu=m}^{k} (-1)^{\nu-m} \binom{k-m}{\nu-m}. \tag{5.5}$$

При $k=m$ это выражение сводится к 1. При $k>m$ сумма является разложением бинома $(1-1)^{k-m}$ и, следовательно, равна нулю. Поэтому подстановка (5.3) превращает (3.1) в тождество $P_{[m]} = P_{[m]}$. Единственность решения системы (3.1) следует из того, что с каждым уравнением вводится только одно новое неизвестное, так что S_ν могут быть вычислены *последовательно*. Правильность формулы (5.4) может быть доказана аналогичным образом.

3. Неравенства Бонферрони. Ряд неравенств для величин P_m и $P_{[m]}$ может быть получен следующим путем. *Если либо в формуле* (3.1), *либо в формуле* (5.2) *сохранены лишь члены* S_m, S_{m+1}, ..., S_{m+r-1}, а S_{m+r}, S_{m+r+1}, ..., S_N *отброшены, то ошибка* (*т. е. истинное значение минус приближение*) *имеет знак первого опущенного слагаемого* [*именно* $(-1)^r$] *и меньше его по абсолютной величине*. Таким образом, при $r=1$ и $r=2$

$$S_m - (m+1) S_{m+1} \leqslant P_{[m]} \leqslant S_m \tag{5.6}$$

и

$$S_m - m S_{m+1} \leqslant P_m \leqslant S_m. \tag{5.7}$$

Идея доказательства. Чтобы доказать утверждение для (3.1), нужно показать, что

$$\sum_{\nu=t}^{N} (-1)^{\nu-t} \binom{\nu}{m} S_\nu \geqslant 0 \tag{5.8}$$

для любого t. Воспользуемся формулой (5.3) для представления левой части в виде линейной комбинации вероятностей $P_{[k]}$. При $t \leqslant k \leqslant N$ коэффициент при $P_{[k]}$ равен

$$\sum_{\nu=t}^{k}(-1)^{\nu-t}\binom{\nu}{m}\binom{k}{\nu}=\binom{k}{m}\sum_{\nu=t}^{k}(-1)^{\nu-t}\binom{k-m}{\nu-m}.$$

Последняя сумма равна $\binom{k-m-1}{t-m-1}$ и, следовательно, положительна. [Задача (12.13) гл. II.]

За дальнейшими неравенствами отсылаем читателя к монографии Фреше, цитированной в начале главы.

§ 6. Задачи

Замечание. *Предполагается, что в каждом случае все возможные комбинации равновероятны.*

1. В шкафу находятся десять пар туфель. Случайно выбираются четыре туфли. Найти вероятность того, что среди выбранных туфель окажется хотя бы одна пара.

2. Бросают пять костей. Найти вероятность того, что хотя бы на трех из них выпадет одинаковое число очков. (Проверить методами § 5 гл. II.)

3. Найти вероятность того, что при пяти бросаниях монеты герб выпадет хотя бы три раза подряд.

4. Решить задачу 3 для пяти последовательных появлений герба при десяти бросаниях.

5. Решить задачи 3 и 4 для выпадений единицы при бросании кости.

6. Две кости бросают r раз. Найти вероятность того, что каждая из шести комбинаций (1.1), ..., (6.6) появится хотя бы один раз.

7. При игре в бридж среди тринадцати карт одного игрока могут содержаться 0, 1, 2 или 3 четверки карт одного значения (например, четыре туза или четыре десятки и т. п.) Вычислить соответствующие вероятности.

8. Выбор с возвращением. Из генеральной совокупности n человек берется выборка объема r. Найти вероятность u_r того, что каждый из N заданных людей войдет в состав выборки. [Эта задача совпадает с задачей (11.12) гл. II.]

9. Выбор без возвращения. Решить задачу 8 для этого случая. Установить предельное соотношение $u_r \to p^N$, если $n \to \infty$ и $r \to \infty$ так, что $\frac{r}{n} \to p$. [Эта задача совпадает с задачей (11.3) гл. II, но метод настоящей главы приводит к совершенно другой формуле.]

10. Доказать, что в разложении определителя порядка N число членов, содержащих один или более диагональных элементов, равно $N!P_1$, где P_1 определяется по формуле (1.7).

11. Доказать, что число способов, которыми можно расставить восемь ладей на шахматной доске так, чтобы они не били друг друга и ни одна из них не стояла на белой главной диагонали, равно $8!(1-P_1)$, где P_1 находится по формуле (1.7) при $N=8$.

12. Задача о выборе (коллекция купонов). Колода карт состоит из s одинаковых серий. В каждой серии имеется n карт, пронумерованных числами 1, 2, ..., n, Выберем из данной колоды случайным образом (без возвращения) набор из $r \geqslant n$ карт. Вычислить вероятность u_r того, что

в выборке представлены все номера. (При $s=4$, $n=13$ наша задача эквивалентна следующей. Сдается колода карт для игры в бридж. Какова вероятность того, что набор карт первого игрока содержит карты всех 13 возможных значений.)

13. Продолжение. Показать, что $u_r \to p_0(r, n)$ при $s \to \infty$, [$p_0(r, n)$ определено в (2.3)]. Это означает, что в пределе наша случайная выборка становится выборкой с возвращением из совокупности чисел 1, 2, ..., n.

14. Продолжение. Опираясь на результаты задачи 12, доказать, что

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} (ns-ks)_r = 0$$

при $r < n$, а для $r = n$

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} (ns-ks)_n = s^n n!.$$

Проверить это, вычисляя в нуле r-ю производную функции при $x=0$

$$\frac{1}{(1-x)^{ns-r+1}} \left\{1-(1-x)^s\right\}^n.$$

15. В задаче 12 о выборке найти вероятность того, что потребуется ровно r проб для получения выборки, содержащей все номера. Перейти к пределу при $s \to \infty$.

16. Клетка содержит N хромосом, любые две из которых могут обмениваться своими частями. Предположим, что произошло r обменов (это может осуществиться $\binom{N}{2}^r$ различными способами). Найти вероятность того, что ровно m хромосом принимали участие в этом процессе [1]).

17. Производится сдача карт для игры в покер. Найти вероятность того, что в наборе карт первого игрока окажется k различных мастей.

18. Найти вероятность того, что при игре в бридж первый игрок получит ровно k пар «туз — король» одной масти.

19. Кратные совпадения. Случайное расположение карт в двух одинаковых колодах по N карт каждая сравнивается с расположением карт в третьей такой же колоде. Найти вероятность u_m наличия m двойных совпадений. Показать, что $u_0 \to 1$ при $N \to \infty$ (откуда вытекает, что $u_m \to 0$ при $m \geqslant 1$).

20. Сложные совпадения. Процедура предыдущей задачи видоизменяется следующим образом. Из $2N$ карт случайно выбираются N карт, и расположение этих N карт сравнивается с расположением карт третьей колоды. Найти вероятность отсутствия совпадений. Доказать, что она стремится к $1/e$ при $N \to \infty$.

21. Сложные совпадения. Решить задачу 20 при условии, что вместо двух берется r колод.

---

[1]) Для $N=6$ см. D. G. Catcheside, D. E. Lea, J. M. Thoday, Types of chromosome structural change introduced by the irradiation of tradescantia microspores, *Journal of Genetics*, **47** (1945—1946), 113—149.

22. В классической задаче о размещении вероятность $P_{[m]}(k)$ того, что в m ячейках окажется ровно по k предметов, равна

$$P_{[m]}(k) = \frac{(-1)^m n!r!}{m!n^r} \sum_j (-1)^j \frac{(n-j)^{r-jk}}{(j-m)!(n-j)!(r-jk)!(k!)^j},$$

где суммирование распространяется на все значения индекса j, удовлетворяющие неравенствам $j \geqslant m$, $j \leqslant n$ и $kj \leqslant r$.

23. Доказать последнее утверждение § 5 для случая $k = 1$.

24. Воспользовавшись (3.1), вывести вероятность того, что в случае статистики Бозе — Эйнштейна ровно m ячеек окажутся пустыми.

25. Проверить тождественность формулы, полученной в задаче 24, с формулой (5.3), (11.14) гл. II.

26. Доказать формулу (1.5) индукцией по N.

ГЛАВА V

УСЛОВНАЯ ВЕРОЯТНОСТЬ. НЕЗАВИСИМОСТЬ

§ 1. Условная вероятность

Понятие условной вероятности является одним из основных рабочих инструментов теории вероятностей. Последующие рассуждения естественным образом приведут к формальному определению.

Предварительные примеры. Допустим, что из общего числа N человек N_A страдают дальтонизмом, т. е. не различают некоторых цветов (большей частью красного и зеленого), и N_H — женщины. Пусть A и H означают события, состоящие, соответственно, в том, что случайно выбранное лицо страдает дальтонизмом или является женщиной. Тогда (см. определение случайного выбора §2 гл. II)

$$P\{A\}=\frac{N_A}{N};\ P\{H\}=\frac{N_H}{N}. \tag{1.1}$$

Может случиться, что интерес представляет не все множество N людей, а подмножество одних женщин, и нужно найти вероятность того, что случайно выбранная женщина страдает дальтонизмом. Эта вероятность равна N_{HA}/N_H, где N_{HA} — число женщин, страдающих дальтонизмом. Здесь не вводится новых понятий, но необходимы новые обозначения, которые указывали бы, какое частное подмножество рассматривается. Наиболее широко употребляется символ $P\{A|H\}$; в нашем примере его следует читать: «вероятность события A (дальтонизм) при условии, что произошло событие H (выбрана женщина)».

Итак, имеем формулу

$$P\{A|H\}=\frac{N_{AH}}{N_H}=\frac{P\{AH\}}{P\{H\}}. \tag{1.2}$$

Конечно, каждое подмножество можно рассматривать как новую генеральную совокупность. Говоря о подмножестве мы желаем подчеркнуть, что все время помним о существовании более широкого пространства элементарных событий. Для страховой компании может представлять интерес вопрос о том, как часто молнии причиняют убыток определенной величины (событие A). Возможно, что эта компания обслуживает несколько различных категорий объектов, например индустриальные объекты и жилые объекты в городе и в деревне. Отдельно изучать убытки индустриальных объектов означает изучать событие A в соединении с событием H — «убыток причинен индустриальному объекту». Формула (1.2) применима очевидным образом. Заметим, однако, что для страховой компании, специализирующейся на индустриальных объектах, категория H совпадает со всем пространством элементарных событий, и $P\{A|H\}$ сводится к $P\{A\}$.

Наконец, рассмотрим одного из игроков в бридж, скажем, первого. Когда колода сдана, он знает те карты, которые имеет на руках, и интересуется распределением лишь оставшихся 39 карт. Вполне законно рассматривать множество всех возможных распределений этих 39 карт как новое пространство элементарных событий, но, очевидно, более удобно исследовать его в соединении с 13 картами первого игрока (событие H) и говорить о вероятности события A (скажем, третий игрок имеет два туза) при условии, что произошло событие H. Формула (1.2) опять применима.

По аналогии с (1.2) мы введем теперь формальное

Определение. *Пусть H — событие, имеющее положительную вероятность, и A — произвольное событие. Положим*

$$P\{A|H\}=\frac{P\{AH\}}{P\{H\}}; \tag{1.3}$$

определенную таким образом величину будем называть условной вероятностью события A при условии H (при гипотезе H).

Если все элементарные события равновероятны, то условная вероятность $P\{A|H\}$ равна отношению N_{AH}/N_H числа элементарных событий, содержащихся в пересечении событий A и H, к числу элементарных событий, содержащихся в H.

Условные вероятности остаются неопределенными, когда гипотеза имеет нулевую вероятность. Это не имеет значения для дискретных пространств элементарных событий, но важно в общем случае.

Хотя символ $P\{A|H\}$ сам по себе и удобен, трудно дать его точное словесное выражение. Поэтому обычно пользуются менее формальными описаниями. Так, в изложенном примере мы говорили о вероятности того, что женщина страдает дальтонизмом, вместо того чтобы говорить об «условной вероятности того, что случайно выбранное лицо страдает дальтонизмом при условии, что это лицо — женщина». Часто слова «при условии H» заменяют словами, «если известно, что H произошло». Короче говоря, наши формулы и символы не допускают никакой двусмысленности, но словесные выражения часто недостаточно четки и требуют точного истолкования.

В противоположность условным вероятностям говорят для ясности о *безусловных вероятностях*. Строго говоря, эпитет «безусловная» является лишним, и его можно опускать.

Рассматривая условные вероятности различных событий при одной и той же частной гипотезе H, мы приходим к возможности выбрать H в качестве нового пространства элементарных событий; нужно лишь все вероятности умножить на постоянный множитель $1/P\{H\}$, чтобы сделать вероятность всего нового пространства равной единице. Это показывает, что *все общие теоремы о вероятностях верны также и для условных вероятностей*, если эти условные вероятности берутся при одном и том же условии H. В качестве примера приведем основное соотношение для вероятности появления либо события A, либо события B, либо обоих этих событий:

$$P\{A \cup B|H\} = P\{A|H\} + P\{B|H\} - P\{AB|H\}. \qquad (1.4)$$

Аналогичным образом можно было бы перенести на условные вероятности и все теоремы гл. IV, касающиеся осуществления m из N событий, но это нам не понадобится.

Формулу (1.3) часто применяют в виде

$$P\{AH\} = P\{A|H\} P\{H\}. \qquad (1.5)$$

Это так называемая теорема умножения вероятностей. Чтобы обобщить ее на случай трех событий A, B, C, примем сначала за гипотезу $H = BC$, а затем еще раз применим (1.5); тогда получим

$$P\{ABC\} = P\{A|BC\} P\{B|C\} P\{C\}. \qquad (1.6)$$

Дальнейшие обобщения для случаев четырех и более событий очевидны.

В заключение приведем простую формулу, часто оказывающуюся полезной. Пусть $H_1, \ldots, H_n$ — совокупность взаимно исключающих друг друга событий, одно из которых обязательно происходит (т. е. объединение событий $H_1, \ldots, H_n$ совпадает со всем пространством элементарных событий). Тогда любое событие A может произойти лишь в сочетании с таким-то событием H_j, или, в наших обозначениях,

$$A = AH_1 \cup AH_2 \cup \ldots \cup AH_n. \qquad (1.7)$$

Так как события AH_j попарно несовместны, то их вероятности складываются. Применяя (1.5) к $H = H_j$ и суммируя по j, получим

$$P\{A\} = \sum P\{A|H_j\} P\{H_j\}. \qquad (1.8)$$

Эта формула бывает полезной в тех случаях, когда оценка условных вероятностей $P\{A|H_j\}$ легче, чем прямое вычисление вероятности $P\{A\}$.

Примеры. а) *Выбор без возвращения.* Из совокупности n элементов $1, 2, \ldots, n$ извлекают последовательно по одному элементу без возвращения. Пусть i и j — два различных элемента. Какова условная вероятность того, что вторым извлеченным элементом будет j (событие A), при условии, что сначала был извлечен элемент i событие H). Ясно, что $P\{AH\} = \frac{1}{n(n-1)}$ и $P\{A|H\} = \frac{1}{n-1}$. Эта формула выражает тот факт, что второй выбор производится из набора $n-1$ элементов, причем каждый из этих элементов может быть извлечен с вероятностью $\frac{1}{n-1}$. В действительности, самым естественным *определением* случайного выбора без возвращения является следующее: «*Если извлечены* r *шаров, то вероятность получить при следующем испытании любой из оставшихся* $n-r$ *шаров равна* $\frac{1}{n-r}$.» Это определение эквивалентно определению из гл. II, но мы не могли его сформулировать ранее, так как оно включает понятие условной вероятности.

б) Четыре шара последовательно размещаются по четырем ящикам, причем все 4^4 размещений равновероятны. Найти условную вероятность того, что найдется ящик, содержащий ровно 3 шара (событие A), при условии, что первые два шара попали в разные ящики (событие H). Если задано H, то A может произойти двумя способами, так что $P\{A|H\}=2\cdot 4^{-2}=\frac{1}{8}$. (Нетрудно проверить прямым подсчетом, что события H и AH состоят соответственно из $12\cdot 4$ и $12\cdot 2$ точек.)

в) *Распределение по признаку пола.* Рассмотрим семьи с 2 детьми. Обозначим мальчика буквой *м* и девочку буквой *д* и условимся первой буквой обозначать старшего ребенка. Каждому из 4 элементарных событий *мм*, *мд*, *дм*, *дд* припишем вероятность $^1/_4$. Какова вероятность того, что оба ребенка — мальчики, если известно, что в семье есть мальчик? Событие AH означает *мм*, H означает или *мм*, или *мд*, или *дм*. Поэтому $P\{A|H\}={}^1/_3$, мы можем ожидать, что примерно одна треть семей со свойством H обладает также свойством A. Любопытно, что многие ожидают ответа $^1/_2$. Это правильный ответ не на наш вопрос, а на следующий: наугад выбранный мальчик оказался из семьи с двумя детьми; какова вероятность того, что второй ребенок мальчик? Разницу в ответах можно наглядно объяснить так: в первой задаче мы имеем дело с совокупностью семей, во второй с совокупностью мальчиков; поэтому в последнем случае каждая семья с двумя мальчиками представлена дважды.

г) *Совокупность, разделенная на классы.* Предположим, что какая-то генеральная совокупность людей состоит из подсовокупностей или классов H_1, H_2, ... Это могут быть национальности, возрастные группы, профессии и т. п. Пусть p_j — вероятность того, что наугад выбранное лицо принадлежит классу H_j. Утверждение «вероятность того, что лицо класса H_j — левша, равна q_j» является кратким выражением того, что «условная вероятность события A (левша), при условии, что лицо принадлежит классу H_j, равна q_j». Безусловная вероятность того, что наугад выбранное лицо окажется левшой, равна $p_1q_1+p_2q_2+p_3q_3+\ \ldots$; это опять частный случай формулы (1.8). Условная вероятность того, что данный человек принадлежит классу H_j при условии, что он левша, равна

$$P\{H_j|A\}=\frac{p_jq_j}{p_1q_1+p_2q_2+\ldots}. \qquad (1.9)$$

§ 2. Вероятности, определяемые через условные вероятности Урновые модели

В предыдущем параграфе мы считали, что вероятности в пространстве элементарных событий уже заданы, и вычисляли некоторые условные вероятности. В приложениях многие опыты описываются указанием определенных условных вероятностей (хотя эпитет «условный» обычно опускается). Теоретически это означает, что вероятности в пространстве элементарных событий можно вычислить, отправляясь от заданных условных вероятностей. Мы уже отмечали (пример 1,а), что выбор без возвращения лучше всего определить следующим образом: если известен результат первых r испытаний, то при $(r+1)$-м испытании с равными вероятностями может быть извлечен любой из оставшихся шаров. Аналогично этому в примере (1.2) наша генеральная совокупность полностью описывается указанием абсолютных вероятностей p_j отдельных классов и условных вероятностей q_j, характеризующих число левшей в каждом классе. Еще несколько примеров разъясняют общую схему лучше, чем прямое описание.

Примеры. а) В примере 5, б гл. I мы рассмотрели игру с тремя игроками a, b, c, описали точки пространства элементарных событий, но не определили вероятностей. Предположим теперь, что игра такова, что в каждой отдельной партии любой из двух партнеров выигрывает с вероятностью $^1/_2$. Это утверждение не содержит слов «условная вероятность», но тем не менее тесно с нею связано. Оно означает, что если игрок a участвует в r-ой партии (событие H), то он выигрывает с вероятностью $^1/_2$. Уравнение (1.5) позволяет теперь заключить, что игрок a выигрывает в первой и второй партиях с вероятностью $^1/_4$, или, используя обозначения, $P\{a, a\}=\frac{1}{4}$. Повторное применение формулы (1.5) показывает, что $P\{acc\}=\frac{1}{8}$, $P\{acbb\}={}^1/_{16}$ и т. д.; т. е. что элементарное событие схемы (*), включающее r букв, имеет вероятность 2^{-r}. Такие же вероятности были использованы в задаче 5 гл. I, но теперь их описание более наглядно. (Продолжение см. в задаче 14.)

б) *Семьи.* Истолкуем следующее утверждение: «вероятность того, что в семье имеется k детей равна $p_k (p_0+p_1+\ldots=1)$. При фиксированном числе детей всевозможные их распределения по признаку пола равновероятны». Пусть по-прежнему *м* — мальчик, *д* — девочка. Наше пространство элементарных событий состоит из точек 0 (детей нет), *м*, *д*, *мм*, *мд*, *дм*, *дд*, *ммм*, ... Второе из наших утверждений более формально можно выразить так: если известно, что в семье имеется n детей, то каждая из 2^n возможных комбинаций полов детей имеет условную вероятность 2^{-n}. Вероятность этой гипотезы мы обозначали через p_n; из (1.5) видно, что безусловная вероятность любой комбинации из n букв *м* и *д* равна $p_n \cdot 2^{-n}$.

Заметим, что это пример генеральной совокупности, разделенной на классы. Семьи, имеющие j детей, образуют класс H_j. Пусть A означает событие «в семье есть мальчики и нет девочек».

Вероятность этого события, очевидно, равна $P\{A\} = p_1 2^{-1} + p_2 2^{-2} + p_3 2^{-3} + \ldots$. Между прочим, это — частный случай формулы (1.8), причем гипотезой H_j является событие «в семье имеется j детей». Найдем (условную) вероятность того, что в семье только один ребенок, при условии, что в семье нет девочек. Здесь гипотезой является событие A. Пусть H — событие «имеется только один ребенок». Тогда AH означает «имеется один ребенок и ни одной девочки» и

$$P\{H|A\} = \frac{P\{AH\}}{P\{A\}} = \frac{p_1 2^{-1}}{p_1 2^{-1} + p_2 2^{-2} + p_3 2^{-3} + \ldots}. \tag{2.1}$$

в) *Урновые модели эффекта последействия.* Рассмотрим для определенности завод, на котором возможны аварии. Каждую аварию можно рассматривать как результат не зависящей от человека игры случая: из урны, содержащей черные и красные шары, через одинаковые промежутки времени случайно извлекается шар, красный шар обозначает аварию. Если вероятность аварии остается постоянной во времени, то состав урны всегда один и тот же. Более реальной, однако, является ситуация, при которой каждая авария приводит к *эффекту последействия*, состоящему в увеличении или уменьшении вероятности следующего несчастного случая. Это соответствует случаю урны, состав которой меняется согласно определенным правилам, зависящим от исхода предыдущих извлечений. Нетрудно придумать много таких правил, которые бы отвечали различным ситуациям, но мы ограничимся рассмотрением лишь одной модели[1]).

Урновая схема. *Урна содержит b черных и r красных шаров. Из урны наугад извлекается шар. Вынутый шар всегда*

---

[1]) Идея использовать урновые схемы для описания повторных эффектов (эпидемий) была, по-видимому, впервые предложена Пойа. Его схема (впервые введенная в работе G. Eggenberger, G. Polya, Über die Statistik verketteter Vorgänge, *Zeitschrift für Angewandte Mathematik und Mechanik*, 7 (1923), 279—289) послужила прототипом для многих других моделей, которым посвящена обширная литература. Модель, описанная в тексте, и три ее специальных случая были предложены B. Friedman, A simple urn model, *Communications on Pure and Applied Mathematics*, 2 (1949), 59—70.

возвращается обратно, причем в урну добавляют c шаров того же цвета и d шаров противоположного цвета. Затем из урны (содержащей теперь $r+b+c+d$ шаров) снова производится случайное извлечение и повторяется тот же процесс. Здесь c и d — произвольные целые числа. Их можно взять и отрицательными, однако в этом случае процесс может окончиться после конечного числа извлечений из-за отсутствия шаров. В частности, положив $c=-1$ и $d=0$, мы приходим к модели *случайного выбора без возвращения*, который кончается через $r+b$ шагов.

Чтобы придать нашему образному описанию точный математический смысл, заметим, что оно определяет условные вероятности, по которым могут быть вычислены некоторые основные вероятности. Точку пространства элементарных событий, соответствующего n извлечениям, можно представить как последовательность n букв K и $Ч$. Событие «первый шар черный» (т. е. множество всех последовательностей, начинающихся с $Ч$) имеет вероятность $\frac{b}{b+r}$. Если первый шар черный, то (условная) вероятность извлечь черный шар при втором испытании равна $(b+c)/(b+r+c+d)$. Безусловная вероятность последовательности «черный, черный» равна, следовательно, по (1.5)

$$\frac{b}{b+r}\cdot\frac{b+c}{b+r+c+d}. \tag{2.2}$$

Вероятность последовательности «черный, черный, черный» равна (2.2), умноженному на $(b+2c)/(b+r+2c+2d)$ и т. д. Ясно, что таким путем можно вычислить все вероятности. (Конечно, в случае отрицательных c и d число n извлечений надо выбирать достаточно малым, чтобы избежать появления отрицательного числа шаров.) Нетрудно проверить по индукции, что сумма вероятностей всех элементарных событий равна единице.

Точные выражения для вероятностей получить нелегко, исключая самый важный и лучше всего изученный частный случай, которым является

Урновая схема Пойа. Эту схему характеризуют значения параметров $d=0$, $c>0$. После каждого испытания число шаров того же цвета, что и извлеченный шар, возрастает, в то время как число шаров противоположного цвета не меняется. Это приводит к эффекту последействия, состоящему в следующем: если мы извлекаем шар какого-то цвета, то вероятность извлечь шар того же цвета при следующем испытании возрастает. Урновая схема Пойа является приближенной моделью таких явлений, как, например, *эпидемии*, при которых осуществление некоторых событий увеличивает вероятность их повторения. Аналитическая простота этой схемы связана со следующим очевидным обстоятельством: любая последовательность n

извлечений, приводящая к n_1 черным и n_2 красным шарам ($n_1 + n_2 = n$), имеет ту же вероятность, что и событие, состоящее в извлечении *сначала* n_1 черных и *затем* n_2 красных шаров, а именно

$$p_{n_1, n} = \frac{b(b+c)(b+2c)\ldots(b+n_1c-c)\,r(r+c)\ldots(r+n_2c-c)}{(b+r)(b+r+c)(b+r+2c)\ldots(b+r+nc-c)}. \quad (2.3)$$

Деля числитель и знаменатель на c и используя обозначения (2.1) гл. II, мы можем переписать эту формулу следующим образом:

$$p_{n_1, n} = \frac{\left(\frac{b}{c}+n_1-1\right)_{n_1}\left(\frac{r}{c}+n_2-1\right)_{n_2}}{\left(\frac{b+r}{c}+n-1\right)_n} = \frac{\left(-\frac{b}{c}\right)_{n_1}\left(-\frac{r}{c}\right)_{n_2}}{\left(-\frac{b+r}{c}\right)_n}. \quad (2.4)$$

(Схема Пойа рассматривается в задачах 18—24.)

Наряду со схемой Пойа урновая модель содержит другой интересный частный случай, а именно:

Модель Эренфестов[1]) *теплового обмена* между двумя изолированными телами. С физической точки зрения здесь мы имеем дело с экспериментом, при котором k молекул распределены по двум резервуарам I и II. Случайным образом выбирается одна молекула и перемещается из своего резервуара в другой. Этот процесс повторяется. Каково распределение частиц через n шагов? Для того чтобы свести наш эксперимент к урновой схеме, достаточно назвать шары, содержащиеся в резервуаре I красными, а в резервуаре II черными. Тогда при каждом испытании вынутый шар заменяется на шар противоположного цвета, так что мы имеем $c = -1$, $d = +1$. Ясно, что в этом случае процесс может продолжаться неограниченно долго (если отсутствуют красные шары, то автоматически вынимается черный шар и заменяется на красный). (С другой точки зрения мы будем обсуждать модель Эренфестов в примере 2,е гл. XV.)

Частный случай $c = 0$, $d > 0$ был предложен Фридманом в качестве модели *службы безопасности*. Всякий раз, когда происходит несчастный случай (т. е. извлекается красный шар), служба безопасности усиливает контроль; если же несчастного случая не произошло, то контроль ослабевает и вероятность катастрофы увеличивается.

г) *Урновые модели при классификации. Ложное заражение.* Развивая предыдущий пример, допустим, что с каждым человеком из некоторой совокупности может произойти несчастный случай. Произойдет или не произойдет этот несчастный случай, определяется

---

[1]) P. und T. Ehrenfest, Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem, *Physikalische Zeitschrift*, 8 (1907), 311—314. Более полное рассмотрение см. M. Kac, Random walk and the theory of Brownian motion, *American Mathematical Monthly*, 54 (1947), 369—391.

извлечением шара из урны. Теперь, однако, мы предположим, что не существует эффекта последействия, так что состав урны остается неизменным на протяжении всего процесса. Зато вероятность несчастного случая может меняться от человека к человеку или от профессии к профессии, и мы предположим, что каждому лицу (или каждой профессии) отнесена своя собственная урна. Для того чтобы напрасно не осложнять дальнейшие рассуждения, допустим, что имеется только два типа людей (две профессии), и их числа в полной совокупности находятся в отношении 1 : 5. Рассмотрим теперь урну I, содержащую r_1 красных и b_1 черных шаров, и урну II, содержащую r_2 красных и b_2 черных шаров. Эксперимент «случайно выбираем человека и наблюдаем сколько несчастных случаев произошло с ним за время n» можно теперь интерпретировать следующим образом: *бросаем кость, если выпало одно очко, то производим n случайных выборов с возвращением из урны I, в противном случае ту же процедуру проводим с урной II.* Наш опыт описывает ситуацию, возникающую при заключении договора между новым клиентом и страховой компанией.

Используя (1.8), нетрудно получить, что вероятность извлечения красного шара при первом испытании равна

$$P\{K\}=\frac{1}{6}\,\frac{r_1}{r_1+b_1}+\frac{5}{6}\,\frac{r_2}{r_2+b_2}, \tag{2.5}$$

а вероятность последовательности «красный, красный» есть

$$P\{KK\}=\frac{1}{6}\left(\frac{r_1}{r_1+b_1}\right)^2+\frac{5}{6}\left(\frac{r_2}{r_2+b_2}\right)^2. \tag{2.6}$$

Хотя математическая сторона вопроса не представляет трудностей, тем не менее наша модель имеет интересную особенность, приводящую в приложениях к неожиданным выводам. Пусть страховая компания регистрирует, что в течение первого года с клиентом произошел несчастный случай, и интересуется вероятностью того, что в течение второго года с ним произойдет еще один случай. Другими словами, дано, что при первом испытании был извлечен красный шар, требуется определить (условную) вероятность последовательности «красный, красный». Ясно, что отношение $P\{KK\}/P\{K\}$ *отлично* от $P\{K\}$. Чтобы пояснить это, предположим, что известны значения $r_1/b_1+r_1=0{,}6$ и $r_2/b_2+r_2=0{,}06$. Вероятность извлечения красного шара при любом испытании равна 0,15, однако условная вероятность равна 0,42. Отметим, что в полной совокупности *эффект последействия отсутствует,* однако несчастный случай с человеком, выбранным случайно, увеличивает вероятность того, что *с тем же человеком* произойдет второй несчастный случай. Здесь имеет место эффект выбора, появление несчастного случая не оказывает никакого действительного влияния, а только указывает на то, что выбранное лицо более предрасположено к несчастным случаям.

В статистической литературе стало привычкой употреблять слово *заражение* вместо эффект последействия. *Кажущийся* эффект последействия выбора сначала неправильно интерпретировался как эффект действительного заражения, так что статистики теперь говорят о заражении (или о вероятностном распределении при заражении) в неясном и вводящем в заблуждение смысле. Возьмем, например, натуралиста, который ловит насекомых в поле. Если после периода неудач он находит нужное насекомое, то он вправе заключить, что, вероятно, где-то близко находится целый выводок, и поэтому его шансы найти еще одно насекомое увеличились. Конечно, тут не наблюдается никакого эффекта последействия, и все же статистики говорят о заражении.

д) Следующий пример широко известен и поучителен, но несколько искусствен. Представим себе $N+1$ урн, каждая из которых содержит N шаров; урна с номером k содержит k красных и $N-k$ белых шаров ($k=0, 1, 2, \ldots, N$). Из наугад выбранной урны n раз наудачу извлекают шары, причем вынутый шар каждый раз возвращают обратно. Предположим, что все n раз извлеченных шаров оказались красными (событие A). Найдем (условную) вероятность того, что следующее испытание тоже даст красный шар (событие B). Условная вероятность извлечь подряд n красных шаров из урны с номером k равна $(k/N)^n$. Отсюда, согласно (1.8),

$$P\{A\} = \frac{1^n+2^n+\ldots+N^n}{N^n(N+1)}. \tag{2.7}$$

Событие AB означает, что $n+1$ испытаний дали красные шары и поэтому

$$P\{AB\} = P\{B\} = \frac{1^{n+1}+2^{n+1}+\ldots+N^{n+1}}{N^{n+1}(N+1)}. \tag{2.8}$$

Искомая вероятность равна $P\{B \mid A\} = P\{B\}/P\{A\}$.

Суммы в (2.7) и (2.8) можно рассматривать как римановы интегральные суммы, и поэтому при большом N

$$N^{-1}\sum_{k=1}^{N}\left(\frac{k}{N}\right)^n \approx \int_0^1 x^n\,dx = \frac{1}{n+1}. \tag{2.9}$$

Следовательно, при большом N приближенно

$$P\{B|A\} \approx \frac{n+1}{n+2}. \tag{2.10}$$

Этой формуле можно дать следующее грубое истолкование: если все возможные количества красных и белых шаров в урне равновероятны и если n испытаний дали красные шары, то вероятность появления красного шара при следующем испытании равна $(n+1)/(n+2)$. Это — так называемый закон следования Лапласа (1812).

До развития современной теории понятие равновероятности часто отождествляли с «отсутствием предварительных сведений». Сам Лаплас для иллюстрации пользы формулы (2.10) вычислял вероятность того, что на следующий день взойдет Солнце, в предположении, что оно всходило ежедневно 5000 лет (или $n = 1\,826\,213$ дней подряд). Передают, что Лаплас готов был ставить 1 826 214 против 1 за то, что Солнце не изменит своего поведения; в наше время следовало бы увеличить ставку, поскольку регулярное движение Солнца наблюдалось в течение еще одного столетия. Чтобы отдать должное Лапласу и понять его намерения, потребовалось бы историческое исследование. Последователи Лапласа, однако, использовали аналогичные доводы в повседневной работе и рекомендовали физикам и инженерам применять такие методы в случаях, в которых формулы не имеют никакого действительного смысла. Мы бы отвергли этот метод, даже если бы на минуту согласились допустить, что наша вселенная выбрана наугад из «множества вселенных», в котором все мыслимые возможности равновероятны. Действительно, предполагаемый восход Солнца 5 февраля 3123 года до н. э. ничуть не более достоверен, чем то, что Солнце взойдет завтра; у нас совершенно одинаковые основания верить в оба эти события.

Замечание о формуле Байеса. В (1.9) и (2.2) мы вычислили некоторые условные вероятности, исходя прямо из определения. Начинающему рекомендуется поступать так всегда и не запоминать формулу (2.12), которую мы сейчас выведем. Эта формула воспроизводит в общем виде то, что мы делали в частных примерах, и является лишь иначе записанной формулой (1.3). В наших примерах мы имели группу взаимно исключающих друг друга и исчерпывающих все пространство событий $H_1, H_2, \ldots$, т. е. таких событий, что каждое элементарное событие принадлежит одному и только одному из событий H_j. Мы пользовались формулой

$$P\{H_k|A\} = \frac{P\{AH_k\}}{P\{A\}}. \tag{2.11}$$

Подставляя в (2.11) выражения (1.5) и (1.8), придадим этой формуле вид

$$P\{H_k|A\} = \frac{P\{A|H_k\}\,P\{H_k\}}{\sum\limits_{j=1}^{n} P\{A|H_j\}\,P\{H_j\}}. \tag{2.12}$$

Если называть события H_k гипотезами, то (2.12) будет выражать собой так называемую «формулу Байеса для вероятностей гипотез» В математическом отношении (2.12) есть просто специальный способ записи формулы (1.3). Выведенная формула полезна во многих статистических приложениях, подобных описанным в примерах (б) — (д). К сожалению, формула Байеса была дискредитирована метафизическими приложениями вроде описанного в примере (д). В практической работе подобные рассуждения могут привести к опасным последствиям. Инженер по контролю качества продукции имеет дело с одной конкретной машиной, а не с бесконечным множеством машин, из которого наугад выбрана одна. Между тем ему советуют пользоваться формулой Байеса на том основании, что она логически приемлема и соответствует нашей манере мышления. Такими доводами Платон пользовался для доказательства существования Атлантиды; некоторые философы применяют эти же доводы для доказательства нелепости механики Ньютона. В нашем случае эти доводы не учитывают того обстоятельства, что инженер желает достичь определенного успеха и что он поступит лучше, оценивая и доводя до минимума источники различного рода ошибок в своих предвидениях и догадках. Современные методы статистической проверки и оценки менее наглядны, но зато более реалистичны. Их можно не только оправдывать рассуждениями, но и применять на деле.

§ 3. Независимость

В приведенных выше примерах условная вероятность $P\{A|H\}$, вообще говоря, не была равна безусловной вероятности $P\{A\}$. Говоря грубо, знание того, что произошло событие H, изменяло нашу оценку шансов появления события A. Только в том случае, когда $P\{A|H\}=P\{A\}$, это знание не оказывает никакого влияния на оценку шансов появления события A. Мы будем говорить, что в этом случае событие A не зависит от события H. Далее, из формулы (1.5) следует, что условие $P\{A|H\}=P\{A\}$ можно записать в этом случае в форме

$$P\{AH\}=P\{A\}\cdot P\{H\}. \tag{3.1}$$

Это равенство симметрично относительно A и H и показывает, что если A не зависит от H, то и H не зависит от A. Поэтому мы предпочтем дать следующее симметричное.

Определение 1. *Два события A и H называются независимыми*[1], *если они удовлетворяют соотношению* (3.1). Это определение применимо и в случае $P\{H\}=0$, когда условная вероятность $P\{A|H\}$ не определена.

Примеры. а) Из колоды игральных карт вытаскивают наугад одну карту. Из соображений симметрии мы склонны ожидать, что события «трефа» и «туз» независимы. Действительно, их вероятности равны 1/4 и 1/13, а вероятность их одновременного осуществления равна 1/52.

б) Бросают две правильные игральные кости. Проверим, что события «одно очко на первой кости» и «четное число очков на второй кости» независимы; действительно, вероятность их одновременного

[1]) В дословном переводе «статистически независимыми». Слово «статистически» подчеркивает, что речь идет о новом понятии, специфическом для теории вероятностей и не совпадающем с обыденным пониманием независимости. Однако в русской литературе общеупотребителен более короткий термин «независимость» без всякого дополнительного эпитета. — *Прим. ред.*

осуществления $3/36 = 1/12$ равна произведению их вероятностей именно $1/6$ и $1/2$.

в) При случайной перестановке четырех букв (a, b, c, d) события «a предшествует b» и «c предшествует d» независимы. Это интуитивно ясно и может быть легко проверено.

г) Возвратимся к примеру 1, в, но рассмотрим теперь семьи с тремя детьми. Допустим, что каждая из восьми возможностей *ммм*, *ммд*, ..., *ддд* имеет вероятность $1/8$. Пусть H означает событие «в семье есть дети обоего пола», а A — событие «в семье не более одной девочки». Тогда $P\{H\} = 6/8$ и $P\{A\} = 4/8$. Одновременное осуществление событий H и A означает одну из возможностей *ммд*, *мдм*, *дмм*, и поэтому $P\{AH\} = 3/8 = P\{A\} \cdot P\{H\}$. Таким образом, в семьях с тремя детьми события A и H независимы. Заметим, что для семей с двумя или четырьмя детьми это неверно. Это показывает, что зависимость или независимость событий не всегда является очевидной.

Если событие H происходит, то противоположное событие не происходит, и наоборот. Независимость событий A и H означает, что из факта появления события H нельзя извлечь никаких следствий относительно появления события A; поэтому независимость событий A и H должна быть эквивалентна независимости событий A и $\overline{H}$ (и, вследствие симметрии, также событий $\overline{A}$ и H и $\overline{A}$ и $\overline{H}$). Это легко проверить, пользуясь соотношением $P\{\overline{H}\} = 1 - P\{H\}$. Если (3.1) выполнено, то (поскольку $A\overline{H} = A - AH$):

$$P\{A\overline{H}\} = P\{A\} - P\{AH\} = P\{A\} - P\{A\} \cdot P\{H\} = P\{A\} \cdot P\{\overline{H}\}, \tag{3.2}$$

как и предполагалось.

Допустим теперь, что три события A, B и C попарно независимы, так что

$$\begin{aligned} P\{AB\} &= P\{A\} \cdot P\{B\}; \\ P\{AC\} &= P\{A\} \cdot P\{C\}; \\ P\{BC\} &= P\{B\} \cdot P\{C\}. \end{aligned} \tag{3.3}$$

Может показаться, что из этих соотношений всегда вытекает независимость таких пар событий, как AB и C. К сожалению, это не обязательно так. Мы приведем пример, в котором соотношения (3.3) выполнены, но одновременное осуществление всех трех событий A, B и C невозможно, так что события AB и C не могут быть независимы.

д) Бросают две кости, и три события определяются следующим образом: A означает «нечетная грань на первой кости», B означает «нечетная грань на второй кости»; наконец, C означает «нечетная сумма» (одна грань четная, другая — нечетная). Если каждое из 36 элементарных событий имеет вероятность $1/36$, то любые два события из числа A, B и C независимы. Вероятность каждого события равна $1/2$, и такова же его условная вероятность при условии, что *одно* из двух других событий произошло. Тем не менее одновременно все три события не могут произойти. Знание того, что произошло событие A и не произошло B, делает достоверным событие C; аналогичные утверждения верны и для других комбинаций.

Желательно сохранить термин «независимость» для случая, когда икакое подобное влияние невозможно. Для этого необходимо выполнение соотношений (3.3) и, кроме того, выполнение условия

$$P\{ABC\}=P\{A\}P\{B\}P\{C\}. \qquad (3.4)$$

Последнее равенство обеспечивает независимость событий A и BC, B и AC, C и AB. Более того, теперь можно доказать, что события $A\cup B$ и C независимы. Действительно, согласно основному соотношению (7.4) гл. I, имеем

$$P\{(A\cup B)C\}=P\{AC\}+P\{BC\}-P\{ABC\}. \qquad (3.5)$$

Пользуясь равенствами (3.3) и (3.4), вынесем за скобки $P\{C\}$. Оставшийся множитель будет равен $P\{A\}+P\{B\}-P\{AB\}=P\{A\cup B\}$, так что

$$P\{(A\cup B)C\}=P\{(A\cup B)\}P\{C\}. \qquad (3.6)$$

Это делает правдоподобным предположение, что условий (3.3) и (3.4) вместе достаточно для того, чтобы избежать затруднений: любое событие, которое можно выразить через A и B, будет независимо от события C.

В общем случае n событий оказывается достаточным следующее.

Определение 2. *События $A_1, A_2, \ldots, A_n$ называются взаимно независимыми, если для всех комбинаций индексов $1\leqslant i<j<k<\ldots\leqslant n$ справедливы следующие правила умножения:*

$$\begin{aligned} P\{A_iA_j\}&=P\{A_i\}P\{A_j\};\\ P\{A_iA_jA_k\}&=P\{A_i\}P\{A_j\}P\{A_k\}; \qquad (3.7)\\ &\ldots\ldots\ldots\ldots\ldots\\ P\{A_1A_2\ldots A_n\}&=P\{A_1\}P\{A_2\}\ldots P\{A_n\}. \end{aligned}$$

Здесь в первой строке имеется $\binom{n}{2}$ уравнений, во второй $\binom{n}{3}$ и т. д. Имеется, следовательно, всего $\binom{n}{2}+\binom{n}{3}+\ldots+\binom{n}{n}=(1+1)^n-\binom{n}{1}-\binom{n}{0}=2^n-n-1$ условий, которые должны быть выполнены. С другой стороны, $\binom{n}{2}$ условий первой строки достаточно для обеспечения *попарной независимости*. Вся система условий (3.7) кажется сложной, но, как будет ясно в дальнейшем, то, что эти условия выполняются, обычно бывает очевидным и не требует проверки. Нетрудно проверить по индукции (начиная с $n=2$ и используя (3.2)), что *в определении 2 систему* (3.7) *можно заменить системой* 2^n *уравнений, получаемых из последнего уравнения* (3.7) *заменой произвольного множества событий* A_j *их дополнениями* A'_j.

Различие между взаимной и попарной независимостью имеет скорее теоретический, чем практический интерес. Практически важных примеров попарно независимых событий, не являющихся взаимно независимыми, по-видимому, не существует. Однако, вообще говоря, такие примеры, как показал Бернштейн[1]), возможны.

§ 4. Повторные испытания

Понятие независимости позволяет, наконец, дать аналитическую формулировку наглядного представления об опытах, «повторяющихся при неизменных условиях».

Рассмотрим пространство элементарных событий $\mathfrak{S}$, представляющее некоторый мыслимый опыт. Пусть $E_1, E_2, \ldots$ — точки пространства $\mathfrak{S}$; обозначим их вероятности через $p_1, p_2, \ldots$. Возможными исходами двух последовательно проведенных опытов являются пары, (E_j, E_k), которые и образуют новое пространство элементарных событий. Вероятности в этом пространстве могут быть определены различными способами. Однако когда экспериментатор говорит, что два измерения совершаются при одинаковых условиях, то он подразумевает независимость: исход первого измерения не должен влиять на исход второго. Это значит, что события «исход первого опыта E_j» и «исход второго опыта E_k» должны быть независимы, т. е.

$$P\{(E_j, E_k)\} = p_j p_k. \tag{4.1}$$

Это равенство определяет вероятность любой пары (E_j, E_k) возможных исходов. Прежде чем можно будет использовать (4.1) как определение вероятностей в новом пространстве элементарных событий, нужно показать, что величины $p_j p_k$ дают в сумме единицу. Для этого

---

[1]) С. Н. Бернштейн предлагает следующий пример попарно независимых событий, не являющихся взаимно независимыми. Рассмотрим тетраэдр, три грани которого окрашены соответственно в красный, синий и зеленый цвета, а четвертая грань окрашена частично красным, частично синим и частично зеленым цветом. При выборе наудачу одной из четырех граней тетраэдра рассмотрим события *К*, *С* и *З*, означающие соответственно обнаружение на выбранной грани красного, синего или зеленого цвета. Легко проверить, что $P\{К\} = P\{С\} = P\{З\} = {}^1/_2$; $P\{КС\} = P\{КЗ\} = P\{СЗ\} = {}^1/_4$, так что события *К*, *С* и *З* попарно независимы. С другой стороны, $P\{КСЗ\} = {}^1/_4 \neq P\{К\}P\{С\}P\{З\}$, так что события *К*, *С* и *З* не являются взаимно независимыми. — *Прим. ред.*

заметим, что в сумме $\sum\sum p_j p_k$ каждое слагаемое встречается один и только один раз, так что

$$\sum\sum p_j p_k = (p_1 + p_2 + \ldots)(p_1 + p_2 + \ldots) = 1.$$

Поэтому (4.1) можно принять за определение вероятностей.

Пусть A и B — два произвольных события в первоначальном пространстве $\mathfrak{S}$. Событие «A появилось при первом испытании, B — при втором» обозначим (A, B). Допустим, что A содержит точки E_{a1}, E_{a2}, ..., а B — точки E_{b1}, E_{b2}, ... Тогда (A, B) будет объединением всех пар $\left(E_{a_j}, E_{b_k}\right)$ и, так же как и раньше,

$$P\{(A, B)\} = \sum\sum p_{a_j} p_{b_k} = \left(\sum p_{a_j}\right)\left(\sum p_{b_k}\right) = P\{A\}\, P\{B\}. \qquad (4.2)$$

Поэтому события A и B независимы. Мы видим, что определение (4.1) ведет к тому, что любое событие, связанное со вторым испытанием, оказывается независимым от любого события, связанного с первым испытанием. Это определение описывает «одинаковые опыты».

Все сказанное, очевидно, применимо также к последовательности из r испытаний и приводит к следующему определению.

Определение 1. *Пусть $\mathfrak{S}$ — пространство элементарных событий с точками E_1, E_2, ... и соответствующими вероятностями p_1, p_2, ... Под r независимыми испытаниями, соответствующими пространству $\mathfrak{S}$, мы понимаем пространство элементарных событий, точками которого являются группы из r исходов $\left(E_{j_1}, E_{j_2}, \ldots, E_{j_r}\right)$ с отнесенными им вероятностями*

$$P\left\{\left(E_{j_1}, E_{j_2}, \ldots, E_{j_r}\right)\right\} = p_{j_1} p_{j_2} \ldots p_{j_r}. \qquad (4.3)$$

Другими словами, каждая точка нового пространства есть выборка объема r (с возможными повторениями) из точек первоначального пространства, а вероятности определяются по правилу умножения (4.3). Напоминаем читателю, что (4.3) *не* есть единственно возможное определение вероятностей. Иначе говоря, повторные испытания не обязательно независимы. Например, в урновой схеме Пойа (пример 2, в) мы имеем дело с зависимыми испытаниями. Соотношение (4.3) определяет независимые испытания, или, в терминах физики, испытания, повторяющиеся при неизменных условиях.

Рассуждения, аналогичные тем, которые привели нас к формуле (4.2), показывают, что вообще верна следующая

Теорема. *Допустим, что система событий A_1, A_2, ..., A_r такова, что появление события A_j зависит только от j-го испытания; тогда события A_1, ..., A_r взаимно независимы, если независимы испытания, т. е. если выполнены условия* (4.3).

Если пространство $\mathfrak{S}$ содержит конечное число N точек, то мы имеем N^r элементарных событий $\left(E_{j_1}, \ldots, E_{j_r}\right)$. Если каждая точка пространства $\mathfrak{S}$ имеет вероятность $1/N$, то по формуле (4.3) каждая точка $(E_{j_1}, \ldots, E_{j_r})$ будет иметь вероятность N^{-r}. Наш путь предпочтительнее, чем формальное объявление всех элементарных событий равновероятными, так как он применим к пространствам элементарных событий с неравными вероятностями, а также к бесконечным пространствам элементарных событий. Этот путь обязателен в общей теории, где даже одно-единственное испытание рассматривается как первое в возможной бесконечной последовательности испытаний. Мы имеем тогда дело только с бесконечными последовательностями $(E_{j1}, E_{j2}, \ldots)$ возможных исходов, и вероятности в таком новом пространстве определяются в согласии с формулой (4.3) К сожалению, это уводит нас за пределы теории дискретных пространств элементарных событий, излагаемой в настоящей книге. Мы развиваем более простую теорию, но расплачиваемся за это необходимостью для каждого числа испытаний рассматривать свое пространство элементарных событий.

До сих пор рассматривались лишь повторения одного и того же опыта, но точно так же можно поступать с последовательностями неодинаковых опытов. Когда мы сначала бросаем монету, а потом кость, то мы, естественно, считаем, что эти два опыта независимы. Это значит, что вероятности определяются по правилу умножения. Так, $P\{$герб, одно очко$\} = {}^1/_2 \cdot {}^1/_6$ и т. д. В данном частном случае можно просто объявить равновероятными все 12 элементарных событий, но вообще нужно поступать согласно (4.3).

Определение 2. *Пусть $\mathfrak{S}'$ и $\mathfrak{S}''$ — два пространства элементарных событий, точки которых обозначим E_1', E_2' ... и E_1'', E_2'', ... Пусть соответствующие вероятности равны p_1', p_2', ... и p_1'', p_2'', ... Последовательность из двух опытов описывается пространством с точками (E_j', E_k''). Сказать, что два опыта независимы, значит определить вероятности новых элементарных событий по формуле*

$$P\{(E_j', E_k'')\} = p_j' p_k''. \tag{4.4}$$

(Понятие, которое только что было введено, никоим образом не является специфическим понятием теории вероятностей. Если даны два пространства $\mathfrak{S}'$ и $\mathfrak{S}''$, состоящие соответственно из точек E' и E'', *множество всех пар* (E', E'') *называется прямым произведением* $\mathfrak{S}'$ *и* $\mathfrak{S}''$. Например, декартова плоскость, т. е. множество пар (x, y), является прямым произведением координатных осей x и y. [Трехмерное пространство можно также рассматривать как произведение трех осей или как прямое произведение плоскости (x, y) на ось z.] Уравнение (4.4) определяет то, что обычно называют *прямым произведением вероятностных мер*, заданных на $\mathfrak{S}'$ и $\mathfrak{S}''$. Для обозначения пространства элементарных событий с вероятностью, определенной на нем, мы использовали слово *испытание*. Подобно этому *последовательность двух независимых испытаний есть просто сокращенное выражение для прямого произведения соответствующих пространств элементарных событий, в котором вероятности определяются по формуле* (4.4).

Эти понятия естественным образом переносятся на случай любого числа пространств. Например, в (4.3) фигурирует r-кратное прямое произведение $\mathfrak{S}$ на само себя. Там, где студент, изучающий теорию вероятностей, говорит о первом, втором ... испытаниях, другие математики говорят о первом, втором ... сомножителях в произведении пространств. (Событие, которое определяется исходом только первого испытания, называют также цилиндрическим множеством с основанием в первом пространстве-сомножителе.)

Множество всех пар (i, j), где i и j — произвольные положительные целые числа, заключенные между 1 и n, можно рассматривать как произведение множества целых чисел $1, 2, \ldots, n$ на само себя. Пространство элементарных событий, соответствующее выбору без возвращения двух шаров, не содержит точек вида (i, i), и поэтому его нельзя *непосредственно* представить в виде прямого произведения. Тем не менее последующие примеры покажут, что этот опыт можно различными способами разложить в последовательность независимых испытаний. Аналогичный метод применим и в более сложных случаях.

Примеры. а) *Перестановки.* Мы рассматривали $n!$ перестановок из элементов $a_1, a_2, \ldots, a_n$ как элементарные события и связывали с каждой перестановкой вероятность $1/n!$. *То же самое пространство элементарных событий* можно рассматривать как представляющее $(n-1)$ последовательных опытов. Начнем с того, что выпишем a_1. Первый опыт состоит в том, чтобы поместить a_2 перед или после a_1. Для a_3 после этого имеется три места, и третий опыт состоит в выборе одного из этих трех мест. Этим устанавливается относительный порядок элементов a_1, a_2 и a_3. Для a_4 мы получаем четыре возможных места. Вообще, когда $a_1, \ldots, a_k$ уже размещены в каком-то определенном порядке, проводится опыт с номером k, состоящий в выборе одного из $k+1$ мест для a_{k+1}. Иными словами, мы имеем последовательность из $(n-1)$ опытов, причем k-й опыт состоит в выборе одного из $(k+1)$ мест (элементарных событий), каждое из которых имеет вероятность $1/k+1$. Опыты независимы, т. е. вероятности перемножаются. Каждая перестановка из n элементов имеет вероятность $\frac{1}{2} \cdot \frac{1}{3} \cdots \frac{1}{n}$, что совпадает с начальным определением.

б) *Выбор без возвращения.* Пусть $(a_1, \ldots, a_n)$ — генеральная совокупность. При выборе без возвращения с каждым извлечением удаляется один элемент. После k шагов остается $n-k$ элементов, и следующий выбор можно описать заданием номера ν места, занимаемого выбранным элементом $(\nu = 1, 2, \ldots, n-k)$. Таким образом, выборка объема r без возвращения превращается в последовательность из r опытов, причем первый опыт имеет n возможных исходов, второй опыт $n-1$, третий опыт $n-2$ и т. д. Предположим, что все исходы индивидуальных опытов равновероятны, и постулируем независимость всех r опытов. Это означает, что каждой выборке приписывается вероятность $1/(n)_r$, что соответствует нашему определению случайного выбора. [Заметим, что при $n = 100$, $r = 3$ выборка (a_{13}, a_{40}, a_{81}) означает соответственно выбор номеров 13, 39, 79. Мы должны сказать, что при третьем опыте выбран 79-й элемент сокращенной *генеральной совокупности* из $n-2$ элементов, потому что нумерация исходов третьего опыта зависит, помимо первоначальной нумерации элементов, еще и от результатов первых двух извлечений] Мы видим, что понятие повторных независимых опытов позволяет рассматривать выборку как последовательность отдельных операций.

§ 5*) Приложения к генетике

Теория наследственности, ведущая начало от Г. Менделя (1822—1884), представляет поучительную иллюстрацию применения элементарных вероятностных моделей. Мы ограничимся рассмотрением самых простых задач. Биологические представления будут по необходимости излагаться в сильно упрощенном виде, и мы сосредоточим внимание на фактах, допускающих простую математическую трактовку.

Наследование любых признаков зависит от специальных носителей, называемых *генами*. Все клетки тела, исключая половые клетки, или гаметы, несут один и тот же комплекс генов. В каждой такой клетке имеется по паре генов каждого типа. Читатель может представлять их как совокупность огромного количества бусинок или коротких отрезков длинных нитей-хромосом. Для каждой хромосомы в клетке есть парная ей, и парные гены занимают одинаковое положение в парных хромосомах. В простейшем случае каждый ген отдельной пары может находиться в одной из двух форм (аллелей) A или a. Тогда можно образовывать три различные пары, и в соответствии с этим организм имеет один из трех *генотипов* AA, Aa и aa (между парами Aa и aA различия нет). Например, горох несет пару генов, таких, что ген A обусловливает красную окраску цветка, а a—белую окраску. Три генотипа соответствуют в этом случае красному, розовому и белому цветкам.

Каждая пара генов определяет один наследственный признак, но большинство наблюдаемых свойств организма зависит от нескольких факторов. Для некоторых характеристик (таких, как цвет глаз или свойство быть левшой) преобладающим является влияние одной пары генов, и наблюдаемые в этом случае эффекты подчиняются законам Менделя. Другие характеристики, например рост, определяются совместным действием очень большого числа генов (см. пример 5 в гл. X). Здесь мы будем изучать генотипы и наследственность лишь для одной пары генов, по отношению к которой возможны три *генотипа* AA, Aa и aa. Более общим является случай, когда каждый ген может находиться в одной из N форм A_1, A_2, ..., A_N и соответственно

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

этому имеется $\binom{N+1}{2}$ генотипов $A_1A_1, A_1A_2, \ldots, A_NA_N$. С очевидными видоизменениями вся теория применима и в этой ситуации (см. задачу 27). Последующие вычисления применимы также к случаю, когда ген A является *доминантным*, а ген a — *рецессивным*. Это означает, что особи Aa имеют те же наблюдаемые свойства, что и AA, так что влияние гена a проявляется лишь у особей aa. В природе встречаются все степени частичного доминирования. Типичными примерами частично рецессивных свойств является голубой цвет глаз или свойство быть левшой.

В процессе образования половых клеток, или гамет, происходит расщепление, в результате чего каждая гамета содержит лишь по одному гену каждого типа. Поэтому организмы чистых генотипов AA или aa (или гомозиготы) производят гаметы одного сорта, в то время как организмы Aa (гибриды или гетерозиготы) производят в равном количестве гаметы A и a. Новый организм развивается из двух родительских гамет, от которых он и получает свои гены. Поэтому каждая пара генов получает один ген от матери и один от отца; это позволяет установить происхождение любого гена у данной особи, какое бы число поколений не отделяло ее от исходной предковой особи. Генотип потомка зависит от случайного процесса. При любых обстоятельствах каждый родительский ген может передаться с вероятностью $^1/_2$, и последовательные испытания независимы. Другими словами, мы представляем генотипы n потомков как результат n независимых опытов, каждый из которых эквивалентен бросанию пары монет. Например, при скрещивании особей Aa ($Aa \times Aa$) могут появиться генотипы AA, Aa и aa с вероятностями, равными соответственно $^1/_4$, $^1/_2$, $^1/_4$. Скрещивание $AA \times aa$ может привести к появлению лишь особей Aa и т. п.

Рассматривая популяцию в целом, мы представляем образование родительской пары как результат второго случайного процесса. Будем исследовать лишь так называемое *случайное скрещивание*, которое определяется следующим условием: если случайно выбрать r особей в первом поколении, то их родители образуют случайную выборку с возвращением объема r из множества всех возможных родительских пар. Другими словами, каждую особь можно рассматривать как результат скрещивания выбранной пары, и все скрещивания взаимно независимы. Случайное скрещивание является идеализированной моделью условий, которые преобладают во многих встречающихся в природе популяциях и в полевых экспериментах. Однако если один угол поля засеян красным горохом, а другой белым, то пары одного типа будут скрещиваться чаще, чем это допускается гипотезой о случайности. Выбор с предпочтением (например, если блондины предпочитают блондинок) также нарушает условия нашей модели. Самоопыляющиеся растения и искусственное оплодотворение представляют полную противоположность случайному скрещиванию. Некоторые системы со специальными законами скрещивания будут в дальнейшем анализироваться математически, но все же основное внимание будет сосредоточено на модели случайного скрещивания.

Генотип потомка есть результат четырех независимых испытаний. Генотипы двух родителей можно выбрать 3×3 способами, их гены — 2×2 способами. Однако мы можем объединить эти два выбора и описать процесс как двойной выбор следующим образом: отцовский и материнский гены выбираются случайно и независимо один от другого из множества всех генов, которые несут мужские или женские особи родительской популяции. Допустим, что три генотипа AA, Aa и aa встречаются среди мужских и женских особей в одной и той же пропорции $u:2v:w$. Предположим, что $u+2v+w=1$; назовем u, $2v$ и w *частотами генотипов*. Положим

$$p = u + v, \quad q = v + w. \tag{5.1}$$

Ясно, что число генов A относится к числу генов a как $p:q$, и так как $p+q=1$, то мы будем называть p и q частотами генов A и a. При каждом из двух случайных выборов ген A появляется с вероятностью p, и в силу независимости этих выборов вероятность того, что потомок имеет генотип AA, равна p^2. Генотип Aa может появиться двумя способами, и вероятность этого события равна $2pq$. Если выполняется гипотеза о случайном скрещивании, *то потомок имеет один из генотипов AA, Aa или aa с вероятностями*

$$u_1 = p^2, \quad 2v_1 = 2pq, \quad w_1 = q^2. \tag{5.2}$$

Примеры. а) Все родители типа Aa (гетерозиготы). В этом случае $u = w = 0$, $2v = 1$ и $p = q = 1/2$. б) Родительские особи AA и aa смешаны в равном отношении; тогда $u = w = 1/2$, $v = 0$ и опять $p = q = 1/2$. в) Наконец, $u = w = 1/4$, $2v = 1/2$; опять $p = q = 1/2$. Во всех трех случаях распределение генотипов потомков одно и то же: $u_1 = 1/4$, $2v_1 = 1/2$, $w_1 = 1/4$.

Для лучшего понимания сущности соотношений (5.2) фиксируем частоты генов p и q $(p + q = 1)$ и рассмотрим все системы, для которых частоты генотипов u, $2v$ и w удовлетворяют условиям $u + v = p$, $v + w = q$. Во всех системах распределение генотипов в первом поколении потомков дается той же формулой (5.2). Выделим частную систему, для которой

$$u = p^2, \quad 2v = 2pq, \quad w = q^2. \tag{5.3}$$

Если начальные частоты генотипов u, v и w в этой системе сравнить с частотами генотипов в первом поколении, то получим — как и в примере (в), — что $u = u_1$, $v = v_1$ и $w = w_1$. Поэтому распределение генотипов (5.3) можно назвать стационарным. Каждому отношению $p:q$ отвечает стационарное распределение, или положение равновесия.

Уравнения (5.2) устанавливают вероятности появления различных генотипов для случайно выбранной особи во втором поколении. Если объем популяций достаточно велик, то мы вправе ожидать, что действительные частоты генотипов близки к теоретическому распределению[1]). Итак, каково бы ни было распределение генотипов $u:2v:w$ в родительской популяции, уравнения (5.2) определяют стационарное распределение; соответствующие частоты генов A и a [см. (5.1)] равны $u_1+v_1=u+v=p$ и $v_1+w_1=v+w=q$. Другими словами, если наблюдаемые частоты в точности совпадают с вычисленными вероятностями, то распределение генотипов уже в первом поколении потомков становится стационарным и повторяется без изменений в последующих поколениях. На практике возможны отклонения, но для больших популяций мы можем сказать: *каков бы ни был состав родительской популяции, при гипотезе о случайном скрещивании уже в первом поколении получается приблизительно стационарное распределение генотипов с неизмененными частотами генов*. Начиная со второго поколения не существует тенденции к систематическим изменениям; система становится генетически устойчивой уже в первом поколении. Этот факт был впервые отмечен Г. Харди[2]), который тем самым преодолел некоторые трудности, связанные с применением законов Менделя. Отсюда, в частности, вытекает, что при случайном скрещивании частоты трех генотипов должны установиться в отношении $p^2:2pq:q^2$. Это обстоятельство в свою очередь может быть использовано для проверки гипотезы о случайном скрещивании.

Харди также отмечал, что следует сделать ударение на слове «приблизительно». Даже для стационарного распределения мы должны ожидать от поколения к поколению небольших изменений, которые приводят к следующей картине. Какова бы ни была родительская популяция, после случайного скрещивания мы уже в первом поколении приходим к стационарному распределению (5.3). В дальнейшем не существует тенденций к систематическим изменениям, однако частоты генов p и q подвержены случайным колебаниям

---

[1]) Без этого предположения наша вероятностная модель оказалась бы неприменимой. Утверждение можно уточнить, опираясь на закон больших чисел и центральную предельную теорему, которые позволяют оценить эффект случайных колебаний.

[2]) G. H. Hardy, Mendelian proportions in a mixed population, Letter to the Editor, *Science*, N. S., 28 (1908), 49—50. Используя терминологию гл. IX и XV, мы можем описать ситуацию следующим образом. Частоты трех генотипов в n-м поколении являются случайными величинами. Математические ожидания определяются по формуле (5.2) и не зависят от n. Их действительные значения меняются от поколения к поколению и представляют собой случайный процесс марковского типа.

от поколения к поколению и генетическая структура популяций медленно, но меняется. Не существует никаких сил, стремящихся восстановить начальные частоты. Наоборот, наша упрощенная модель приводит к заключению (см. пример 2,л гл. XV), что в ограниченных популяциях в конечном счете один из двух генов изчезает, так что в конечном счете популяция будет состоять лишь из организмов чистых генотипов AA или aa. В природе этот процесс происходит не всегда, так как мутации, селекция и другие эффекты приводят к образованию новых генов. Изучение этих эффектов требует более тонких математических средств (цепи Маркова, теория диффузии).

Часто считают, что из теорем Харди вытекает строгая устойчивость в течение неограниченного времени. Обычной ошибкой является убеждение в том, что закон больших чисел действует как наделенная памятью сила, стремящаяся вернуть систему к исходному состоянию. Такого рода представления породили много неправильных заключений. (Биологические процессы, рассматриваемые здесь, являются типичным и важным классом марковских процессов, изучению которых посвящена гл. XV.) Отметим, что закон Харди неприменим к распределению двух пар генов (например, обусловливающих голубой цвет глаз и свойство быть левшой) с девятью генотипами $AABB$, $AABb, \ldots, aabb$. Здесь все еще имеется тенденция к установлению стационарного распределения, но положение равновесия не достигается в первом поколении.

§ 6*. Сцепленные с полом признаки

Во введении к предыдущему параграфу отмечалось, что гены являются составными частями хромосом. Хромосомы в клетках тела парные и переносятся таким образом, что все гены, лежащие в одной хромосоме, остаются соединенными вместе [1]). Наша схема генной передачи наследственности применима поэтому также и к хромосомам, рассматриваемым как единицы наследственности. Пол определяется парой хромосом, для женщин это XX, для мужчин XY. Материнский организм необходимо переносит хромосому X, и пол потомка определяется хромосомой, переносимой отцовским организмом. Соответственно этому мужские и женские гаметы производятся в равном количестве. Разница в частотах рождения мальчиков и девочек объясняется тем, что они имеют различные шансы выжить в предродовой период.

---

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

[1]) Эта картина несколько осложняется явлениями разрывов и воссоединений хромосом [см. задачу (10.12) гл. II].

Мы уже говорили, что как гены, так и хромосомы встречаются парами Однако имеются исключения, связанные с тем, что гены, лежащие в хромосоме X, не имеют парных им генов в хромосоме Y. Женщины имеют две хромосомы X и, следовательно, двойной набор генов X, но мужчины имеют лишь одинарный набор генов X. Типичными генами, сцепленными с полом, являются гены, обусловливающие дальтонизм и гемофилию. По отношению к ним женщины могут иметь три генотипа AA, Aa и aa, а мужчины, организм которых несет лишь *один* ген, два генотипа A и a. Заметим, что сын всегда наследует отцовскую хромосому Y, так что сцепленные с полом признаки не могут передаваться от отца к сыну. Однако они могут быть переданы от отца дочери, а от дочери — внуку.

Теперь мы обобщим результаты предыдущего параграфа. Опять примем гипотезу о случайном скрещивании и допустим, что частоты генотипов AA, Aa и aa среди особей *женского* пола равны соответственно u, $2v$ и w. Как и раньше, положим $p=u+v$, $q=v+w$. Частоты мужских генотипов A и a обозначим через p' и q' $(p'+q'=1)$. Тогда p' и p будут частотами генов A среди особей мужского и женского пола соответственно. Потомки мужского пола наследуют материнские хромосомы X, и поэтому

$$p'_1=p, \quad q'_1=q. \tag{6.1}$$

Повторяя рассуждения § 5, для трех женских генотипов получаем:

$$u_1=pp', \quad 2v_1=pq'+qp', \quad w_1=qq'. \tag{6.2}$$

Отсюда

$$p_1=u_1+v_1={}^1/_2(p+p'), \quad q_1=v_1+w_1={}^1/_2(q+q'). \tag{6.3}$$

Эти формулы можно интерпретировать следующим образом. Среди потомков мужского пола гены A и a встречаются приблизительно с теми же частотами p и q, что и в материнской популяции. Аналогичные частоты p_1 и q_1 для потомков женского пола равны среднему арифметическому частот генов A и a в отцовской и материнской популяции. Мы наблюдаем тенденцию к сближению частот генов. Действительно, из (6.1) и (6.3) следует, что

$$p'_1-p_1={}^1/_2(p-p'), \quad q'_1-q_1={}^1/_2(q-q'). \tag{6.4}$$

Это означает, что при случайном скрещивании разница в частотах генов A и a среди мужчин и женщин уже в первом поколении сокращается в два раза. Однако эта разница полностью не исчезает, и существует тенденция к дальнейшему сближению. В отличие от ситуации, описываемой законом Харди, распределение генотипов в первом и последующих поколениях не является стационарным. Однако от поколения к поколению можно проследить за систематической компонентой, пренебрегая случайными колебаниями и определяя теоретические вероятности по формулам (6.2) и (6.3), в соответствии с действительными частотами в первом поколении[1]). Для частот A и генов a во втором поколении получим выражения

$$p_2 = {}^1/_2(p_1 + p_1') = {}^3/_4 p + {}^1/_4 p', \quad q_2 = \frac{1}{2}(q_1 + q_1') = \frac{3}{4}q + \frac{1}{4}q' \tag{6.5}$$

и, конечно, $p_2' = p_1$ и $q_2' = q_1$. Повторяя эти рассуждения несколько раз, можно установить общие формулы для вероятностей p_n и q_n, характеризующих распределение генов A и a среди особей женского пола в n-м поколении. Положим

$$\alpha = \frac{1}{3}(2p + p'), \quad \beta = \frac{1}{3}(2q + q'). \tag{6.6}$$

(Заметим, что $\alpha + \beta = 1$.) Тогда

$$\begin{gathered} p_n = \frac{p_{n-1} + p_{n-1}'}{2} = \alpha + (-1)^n \frac{p - p'}{3 \cdot 2^n}, \\ q_n = \frac{q_{n-1} + q_{n-1}'}{2} = \beta + (-1)^n \frac{q - q'}{3 \cdot 2^n} \end{gathered} \tag{6.7}$$

и $p_n' = p_{n-1}$, $q_n' = q_{n-1}$. Следовательно,

$$p_n \to \alpha, \quad p_n' \to \alpha, \quad q_n \to \beta, \quad q_n' \to \beta. \tag{6.8}$$

Частоты генотипов в популяции особей женского пола, согласно (6.2), равны

$$\begin{gathered} u_n = p_{n-1}p_{n-1}', \quad 2v_n = p_{n-1}q_{n-1}' + q_{n-1}p_{n-1}', \\ w_n = q_{n-1}q_{n-1}'. \end{gathered} \tag{6.9}$$

Отсюда следует

$$u_n \to \alpha^2, \quad 2v_n \to 2\alpha\beta, \quad w_n \to \beta^2. \tag{6.10}$$

Эти формулы показывают, что существует сильная систематическая тенденция, сводящая систему к устойчивому состоянию, которое характеризуется тем, что частоты мужских генотипов A и a равны α и β, а частоты женских генотипов AA, Aa и aa равны соответственно α^2, $2\alpha\beta$ и β^2. Формулы (6.7) показывают, что сходимость очень быстрая. На практике распределение генотипов в третьем или четвертом поколениях уже можно считать устойчивым. Конечно, случай-

---

[1]) Используя терминологию, введенную в примечании 2 на стр. 141, мы можем интерпретировать p_n и q_n как математические ожидания частот генов в n-м поколении. При такой интерпретации формулы для p_n и q_n становятся уже не приближенными, а точными.

ные колебания будут слегка искажать нарисованную картину, но последняя правильно отражает основную тенденцию развития.

Главный вывод из всего сказанного состоит в том, что при случайном скрещивании генотипы A и a среди особей мужского пола и генотипы AA, Aa и aa среди особей женского пола должны встречаться приблизительно с частотами α, β, α^2, $2\alpha\beta$, β^2 соответственно, причем $\alpha+\beta=1$.

Примечание. Многие гены, сцепленные с полом, например ген дальтонизма, являются рецессивными и вызывают дефекты. Пусть a — такой ген. Тогда дефект имеют все мужчины a и женщины aa. Женщины типа Aa дефекта не имеют, но могут передавать его своим потомкам. Поэтому мы можем ожидать, что *рецессивный, сцепленный с полом дефект, который встречается у мужчин с вероятностью* α, *у женщин встречается с вероятностью* α^2. Если один из ста мужчин — дальтоник, то женщина с таким же недостатком встречается одна на 10 000.

§ 7.* Селекция

В качестве типичного примера влияния селекции мы исследуем тот случай, когда особи aa не могут размножаться. Это может случиться, если гены a являются рецессивными летальными, так что появившиеся на свет индивидуумы aa сразу же умирают. Кроме того, этого можно добиться искусственным изменением условий размножения или законом, запрещающим скрещивание особей aa.

Предположим, что гипотеза о случайном скрещивании применима к совокупности всех особей AA и Aa, но нет скрещивания индивидуумов aa. Пусть частоты генотипов AA, Aa и aa в *начальной* популяции равны u, $2v$ и w соответственно. Аналогичные частоты в совокупности *родителей* равны

$$u^*=\frac{u}{1-w},\quad 2v^*=\frac{2v}{1-w},\quad w^*=0. \tag{7.1}$$

В дальнейшем можно рассуждать, как в § 5, используя, однако, вместо величин u, $2v$, w величины (7.1). Соотношения (5.1) заменяются другими соотношениями:

$$p=\frac{u+v}{1-w},\qquad q=\frac{v}{1-w}. \tag{7.2}$$

Вероятности трех генотипов в первом поколении опять определяются формулами (5.2), т. е. $u_1=p^2$, $2v_1=2pq$, $w_1=q^2$.

Как и раньше, для того чтобы проследить систематические изменения от поколения к поколению, мы должны заменить u, v, w

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

на u_1, v_1, w_1 и получить вероятности генотипов во втором поколении u_2, v_2, w_2 и т. д. Вообще, исходя из (7.2), получим

$$p_n = \frac{u_n + v_n}{1 - w_n}, \qquad q_n = \frac{v_n}{1 - w_n} \tag{7.3}$$

и

$$u_{n+1} = p_n^2, \quad 2v_{n+1} = 2p_n q_n, \quad w_{n+1} = q_n^2. \tag{7.4}$$

Сопоставляя (7.3) и (7.4), найдем, что

$$p_{n+1} = \frac{u_{n+1} + v_{n+1}}{1 - w_{n+1}} = \frac{p_n}{1 - q_n^2} = \frac{1}{1 + q_n} \tag{7.5}$$

и аналогично

$$q_{n+1} = \frac{v_{n+1}}{1 - w_{n+1}} = \frac{q_n}{1 + q_n}. \tag{7.6}$$

Используя (7.6), нетрудно вычислить q_n в явном виде. Действительно

$$\frac{1}{q_{n+1}} = 1 + \frac{1}{q_n}, \tag{7.7}$$

и отсюда последовательно

$$\frac{1}{q_1} = 1 + \frac{1}{q}; \quad \frac{1}{q_2} = 2 + \frac{1}{q}; \quad \ldots; \quad \frac{1}{q_n} = n + \frac{1}{q} \tag{7.8}$$

или

$$q_n = \frac{q}{1 + nq}, \qquad w_{n+1} = \left(\frac{q}{1 + qn}\right)^2. \tag{7.9}$$

Мы видим, что генотип, не способный к воспроизведению (или нежелательный), постепенно исчезает, но процесс этот крайне медленный. При $q = 0{,}1$ требуется десять поколений, чтобы наполовину сократить частоту генов a; при этом частота особей, имеющих генотип aa изменяется с 1 до 0,25%. (Если ген a сцеплен с полом, то процесс вымирания значительно быстрее, как показано в задаче 29; обобщенной схеме селекции посвящена задача 30 [1].)

§ 8. Задачи

1. Бросают три кости. Какова вероятность того, что хотя бы на одной из них выпадет одно очко, если на всех трех костях выпали разные грани?

2. Известно, что при бросании 10 костей появилась по крайней мере одна единица. Какова вероятность того, что появилось две или более единицы?

---

[1]) Дальнейший анализ разнообразных евгенических эффектов (которые часто отличны от того, что думают горячие сторонники законов о стерилизации) см. G. Dahlberg, Mathematical methods for population genetics, New York and Basel, 1948.

3. Бридж. При сдаче карт для игры в бридж первый из игроков не получил тузов. Какова вероятность того, что его партнер а) не имеет тузов, б) имеет не менее двух тузов?

4. Бридж. Второй и четвертый игроки имеют вместе десять козырей (козыри — карты определенной масти). а) Найти вероятность того, что все три оставшихся козыря находятся у одного игрока (т. е. либо первый, либо третий игрок не имеет козырей). б) Известно, что среди трех оставшихся козырей имеется король. Какова вероятность того, что он является единственным козырем у игрока, имеющего его на руках?

5. Обсудить задачу о ключах (пример 7, б гл. II) в терминах условных вероятностей, используя метод примера 2, б.

6. На фабрике, изготовляющей болты, машины A, B, C производят соответственно 25, 35 и 40% всех изделий. В их продукции брак составляет соответственно 5, 4 и 2%. Случайно выбранный из продукции болт оказался дефектным. Каковы вероятности того, что он был произведен машинами A, B, C?

7. Предположим, что 5% всех мужчин и 0,25% всех женщин дальтоники. Наугад выбранное лицо страдает дальтонизмом. Какова вероятность того, что это мужчина? (Считать, что мужчин и женщин одинаковое число.)

8. Семь шаров случайно распределяются по семи ящикам; вероятности различных размещений представлены в табл. 1 § 5 гл. II. Используя эту таблицу, доказать, что условная вероятность события «найдется ящик, содержащий три шара» при условии, что имеется ровно два пустых ящика, равна $^1/_4$.

9. Кость бросается до тех пор, пока не выпадет очко. Предполагая, что при первом испытании очко не выпало, найти вероятность того, что потребуется не менее трех бросаний.

10. Продолжение. Допустим, что число n испытаний четно. Какова вероятность того, что $n=2$.

11. Пусть[1]) вероятность p_n того, что в семье n детей, равна αp^n при $n \geqslant 1$, и $p_0 = 1 - \alpha p\,(1 + p + p^2 + \ldots)$. Допустим, что все комбинации полов n детей равновероятны. Показать, что при $k \geqslant 1$ вероятность того, что в семье k мальчиков, равна

$$\frac{2\alpha p^k}{(2-p)^{k+1}}.$$

12. Продолжение. Пусть известно, что в семье есть по крайней мере один мальчик. Какова вероятность того, что в ней два или более мальчиков?

13. Кость A имеет четыре красных и две белых грани, в то время как кость B имеет, наоборот, две красных и четыре белых грани. Бросаем монету. Если выпадет герб, то продолжение игры состоит в том, что один раз бросаем кость A, если выпадет решетка, то один раз бросаем кость B. а) Доказать, что вероятность выпадения красной грани при каком-либо испытании равна $^1/_2$. б) Первые два испытания привели к выпадению красных граней. Какова вероятность того, что и при третьем испытании выпадет красная грань? в) Первые n испытаний привели к выпадению только красных граней. Какова условная вероятность того, что хотя бы раз бросалась кость A? г) Какой урновой модели отвечает эта схема?

---

[1]) Согласно A. J. Lotka, американская семейная статистика удовлетворяет этой гипотезе с $p = 0{,}7358$. См. Théorie analytique des associations biologiques II. *Actualités scientifiques et industrielles.* № 780, Paris, 1939.

14. В примере 2, а обозначим через x_n вероятность того, что игрок, победивший в n-й встрече, выиграет и всю игру, через y_n и z_n соответственно обозначим вероятности выиграть всю игру для игрока, потерпевшего поражение в n-й встрече и не принимавшего в ней участия. а) Показать, что

$$x_n = \frac{1}{2} + \frac{1}{2} y_{n+1}, \quad y_n = \frac{1}{2} z_{n+1}, \quad z_n = \frac{1}{2} x_{n+1}. \tag{$*$}$$

б) Непосредственно из рассмотрения пространства элементарных событий установить, что в действительности независимо от n $x_n = x$, $y_n = y$, $z_n = z$. в) Доказать, что вероятность того, что игрок a выигрывает игру, равна $^5/_{14}$ (в соответствии с задачей 5 гл. I). г) Показать, что $x_n = {}^4/_7$, $y_n = {}^1/_7$, $z_n = {}^2/_7$ есть единственное ограниченное решение системы ($*$).

15. Пусть события $A_1, A_2, \ldots, A_n$ независимы, и $P\{A_k\} = p_k$. Найти вероятность p непоявления ни одного из этих событий.

16. Продолжение. Показать, что всегда $p < e^{-\Sigma p_k}$.

17. Продолжение. Вывести из неравенства Бонферрони, что вероятность одновременного осуществления k или более событий из $A_1, A_2, \ldots$ $\ldots, A_n$ меньше, чем $(p_1 + \ldots + p_n)^k/k!$.

18. К урновой схеме Пойа, пример 2, в). Какова вероятность того, что первый шар был черным, если второй шар оказался черным?

19. К урновой схеме Пойа, пример 2, в). Показать с помощью индукции, что вероятность извлечения черного шара при каком-либо испытании равна $b/(b+r)$.

20. Продолжение. Доказать по индукции: для любых $m < n$ вероятности того, что m-е и n-е испытания дадут комбинации (ч, ч) или (ч, к), равны соответственно

$$\frac{b(b+c)}{(b+r)(b+r+c)}, \quad \frac{br}{(b+r)(b+r+c)}.$$

Обобщить на случай более чем двух испытаний.

21. Симметрия по времени в схеме Пойа. Пусть A и B означают каждое или черный, или красный цвет (так что AB может быть любой из четырех комбинаций). Показать, что вероятность цвета A при n-м испытании при условии, что m-е испытание дало цвет B, равна вероятности цвета A при m-м испытании при условии, что n-е испытание дало цвет B.

22. Пусть в схеме Пойа $p_k(n)$ — вероятность извлечения k черных шаров в первых n испытаниях. Доказать рекуррентное соотношение

$$p_k(n+1) = p_k(n)\frac{r+(n-k)c}{b+r+nc} + p_{k-1}(n)\frac{b+(k-1)c}{b+r+nc},$$

где $p_{-1}(n)$ принимается равным нулю. Использовать это соотношение для нового доказательства формулы (2.5).

23. Распределение Пойа. В (2.4) положим для краткости

$$\frac{b}{b+r} = p, \quad \frac{r}{b+r} = q, \quad \frac{c}{b+r} = \gamma. \tag{8.1}$$

Показать, что выражение

$$p_{n_1, n} = \frac{\left(-\frac{p}{\gamma}\right)_{n_1}\left(-\frac{q}{\gamma}\right)_{n_2}}{(-1/\gamma)_n}, \quad n_1 + n_2 = n, \tag{8.2}$$

имеет смысл для произвольных (не обязательно рациональных) постоянных $p > 0$, $q > 0$, $\gamma > -1$, таких, что $p + q = 1$. Проверить, что $p_{n_1, n} > 0$ и что

$$\sum_{\nu=0}^{n} p_{\nu, n} = 1.$$

Уравнение (8.2) определяет распределение Пойа.

24. Предельная форма распределения Пойа. Если $n \to \infty$, $p \to 0$, $\gamma \to 0$ так, что $np \to \lambda$, $n\gamma \to \rho^{-1}$, то

$$p_{n_1, n} \to \binom{\lambda\rho + n_1 - 1}{n_1} \left(\frac{\rho}{\rho + 1}\right)^{\lambda\rho} \left(\frac{1}{1 + \rho}\right)^{n_1}.$$

Проверить это и доказать, что при фиксированных λ и ρ члены справа дают в сумме единицу. (Правая часть определяет так называемое отрицательное биномиальное распределение; см. § 8 гл. VI и задачу 37 гл. VI.)

25. Интерпретировать уравнение (11.8) гл. II в терминах условных вероятностей.

Приложения к биологии

26. При случайном скрещивании менее половины особей, входящих в популяцию, принадлежит генотипу *Aa*.

27. Обобщить результаты § 5 на случай, когда каждый ген может находиться в одной из n форм $A_1, A_2, \ldots, A_k$, так что существует $\binom{k+1}{2}$ генотипов, а не три, как это предполагалось при рассмотрениях § 5.

28. Из некоторой совокупности, в которой генотипы *AA*, *Aa* и *aa* встречаются с частотами u, $2v$ и w, случайно выбираются две особи. Эти особи скрещиваются, и с их потомством повторяется та же процедура. Найти вероятности того, что оба родителя в первом, втором, третьем поколениях принадлежат генотипу *AA* (см. примеры 2, м гл. XV и 4, б гл. XVI).

29. Селекция. Пусть a — рецессивный, сцепленный с полом ген. Предположим, что процесс селекции делает невозможным скрещивание двух мужских особей a. Если генотипы *AA*, *Aa* и *aa* в начальной совокупности встречаются с частотами u, $2v$, w, то показать, что в первом поколении $u_1 = u + v$, $2v_1 = v + w$, $w_1 = 0$, так что $p_1 = p + q\frac{1}{2}$, $q_1 = \frac{1}{2}q$. Короче говоря, частота генов a уменьшается в два раза.

30. Обобщим задачу о селекции § 7, предполагая, что уничтожается лишь некоторая часть λ $(0 < \lambda \leqslant 1)$ особей *aa*.

Показать, что

$$p = \frac{u + v}{1 - \lambda w}, \quad q = \frac{v + (1 - \lambda) w}{1 - \lambda w}.$$

Вообще вместо (7.3) можно записать

$$p_{n+1} = \frac{p_n}{1 - \lambda q_n^2}, \quad q_{n+1} = q_n \frac{1 - \lambda q_n}{1 - \lambda q_n^2}.$$

(Общее решение этих уравнений, по-видимому, неизвестно.)

31. Рассмотрим одновременно две пары генов с возможными формами (A, a) и (B, b). Любая особь передает своим прямым потомкам один ген из каждой пары, так что каждая из четырех возможных комбинаций имеет вероятность $\frac{1}{4}$ (это предположение справедливо лишь в том случае, если

разные гены принадлежат разным хромосомам, иначе имеет место строгая зависимость). Существует девять генотипов, и мы обозначаем их частоты в полной популяции через U_{AABB}, U_{aaBB}, U_{AAbb}, U_{aabb}, $2U_{AaBB}$, $2U_{Aabb}$, $2U_{AABb}$, $2U_{aaBb}$, $4U_{AaBb}$. Положим $p_{AB}=U_{AABB}+U_{AABb}+U_{AaBB}+U_{AaBb}$, $p_{Ab}=U_{AAbb}+U_{Aabb}+U_{AABb}+U_{AaBb}$, $p_{aB}=U_{aaBB}+U_{aaBb}+U_{AaBB}+$ $+U_{AaBb}$, $p_{ab}=U_{aabb}+U_{Aabb}+U_{aaBb}+U_{AaBb}$. Вывести частоты, характеризующие распределение генотипов в первом поколении. Показать, что $p_{AB}^{(1)}=p_{AB}-\delta$, $p_{Ab}^{(1)}=p_{Ab}+\delta$, $p_{aB}^{(1)}=p_{aB}+\delta$, $p_{ab}^{(1)}=p_{ab}-\delta$, где $2\delta=$ $=p_{AB}p_{ab}-p_{Ab}p_{aB}$. Стационарное распределение определяется системой равенств $p_{AB}-2\delta=p_{Ab}+2\delta$ и т. д. (Отметим, что закон Харди неприменим; композиция меняется от поколения к поколению.)

32. Предположим, что частоты генотипов в совокупности равны $u=p^2$, $2v=2pq$, $w=q^2$. Известно, что некто имеет генотип Aa, тогда вероятность (условная) того, что генотип его брата тоже Aa, равна $(1+pq)/2$.

Замечание[1]). *Последующие задачи очень близки. Они связаны с понятием степени родства. Каждая задача продолжает предыдущую. Предположение о случайном скрещивании и обозначения § 5 сохраняется. В сущности, здесь мы имеем дело со специальным случаем цепей Маркова (см. гл. XVI). Употребление матриц упрощает запись.*

33. Занумеруем генотипы AA, Aa и aa числами 1, 2, 3 соответственно. Пусть p_{ik} (i, $k=1, 2, 3$) — условная вероятность того, что особь имеет генотип k, при условии, что одна из родительских особей (мужская или женская) имеет генотип i. Вычислить все девять вероятностей p_{ik}, предполагая, что другая родительская особь имеет один из генотипов 1, 2, 3 с вероятностями p^2, $2pq$ и q^2 соответственно.

34. Показать, что p_{ik} является в то же время условной вероятностью того, что родительская особь имеет генотип k, при условии, что заданный потомок имеет генотип i.

35. Показать, что условная вероятность того, что внук (дедушка) имеет генотип k, при условии, что дедушка (внук) имеет генотип i, определяется соотношением

$$p_{ik}^{(2)}=p_{i1}p_{1k}+p_{i2}p_{2k}+p_{i3}p_{3k}.$$

[Матрица $\left(p_{ik}^{(2)}\right)$ является квадратом матрицы (p_{ik}).]

36. Показать, что $p_{ik}^{(2)}$ является также условной вероятностью того, что человек имеет генотип k, при условии, что некий двоюродный брат имеет генотип i.

37. Показать, что условная вероятность того, что некто имеет генотип k, при условии, что определенный его прадедушка (правнук) имеет генотип i, определяется формулой

$$p_{ik}^{(3)}=p_{i1}^{(2)}p_{1k}+p_{i2}^{(2)}p_{2k}+p_{i3}^{(2)}p_{3k}=p_{i1}p_{1k}^{(2)}+p_{i2}p_{2k}^{(2)}+p_{i3}p_{3k}^{(2)}.$$

[Матрица $\left(p_{ik}^{(3)}\right)$ — куб матрицы $\left(p_{ik}\right)$. Эти задачи позволяют составить мнение о степени родства в зависимости от поколения.]

---

[1]) C. C. Li and L. Sacks, The derivation of the joint distribution and correlation between relatives by the use of stochastic matrices, *Biometrica*, 40 (1954), 347—360.

38. Вообще определим вероятности $p_{ik}^{(n)}$ того, что потомок в n-м поколении имеет генотип k, если некий предок имеет генотип i. Доказать по индукции, что $p_{ik}^{(n)}$ являются элементами следующей матрицы:

$$\begin{pmatrix} p^2+pq/2^{n-1} & 2pq+q(q-p)/2^{n-1} & q^2-q^2/2^{n-1} \\ p^2+p(q-p)/2^n & 2pq+(1-4pq)/2^n & q^2+q(p-q)/2^n \\ p^2-p^2/2^{n-1} & 2pq+p(p-q)/2^{n-1} & q^2+pq/2^{n-1} \end{pmatrix}$$

Отсюда следует, что влияние предка убывает от поколения к поколению в два раза.)

39. Рассмотреть задачу 36, заменив двоюродного брата на брата. Показать, что соответствующая матрица имеет вид

$$\begin{Vmatrix} \frac{1}{4}(1+p)^2 & \frac{1}{2}q(1+pq) & \frac{1}{4}q^2 \\ \frac{1}{4}p(1+p) & \frac{1}{2}(1+pq) & \frac{1}{4}q(1+q) \\ \frac{1}{4}p^2 & \frac{1}{2}p(1+q) & \frac{1}{4}(1+q)^2 \end{Vmatrix}.$$

40. Показать, что степень родства между дядей и племянником такая же как между внуком и дедушкой.

ГЛАВА VI

БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ И РАСПРЕДЕЛЕНИЕ ПУАССОНА

§ 1. Испытания Бернулли[1])

Повторные независимые испытания называются испытаниями Бернулли, если при каждом испытании имеется только два возможных исхода и вероятности этих исходов остаются неизменными для всех испытаний. Принято называть исходы «*успехом*» $У$, и «*неудачей*» $Н$, а их вероятности обозначать соответственно буквами p и q. Ясно, что p и q должны быть неотрицательными, причем

$$p+q=1. \tag{1.1}$$

Пространство элементарных событий для каждого отдельного испытания состоит из двух точек $У$ и $Н$. Пространство элементарных событий для n испытаний Бернулли содержит 2^n точек или последовательностей из n символов $У$ и $Н$; каждая точка представляет один возможный исход составного опыта. Поскольку опыты независимы, вероятности перемножаются. Иными словами, *вероятность какой-либо определенной последовательности равна произведению, полученному из этой последовательности соответственной заменой букв $У$ и $Н$ на p и q. Так, $P\{УУНУН\ldots$ $\ldots ННУ\} = ppqpq\ldots qqp$.*

Примеры. Наиболее известным примером испытаний Бернулли являются последовательные бросания правильной, т. е. симметричной монеты; здесь $p=q=1/2$. В случае несимметричной монеты мы по-прежнему считаем последовательные бросания независимыми, так что имеем модель испытаний Бернулли, в которой вероятность успеха p может быть произвольной.

Повторные случайные извлечения из урны, содержащей при каждом опыте r красных и b черных шаров, являются испытаниями Бернулли с

$$p=\frac{r}{r+b}.$$

Часто бывает так, что существует несколько возможных исходов, но рассматривать каждый из них отдельно нежелательно, и удобнее

[1]) Яков Бернулли (1654—1705). Его основная работа „Ars Conjectandi“ была опубликована в 1713 г.

разбить эти исходы на группу «положительных» и группу «отрицательных» исходов. Так, в случае правильной кости разбиение всех возможностей на выпадение одного очка ($У$) и невыпадение одного очка ($Н$) приводит к испытаниям Бернулли с $p = 1/6$, а разбиение на выпадение четного и нечетного числа очков приводит к испытаниям Бернулли с $p = 1/2$. Если кость несимметрична,то последовательные бросания по-прежнему представляют собой испытания Бернулли, но соответствующие вероятности p изменят свои значения.

«Успехом» можно назвать также получение первым игроком десятки, валета, дамы, короля и туза одной масти при сдаче карт для игры в покер или выпадение двух единиц при бросании двух костей; называя все прочие исходы «неудачей», мы получим испытания Бернулли с $p = 1/649740$ и $p = 1/36$. Схемы такого рода часто удобны в статистических приложениях. Например, шайбы, изготовляемые при массовом производстве, могут непрерывно меняться по толщине, но при проверке они классифицируюся на годные ($У$) и дефектные ($Н$) в зависимости от того, находится ли толщина в предписанных границах.

Испытания Бернулли—теоретическая схема, и только практика может показать, годна ли схема для описания данного физического опыта. То, что последовательные бросания монеты подходят под схему Бернулли, выведено опытным путем. Марбе[1]) полагает, что после серии из 17 выпадений герба выпадение решетки становится более вероятным. При этом он наделяет природу памятью или, по нашей терминологии, отрицает независимость последовательных испытаний. Теория Марбе должна быть отвергнута, так как она не подтверждается опытами.

Для промышленного контроля качества продукции и т. п. схема испытаний Бернулли дает идеальный стандарт, несмотря даже на то, что этот стандарт никогда не достигается вполне точно. Так, в приведенном выше примере производства шайб продукция по многим причинам не может вполне соответствовать схеме Бернулли. Машины подвержены изменениям, и поэтому вероятности не остаются одними и теми же; в режиме работы машин имеется некоторое постоянство, в результате чего длинные серии одинаковых отклонений оказываются более вероятными, чем это было бы при действительной независимости испытаний. Однако с точки зрения контроля качества продукции желательно, чтобы процесс соответствовал схеме Бернулли, и важно то, что в некоторых пределах этого можно добиться. Целью текущего контроля является обнаружение уже на ранней стадии существенных отступлений от идеальной схемы и использование их как указаний на угрожающее нарушение правильности работы машины.

---

[1]) K. Marbe, Die Gleichförmigkeit in der Welt, Munich, 1916. Существует обширная критическая литература, посвященная теории Марбе.

§ 2. Биномиальное распределение

Часто представляет интерес лишь суммарное число успехов, достигнутых в последовательности из n испытаний Бернулли, независимо от порядка их следования. Число успехов может быть равно $0, 1, \ldots, n$, и нашей первой задачей является определить соответствующие вероятности. Событие «n испытаний привели k раз к успеху и $n-k$ раз к неудаче» содержит столько элементарных событий, сколькими способами можно k букв $У$ распределить по n местам. Это значит, что наше событие содержит $\binom{n}{k}$ точек, а каждая точка, по определению, имеет вероятность $p^k q^{n-k}$. Этим доказана следующая

Теорема. *Пусть $b(k; n, p)$ — вероятность того, что n испытаний Бернулли с вероятностью успеха p и вероятностью неудачи $q = 1 - p$ привели k раз к успеху и $n-k$ раз к неудаче $(0 \leqslant k \leqslant n)$. Тогда*

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k}. \tag{2.1}$$

В частности, вероятность не иметь ни одного успеха равна q^n, а вероятность иметь по крайней мере один успех равна $1 - q^n$.

Считая p постоянной, обозначим число успехов n в испытаниях через S_n; тогда $b(k; n, p) = P\{S_n = k\}$. Согласно общей терминологии, S_n является *случайной величиной*, а функция (2.1) определяет «распределение» этой случайной величины; мы будем называть это распределение *биномиальным распределением*. Эпитет «биномиальное» связан с тем, что (2.1) представляет собой k-й член биномиального разложения $(p+q)^n$. Это замечание показывает также, что

$$b(0; n, p) + b(1; n, p) + \ldots + b(n; n, p) = (p+q)^n = 1,$$

как того и требует понятие вероятности. Биномиальное распределение табулировано [1]).

а) *Данные Уэлдона о бросании костей.* Рассмотрим опыт, состоящий в бросании 12 костей, причем для каждой кости «успехом» считается выпадение пяти или шести очков. Если кости правильные, то вероятность успеха $p = 1/3$, и число успехов должно следовать биномиальному распределению $b(k; 12, 1/3)$. В табл. 1 приведены эти вероятности вместе с соответствующими средними частотами, наблюденными при 26 306 действительных опытах. Данные кажутся вполне согласованными, но на самом деле для столь большого числа опытов

[1]) Для $n \leqslant 50$ см. National Bureau of Standards, Tables of the binomial probability distribution, *Applied Math. Series*, 6 (1950). Для $50 \leqslant n \leqslant 100$ см. Roming H.C., 50—100 Binomial tables, John Wiley and Sons, 1953. Для более широкого класса см. Tables of the cumulative binomial probability distribution, by the Harvard Computation Laboratory, 1955 и Tables of the cumulative binomial probabilities, by the Ordnance Corps, ORDP 20—11 (1952).

согласие очень плохое. Статистики обычно судят о степени согласия с помощью критерия χ^2. Согласно этому критерию, при правильной кости такое большое отклонение, как наблюденное, происходило бы лишь в четырех из 10 000 случаев. Поэтому законно допустить, что кости не были симметричны. Асимметрия, приводящая к вероятности успеха $p = 0{,}3377$, согласуется с наблюдениями[1]).

б) В § 4 гл. IV мы пользовались биномиальным распределением в задаче об угадывании карт; в столбцах табл. 3 стоят члены биномиального распределения при $n = 3,\ 4,\ 5,\ 6,\ 10$ и $p = 1/n$. В задаче о случайном размещении шаров по ящикам (4.в) гл. II мы получили формулу (4.5) гл. II, которая является другим частным случаем биномиального распределения.

в) Сколько испытаний потребуется для того, чтобы сделать вероятность хотя бы одного успеха не менее 1/2, если вероятность успеха при одном испытании равна 0,01? Здесь мы ищем наименьшее целое положительное n, для которого $1-(0{,}99)^n \geqslant 1/2$ или $-n \log 0{,}99 \geqslant \log 2$; отсюда $n \geqslant 70$.

Таблица 1

Данные Уэлдона о бросании костей

k	$b(k;\ 12,\ 1/3)$	Наблюденная частота	$b(k;\ 12,\ 0{,}3377)$
0	0,007707	0,007033	0,007123
1	0,046244	0,043678	0,043584
2	0,127171	0,124116	0,122225
3	0,211952	0,208127	0,207736
4	0,238446	0,232418	0,238324
5	0,190757	0,197445	0,194429
6	0,111275	0,116589	0,115660
7	0,047689	0,050597	0,050549
8	0,014903	0,015320	0,016109
9	0,003312	0,003991	0,003650
10	0,000497	0,000532	0,000558
11	0,000045	0,000152	0,000052
12	0,000002	0,000000	0,000002

г) *Задача о снабжении энергией.* Допустим, что $n = 10$ рабочих время от времени используют электрическую энергию. Чтобы получить грубое представление об ожидаемой нагрузке, представим себе,

[1]) R. A. Fisher, Statistical methods for research workers, Edinburgh—London, 1932, 66 или T. C. Fry, Probability and its engineering uses, N. Y., 1928, 303.

что в любой момент времени каждому рабочему с одной и той же вероятностью p может потребоваться единица энергии. Если они работают независимо, то вероятность того, что энергия потребуется одновременно k рабочим, равна $b(k;\ n,\ p)$. Если один рабочий потребляет энергию в среднем 12 *мин* в течение часа, то следует положить $p=\frac{1}{5}$. В этом случае вероятность того, что не менее семи рабочих будут одновременно использовать электрическую энергию, равна

$$b(7;\ 10,\ 0,2)+\ldots+b(10;\ 10,\ 0,2)=0{,}0008643584.$$

Другими словами, если снабжение рассчитано на шесть единиц энергии, то вероятность перегрузки равна 0,00086. Это означает, что одна перегрузка приходится в среднем на 1157 *мин*, т.е. приблизительно на двенадцать часов рабочего времени. Вероятность того, что электроэнергию будут одновременно потреблять не менее восьми рабочих, равна всего лишь 0,0000779264, или почти в одиннадцать раз меньше.

д) *Проверка вакцин*[1]. Допустим, что нормальная частота заболеваний определенной болезнью среди крупного рогатого скота 25%. Для проверки новой вакцины выбирают здоровых животных и делают прививки. Как оценить результаты эксперимента? Если вакцина абсолютно недейственна, то вероятность того, что ровно k из n испытуемых животных не заболеют, равна $b(k;\ n,\ 0{,}75)$. При $k=n=10$ эта вероятность равна приближенно 0,056, а при $k=n=12$ — только 0,032. Поэтому отсутствие заболеваний в партии из десяти или двенадцати испытуемых животных можно расценивать как указание на эффективность вакцины, хотя, конечно, такое „доказательство“ нельзя считать окончательным. Отметим, что вероятность того, что среди семнадцати животных, не получивших прививки, заболеет не более одного, равна приблизительно 0,0501. Отсюда вытекает, что одно заболевание в партии из семнадцати животных является *более сильным свидетельством* в пользу вакцины, чем отсутствие заболеваний в группе из десяти животных. Для $n=23$ вероятность не более двух заболеваний приближенно равна 0,0492, и поэтому два заболевания среди двадцати трех животных опять являются лучшим свидетельством эффективности вакцины, чем одно среди семнадцати или отсутствие заболеваний среди десяти.

е) *Другой статистический критерий*. Допустим, что у человека измеряется кровяное давление сначала до введения некоторого лекарства, а затем после этого. Пусть наблюдения дали результаты $x_1, \ldots, x_n$ и $x'_1, \ldots, x'_n$. Будем говорить что i-е испытание

---

[1] P. V. Sukhatme, V. G. Panse, Size of experiments for testing sera or vaccines, *Indian Journal of Veterinary Science and Animal Husbandry*, 13 (1943), 75—82.

привело к успеху, если $x_i < x'_i$, и к неудаче, если $x_i > x'_i$. (Можно считать, что не существует двух в точности одинаковых результатов.) Если лекарство не приводит к эффекту, то наши испытания соответствуют схеме Бернулли с вероятностью успеха $p = {}^1/_2$. Большое число успехов свидетельствует о действенности лекарства.

§ 3. Максимальная вероятность в биномиальном распределении

Из формулы (2.1) видно, что

$$\frac{b(k;\, n,\, p)}{b(k-1;\, n,\, p)} = \frac{(n-k+1)\,p}{kq} = 1 + \frac{(n+1)\,p - k}{kq}. \tag{3.1}$$

В соответствии с этим вероятность $b(k;\, n,\, p)$ больше предшествующей вероятности $b(k-1;\, n,\, p)$, если $k < (n+1)\,p$, и меньше ее, если $k > (n+1)\,p$. Если $(n+1)\,p = m$ — целое число, то $b(m;\, n,\, p) = b(m-1;\, n,\, p)$. Имеется только одно целое число m, такое, что

$$(n+1)\,p - 1 < m \leqslant (n+1)\,p, \tag{3.2}$$

и мы приходим к следующей теореме.

Теорема 1. *Если k изменяется от 0 до n, то вероятность $b(k;\, n,\, p)$ сначала монотонно возрастает, а затем монотонно убывает, достигая своего наибольшего значения при $k = m$, исключая тот случай, когда при $m = (n+1)\,p$ имеет место равенство $b(m-1;\, n,\, p) = b(m;\, n,\, p)$.*

Мы будем называть $b(m;\, n,\, p)$ *максимальной* вероятностью. Часто m называют «наиболее вероятным числом успехов». Следует, однако, помнить, что при больших значениях n все вероятности $b(k;\, n,\, p)$, включая и максимальную, малы. При ста бросаниях правильной монеты наиболее вероятное число выпадений герба равно 50, но вероятность такого события меньше, чем 0,08. В следующей главе мы увидим, что $b(m;\, n,\, p) \approx (2\pi npq)^{-1/2}$.

Очевидно, что отношение, стоящее в правой части формулы (3.1), монотонно убывает с возрастанием k; таким образом, при $k \geqslant r+1$

$$\frac{b(k;\, n,\, p)}{b(k-1;\, n,\, p)} < \frac{(n-r)\,p}{(r+1)\,q}. \tag{3.3}$$

Полагая последовательно $k = r+1, \ldots, r+\nu$ и перемножая ν неравенств, получаем

$$\frac{b(r+\nu;\, n,\, p)}{b(r;\, n,\, p)} < \left\{\frac{(n-r)\,p}{(r+1)\,q}\right\}^{\nu}. \tag{3.4}$$

При $r \geqslant np$ дробь в скобках меньше, чем единица, и суммирование по ν приводит к конечной геометрической прогрессии со знаменателем $(n-r)\,p/(r+1)\,q$. Отсюда вытекает, что при $s \leqslant np$

$$\sum_{\nu=0}^{n-r} b(r+\nu;\ n,\ p) < b(r;\ n,\ p)\,\frac{(r+1)\,q}{r+1-(n+1)\,p} \tag{3.5}$$

эта формула позволяет оценивать правый «хвост» биномиального распределения, а именно вероятность не менее чем r успехов. Аналогичные рассуждения в применении к левому «хвосту» показывают, что при $s \leqslant np$

$$\sum_{\rho=0}^{s} b(\rho;\ n,\ p) < b(s;\ n,\ p)\,\frac{(n-s+1)\,p}{(n+1)\,p-s}. \tag{3.6}$$

Итак, доказана

Теорема 2. *Если $r \geqslant np$, то вероятность не менее чем r успехов удовлетворяет неравенству* (3.5); *если $s \leqslant np$, то вероятность не более чем s успехов удовлетворяет* (3.6).

Другое доказательство дает задача 39, а.

§ 4. Закон больших чисел

Мы несколько раз отмечали, что наше интуитивное представление о вероятности основано на следующем предположении. Если при n одинаковых испытаниях событие A произошло ν раз, и если n очень велико, тогда ν/n должно быть близко к вероятности p события A. Ясно, что формальная математическая теория не может относиться непосредственно к действительной жизни, но она должна по крайней мере указывать теоретическую модель явления, которое она пытается объяснить. Поэтому мы требуем, чтобы неясное вводное замечание стало точной теоремой.

Если S_n — число успехов в n испытаниях, то средняя доля успехов S_n/n должна быть близка к p. Нетрудно теперь придать этому утверждению точный смысл. Рассмотрим, например, вероятность того, что S_n/n превосходит $p+\varepsilon$, где $\varepsilon > 0$ произвольно малое, но фиксированное число. Эта вероятность совпадает с $P\{S_n > n(p+\varepsilon)\}$ и равна левой части (3.5), где r представляет наименьшее целое число, превосходящее $n(p+\varepsilon)$. Неравенство (3.5) позволяет заключить, что

$$P\{S_n > n(p+\varepsilon)\} < b(r;\ n,\ p)\,\frac{n(p+\varepsilon)+q}{n\varepsilon+q}. \tag{4.1}$$

При возрастании n дробь в правой части этой формулы остается ограниченной. В то же время $b(r;\ n,\ p) \to 0$, так как $b(r;\ n,\ p) < b(k;\ n,\ p)$ для любого $k(n+1)\,p \leqslant k < r$, а таких членов имеется приблизительно $n\varepsilon$. Отсюда вытекает, что при неограниченном возрастании n имеет место предельное соотношение $P\{S_n > n(p+\varepsilon)\} \to 0$. Используя формулу (3.6), таким же образом можно доказать, что $P\{S_n < n(p-\varepsilon)\} \to 0$. Объединяя эти неравенства, получаем, что

$$P\left\{\left|\frac{S_n}{n} - p\right| < \varepsilon\right\} \to 1. \tag{4.2}$$

Другими словами, вероятность того, что средняя доля успехов отклоняется от p больше, чем на любое наперед заданное число ε, стремится к нулю с возрастанием n. Это одна из форм *закона больших чисел*, которая и служит основой для интуитивного представления о вероятности как мере действительной частоты. Для практических приложений эту формулировку надо дополнить более точной оценкой вероятности в левой части (4.2); такую оценку позволяет получить нормальное приближение для биномиального распределения (типичным в этом смысле является пример (3.2) гл. VII). Фактически формула (4.2) является простым следствием этого приближения (см. задачу 18, гл. VII).

Утверждение (4.2) является формулировкой классического закона больших чисел, который не представляет большого интереса. Более тонкой и значительно более интересной теоремой является *усиленный закон больших чисел* (см. § 4 гл. VIII).

Предостережение. Стало обычным обнаруживать в законе больших чисел факты, которые определенно из него не следуют. Если Петр и Павел по очереди бросают правильную монету 10 000 раз, то принято считать, что Петр будет в выигрыше приблизительно половину времени. *Но это не так.* Закон арксинуса (§ 5 гл. III) устанавливает, что такое равновесие менее всего вероятно. Вероятность того, *что Петр будет в выигрыше не более* 20 *раз, несравненно больше вероятности того, что число игр, после которых он будет впереди, заключено между* 4990 *и* 5010. Не существует никакой тенденции к выравниванию периодов лидерства. Закон больших чисел устанавливает *только* то, что для большого числа *различных* игр с бросанием монеты доля тех из них, в которых в данный момент в выигрыше находится герб, близка к $^1/_2$. Здесь ничего не говорится о колебаниях лидерства в *отдельной* игре.

§ 5. Приближенная формула Пуассона[1]

Во многих приложениях встречаются испытания Бернулли, в которых n относительно велико, p относительно мало, а произведение

$$\lambda = np \tag{5.1}$$

не мало, но и не велико. В таких случаях удобна приближенная формула Пуассона для $b(k;\, n,\, p)$, которую мы сейчас выведем.

[1] Симон Д. Пуассон (1781—1840). Его книга Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précedées des règles générales du calcul des probabilités появилась в 1837 г.

Согласно (2.1), $b(0;\ n,\ p)=(1-p)^n$, или, если подставить p из (5.1),

$$b(0;\ n,\ p)=\left(1-\frac{\lambda}{n}\right)^n. \qquad (5.2)$$

Логарифмируя и разлагая в ряд Тейлора [гл. II, (8.10)], получаем

$$\log b(0;\ n,\ p)=n\log\left(1-\frac{\lambda}{ae}\right)=-\lambda-\frac{\lambda^2}{2n}-\ \ldots, \qquad (5.3)$$

так что для больших n $b(0;\ n,\ p)\approx e^{-\lambda}$, где знак $\approx$ используется для обозначения приближенного равенства (в данном случае с точностью до членов порядка n^{-1}). Далее, из (3.1) следует, что при любом фиксированном k и достаточно большом n

$$\frac{b(k;\ n,\ p)}{b(k-1,\ n,\ p)}=\frac{\lambda-(k-1)p}{kq}\approx\frac{\lambda}{k}. \qquad (5.4)$$

При $k=1$ из (5.4) получаем, что $b(1;\ n,\ p)\approx\lambda e^{-\lambda}$. При $k=2$ получаем, что $b(2;\ n,\ p)\approx\frac{\lambda^2}{2}e^{-\lambda}/2$. Продолжая эти рассуждения, убеждаемся по индукции, что вообще

$$b(k;\ n,\ p)\approx\frac{\lambda^k}{k!}e^{-\lambda}. \qquad (5.5)$$

Это—*знаменитое пуассоновское приближение для биномиального распределения.* (См. задачи 30—34, где проводится оценка ошибки и доказывается, что приближение в формуле (5.5) равномерно, если $n\to\infty$, $p\to 0$, таким образом, что $\lambda=np$ меняется в конечном интервале.) Введем специальное обозначение для правой части (5.6):

$$p(k;\ \lambda)=e^{-\lambda}\frac{\lambda^k}{k!}. \qquad (5.6)$$

При таком обозначении $p(k;\ \lambda)$ будет приближенным выражением для вероятности $b(k;\ n,\ \lambda/n)$, когда n достаточно велико.

Примеры. а) Величины p_m в последнем столбце табл. 3 гл. IV представляют собой значения $p(m;\ 1)$. Величины b_m в предшествующих столбцах совпадают с $b(m;\ N,\ 1/N)$. Таблица дает возможность сравнить распределение Пуассона $p(m;\ 1)$ с биномиальным распределением при $p=1/n$ и $n=3,\ 4,\ 5,\ 6,\ 10$. Мы видим, что, несмотря на небольшие значения n, приближение поразительно хорошее.

б) В табл. 2 сравниваются значения $p(k;\ 1)$ с биномиальным распределением при $n=100$, $p=1/100$. Таблица показывает, что приближение является удовлетворительным для многих целей. В качестве примера рассмотрим число появлений комбинации (7,7) среди 100 пар случайных цифр, подчиненное биномиальному распределению $b(k;\ 100,\ 1/100)$. В последнем столбце табл. 2 [1]) приведены результаты действительного подсчета для 100 групп по 100 случайных цифр

---

[1]) H. G. Kendall, B. Smith, Tables of random sampling numbers, Tracts for Computers № 24, Cambridge, 1940.

каждая. Чтобы получить относительные частоты, нужно все величины последнего столбца разделить на 100. Эти частоты вполне согласуются с теоретическими вероятностями. (Оценка по критерию χ^2 показывает, что в 75 из 100 аналогичных случаев случайные колебания должны были бы вызвать бо́льшие отклонения наблюдаемых частот от теоретических вероятностей.)

Таблица 2

Пример использования приближенной формулы Пуассона

k	$b(k;\ 100,\ 1/100)$	$p(k;\ 1)$	N_k
0	0,366032	0,367879	41
1	0,369730	0,367879	34
2	0,184865	0,183940	16
3	0,060999	0,061313	8
4	0,014942	0,015328	0
5	0,002898	0,003066	1
6	0,000463	0,000511	0
7	0,000063	0,000073	0
8	0,000007	0,000009	0
9	0,000001	0,000001	0

Первые столбцы иллюстрируют пуассоновское приближение для биномиального распределения. В последнем столбце даны числа групп по 100 пар случайных цифр, в которых комбинация (7,7) встречается ровно k раз.

в) *Дни рождения.* Какова вероятность p_k того, что в обществе из 500 человек k человек родились в день нового года? Если эти 500 человек выбраны наугад, то можно применить схему из 500 испытаний Бернулли с вероятностью успеха $p = 1/365$. Тогда $p_0 = (364/365)^{500} = 0{,}2537\ \ldots$. Для использования приближенной формулы Пуассона полагаем $\lambda = 500/365 = 1{,}3699\ldots$. Тогда $p(0;\ \lambda) = 0{,}2541\ldots$ и содержит ошибку лишь в четвертом десятичном знаке. Для $k = 1,\ 2,\ \ldots$ точные значения p_k, вычисленные по биномиальной формуле, равны $p_1 = 0{,}3484\ \ldots$, $p_2 = 0{,}2388\ \ldots$, $p_3 = 0{,}1089\ \ldots$, $p_4 = 0{,}0372\ \ldots$, $p_5 = 0{,}0101\ \ldots$, $p_6 = 0{,}0023\ \ldots$ Соответствующие приближения Пуассона равны $p(1;\lambda) = 0{,}3481\ldots$, $p(2;\lambda) = 0{,}2385\ldots$, $p(3;\ \lambda) = 0{,}1089\ldots$, $p(4;\ \lambda) = 0{,}0373\ldots$, $p(5;\ \lambda) = 0{,}0102\ldots$, $p(6;\ \lambda) = 0{,}0023\ldots$. Все ошибки лишь в четвертом десятичном знаке.

г) *Бракованные изделия.* Рассмотрим массовое производство винтов с налаженной системой статистического контроля качества продукции и предположим, что применение схемы испытаний Бернулли законно. Если вероятность того, что винт бракованный, равна $p=0{,}015$, то вероятность того, что коробка из 100 винтов не содержит брака, равна $(0{,}985)^{100}=0{,}22061$. Подсчет по приближенной формуле Пуассона дает $e^{-1{,}5}=0{,}22313$, что достаточно точно для большинства практических целей. Поставим следующий вопрос: сколько винтов должна содержать коробка для того, чтобы вероятность найти в ней по крайней мере 100 исправных винтов была равна 0,8 или более? Пусть $100+x$ — искомое число винтов, где x — малое целое положительное число. Чтобы применить приближенную формулу Пуассона для $n=100+x$ испытаний, нужно было бы положить $\lambda=np$; но np приближенно равно $100p=1{,}5$. Задача сводится к нахождению наименьшего целого положительного x, при котором

$$e^{-1{,}5}\left\{1+\frac{1{,}5}{1}+\ldots+\frac{(1{,}5)^x}{x!}\right\}\geqslant 0{,}8. \tag{5.7}$$

По таблицам[1]) находим, что при $x=1$ левая часть приближенно равна 0,56, а при $x=2$ она равна 0,809. Таким образом, согласно приближенной формуле Пуассона, необходимо иметь 102 винта. Поскольку величина 0,809 слишком близка к заданному предельному значению 0,8, число 103 надежнее. На самом деле вероятность найти по крайней мере 100 исправных винтов в коробке со 102 винтами равна

$$(0{,}985)^{102}+\binom{102}{1}(0{,}985)^{101}(0{,}015)+$$
$$+\binom{102}{2}(0{,}985)^{100}(0{,}015)^2=0{,}8022\ldots$$

д) *Столетние старики.* Любое конкретное лицо в момент рождения имеет мало шансов прожить 100 лет; число же ежегодных рождений в большом обществе велико. Благодаря войнам, эпидемиям и т. п. сроки жизни различных людей не являются независимыми, но в первом приближении можно n рождений сравнить с n испытаниями Бернулли, в которых успехом является дожитие до 100 лет. Следует ожидать, что в устойчивом обществе, где ни численность населения, ни смертность заметно не изменяются, частота тех лет, когда умирает ровно k столетних стариков, приближенно равна $p(k;\lambda)$ (при этом λ зависит от численности общества и от организации здравоохранения). Данные по Швейцарии подтверждают это заключение[2]).

е) *Опечатки и т. п.* Если при наборе книги существует постоянная вероятность того, что любая буква будет набрана неправильно, и если условия набора остаются неизменными, то мы имеем

---

[1]) E. C. Molina, Poisson's Exponential Binomial Limit, New York, 1942. В этих таблицах даны числа $p(k;\lambda)$ и $p(k;\lambda)+p(k+1;\lambda)+\ldots$ для значений k от 0 до 100.

[2]) E. J. Gumbel, Les centenaires, Aktuárske Vêdy, Prague, 7 (1937), 1—8.

столько испытаний Бернулли, сколько букв. Частота страниц, содержащих ровно k опечаток, при этих условиях приближенно равна $p(k;\lambda)$, где λ — характеристика наборщика. Случайная усталость наборщика, трудные места текста и т. п. увеличивают вероятность ошибки и могут привести к обилию опечаток. Формула Пуассона, таким образом, может быть использована для обнаружения существенных отклонений от нормы или от требований статистического контроля. То же самое применимо и ко многим другим случаям Например, если в тесто кладут большое количество изюма, то следует ожидать, что благодаря размешиванию частота булок с k изюминками будет приблизительно равна $p(k;\lambda)$, где λ — плотность изюма в тесте.

§ 6. Распределение Пуассона

В предыдущем параграфе формула Пуассона (5.7) служила лишь удобным приближением для биномиального распределения в случае большого n и малого p. В связи с задачей о совпадениях и задачей о размещении гл. IV мы рассматривали совсем другие распределения вероятностей, которые также привели в пределе к формуле Пуассона $p(k;\lambda)$. Мы сталкиваемся здесь с частным случаем того замечательного факта, что существует несколько распределений большой общности, встречающихся в удивительно разнообразных задачах. Тремя главными распределениями, встречающимися во многих задачах теории вероятностей, являются биномиальное распределение, нормальное распределение (которое мы введем в следующей главе) и *распределение Пуассона*

$$p(k;\lambda)=e^{-\lambda}\frac{\lambda^k}{k!}, \tag{6.1}$$

которое мы теперь рассмотрим детально.

Заметим прежде всего, что, складывая равенства (6.1) при $k=0, 1, 2, \ldots$, мы получим в правой части ряд Тейлора для $e^{-\lambda}$, умноженный на e^{λ}. Поэтому при любом фиксированном λ величины $p(k;\lambda)$ дают в сумме единицу. Следовательно, можно представить себе идеальный опыт, в котором $p(k;\lambda)$ является вероятностью k успехов. Сейчас мы покажем, почему многие физические опыты и статистические наблюдения приводят к такому истолкованию распределения (6.1). Общность и важность применений распределения (6.1) иллюстрируются примерами следующего параграфа. Истинная природа распределения Пуассона станет ясной лишь в связи с теорией стохастических процессов (см. гл. XVII, где дана еще одна интерпретация распределения Пуассона).

Рассмотрим последовательность появляющихся во времени случайных событий, как, например, радиоактивный распад или вызовы, поступающие на телефонную станцию. Каждое событие можно изобразить точкой на оси времени, и мы имеем дело со случайным распределением этих точек. Существует много различных типов таких распределений, но их изучение принадлежит теории непрерывных распределений вероятностей, которую мы отложили до второй книги. Здесь мы лишь покажем, что простейшие физические предположения приводят к величине $p(k;\ \lambda)$, как к вероятности иметь k точек (событий) внутри фиксированного интервала определенной длины. Наши методы по необходимости грубы, и мы вернемся к этой же самой задаче с более удобными методами в гл. XVII.

Физические предположения, которые мы хотим выразить математически, состоят в том, что условия опыта остаются неизменными во времени, и неперекрывающиеся интервалы времени независимы в том смысле, что сведения, касающиеся числа событий в одном интервале, не дают никакой возможности судить о числе событий в другом интервале. Рассмотрение непрерывных распределений вероятностей дает возможность выразить эти положения непосредственно; мы, будучи ограничены дискретными распределениями вероятностей, вынуждены пользоваться приближенной конечной моделью и переходить к пределу.

Представим себе единичный интервал времени, разделенный на большое число n интервалов, каждый длины $1/n$. Произвольный фиксированный интервал разбиения либо пуст, либо содержит по крайней мере одну из наших случайных точек; условимся называть эти две возможности соответственно неудачей и успехом. Вероятность успеха p_n должна быть одной и той же для всех интервалов разбиения, так как они имеют одинаковую длину. В силу предположенной независимости неперекрывающихся интервалов мы имеем тогда n испытаний Бернулли, и вероятность k успехов оказывается равной $b(k;\ n,\ p_n)$. Число успехов не обязательно равно числу случайных точек, так как интервал разбиения может содержать несколько случайных точек. Однако естественно ввести дополнительное допущение, состоящее в том, что вероятностью появления двух или более событий в течение очень короткого промежутка времени можно в пределе пренебречь [1]). В этом случае вероятность иметь в единичном интервале времени k случайных точек получается как предел $b(k;\ n,\ p_n)$ при $n \to \infty$. Далее, разделив каждый из интервалов

[1]) Это допущение уже подразумевается, когда мы пользуемся наглядным представлением об изолированных случайных точках на оси времени. Однако необходимо исключить возможность того, что наши события появляются сразу парами. Например, если событиями являются автомобильные катастрофы, то вероятность осуществления двух таких событий в течение короткого промежутка времени исчезающе мала по сравнению с вероятностью осуществления одного события. С другой стороны, в катастрофе вероятнее всего участвуют сразу две машины, и если под событием понимать разбитую машину, то вполне вероятно появление событий парами, и наше допущение неприменимо.

разбиения пополам, мы найдем, что $p_n = 2p_{2n} - p_{2n}^2$; это равенство выражает тот факт, что успех в интервале длины $1/n$ означает успех либо в его левой половине, либо в правой, либо в обеих частях. Отсюда следует, что $p_n < 2p_{2n}$; это подсказывает, что np_n монотонно возрастает (последнее может быть строго доказано). Если $np_n \to \lambda$, то $b(k;\ n,\ p_n) \sim b(k;\ n,\ \lambda/n) \to p(k;\ \lambda)$, и мы приходим к (6.1), как к вероятности того, что в нашем единичном интервале содержится в общей сложности k случайных точек. Допущение $np_n \to \infty$ не приводит ни к какому разумному результату, так как оно влечет за собой появление бесконечного числа случайных точек в сколь угодно малом интервале.

Если вместо единичного интервала мы возьмем произвольный интервал длины t и снова воспользуемся разбиением на интервалы длины $1/n$, то мы получим испытания Бернулли с той же самой вероятностью успеха p_n. Но количество испытаний будет равно уже не n, а ближайшему к nt целому числу. Переход к пределу таков же, только вместо λ мы получим λt. Это приводит к истолкованию величины

$$p(k;\ \lambda t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \tag{6.2}$$

как вероятности иметь k точек на фиксированном интервале длины t. В частности, вероятность того, что на интервале длины t не будет ни одной точки, равна

$$p(0;\ \lambda t) = e^{-\lambda t}, \tag{6.3}$$

а вероятность «не менее чем одной точки» равна, следовательно, $1 - e^{-\lambda t}$.

Параметр λ — физическая константа, определяющая плотность точек на оси t. Чем больше λ, тем меньше вероятность (6.3) не иметь ни одной точки. Предположим, что некоторый физический опыт повторяется большое число N раз и что каждый раз мы подсчитываем число событий на некотором интервале фиксированной длины t. Пусть N_k — число опытов, при которых наблюдалось ровно k событий. Тогда

$$N_0 + N_1 + N_2 + \ldots = N. \tag{6.4}$$

Общее число наблюденных в N опытах точек равно

$$N_1 + 2N_2 + 3N_3 + \ldots = T; \tag{6.5}$$

среднее число точек на один опыт равно T/N. Если N велико, то естественно ожидать, что

$$N_k \approx Np(k;\ \lambda t). \tag{6.6}$$

(Такое утверждение является основой всех применений понятия вероятности; оно будет оправдано и уточнено законом больших чисел в гл. X.) Подставляя (6.6) в (6.5), получим

$$T \approx N[p(1;\ \lambda t)+2p(2;\ \lambda t)+3p(3;\ \lambda t)+\ \ldots]=$$
$$= Ne^{-\lambda t}\lambda t\left[1+\frac{\lambda t}{1}+\frac{(\lambda t)^2}{2!}+\ldots\right]=N\lambda t, \tag{6.7}$$

откуда

$$\lambda t \approx \frac{T}{N}. \tag{6.8}$$

Это соотношение дает способ оценки λ по наблюдениям и способ сравнения теории с опытами. Иллюстрации излагаются в примерах следующего параграфа.

Пространственные распределения. Те же самые рассуждения, которыми мы пользовались при рассмотрении распределения случайных событий или точек на оси t, применимы также к распределению точек на плоскости или в пространстве. Вместо интервалов длины t мы имеем здесь области с площадью t (или объемом t), и основное допущение состоит в том, что вероятность иметь k точек в какой-либо определенной области зависит только от площади (или объема) области, но не от ее формы. Кроме того, сохраним те же предположения, что и раньше: 1) при малом t вероятность иметь более одной точки в области объема t мала по сравнению с t; 2) неперекрывающиеся области взаимно независимы. Чтобы найти вероятность того, что в области объема t содержится k случайных точек, разобьем ее на n частичных областей и приближенно заменим искомую вероятность вероятностью k успехов при n испытаниях. Это означает, что мы пренебрегаем вероятностью иметь более одной точки в одной и той же частичной области; из нашего предположения 1 следует, что совершаемая при этом ошибка стремится к нулю при $n \to \infty$. В пределе мы опять получаем распределение Пуассона (6.2). Звезды в пространстве, изюм в кексе, семена сорняков среди семян злака, выводки животных в поле распределены согласно закону Пуассона. Численные примеры см. в примерах б и д § 7.

§ 7. Примеры схем, приводящих к распределению Пуассона

а) Радиоактивный распад. Радиоактивное вещество испускает α-частицы. Событие, заключающееся в том, что число α-частиц, достигающих в течение времени t заданного участка пространства, принимает фиксированное значение k, является широко известным примером случайного события, подчиняющегося закону Пуассона. Конечно, вещество непрерывно разрушается, и в конце концов плотность потока α-частиц уменьшается. Однако для радия, например, потребуются годы, прежде чем уменьшение плотности потока α-частиц станет заметным; при сравнительно коротких промежутках времени условия должны рассматриваться как постоянные, и мы имеем идеальное осуществление гипотез, приведших к распределению Пуассона.

В известном опыте Резерфорда, Чэдвика и Эллиса [1]) радиоактивное вещество наблюдали в течение $N = 2608$ промежутков времени, каждый длиной 7,5 секунды; для каждого интервала регистрировалось число частиц, достигших счетчика. В табл. 3 приведены числа N_k промежутков времени, в течение которых наблюдалось ровно k частиц. Общее число всех частиц равно $T = \sum kN_k = 10094$, среднее число частиц равно $T/N = 3870$. Теоретические значения $Np(k;\ 3870)$ представляются довольно близкими к наблюденным числам N_k. Чтобы сделать выводы о согласованности наблюдений с теорией, нужно оценить вероятные значения случайных отклонений. Статистики пользуются для этого критерием χ^2. Согласно этому критерию, следует ожидать, что при идеальных условиях примерно в 17 из 100 подобных случаев согласованность окажется худшей, чем в табл. 3.

Таблица 3

Пример „а"; радиоактивный распад

k	N_k	$Np(k;\ 3870)$
0	57	54,399
1	203	210,523
2	383	407,361
3	525	525,496
4	532	508,418
5	408	393,515
6	273	253,817
7	139	140,325
8	45	67,882
9	27	29,189
$k \geqslant 10$	16	17,075
Итого	2608	2608,000

б) Разрывы самолетов-снарядов в Лондоне. В качестве примера пространственного распределения случайных точек рассмотрим

[1]) Rutherford, Chadwick and Ellis, Radiations from radioactive substances (Cambridge, 1920), стр. 172. Содержащиеся в тексте табл. 3 и оценка с помощью критерия χ^2 заимствованы из книги Г. Крамера „Математические методы статистики", ИЛ., М., 1948, стр. 472.

статистику падений самолетов-снарядов в южной части Лондона во время второй мировой войны. Вся территория разделена на $N=576$ малых участков площадью $t=1/4$ *кв. км* каждый. В табл. 4 приведены числа N_k участков, на которые приходилось по k падений снарядов[1]) Общее число снарядов $T=\sum kN_k=537$; среднее число $\lambda t=T/N=0{,}9323$. Согласие с распределением Пуассона удивительно хорошее: по критерию χ^2 при идеальных условиях примерно 88% подобных наблюдений должны были бы дать худшее согласие. Интересно отметить, что большинство населения верило в тенденцию точек падения к группировке. Если бы это было так, то мы имели бы большую частоту участков со многими попаданиями или вообще без попаданий и меньшее число промежуточных участков. Табл. 4 показывает, что точки падения были совершенно случайными, все участки равноправны; мы имеем здесь поучительную иллюстрацию того факта, что неискушенному глазу случайность кажется регулярностью или тенденцией к группировке.

Таблица 4

Пример „б"; падение самолетов-снарядов в Лондоне

k	0	1	2	3	4	5 и больше
N_k	229	211	93	35	7	1
$p(k;\ 0{,}9323)$	226,74	211,39	98,54	30,62	7,14	1,57

в) Изменение хромосом в клетках. Облучение живых клеток рентгеновскими лучами приводит к тому, что отдельные хромосомы в клетках изменяются. Пока клетки подвергаются воздействию лучей, вероятность изменения остается постоянной, и, согласно общей теории, число N_k клеток с k изменениями должно следовать распределению Пуассона. Теория в состоянии предсказать зависимость параметра λ от интенсивности облучения, температуры и т. п., но мы не будем входить в такие детали. В табл. 5 приведены данные одиннадцати различных серий экспериментов[2]). Данные расположены в соответствии с тем, в какой мере они отвечают теоретической

---

[1]) Данные заимствованы из статьи R. D. Clarke, An Application of the Poisson Distribution, *Journal of the Institute of Actuaries*, 72 (1946), 48.

[2]) D. G. Catcheside, D. E. Lea и I. M. Thoday, Types of chromosome structural change induced by the irradiation of Tradescontia microspores. *Journal of Genetics*, 47 (1945—46) 113—136. Наша таблица совпадает с табл. 9 этой работы, если не учитывать, что мы дополнительно подсчитали уровни χ^2 с одной степенью свободы.

модели. В последнем столбце указано, в какой доле аналогичных опытов, проведенных при идеальных условиях, следовало бы ожидать (по критерию χ^2) худшего согласия с распределением Пуассона. Степень согласия между теорией и опытными данными поразительна.

Таблица 5

Пример „в"; изменение хромосом, вызванное рентгеновским облучением

Номер эксперимента		Клетки с k изменениями				Общее число N	Уровни χ^2 в процентах
		0	1	2	$\geqslant 3$		
1	Наблюденное N_k	753	266	49	5	1073	95
	$Np(k;\ 0{,}35508)$	752,3	267,1	47,4	6,2		
2	Наблюденное N_k	434	195	44	9	682	85
	$Np(k;\ 0{,}45601)$	432,3	197,1	44,9	7,7		
3	Наблюденное N_k	280	75	12	1	368	65
	$Np(k;\ 0{,}27717)$	278,9	77,3	10,7	1,1		
4	Наблюденное N_k	2278	273	15	0	2566	65
	$Np(k;\ 0{,}11808)$	2280,2	269,2	15,9	0,7		
5	Наблюденное N_k	593	143	20	3	759	45
	$Np(k;\ 0{,}25296)$	589,4	149,1	18,8	1,7		
6	Наблюденное N_k	639	141	13	0	793	45
	$Np(k;\ 0{,}21059)$	642,4	135,3	14,2	1,1		
7	Наблюденное N_k	359	109	13	1	482	40
	$Np(k;\ 0{,}28631)$	362,0	103,6	14,9	1,5		
8	Наблюденное N_k	493	176	26	2	697	35
	$Np(k;\ 0{,}33572)$	498,2	167,3	28,1	3,4		
9	Наблюденное N_k	793	339	62	5	1199	20
	$Np(k;\ 0{,}39867)$	804,8	320,8	64,0	9,4		
10	Наблюденное N_k	579	254	47	3	883	20
	$Np(k;\ 0{,}40544)$	588,7	238,7	48,4	7,2		
11	Наблюденное N_k	444	252	59	1	756	5
	$Np(k;\ 0{,}49339)$	461,6	227,7	56,2	10,5		

г) Телефонные соединения с неправильным номером. В табл. 6 приведена статистика телефонных соединений с неправильным номером [1]). Наблюдение велось за общим числом $N = 267$ номеров; N_k обозначает число номеров, имевших k неправильных соединений. Распределение Пуассона опять подходит самым блестящим обра-

[1]) Данные заимствованы из статьи F. Thorndike, Applications of Poisson's Probability Summation. *The Bell System Technical Journal*, 5 (1926), 604—624. Статья содержит графический анализ 32 различных статистик.

зом. (Согласно критерию χ^2, отклонения близки к своей медиане [1]).) Сведения о других телефонных статистиках, подчиненных закону Пуассона, читатель может найти в статье Торндайка. В некоторых случаях (таких, как вызовы, поступающие от группы абонентов, обслуживаемых коммутатором; вызовы из телефона-автомата и т. п.) между событиями существует явная зависимость, и распределение Пуассона уже не годится.

Таблица 6

Пример „г"; телефонные соединения с неправильным номером

k	N_k	$Np(k;\ 8{,}74)$
0—2	1	2,05
3	5	4,76
4	11	10,39
5	14	18,16
6	22	26,45
7	43	33,03
8	31	36,09
9	40	35,04
10	35	30,63
11	20	24,34
12	18	17,72
13	12	11,92
14	7	7,44
15	6	4,33
$\geqslant 16$	2	4,65
—	267	267,00

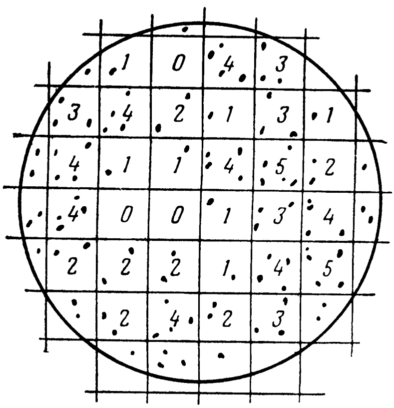


Рис. 1. Бактерии на чашке Петри.

д) Бактерии и кровяные тельца. Рис. 1 воспроизводит фотографию чашки Петри с колониями бактерий, которые кажутся под микроскопом темными пятнышками. Чашка разделена на маленькие квадраты. В табл. 7 приведены наблюденные в восьми опытах числа квадратов с k темными пятнышками, причем во всех опытах брались разные виды бактерий [2]). В последнем столбце указано, в каком проценте аналогичных опытов, проведенных при идеальных условиях, следовало бы ожидать по критерию χ^2 худшего согласия с распре-

[1]) То есть при идеальных условиях худшее согласие наблюдалось бы приблизительно столь же часто, как и лучшее. — *Прим. ред.*

[2]) Таблица заимствована из книги J. Neyman, Lectures and conferences on mathematical statistics (отпечатано на мимеографе), Департамент сельского хозяйства, Вашингтон, 1938. Исходные данные опубликованы T. Matuszewski, J. Supinska и J. Neyman в „Zentralblatt für Bakteriologie, Parasitenkunde und Infektionskrankheiten", II Abt., 95 (1936).

Таблица 7

Пример „д"; числа бактерий

k	0	1	2	3	4	5	6	7	χ^2-уровень
Наблюденное N_k	5	19	26	26	21	13	8		97
Пуассоновское	6,1	18,0	26,7	26,4	19,6	11,7	9,5		
Наблюденное N_k	26	40	38	17	7				66
Пуассоновское	27,5	42,2	32,5	16,7	9,1				
Наблюденное N_k	59	86	49	30	20				26
Пуассоновское	55,6	82,2	60,8	30,0	15,4				
Наблюденное N_k	83	134	135	101	40	16	7		63
Пуассоновское	75,0	144,5	139,4	89,7	43,3	16,7	7,4		
Наблюденное N_k	8	16	18	15	9	7			97
Пуассоновское	6,8	16,2	19,2	15,1	9,0	6,7			
Наблюденное N_k	7	11	11	11	7	8			53
Пуассоновское	3,9	10,4	13,7	12,0	7,9	7,1			
Наблюденное N_k	3	7	14	21	20	19	7	9	85
Пуассоновское	2,1	8,2	15,8	20,2	19,5	15	9,6	9,6	
Наблюденное N_k	60	80	45	16	9				78
Пуассоновское	62,6	75,8	45,8	18,5	7,3				

Последнее значение N_k в каждой строке равно числу квадратов поля микроскопа с k и более колониями бактерий.

делением Пуассона. Мы имеем здесь пример важного практического приложения распределения Пуассона к пространственным распределениям случайных точек.

§ 8. Время ожидания. Отрицательное биномиальное распределение

Рассмотрим последовательность n испытаний Бернулли с вероятностью успеха p. Нас интересует вопрос о том, сколько испытаний предшествуют r-му успеху. Здесь r — фиксированное положительное число. Полное число успехов в n испытаниях может быть, конечно, меньше, чем r, но вероятность того, что r-й успех осуществится при ν-м испытании (где $\nu \leqslant n$), очевидно, не зависит только от ν, r и p. Так как $\nu \geqslant r$, то удобно писать $\nu = k + r$. *Вероятность того, что r-й успех случится при $r+k$-м испытании (где $k = 0, 1, \ldots$) будет обозначаться $f(k, r, p)$.* ***Она равна вероятности того, что r-му успеху предшествует ровно k неудач.*** Это событие осуществляется тогда и только тогда, когда среди $r + k - 1$ испытаний ровно k привели к неудаче, а следующее $(r+k)$-е испытание привело к успеху; соответствующие вероятности равны $\binom{r+k-1}{k} \cdot p^{r-1}q^k$ и p, так что

$$f(k;\ r,\ p) = \binom{r+k-1}{k} \cdot p^r q^k. \tag{8.1}$$

Переписывая биномиальный коэффициент в соответствии с формулой (12.4) гл. II, мы получим эквивалентную форму записи:

$$f(k;\ r,\ p) = \binom{-r}{k} p^r (-q)^k \qquad k = 0,\ 1,\ 2,\ \ldots . \tag{8.2}$$

Допустим теперь, что *испытания продолжаются до тех пор, пока не осуществится r-й успех.* Типичное элементарное событие представляет последовательность, состоящую из произвольного числа k, букв H и равно r букв $У$, последовательность оканчивается на $У$; вероятность, соответствующая такому элементарному событию, равна, по определению $p^r q^k$. Мы должны поставить вопрос, возможно ли, что последовательность испытаний не оборвется, т. е. возможно ли, что бесконечная последовательность испытаний содержит менее чем r успехов. Так как $\sum_{k=0}^{\infty} f(k;\ r,\ p)$ есть вероятность того, что r-й успех осуществится после конечного числа испытаний, то бесконечными последовательностями, содержащими менее чем r успехов, можно пренебречь тогда и только тогда, когда

$$\sum_{k=0}^{\infty} f(k;\ r,\ p) = 1. \tag{8.3}$$

Для проверки соотношения (8.3) достаточно заметить, что по формуле бинома Ньютона

$$\sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k = (1-q)^{-r} = p^{-r}. \tag{8.4}$$

Умножая обе части равенства (8.4) на p^r, получаем (8.3).

В нашей задаче о времени ожидания r обязательно целое число, но величины, определяемые формулами (8.1) или (8.2), неотрицательны, и соотношение (8.3) выполняется при любом положительном r. *Для произвольного, но фиксированного действительного числа $r > 0$ и $0 < p < 1$ последовательность $\{f(k; r, p)\}$ определяет отрицательное биномиальное распределение.* Оно встречается во многих приложениях (мы сталкивались с ним в задаче 24 гл. V как с предельной формой распределения Пойа). Если r — целое, то $\{f(k; r, p)\}$ можно интерпретировать как *распределение вероятностей времени ожидания r-го успеха*; в такой форме оно также называется распределением Паскаля. При $r = 1$ оно сводится к *геометрическому распределению* $\{pq^k\}$.

Пример. *Задача Банаха о спичечных коробках*[1]) Некий математик всегда носит с собой две коробки спичек; каждый раз, когда он хочет достать спичку, он выбирает наугад одну из коробок. Неизбежно наступит момент, когда он впервые вынет пустую коробку. В этот момент в другой коробке может быть $r = 0, 1, 2, \ldots$

Таблица 8

Вероятности (8.5)

r	u_r	U_r	r	u_r	U_r
0	0,079589	0,079589	15	0,023171	0,917941
1	0,079589	0,159178	16	0,019081	0,937022
2	0,078785	0,237963	17	0,015447	0,952469
3	0,077177	0,315140	18	0,012283	0,964752
4	0,074790	0,389931	19	0,009587	0,974338
5	0,071674	0,461605	20	0,007338	0,981676
6	0,067902	0,529506	21	0,005504	0,987180
7	0,063568	0,593073	22	0,004041	0,991220
8	0,058783	0,651855	23	0,002901	0,994121
9	0,053671	0,705527	24	0,002034	0,996155
10	0,048363	0,753890	25	0,001392	0,997547
11	0,042989	0,796879	26	0,000928	0,998475
12	0,037676	0,834555	27	0,000602	0,999077
13	0,032538	0,867094	28	0,000379	0,999456
14	0,027676	0,894770	29	0,000232	0,999688

u_r есть вероятность того, что к моменту, когда первая коробка окажется пустой, вторая содержит ровно r спичек, предполагая, что вначале каждая коробка содержала по 50 спичек. $U_r = u_0 + u_1 + \ldots + u_r$ означает вероятность того, что вторая коробка содержит не более чем r спичек.

[1]) Сообщена Г. Штейнгаузом.

спичек, и задача состоит в том, чтобы найти соответствующие вероятности u_r. Допустим, что вначале в каждой коробке N спичек. Если вынутая коробка пуста, а другая коробка содержит r спичек, то это означает, что спички брались всего $2N-r$ раз, причем N раз из коробки, оказавшейся пустой. Это соответствует N успехам при $2N-r$ испытаниях, и поэтому

$$u_r = \binom{2N-r}{N} \frac{1}{2^{2N-r}} \tag{8.5}$$

(см. табл. 8. Рассмотрение задачи продолжено в примере 3, е гл. IX). Числовые значения для случая $N=50$ даны в табл. 8 (См. также задачи 21—23 и пример 3, е гл. IX.)

§ 9. Полиномиальное распределение

Биномиальное распределение можно легко обобщить на случай n повторных независимых испытаний, каждое из которых может иметь несколько исходов. Обозначим возможные исходы любого испытания через $E_1, \ldots, E_r$ и допустим, что вероятность осуществления исхода E_i в каждом испытании равна $p_i (i = 1, \ldots, r)$.

При $r=2$ получаем испытания Бернулли; в общем случае числа p_i подчинены единственному условию:

$$p_1 + \ldots + p_r = 1, \quad p_i \geqslant 0. \tag{9.1}$$

Результатом n испытаний является последовательность, подобная $E_3E_1E_2\ldots$. *Вероятность того, что в n испытаниях исход E_1 появится k_1 раз, исход E_2 появится k_2 раз и т. д., равна*

$$\frac{n!}{k_1!\,k_2!\ldots k_r!}\, p_1^{k_1} \ldots p_r^{k_r}. \tag{9.2}$$

Здесь k_i — произвольные неотрицательные числа, подчиненные очевидному условию

$$k_1 + k_2 + \ldots + k_r = n. \tag{9.3}$$

При $r=2$ формула (9.2) сводится к биномиальному распределению с $p_1 = p$, $p_2 = q$, $k_1 = k$, $k_2 = n-k$. Доказательство в общем случае опирается на формулу (4.7) гл. II и проводится так же, как и для биномиального распределения.

Формула (9.2) задает так называемое *полиномиальное распределение*; название объясняется тем, что выражение (9.2) является общим членом разложения многочлена $(p_1 + \ldots + p_r)^n$. Полиномиальное распределение применяется главным образом к *выбору с возвращением*, когда элементы разделяются на более чем две категории (например, распределение людей по профессиям).

Примеры. а) Какова вероятность того, что при бросании 12 костей каждая грань выпадает дважды? Здесь $E_1, \ldots, E_6$ представляют шесть граней, все k_i равны 2 и все p_i равны $^1/_6$. Следовательно, ответом будет $(12!)(2)^{-6}(6)^{-12} = 0{,}0034\ldots$.

б) *Выбор.* Допустим, что совокупность N предметов разделена на r классов $E_1, \ldots, E_r$ объемов $Np_1, Np_2, \ldots, Np_r$ соответственно. Полиномиальное распределение задает вероятности любого исхода опыта, состоящего в выборе с возвращением n предметов из этой совокупности.

в) *Сложные испытания Бернулли.* Две последовательности испытаний Бернулли с соответственными вероятностями успеха и неудачи p_1, q_1 и p_2, q_2 можно рассматривать как один опыт с четырьмя возможными исходами $(У, У)$, $(У, Н)$, $(Н, У)$, $(Н, Н)$ в каждом испытании. При этом предположение, что две первоначальные последовательности испытаний были независимы, переходит в утверждение, что вероятности четырех исходов равны соответственно p_1p_2, p_1q_2, q_1p_2, q_1q_2. Пусть k_1, k_2, k_3, k_4 — четыре целых неотрицательных числа, дающие в сумме n; вероятность того, что при n испытаниях комбинация $(У, У)$ появится k_1 раз, комбинация $(У, Н)$ — k_2 раз и т. д., равна

$$\frac{n!}{k_1!\,k_2!\,k_3!\,k_4!}\,p_1^{k_1+k_2}q_1^{k_3+k_4}p_2^{k_1+k_3}q_2^{k_2+k_4}. \tag{9.4}$$

Частный случай такой схемы появляется при *выборочном контроле.* Изделие годно с вероятностью p и дефектно с вероятностью q; с вероятностью p' оно проверяется, с вероятностью q' не проверяется. При выборе изделий для контроля заранее не известно, какие из них дефектны, и поэтому испытания независимы. (См. задачи, 25, 26 и 12 гл. IX.)

§ 10. Задачи

1. Предположим, что все комбинации полов детей равновероятны. Какую примерную долю всех семей с шестью детьми составляют семьи с тремя мальчиками и тремя девочками?

2. Игрок в бридж при трех последовательных сдачах колоды карт не получал тузов. Есть ли у него основания жаловаться на невезение?

3. Сколько нужно взять случайных цифр для того, чтобы вероятность появления среди них цифры 7 была не менее 9/10?

4. Сколько раз необходимо сдать колоду карт для игры в бридж (сдачи независимы) для того, чтобы вероятность заранее указанному игроку иметь хотя бы один раз четырех тузов была не меньше 1/2? Ответить на этот же вопрос для события «все четыре туза на руках у одного игрока (который заранее не фиксируется)».

5. Какова вероятность не менее 2 раз попасть в цель, если вероятность попадания равна 1/5 и производится 10 независимых выстрелов?

6. В задаче 5 найти условную вероятность хотя бы двух попаданий при условии, что одно попадание произошло.

7. Найти вероятность того, что среди наугад выбранных из полной колоды 13 карт содержатся 2 карты красной масти; сравнить эту вероятность с соответствующей вероятностью для испытаний Бернулли с $p=1/2$.

8. Какова вероятность того, что все дни рождения шести людей приходятся на два календарных месяца, оставляя ровно 10 месяцев свободными? (Допустить независимость и равновероятность всех месяцев.)

9. Пусть бросаются 6 правильных костей. Какова при этом вероятность выпадения: а) хотя бы одной, б) ровно одной, в) ровно двух единиц? Сравнить с расчетом по приближенной формуле Пуассона.

10. Каковы шансы на то, что среди 200 человек окажется четверо левшей, если в среднем левши составляют 1%?

11. Книга в 500 страниц содержит 500 опечаток. Оценить вероятность того, что на странице не меньше трех опечаток.

12. В некотором обществе имеется 1% дальтоников. Каков должен быть объем случайной выборки (с возвращением), чтобы вероятность встретить в ней хотя бы одного дальтоника была не менее 0,95?

13. Какова в предыдущем примере вероятность того, что выборка из 100 человек содержит: а) нуль, б) двух или более дальтоников?

14. Сколько изюма должны содержать в среднем сдобные булочки, для того чтобы вероятность иметь хотя бы одну изюминку в булке была не меньше 0,99?

15. Вероятность того, что при сдаче карт для игры в покер первый игрок получит десятку, валета, даму, короля и туза одной масти равна 1/649740. Сколь бо́льшим должно быть n для того, чтобы вероятность не получить ни одного упомянутого набора карт в n испытаниях была меньше $1/e \approx 1/3$. (Замечание. Решение не требует никаких вычислений.)

16. Книга в n страниц содержит в среднем λ опечаток на страницу. Оценить вероятность того, что по крайней мере на одной странице имеется более чем k опечаток.

17. Предположим, что существует два типа звезд (или изюминок в кексе, или дефектов в материале). Вероятность того, что заданный участок пространства содержит j звезд первого типа, равна $p(j;\, a)$, а вероятность того, что он содержит k звезд второго типа равна $p(k;\, b)$; эти два события предполагаются независимыми. Доказать, что вероятность того, что тот же участок пространства содержит в целом n звезд, равна $p(n;\, a+b)$. (Сформулировать результат и условия задачи в абстрактных терминах.)

18. Задача о движении городского транспорта. Движение транспорта через определенный уличный переход описывается предположением, что вероятность прохождения автомобиля через переход в любую данную секунду равна постоянной p и что между прохождениями автомобилей в различные секунды нет никакой зависимости. Считая секунду неделимой единицей времени, приходим к схеме испытаний Бернулли. Допустим, что пешеход может переходить улицу только при условии, что в последующие три секунды переход будет свободным. Найти вероятность того, что ему придется ждать $k=0, 1, 2, 3, 4$ секунд. (Соответствующие общие формулы не очевидны и будут выведены в § 7 гл. XIII в связи с теорией серий успехов.)

19. Двое бросают правильную монету n раз каждый. Найти вероятность того, что у них выпадет одинаковое число гербов.

20. Найти вероятность того, что в последовательности испытаний Бернулли с вероятностью успеха p a успехов встречается раньше, чем b неудач. (Замечание. Результат выясняется самое большее через $a+b-1$ испытаний. Эта задача имеет значение для классической теории азартных игр в связи с вопросом о том, как делить ставку, если игра прекращена к тому моменту, когда один игрок выиграл a партий, другой b партий.)

21. В задаче Банаха о спичечных коробках (§ 8) найти вероятность того, что в момент, когда впервые одна из коробок оказалась пустой (этот момент не совпадает с моментом, когда впервые была вынута пустая коробка), другая содержит r спичек ($r = 1, 2, \ldots, N$).

22. Продолжение. Используя результат предыдущей задачи, найти вероятность того, что отсутствие спичек было обнаружено впервые не в той коробке, которая опустела первой. Преобразовать полученное выражение к виду $x = \binom{2N}{N} 2^{-2N-1}$, или приближенно $\frac{1}{2}(N\pi)^{-1/2}$.

23. Корректуры некоторой книги независимо друг от друга прочитаны двумя корректорами, нашедшими соответственно k_1 и k_2 ошибок, причем k_{12} ошибок были найдены обоими. Дать приемлемую оценку неизвестного числа n всех ошибок в корректурах. (Допустить, что проверка книги отвечает схеме испытаний Бернулли, в которой корректоры обнаруживают каждую ошибку с вероятностями соответственно p_1 и p_2. Воспользоваться законом больших чисел.)

Замечание. Эта задача отображает в простых терминах экспериментальный прием, использованный Резерфордом для подсчета сцинцилляций.

24. Для того чтобы оценить объем популяции животных с помощью отлова [1]), последовательно r раз расставляют ловушки. Предполагая, что каждое животное может быть поймано с вероятностью q, что первоначально было n животных и что единственная разница в ситуациях между последовательными отловами состоит в изменении численности стада (т. е. уже пойманные животные выходят из рассмотрения), найти вероятность того, что r последовательных отловов дали соответственно результаты $n_1, n_2, \ldots, n_r$.

25. Сложные испытания Бернулли. Найти в примере 9, в условные вероятности p и q событий ($У$, $Н$) и ($Н$, $У$), при условии, что имела место одна из этих комбинаций. Показать, что $p > 1/2$ в случае $p_1 > p_2$ и $p < 1/2$ в случае $p_2 < p_1$.

26. Продолжение [2]). Пусть среди n пар испытаний m пар привели к одной из комбинаций ($У$, $Н$) или ($Н$, $У$). Показать, что при этом условии вероятность появления комбинации ($У$, $Н$) ровно k раз равна $b(k; m, p)$.

27. Комбинация биномиального и пуассоновского распределений. Допустим, что некоторое насекомое с вероятностью $p(r; \lambda)$ кладет r яиц, а вероятность развития насекомого из яйца равна p. Предполагая взаимную независимость развития яиц, показать, что вероятность того, что у насекомого будет в общей сложности k потомков, получается из распределения Пуассона с параметром λp.

Замечание. Приведем еще один пример с аналогичной ситуацией. Вероятность расщепления k хромосом равна $p(k, \lambda)$, вероятность восстановления — p (другие примеры такого рода см. примеры 1, г гл. IX и § 1 гл. XII.)

---

[1]) P. A. P. Moran, A mathematical theory of animal trapping, *Biometrica*, **38** (1951), 307—311.

[2]) A. Wald, Sequential tests of statistical hypotheses, *Ann. Math. Statistics*, **16** (1945), 166. Вальд использует приведенные выше результаты, чтобы создать метод практического сравнения двух эмпирически полученных последовательностей испытаний (например, результаты работы двух машин) в целях выбора той, которая обещает большую вероятность успеха. Он сводит задачу к более простой задаче нахождения того, будет ли в серии испытаний Бернулли частота успеха отличаться от $^1/_2$.

28. Показать, что для максимальной вероятности полиномиального распределения выполняются неравенства [1])

$$np_i - 1 < k_i \leqslant (n+r-1)\,p_i, \qquad i = 1, 2, \ldots, r. \tag{10.1}$$

(Указание. Доказать сначала, что вероятность максимальна тогда и только тогда, когда $p_i k_j \leqslant p_j (k_i + 1)$ для любой пары (i, j). Просуммировать эти неравенства по j и $i \neq j$.)

29. Показать, что в распределении Пуассона вероятность $p(k; \lambda)$ достигает своего наибольшего значения, если k — наибольшее целое число, не превосходящее λ.

Замечание. *Задачи 30—34 относятся к пуассоновскому приближению для биномиального распределения. В дальнейшем приняты обозначения:* $\lambda = np$ *и* m *— наибольшее целое число, не превосходящее* $(n+1)\,p$ *(другими словами,* m *— индекс максимальной вероятности).*

30. Показать, что если k меняется от 0 до ∞, то отношение $a_k = b(k; n, p)/p(k, \lambda)$ сначала возрастает, затем убывает, достигая максимума при $k = m$.

Если k возрастает, то вероятность $b(k; n, p)$ сначала меньше, затем больше и затем опять меньше, чем $p(k; \lambda)$.

31. Если $n \to \infty$, $p \to 0$, так что $np = \lambda$ остается постоянной, то

$$b(k; n, p) \to p(k; \lambda)$$

равномерно по k.

32. Показать, что

$$\frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^{n-k} \geqslant b(k; n, p) \geqslant \frac{\lambda^k}{k!}\left(1 - \frac{k}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n. \tag{10.2}$$

33. Вывести из (10.2), используя неравенства (8.12) гл. II, что

$$p(k; \lambda)\, e^{\frac{k\lambda}{n}} > b(k; n, p) > p(k; \lambda)\, e^{-\frac{k^2}{(n-k) - \lambda^2/(n-\lambda)}}. \tag{10.3}$$

Замечание. Хотя неравенство (10.2) является очень грубым, соотношение (10.3) обеспечивает превосходную оценку ошибки. Нетрудно уточнить (10.3) с помощью рассуждений, аналогичных рассуждениям § 9 гл. II. Между прочим, используя результат задачи 30, легко показать, что экспоненту в левой части (10.3) можно заменить на $\frac{m\lambda}{n}$, что, в свою очередь, меньше, чем $(p + n^{-1})\,\lambda$.

Другие предельные теоремы

35. Биномиальное приближение для гипергеометрического распределения. В генеральной совокупности, состоящей из N шаров, имеются красные и черные шары в пропорции $p:q$ $(p+q=1)$. Из этой совокупности выбирается без возвращения группа в n шаров. Вероятность того, что выборка содержит ровно k красных шаров, задается гипергеометрическим распределением § 6 гл. II. Показать, что если $N \to \infty$, то эта вероятность стремится к $b(k; n, p)$.

36. Допустим, что в предыдущей задаче p мало, n велико, а произведение $\lambda = np$ не мало и не велико. В этом случае гипергеометрическое распределение можно аппроксимировать распределением Пуассона $p(k; \lambda)$. Проверить это непосредственно, не используя пуассоновское приближение для биномиального распределения.

---

[1]) В первом издании было сформулировано более слабое утверждение, а именно $|k_i - np_i| \leqslant r$. Настоящее уточнение и его изящное доказательство были даны P. A. P. Moran.

37. Допустим, что в *отрицательном биномиальном распределении* $\{f(k; r, p)\}$, введенном в § 8, $g \to 0$ и $r \to \infty$ таким образом, что $rq = \lambda$ остается постоянной. Показать, что

$$f(k; r, p) \to p(k; \lambda).$$

(Замечание. Тем самым мы приходим к предельной теореме для *распределения Пойа*; см. задачу 24 гл. V.)

38. Многомерное распределение Пуассона. Доказать, что когда n велико, а $np_i = \lambda_i$ не малы и не велики, полиномиальное распределение (9.2) можно аппроксимировать посредством

$$e^{-(\lambda_1 + \ldots + \lambda_r)} \frac{\lambda_1^{k_1} \ldots \lambda_r^{k_r}}{k_1! \ldots k_r!}.$$

Доказать также, что вероятности в этом распределении дают в сумме единицу. (Заметим, что в задаче 17 получается двумерное распределение Пуассона.)

39. а) Вывести (3.6) непосредственно из (3.5), используя очевидное соотношение

$$b(k; n, p) = b(n - k; n, q).$$

б) Вывести биномиальное распределение как по индукции, так и опираясь на общую формулу (3.1) гл. IV.

40. Доказать, что $\sum kb(k; n, p) = np$ и $\sum k^2 b(k; n, p) = n^2p^2 + npq$.

41. Доказать, что

$$\sum k^2 p(k; \lambda) = \lambda^2 + \lambda.$$

42. Проверить тождество

$$\sum_{\nu=0}^{k} b(\nu; n_1, p)\, b(k - \nu; n_2, p) = b(k; n_1 + n_2, p) \tag{10.4}$$

и указать его теоретико-вероятностный смысл.

Указание. Равенство (10.4) — пример *композиций*, которые рассматриваются в гл. XI; другим примером является (10.5).

43. Проверить тождество

$$\sum_{\nu=0}^{k} p(\nu; \lambda_1)\, p(k - \nu; \lambda_2) = p(k; \lambda_1 + \lambda_2). \tag{10.5}$$

44. Пусть

$$B(k; n, p) = \sum_{\nu=0}^{k} b(\nu; n, p) \tag{10.6}$$

есть вероятность получить в n испытаниях не более k успехов. Тогда

$$\begin{aligned} B(k; n+1, p) &= B(k; n, p) - pb(k; n, p); \\ B(k+1; n+1, p) &= B(k; n, p) + qb(k+1; n, p). \end{aligned} \tag{10.7}$$

Получить это а) из вероятностных соображений, б) аналитически.

45. При том же обозначении

$$B(k;\ n,\ p) = (n-k)\binom{n}{k}\int_0^q t^{n-k-1}(1-t)^k\,dt \tag{10.8}$$

и

$$1 - B(k;\ n,\ p) = n\binom{n-1}{k}\int_0^p t^k(1-t)^{n-k-1}\,dt. \tag{10.9}$$

Указание. Интегрировать по частям или дифференцировать обе части по p. Вывести одну формулу из другой.

Замечание. Интеграл в (10.9) называется *неполной бета-функцией*. Таблицы значений $1 - B(k;\ n,\ p)$ с семью десятичными знаками для k и n, меньших 50, и $p = 0{,}01$, $0{,}02$, $0{,}03$ приведены в книге K. Pearson «Tables of the Incomplete Beta Function», London (Biometrika Office), 1934.

46. Доказать, что

$$p(0;\ \lambda) + \ldots + p(n,\ \lambda) = \frac{1}{n!}\int_\lambda^\infty e^{-x}x^n\,dx. \tag{10.10}$$

Замечание. *В последующих задачах мы дадим нижнюю оценку для всех вероятностей в биномиальном распределении. Вычисления довольно несложны, а сам метод, надлежащим образом усовершенствованный, позволяет получить очень простое доказательство предельной теоремы Муавра — Лапласа* (*см. задачи* 19—21 *гл.* VII). *Положим для сокращения*

$$\xi_k = \frac{k-(n+1)p+{}^1/_2}{\{(n+1)pq\}^{1/2}}, \tag{10.11}$$

и пусть m — номер максимальной вероятности, т. е. целое число, удовлетворяющее неравенствам (3.2).

47 Доказать, что при $r \geqslant (n+1)p$

$$b(r;\ n,\ p) \leqslant b(m;\ n,\ p)\,e^{-\frac{1}{2}p\xi_r^2+\delta^2} \tag{10.12}$$

где $\delta = m-(n+1)p+1/2$. Очевидно, $|\delta| < 1/2$. Указание. Записать (3.1) в форме

$$\frac{b(k;\ n,\ p)}{b(k-1;\ n,\ p)} = \frac{(n+1)pq - \{k-(n+1)p\}\,p}{(n+1)pq + \{k-(n+1)p\}\,q}, \qquad k \geqslant (n+1)p. \tag{10.13}$$

Вывести отсюда, что

$$\log\frac{b(k;\ n,\ p)}{b(k-1;\ n,\ p)} < \log\left\{1 - \frac{k-(n+1)p}{(n+1)pq}\,p\right\} < -\frac{k-(n+1)p}{(n+1)pq}, \tag{10.14}$$

после чего требуемое утверждение получается суммированием по k.

48. При $r \leqslant (n+1)p$ неравенство (10.12) выполняется с заменой множителя p в экспоненте на q. Поэтому, если p заменить на pq, то неравенство будет восполнено для всех r.

ГЛАВА VII

НОРМАЛЬНОЕ ПРИБЛИЖЕНИЕ ДЛЯ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ

§ 1. Нормальное распределение

Во избежание дальнейших перерывов в изложении основных идей сделаем здесь отступление и введем две чрезвычайно важные функции.

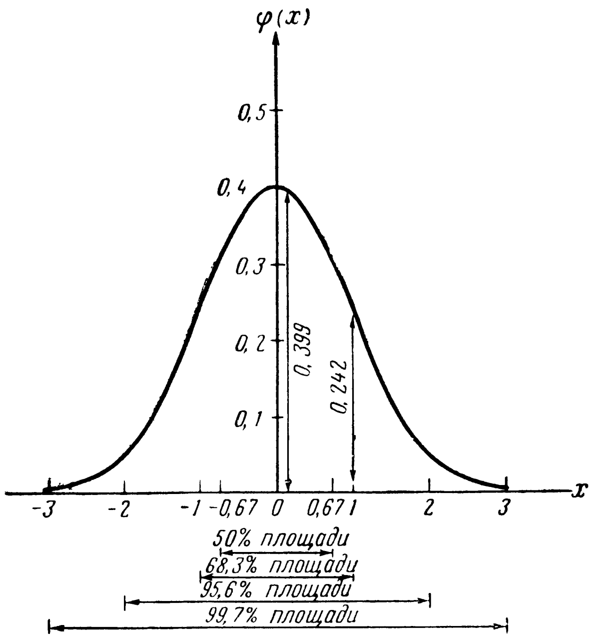


Рис. 1. Плотность вероятности нормального распределения.

Определение. *Функция*

$$\varphi(x) = \frac{1}{(2\pi)^{1/2}} e^{-x^2/2} \tag{1.1}$$

называется плотностью вероятности нормального распределения; ее интеграл

$$\Phi(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{x} e^{-y^2/2}\, dy \tag{1.2}$$

называется нормальной функцией распределения.

График $\varphi(x)$ — симметричная колоколообразная кривая, изображенная на рис. 1. Заметим, что мы пользуемся разными масштабами вдоль осей x и y. Максимум $\varphi(x)$ равен $(2\pi)^{-1/2} \approx 0{,}399$, так что в обычной декартовой системе координат (с одинаковыми масштабами по осям x и y) кривая $y=\varphi(x)$ была бы более приплюснутой.

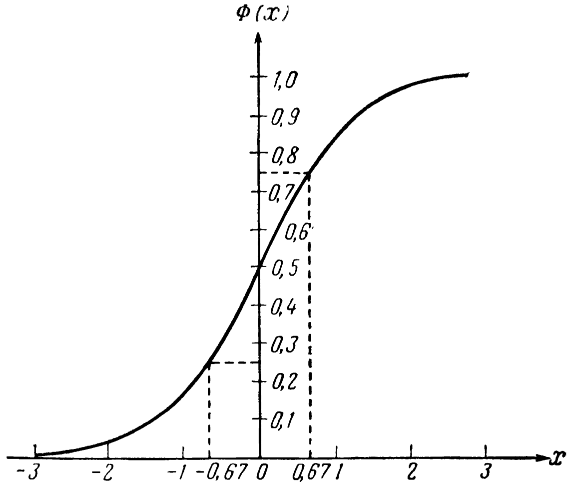


Рис. 2. Нормальная функция распределения.

Лемма 1. *Площадь, ограниченная графиком функции* $\varphi(x)$ *и осью* x, *равна единице, т. е.*

$$\int_{-\infty}^{+\infty} \varphi(x)\,dx = 1. \tag{1.3}$$

Доказательство. Имеем

$$\left\{\int_{-\infty}^{+\infty} \varphi(x)\,dx\right\}^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \varphi(x)\varphi(y)\,dx\,dy =$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2}\,dx\,dy. \tag{1.4}$$

Переходя в этом двойном интеграле к полярным координатам, получим

$$\frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0^{\infty} e^{-r^2/2} r\,dr = \int_0^{\infty} e^{-r^2/2} r\,dr = -e^{-r^2/2}\Big|_0^{\infty} = 1, \tag{1.5}$$

что и доказывает лемму.

Из определения и леммы следует, что *функция* $\Phi(x)$ *моно-*

тонно возрастает от 0 *до* 1. Ее график (рис. 4) — S-образная кривая, причем

$$\Phi(-x) = 1 - \Phi(x). \tag{1.6}$$

В табл. 1 приведены значения[1]) $\Phi(x)$ для положительных x; значения $\Phi(-x)$ получают с помощью формулы (1.6)[2]).

Для многих целей удобно иметь простую оценку для $1-\Phi(x)$ при больших x. Такую оценку дает

Лемма 2[3]). *При* $x \to \infty$

$$1 - \Phi(x) \sim \frac{1}{(2\pi)^{1/2} x} e^{-x^2/2}; \tag{1.7}$$

более точную для любого $x > 0$ *дает неравенство*

$$\frac{1}{(2\pi)^{1/2}} e^{-x^2/2} \left\{ \frac{1}{x} - \frac{1}{x^3} \right\} < 1 - \Phi(x) < \frac{1}{(2\pi)^{1/2}} e^{-x^2/2} \cdot \frac{1}{x} \tag{1.8}$$

(см. задачу 1).

Доказательство. Дифференцированием можно проверить, что

$$\frac{1}{(2\pi)^{1/2}} e^{-x^2/2} \frac{1}{x} = \frac{1}{(2\pi)^{1/2}} \int_x^\infty e^{-y^2/2} \left\{ 1 + \frac{1}{y^2} \right\} dy. \tag{1.9}$$

Подинтегральное выражение правой части больше подинтегрального выражения в интеграле

$$1 - \Phi(x) = \frac{1}{(2\pi)^{1/2}} \int_x^\infty e^{-y^2/2} dy, \tag{1.10}$$

что и доказывает второе из неравенств (1.8). Первое неравенство получается совершенно аналогично, если в качестве подинтегрального выражения взять функцию $e^{-y^2/2}\{1-3/y^4\} < e^{-y^2/2}$.

Замечание о терминологии. Термин *функция распределения* применяется в математической литературе для обозначения любой неубывающей функции $F(x)$, которая стремится к нулю при $x \to -\infty$ и к единице при $x \to \infty$. *Плотность вероятности* — неотрицательная функция $f(x)$, интеграл которой, взятый по всей оси x, равен единице. Интеграл от $-\infty$ до x от любой *плотности вероятности* является функцией распределения.

Нормальную функцию распределения часто называют распределением *Гаусса*, но ею пользовались в теории вероятностей еще Муавр и Лаплас.

---

[1]) Более подробные таблицы см. в Tables of Probability Functions, 2, Nat. Bureau of Standards, New York, 1942. Значения функции $\varphi(x)$ и $\Phi(x) - \Phi(-x)$ приведены в этих таблицах с 15 знаками после запятой. При $0 < x < 1$ значения x берутся через 0,0001, при $x > 1$ — через 0,001.

[2]) В русской литературе вместо функции $\Phi(x)$, определяемой формулой (1.2), чаще табулируется функция $\Psi(x) = \frac{1}{(2\pi)^{1/2}} \int_0^x e^{-y^2/2} dy$, связанная с $\Phi(x)$ очевидным соотношением $\Psi(x) = \Phi(x) - 1/2$. Эта функция удобнее тем, что для нее соотношение (1.6) заменяется более простым соотношением $\Psi(-x) = -\Psi(x)$. — *Прим. ред.*

[3]) Здесь и в дальнейшем знак $\sim$ употребляется для обозначения того, что *отношение* величин, соединенных этим знаком, стремится к единице.

Если изменить начало отсчета и единицу измерения, то $\Phi(x)$ преобразуется в $\Phi((x-a)/b)$; эта функция называется нормальной функцией распределения со средним значением a и дисперсией b^2 (или стандартным отклонением $|b|$). Функцию $2\Phi(x\sqrt{2})-1$ часто называют *функцией ошибок.*

Таблица 1

Нормальное распределение

t	$\varphi(t)$	$\Phi(t)$	t	$\varphi(t)$	$\Phi(t)$
0,0	0,398942	0,500000	2,3	0,028327	0,989276
0,1	0,396952	0,539828	2,4	0,022395	0,991802
0,2	0,391043	0,579260	2,5	0,017528	0,993790
0,3	0,381388	0,617911	2,6	0,013583	0,995339
0,4	0,368270	0,655422	2,7	0,010421	0,996533
0,5	0,352065	0,691462	2,8	0,007915	0,997445
0,6	0,333225	0,725747	2,9	0,005953	0,998134
0,7	0,312254	0,758036	3,0	0,004432	0,998650
0,8	0,289692	0,788145	3,1	0,003267	0,999032
0,9	0,266085	0,815940	3,2	0,002384	0,999313
1,0	0,241971	0,841345	3,3	0,001723	0,999517
1,1	0,217852	0,864334	3,4	0,001232	0,999663
1,2	0,194186	0,884930	3,5	0,000873	0,999767
1,3	0,171369	0,903200	3,6	0,000612	0,999841
1,4	0,149727	0,919243	3,7	0,000425	0,999892
1,5	0,129518	0,933193	3,8	0,000292	0,999928
1,6	0,110921	0,945201	3,9	0,000199	0,999952
1,7	0,094049	0,955435	4,0	0,000134	0,999968
1,8	0,078950	0,964070	4,1	0,000089	0,999979
1,9	0,065616	0,971283	4,2	0,000059	0,999987
2,0	0,053991	0,977250	4,3	0,000039	0,999991
2,1	0,043984	0,982136	4,4	0,000025	0,999995
2,2	0,035475	0,986097	4,5	0,000016	0,999997

§ 2. Предельная теорема Муавра — Лапласа

Пусть S_n означает число успехов в n испытаниях Бернулли с вероятностью успеха p. Тогда $b(k;\, n,\, p)$ является вероятностью события $S_n = k$. На практике нас обычно интересует ***вероятность того, что число успехов находится между заданными пределами*** α ***и*** β. Если α и β — целые неотрицательные числа и $\alpha < \beta$, то интересующее нас событие определяется неравенством $\alpha \leqslant S_n \leqslant \beta$, и его вероятность равна

$$P\{\alpha \leqslant S_n \leqslant \beta\} = b(\alpha;\ n,\ p) + b(\alpha+1;\ n,\ p) + \ldots + b(\beta;\ n,\ p). \quad (2.1)$$

Слагаемых в этой сумме может быть очень много, и прямое вычисление суммы обычно практически невозможно. К счастью, когда n велико, для вывода простых приближенных выражений для вероятности (2.1) может быть использована нормальная функция распределения. Это открытие принадлежит Муавру[1]) и Лапласу[2]). Мы увидим, что значение этого открытия выходит далеко за пределы области численных расчетов.

Наша ближайшая цель — вывести асимптотическую формулу для отдельного слагаемого

$$b(k;\ n,\ p) = \frac{n!}{k!\,(n-k)!}\,p^k q^{n-k}. \quad (2.2)$$

Вероятность p будет оставаться фиксированной, а n мы устремим к бесконечности.

Согласно закону больших чисел [(4.2) гл. VI], вероятность того, что $|S_n - np| > n\varepsilon$ стремится к нулю при любом $\varepsilon > 0$, и поэтому действительный интерес для нас представляют только такие значения k, для которых $|k-np|\,n^{-1} \to 0$. Удобно ввести новую переменную $\delta_k = k - np$. Тогда

$$k = np + \delta_k, \quad n - k = nq - \delta_k. \quad (2.3)$$

В дальнейшем мы будем рассматривать комбинации n и k, при которых $n \to \infty$ и $\frac{\delta_k}{n} \to 0$.

Заменяя входящие в (2.2) факториалы по формуле Стирлинга[3]) [(9.1) гл. II], получаем

$$b(k;\ n,\ p) \sim \left\{\frac{n}{2\pi k\,(n-k)}\right\}^{1/2}\left(\frac{np}{k}\right)^k\left(\frac{npq}{n-k}\right)^{n-k} =$$
$$= \left\{\frac{n}{2\pi\,(np+\delta_k)\,(nq-\delta_k)}\right\}^{1/2} \frac{1}{\left(1+\frac{\delta_k}{np}\right)^{np+\delta_k}\left(1-\frac{\delta_k}{nq}\right)^{nq-\delta_k}}, \quad (2.4)$$

---

[1]) Abraham De Moivre (1667—1754) «The Doctrine of chance», 1718.

[2]) Pierre S. Laplace (1749—1827) «Théorie analytique des probabilites», 1812.

[3]) Напомним, что в гл. II мы не закончили доказательство формулы Стирлинга, а лишь показали, что $r! \sim c r^{\sqrt{r}} e^{-r}$, где $c>0$ — константа. В тексте предполагается, что $c = (2\pi)^{1/2}$. Если мы хотим доказать этот факт, то множитель $(2n)^{1/2}$ в уравнениях (2.4), (2.7) и (2.8) должен быть заменен на c. В этом случае в правые части уравнений (2.11), (2.14) и (2.18) следует вставить множитель $c\,(2\pi)^{-1/2}$. Для доказательства того, что этот множитель равен 1, достаточно положить x_β и $-x_\alpha$ достаточно большими. Правая часть модифицированного уравнения (2.18) тогда сколь угодно близка к $c\,(2\pi)^{-1/2}$, а левая часть близка к 1 по оценкам § 3 гл. VI.

где знак $\sim$ означает, что отношение величин, стоящих в правой и левой части, стремится к единице.

Чтобы оценить величину дроби в (2.7), перейдем к логарифмам. В интервале $|\delta_k| \leqslant npq$ можно воспользоваться формулой Тейлора [(8.9) гл. II] и записать логарифм знаменателя в виде

$$(np+\delta_k)\log\left(1+\frac{\delta_k}{np}\right)+(nq-\delta_k)\log\left(1-\frac{\delta_k}{nq}\right)=$$
$$=(np+\delta_k)\left(\frac{\delta_k}{np}-\frac{\delta_k^2}{2n^2p^2}+\frac{\delta_k^3}{3n^3p^3}-+\ldots\right)-$$
$$-(nq-\delta_k)\left(\frac{\delta_k}{nq}+\frac{\delta_k^2}{2n^2q^2}+\frac{\delta_k^3}{3n^3q^3}+\ldots\right). \quad (2.5)$$

Перегруппировав члены по степеням δ_k, получим

$$\frac{\delta_k^2}{2n}\left(\frac{1}{p}+\frac{1}{q}\right)-\frac{\delta_k^3}{6n^2}\left(\frac{1}{p^2}-\frac{1}{q^2}\right)+\ldots=$$
$$=\frac{\delta_k^2}{2npq}\left\{1+\frac{p-q}{3pq}\frac{\delta_k}{n}+\ldots\right\}. \quad (2.6)$$

В этом выражении член $\frac{\delta_k^2}{2npq}$ является преобладающим, так как $\frac{\delta_k}{n}\to 0$. Предположим теперь, что $\frac{\delta_k^3}{n^2}\to 0$. Тогда все слагаемые в (2.6), кроме первого, стремятся к 0 и формула (2.4) принимает более простой вид

$$b(k;\ n,\ p)\sim\left\{\frac{n}{2\pi(np+\delta_k)(nq-\delta_k)}\right\}^{1/2}e^{-\frac{\delta_k^2}{2npq}}. \quad (2.7)$$

Но $np+\delta_k\sim np$ и $nq-\delta_k\sim nq$, так что (2.7) можно еще более упростить:

$$b(k;\ n,\ p)\sim\frac{1}{(2\pi npq)^{1/2}}e^{-\frac{\delta_k^2}{2npq}}=\frac{1}{(npq)^{1/2}}\varphi\left(\frac{\delta_k}{(npq)^{1/2}}\right). \quad (2.8)$$

Это — искомая асимптотическая формула. Перепишем ее, введя более удобные обозначения. Положим

$$h=\frac{1}{(npq)^{1/2}}, \quad (2.9)$$

и определим функцию x_k независимой переменной k формулой

$$x_k=(k-np)h=\frac{\delta_k}{(npq)^{1/2}}. \quad (2.10)$$

Пользуясь этими обозначениями, можно записать (2.8) в виде

$$b(k;\ n,\ p)\sim h\varphi(x_k). \quad (2.11)$$

При выводе этой формулы мы предполагали, что $n\to\infty$ и $k\to\infty$ таким образом, что $\frac{\delta_k}{n}\to 0$ и также $\frac{\delta_k^3}{n^2}\to 0$. Последнее условие, которое, очевидно, включает первое, можно записать в виде $x_k^3 n^{-1/2} \to 0$. Нами получена, таким образом,

Теорема 1. *Если $n \to \infty$ и $k \to \infty$ меняются таким образом, что $x_k^3 n^{-1/2} \to 0$, то верна асимптотическая формула* (2.11). Точнее

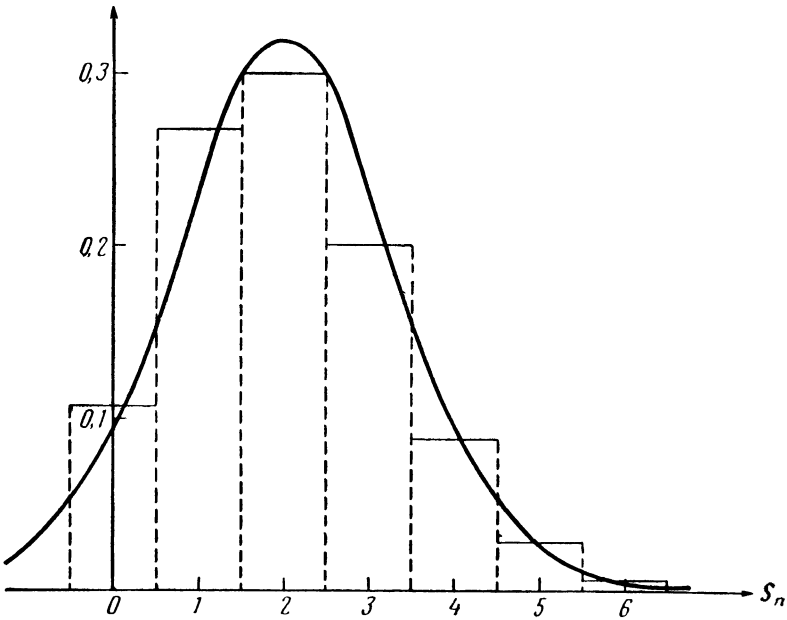


Рис. 3. Нормальное приближение для биномиального распределения. Ступенчатая функция дает вероятности $b(k; 10, 1/5k)$ успехов в 10 испытаниях Бернулли с $p = 1/5$. Непрерывная кривая дает для каждого целого k соответствующее нормальное приближение.

говоря, мы показали, что существуют две постоянные A и B, такие, что

$$\left| \frac{b(k;\ n,\ p)}{h\varphi(x_k)} - 1 \right| < \frac{A}{n} + \frac{B\,|\,x_k\,|^3}{n^{1/2}}. \tag{2.12}$$

(Другой вариант доказательства неравенства (2.11) см. в задачах 19 и 21.)

Рис. 3 иллюстрирует теорему для случая $n = 10$, $p = 0{,}2$, где npq равно всего 1,6. Мы видим, что даже в этом крайне неблагоприятном случае приближение удивительно хорошее[1]).

---

[1]) Значения $b(k,\ 10,\ 0,\ 2)$ при $k = 0, 1, \ldots, 6$ равны 0,1074, 0,2684, 0,3020, 0,2013, 0,0880, 0,0264, 0,0055. Соответствующие приближения $h\Phi(x_k)$ равны 0,0904, 0,2307, 0,3154, 0,2307, 0,0904, 0,0904, 0,0021.

Наша теорема сразу приводит к простому приближенному выражению для сумм (2.1). Если

$$hx_\alpha^3 \to 0 \quad \text{и} \quad hx_\beta^3 \to 0, \tag{2.13}$$

то (2.11) выполняется равномерно для всех членов суммы (2.1) и, следовательно,

$$P\{\alpha \leqslant S_n \leqslant \beta\} \sim h\{\varphi(x_\alpha) + \varphi(x_{\alpha+1}) + \ldots + \varphi(x_\beta)\}. \tag{2.14}$$

Правая часть представляет риманову интегральную сумму, аппроксимирующую интеграл[1]), и мы сначала исследуем степень этого приближения.

По теореме о среднем найдется такое значение ξ_k, что

$$\Phi(x_{k+1/2}) - \Phi(x_{k-1/2}) = h\varphi(\xi_k), \qquad x_k - \frac{1}{2}h < \xi_k < x_k + \frac{1}{2}h. \tag{2.15}$$

Тогда

$$h\varphi(x_k) = e^{1/2\left(\xi_k^2 - x_k^2\right)}\{\Phi(x_{k+1/2}) - \Phi(x_{k-1/2})\}. \tag{2.16}$$

Выберем произвольное $\varepsilon > 0$. Если выполнено (2.13), то для любого $\alpha \leqslant k \leqslant \beta$ и достаточно большого n

$$\frac{1}{2}|\xi_k^2 - x_k^2| = \frac{1}{2}|\xi_k - x_k||\xi_k + x_k| < h\left[|x_k| + \frac{1}{4}h\right] < \varepsilon.$$

Поэтому

$$e^{-\varepsilon}\left\{\Phi\left(x_{k+\frac{1}{2}}\right) - \Phi\left(x_{k-\frac{1}{2}}\right)\right\} < h\varphi(x_k) < e^{\varepsilon}\left\{\Phi\left(x_{k+\frac{1}{2}}\right) - \Phi\left(x_{k-\frac{1}{2}}\right)\right\}. \tag{2.17}$$

Суммируя по k, убеждаемся, что отношение выражения, стоящего в правой части формулы (2.14), к $\Phi(x_{\beta+1/2}) - \Phi(x_{\alpha-1/2})$ стремится к единице. Нами доказана, таким образом,

Предельная теорема Муавра—Лапласа. *Если α и β меняются так, что $hx_\alpha^3 \to 0$ и $hx_\beta^3 \to 0$, то*

$$P\{\alpha \leqslant S_n \leqslant \beta\} \sim \Phi(x_{\beta+1/2}) - \Phi(x_{\alpha-1/2}), \tag{2.18}$$

где $h = (npq)^{-1/2}$ и $x_t = (t - np)h$. Иначе говоря, относительная разность величин $P\{\alpha \leqslant S_n \leqslant \beta\}$ и $[\Phi(x_{\beta+1/2}) - \Phi(x_{\alpha-1/2})]$ стремится к нулю вместе с hx_β^3 и hx_α^3.

---

[1]) Ясно, что $\Phi(x_{k+1/2}) - \Phi(x_{k+1/2})$ представляет площадь трапеции, основанием которой служит отрезок $x_k - \frac{1}{2}h < x < x_k + \frac{1}{2}h$ и которая ограничена сверху касательной к кривой $y = \varphi(x)$ в точке $x = x_k$, а $h\varphi(x_k)$ представляет площадь прямоугольника с тем же основанием.

В частности, (2.18) выполняется, если α и β принимают значения, для которых x_α и x_β остаются внутри фиксированного интервала. [Случай, когда α и β столь велики, что нарушается условие (2.13), будет рассмотрен в § 5 и задаче 14.]

В статистических приложениях формула (2.18) обычно используется для значений α и β, при которых $|x_\alpha|$ и $|x_\beta|$ не превосходят 3 или 4. В теоретических приложениях часто приходится пользоваться формулой (2.18) для интервалов (α, β), находящихся далеко от центральной части биномиального распределения, для которых и x_α и x_β велики. В таких случаях обе части равенства (2.18) малы, и тогда важно знать, что их *отношение* близко к единице, так что не только разность, но и отношение этой разности к уменьшаемому (или вычитаемому) стремится к нулю.

Предельная теорема (2.18) принимает более простой вид, если вместо S_n ввести *нормированное число успехов*, определяемое формулой

$$S_n^* = \frac{S_n - np}{(npq)^{1/2}}. \tag{2.19}$$

Это равносильно выбору $(npq)^{1/2}$ в качестве единицы измерения отклонений S_n от np. *Величина np называется средним значением, а $(npq)^{1/2}$ — стандартным отклонением величины S_n*; эти термины взяты из теории случайных величин (см. гл. IX). Неравенство $\alpha \leqslant S_n \leqslant \beta$ эквивалентно неравенству $x_\alpha \leqslant S_n^* \leqslant x_\beta$, и (2.18) означает, что для произвольных фиксированных $x_\alpha < x_\beta$

$$P\{x_\alpha \leqslant S_n^* \leqslant x_\beta\} \sim \Phi\left(x_\beta + \frac{h}{2}\right) - \Phi\left(x_\alpha - \frac{h}{2}\right), \tag{2.20}$$

где $h = (npq)^{-1/2}$. Далее, $h \to 0$ при $n \to \infty$, и, следовательно, правая часть (2.20) стремится к $\Phi(x_\beta) - \Phi(x_\alpha)$. Мы имеем, таким образом, следующее

Следствие предельной теоремы. *Для любых фиксированных $a < b$*

$$P\{a \leqslant S_n^* \leqslant b\} \to \Phi(b) - \Phi(a). \tag{2.21}$$

Это ослабленное утверждение теоремы (2.18), являющееся, однако, традиционной формой предельной теоремы Лапласа. Отбрасывание $h/2$ в (2.20) вводит ошибку, стремящуюся к нулю при $n \to \infty$, но оказывающую значительное влияние при небольших значениях npq (как в случае примеров „а“ — „в“).

Главным следствием формулы (2.21) является тот факт, что для больших n вероятность $P\{a \leqslant S_n^* \leqslant b\}$ практически не зависит от p. Это позволяет сравнить случайные колебания в различных последовательностях испытаний Бернулли, просто относя их к нашим стандартным единицам.

Теорема (2.24) была исторически первой предельной теоремой теории вероятностей. С современной точки зрения это всего лишь чрезвычайно частный случай *центральной предельной теоремы*, к которой мы вернемся в гл. X, но общее доказательство которой придется отложить до второй книги. Статистики пользуются приближенной формулой (2.21) даже при относительно малом npq, и в таких случаях желательна оценка ошибки.

Оказывается, что в большинстве случаев ошибка формулы (2.11) мала по сравнению с той ошибкой, которую вводит замена суммы в (2.14) интегралом. (К счастью, эту ошибку можно значительно уменьшить, используя формулу суммирования Эйлера — Маклорена.) С. Н. Бернштейн посвятил ряд статей исследованию остаточного члена в общем случае и занимался вопросом о том, как нужно видоизменить определение величины x_i, чтобы улучшить сходимость в формуле (2.18) [1].

Замечание. Необходимо отметить, что наши *предельные теоремы и приближенные формулы справедливы только тогда, когда число испытаний n фиксировано заранее, независимо от исхода испытаний.* Если игрок по правилам игры может кончить игру в благоприятный для него момент, то в целом результат игры не может быть оценен с помощью нормального приближения, так как продолжительность игры зависит теперь от случая. *Для любого фиксированного n* чрезвычайно маловероятно, что S_n^* велико. Однако в длинном ряду испытаний даже самое маловероятное событие обязательно произойдет. Мы увидим, что при непрерывной игре S_n^* практически наверное будет иметь последовательность максимумов порядка величины $(\log\log n)^{1/2}$ (закон повторного логарифма, см. § 5 гл. VIII).

§ 3. Примеры

а) Пусть $p=\frac{1}{2}$, $n=200$, $\alpha=95$, $\beta=105$. В этом случае $x=P\{95\leqslant S_n\leqslant 105\}$ можно рассматривать как вероятность того, что при 200 бросаниях монеты число выпадений герба отклонится от 100 самое большее на 5. Для того чтобы вычислить правую часть (2.18), находим $h=50^{-1/2}=0{,}141421...$, откуда $-x_{\alpha-\frac{1}{2}}=x_{\beta+\frac{1}{2}}=$ $=5{,}5h=0{,}7778$. Из таблиц получаем $\Phi(x_{\beta+1/2})-\Phi(x_{\alpha-1/2})=0{,}56331...$. Точное значение (определяемое опять-таки по таблицам) равно 0,56325. Ошибка меньше, чем следовало бы ожидать, но это объясняется тем,

[1]) Упрощенный вывод результатов С. Н. Бернштейна с некоторыми добавлениями излагается в статье W. Feller, On the Normal Approximation to the Binomial Distribution, *Annals of Math. Statistics*, **16** (1945), 319—329.

[См. С. Н. Бернштейн, Возврат к вопросу о точности предельной формулы Лапласа, *Изв. АН СССР*, серия математич., 7 (1943), 3—16; Ю. В. Линник, О точности приближения к гауссову распределению сумм независимых случайных величин, *Изв. АН СССР*, 11, 1947. — *Прим. ред.*]

что в рассматриваемом интервале интеграл дает верхнюю оценку для суммы в (2.14), а приближенная формула (2.11) оценивает каждый член с недостатком.

б) Положим $p = 1/10$, $n = 500$, $\alpha = 50$, $\beta = 55$. Истинное значение $P\{50 \leqslant S_n \leqslant 55\}$ равно $0{,}317573\ldots$; h здесь равно $(45)^{-1/2} = 0{,}1490712\ldots$, откуда $\Phi(5{,}5h) - \Phi(-0{,}5h) = 0{,}3235\ldots$. Ошибка около 2%.

в) Положим $n = 100$, $p = 0{,}3$. Табл. 2 показывает на типичном примере (для относительно малых n), как ухудшается нормальное приближение по мере удаления интервала (α, β) от центрального члена.

Таблица 2

Сравнение биномиального распределения при $n = 100$, $p = 0{,}3$ и нормального приближения

Число успехов	Вероятность	Нормальное приближение	Ошибка в процентах
$9 \leqslant S_n \leqslant 11$	0,000006	0,00003	+400
$12 \leqslant S_n \leqslant 14$	0,00015	0,00033	+100
$15 \leqslant S_n \leqslant 17$	0,00201	0,00283	+ 40
$18 \leqslant S_n \leqslant 20$	0,01430	0,01599	+ 12
$21 \leqslant S_n \leqslant 23$	0,05907	0,05895	0
$24 \leqslant S_n \leqslant 26$	0,14887	0,14447	—3
$27 \leqslant S_n \leqslant 29$	0,23794	0,23405	—2
$31 \leqslant S_n \leqslant 33$	0,23013	0,23405	+2
$34 \leqslant S_n \leqslant 36$	0,14086	0,14447	+3
$37 \leqslant S_n \leqslant 39$	0,05886	0,05895	—0
$40 \leqslant S_n \leqslant 42$	0,01702	0,01599	6
$43 \leqslant S_n \leqslant 45$	0,00343	0,00283	—18
$46 \leqslant S_n \leqslant 48$	0,00049	0,00033	—33
$49 \leqslant S_n \leqslant 51$	0,00005	0,00003	—40

Примеры. г) Найдем число a, такое, что для больших n неравенство $|S_n^*| > a$ имеет вероятность, близкую к $1/2$. Для этого необходимо, чтобы $\Phi(a) - \Phi(-a) = 1/2$ или $\Phi(a) = 3/4$. По таблицам нормального распределения находим $a = 0{,}6745$. Поэтому неравенства

$$|S_n - np| < 0{,}6745\,(npq)^{1/2} \text{ и } |S_n - np| > 0{,}6745\,(npq)^{1/2} \qquad (3.1)$$

примерно равновероятны. В частности, близка к $1/2$ вероятность того, что при n бросаниях монеты число выпадений герба будет лежать в пределах $n/2 \pm 0{,}337 n^{1/2}$, и аналогично при n бросаниях кости число выпадений одного очка будет лежать внутри интервала $n/6 \pm 0{,}251 n^{1/2}$. Вероятность того, что S_n лежит в пределах $np \pm 2\,(npq)^{1/2}$, равна примерно $\Phi(2) - \Phi(-2) = 0{,}9545\ \ldots$, а для $np \pm 3\,(npq)^{1/2}$ она равна $0{,}9973\ \ldots$.

д) *Задача о конкуренции.* Этот пример иллюстрирует практические применения формулы (2.21). Две конкурирующие железнодорожные компании имеют по одному поезду, курсирующему между Чикаго и Лос-Анжелесом; эти поезда отправляются и прибывают одновременно и оборудованы примерно одинаково. Предположим, что n пассажиров независимо и наугад выбирают себе поезд, так что число пассажиров в каждом поезде определяется исходом n испытаний Бернулли с $p = 1/2$. Если в поезде имеется $s < n$ мест, то существует положительная вероятность $f(s)$ того, что явится больше s пассажиров и мест не хватит. Пользуясь приближением (2.21), найдем

$$f(s) \approx 1 - \Phi\left(\frac{2s - n}{n^{1/2}}\right). \tag{3.2}$$

Если s столь велико, что $f(s) < 0{,}01$, то число мест будет достаточным в 99 случаях из 100. Вообще компания может установить произвольный уровень риска α и определить s так, чтобы $f(s) < \alpha$. Для этой цели достаточно положить

$$s \geqslant \frac{1}{2}\left(n + t_\alpha n^{1/2}\right), \tag{3.3}$$

где t_α — корень уравнения $\alpha = 1 - \Phi(t_\alpha)$, который можно найти по таблицам. Например, если $n = 1000$ и $\alpha = 0{,}01$, то $t_\alpha \approx 2{,}33$ и достаточно $s = 537$ мест. Если обе компании примут уровень риска $\alpha = 0{,}01$, то два поезда будут иметь в общей сложности 1074 места, 74 из которых будут пустыми. Аналогично 514 мест было бы достаточно в 80% всех случаев, а 549 мест — в 999 из 1000 случаев.

Подобные соображения применимы и в других задачах о конкурирующем обслуживании. Например, если m кинотеатров соперничают из-за одних и тех же n зрителей, то следует положить $p = 1/m$, и (3.3) должно быть заменено неравенством $s \geqslant (1/m)\left[n + t_\alpha n^{1/2} (m-1)^{1/2}\right]$. Общее число пустых мест при такой системе равно $ms - n \approx t_\alpha n^{1/2} (m-1)^{1/2}$. Для $\alpha = 0{,}01$, $n = 1000$ и $m = 2, 3, 4$ значения этого числа приближенно равны соответственно 74, 126 и 147.

е) *Случайные числа* (продолжение). В примере 3, б гл. II мы рассматривали событие с вероятностью $p = 0{,}3024$. При $n = 1200$ испытаниях это событие имело среднюю частоту 0,3142. Отклонение частоты от вероятности p равно $\varepsilon = 0{,}0118$. В этом случае $(pq)^{1/2} = 0{,}4593$

и $\varepsilon(n/pq)^{1/2}=0{,}890\ldots$ Вероятность неравенства $|S_n/n-p|>\varepsilon$ равна примерно $0{,}37\ldots$ Это значит, что примерно в 37% всех случаев среднее число успехов отклонялось бы от вероятности p больше, чем в нашем примере.

в) *Выбор*. Доля p членов некоторого общества курит. Допустим, что p неизвестно и что с целью определить p применяют выбор с возвращением. Желательно найти p с ошибкой, не превосходящей 0,005. Каким должен быть объем выборки n? Пусть p' — доля курящих в выборке; мы хотим, чтобы $|p-p'|<0{,}005$. Никакой объем выборки, однако, не может дать абсолютной уверенности в том, что $|p-p'|<0{,}005$; может случиться так, что выборка будет состоять из одних курящих. Поскольку абсолютная уверенность недостижима, мы назначаем произвольный *доверительный уровень*, скажем, $\alpha=0{,}95$, и требуем, чтобы неравенство $|p-p'|<0{,}005$ выполнялось с вероятностью 0,95 или большей. Заметим, что np' есть число успехов в n испытаниях, откуда

$$P\{|p-p'|<0{,}005\}=P\left\{\left|\frac{S_n}{n}-p\right|<0{,}005\right\}.$$

Поэтому n должно быть выбрано настолько большим, чтобы эта величина сделалась больше 0,95. Воспользуемся нормальным приближением. Корень x уравнения $\Phi(x)-\Phi(-x)=0{,}95$ равен $x=1{,}96$, откуда $0{,}005\,(n/pq)^{1/2}\geqslant 1{,}96$. Мы приходим, таким образом, к неравенству $n\geqslant 392^2pq$ или, приближенно, $n>160\,000\,pq$. В это неравенство входит неизвестное p, то так как pq никогда не превосходит 1/4, то объем выборки $n=40\,000$ был бы надежным при всех обстоятельствах; при этом примерно 20 шансов против одного за то, что $|p-p'|<0{,}005$.

§ 4. Связь с приближенной формулой Пуассона

Ошибка нормального приближения мала при большом npq. С другой стороны, при большом n и малом p члены $b(k;\,n,\,p)$ оказываются близкими к вероятностям $p(k;\,\lambda)$ в распределении Пуассона с $\lambda=np$. Если λ мало, то применима лишь приближенная формула Пуассона. Однако при большом λ мы можем пользоваться как нормальным, так и пуассоновским приближением. Это заставляет предположить, что при большом λ распределение Пуассона может быть приближенно заменено нормальным распределением; в примере 1, в гл. X мы увидим, что это действительно так (см. также задачу 9). Здесь мы удовлетворимся иллюстрацией этого положения на одном численном примере и на одном примере, взятом из практики.

Примеры. а) Рассмотрим распределение Пуассона $p(k;\ 100)$. Мы можем считать это распределение, скажем, приближением биномиального распределения с $n = 100\,000\,000$ и $p = 1/1\,000\,000$. Тогда $npq \approx 100$; эта величина хотя и невелика, но достаточна для того, чтобы нормальное распределение давало хорошее приближение по крайней мере в средней части биномиального распределения. Распределение Пуассона $p(k;\ 100)$ совпадает с $b(k;\ 10^8,\ 10^{-6})$ с точностью до многих десятичных знаков, и мы можем сравнить это распределение Пуассона с нормальным приближением для $b(k;\ 10^8,\ 10^{-6})$. Положим, для краткости, $P(a,\ b) = p(a;\ 100) + p(a+1,\ 100) + \ldots$ $\ldots + p(b;\ 100)$, так что $P(a,\ b)$ будет заменять $P\{a \leqslant S_n \leqslant b\}$ и должно оказаться близким к величине $\Phi((b - 99{,}5)/10) -$ $- \Phi((a - 100{,}5)/10)$. Табл. 3 дает представление о степени приближения.

б) *Задача о телефонных линиях.* Следующая задача взята с некоторыми упрощениями из действительной практики[1]. Телефонная станция A, обслуживающая 2000 абонентов, должна соединять их с соседней станцией B. Было бы слишком дорого и бессмысленно проводить 2000 линий из A в B. Достаточно сделать число линий N настолько большим, чтобы при нормальных условиях лишь на один из каждой сотни вызовов не нашлось бы немедленно свободной линии. Предположим, что в течение наиболее напряженного часа дня каждый абонент разговаривает с B в среднем 2 минуты. Мы, естест венно, можем сравнить положение в фиксированный момент наиболее напряженного часа с группой из 2000 испытаний с вероятностью «успеха» $p = 1/30$ (вероятностью того, что потребуется линия). Эти

Таблица 3

	Точные значения	Нормальное значение
$P(85,\ 90)$	0,11384	0,11049
$P(90,\ 95)$	0,18485	0,17950
$P(95,\ 105)$	0,41763	0,41768
$P(90,\ 110)$	0,70652	0,70628
$P(110,\ 115)$	0,10738	0,11049
$P(115,\ 120)$	0,05323	0,05335

[1] E. C. Molina, Probability in engineering. *Electrical Engineering*, 54 (1935), 423—427, или *Bell Telephone System Technical Publication Monograph* B-854, где задача решается используемым нами методом Пуассона, более удобным для инженеров.

испытания при нормальных условиях можно считать независимыми (хотя, конечно, это перестает быть верным, когда такие события, как неожиданный ливень или землетрясение, заставляют многих вызывать такси или звонить в местную газету; теория становится неприменимой; линии будут забиты). Итак, мы имеем 2000 испытаний Бернулли с $p=1/30$; требуется найти наименьшее число N, такое, что вероятность более чем N «успехов» меньше 0,01; в наших обозначениях

$$P\{S_{2000} \geqslant N\} < 0{,}01.$$

При использовании приближенной формулы Пуассона мы должны были бы взять $\lambda = 200/3 \approx 66{,}67$. По таблицам находим, что вероятность 87 или более успехов равна примерно 0,0097, тогда как вероятность 86 или более успехов примерно 0,013. Это показывает, что *было бы достаточно* 87 *линий*. При использовании *нормального приближения* мы прежде всего находим по таблицам корень уравнения $1-\Phi(x)=0{,}01$, равный 2,327. Далее, должно иметь место неравенство $(N-1/2-np)/(npq)^{1/2} \geqslant 2{,}327$. Так как $n=2000$, $p=1/30$, то $N \geqslant 67{,}17+2{,}327 \cdot 8{,}027 \approx 85{,}8$. Таким образом, нормальное приближение показывает, что *достаточно* 86 *линий*.

С точки зрения практики оба ответа совпадают. Тот же метод позволяет получить и другие практические результаты. Быть может, выгоднее разбить 2000 абонентов на две группы по 1000 человек и каждую из этих групп соединить проводами с B? Пользуясь тем же методом, находим, что при этом потребовалось бы 10 дополнительных линий, так что следует предпочесть первоначальную систему.

§ 5. Большие отклонения[1]

Часто желательно иметь оценку вероятности того, что нормированное число успехов S_n^* [см. (2.19)] превосходит заданное число x. В этом случае верхней границей интервала является бесконечность и необходимо особое рассуждение, которое показывало бы, что наша предельная теорема (2.18) по-прежнему применима.

Теорема. *Если* $n \to \infty$ *и* x, *как функция от* n, *меняется так, что* $x \to \infty$, *но* $x^3 h \to 0$, *то*

$$P\{S_n^* > x\} \sim 1-\Phi(x). \tag{5.1}$$

Вследствие (1.7) *это эквивалентно тому, что*

$$P\{S_n^* > x\} \sim \frac{1}{(2\pi)^{1/2} x} e^{-x^2/2}. \tag{5.2}$$

[1]) Этот параграф весьма полезен, но его результаты будут использованы в настоящей книге лишь при доказательстве закона повторного логарифма (гл. VIII).

Доказательство. Выберем в (2.18) α и β так, чтобы x лежало между x_α и $x_{\alpha+1}$, а $x_\beta \approx x + \log x$. Тогда $x_\beta^3 h \to 0$ и (2.18) выполнено. Следовательно,

$$P\{\alpha < S_n < \beta\} \sim \{1 - \Phi(x_\alpha)\} - \{1 - \Phi(x_\beta)\}. \tag{5.3}$$

Из (1.7) и из того, что $x_\beta \approx x_\alpha + \log x_\alpha$, видно, что $1 - \Phi(x_\beta)$ есть величина высшего порядка малости сравнительно с $1 - \Phi(x_\alpha)$, а $1 - \Phi(x_\alpha) \sim 1 - \Phi(x)$. Отсюда

$$P\{\alpha < S_n < \beta\} \sim 1 - \Phi(x). \tag{5.4}$$

С другой стороны, из (2.11) и формулы (3.5) гл. VI следует, что

$$P\{S_n \geqslant \beta\} \leqslant \frac{n}{\beta - np} b(\beta;\, n,\, p) \sim \frac{nh^2}{x_\beta} \varphi(x_\beta). \tag{5.5}$$

Далее, $nh^2 = 1/pq$ — постоянная, и

$$\frac{1}{x_\beta} \varphi(x_\beta) \sim 1 - \Phi(x_\beta). \tag{5.6}$$

Мы видели, что правая часть (5.6) стремится к нулю быстрее, чем $1 - \Phi(x)$; это означает, что $P\{S_n \geqslant \beta\}$ — величина высшего порядка малости по сравнению с $1 - \Phi(x)$. Объединяя этот результат с (5.4), находим, что

$$P\{S_n > \alpha\} \sim 1 - \Phi(x). \tag{5.7}$$

Это и есть утверждение нашей теоремы.

Дальнейшие предельные теоремы для больших отклонений приведены в задачах 12—17.

§ 6. Задачи

1. Обобщая формулу (1.8), доказать, что

$$1 - \Phi(x) \sim \frac{1}{(2\pi)^{1/2}} e^{-x^2/2} \left\{ \frac{1}{x} - \frac{1}{x^3} + \frac{1\cdot 3}{x^5} - \frac{1\cdot 3\cdot 5}{x^7} + - \ldots + (-1)^k \frac{1\cdot 3 \ldots (2k-1)}{x^{2k+1}} \right\} \tag{6.1}$$

и что при $x > 0$ правая часть *превосходит* $1 - \Phi(x)$ при k четном и *меньше* $1 - \Phi(x)$ при k нечетном.

2. Для любого постоянного $a > 0$

$$\left\{1 - \Phi\left(x + \frac{a}{x}\right)\right\} : \{1 - \Phi(x)\} \to e^{-a} \tag{6.2}$$

при $x \to \infty$.

3. Найти вероятность того, что среди 10 000 случайных цифр цифра 7 появится не более 968 раз.

4. Найти приближенное выражение для вероятности того, что число выпадений единицы при 12 000 бросаний кости будет заключено между 1900 и 2150.

5. Найти такое число k, чтобы с вероятностью $\approx 0{,}5$ число выпадений герба при 1000 бросаниях монеты было заключено между 440 и k.

6. Чтобы найти процент женщин в некотором обществе, производится выборка. Найти такой объем выборки, чтобы с вероятностью, не меньшей 0,99, ошибка при выборке была меньше 0,005.

7. При 10 000 бросаний монеты герб выпал 5400 раз. Следует ли считать, что монета несимметрична?

8. Найти приближенное выражение для максимальной вероятности в триномиальном распределении

$$\frac{n!}{k!\,r!\,(n-k-r)!}\,p_1^k p_2^r (1-p_1-p_2)^{n-k-r}.$$

9. Нормальное приближение для распределения Пуассона. Пользуясь формулой Стирлинга, показать, что если $\lambda\to\infty$, то при любых фиксированных $\alpha<\beta$

$$\sum_{\lambda+\alpha\lambda^{1/2}<k<\lambda+\beta\lambda^{1/2}} p(k;\lambda)\to\Phi(\beta)-\Phi(\alpha). \tag{6.3}$$

10. Нормальное приближение для гипергеометрического распределения. Пусть n, m, k — целые положительные числа и

$$\frac{r}{n+m}\to t,\quad \frac{n}{n+m}\to p,\quad \frac{m}{n+m}\to q,\quad h\,\{k-rp\}\to x, \tag{6.4}$$

где $1/h=\{(n+m)\,pqt\,(1-t)\}^{1/2}$. Доказать, что

$$\frac{\binom{n}{k}\binom{m}{r-k}}{\binom{n+m}{r}}\sim h\varphi\,(x). \tag{6.5}$$

Указание. Лучше использовать нормальное приближение для биномиального распределения, а не формулу Стирлинга.

11. Нормальное распределение и комбинаторные серии[1]). В задаче (11.19) гл. II мы нашли, что при размещении n красных и m белых предметов вероятность иметь ровно k серий красных предметов равна

$$\pi_k=\binom{n-1}{k-1}\binom{m+1}{k}:\binom{n+m}{n}. \tag{6.6}$$

Пусть $n\to\infty$ и $m\to\infty$ так, что выполнено (6.4). При фиксированных $\alpha<\beta$ вероятность того, что число красных серий лежит между $npq+\alpha\,(npq)^{1/2}$ и $npq+\beta\,(npq)^{1/2}$, стремится к $\Phi(\beta)-\Phi(\alpha)$.

Замечание. *В следующих задачах* $h^{-2}=npq$, S_n^* — *нормированное число успехов, определяемое формулой* (2.19), *и*

$$F_n(x)=P\left\{S_n^*>x\right\}. \tag{6.7}$$

---

[1]) A. Wald and J. Wolfowitz, On a test whether two samples are from the same population, *Annals of Math. Statistics*, 11 (1940), 147—162. Более общие результаты см. в статье A. H. Mood, The distribution theory of runs, там же, стр. 367—392.

12. Если x, как функция n, меняется так, что $x^{3+a}h \to 0$, но $x \to \infty$, то [1])

$$\frac{F_n(x)}{1-\Phi(x)} = 1 + o(x^a), \tag{6.8}$$

где $o(x^a)$ означает член высшего порядка малости по сравнению с x^a.

13. Если $x^3 h \to 0$, $x \to \infty$, то [2]) для любого постоянного $a > 0$

$$\frac{F_n(x) - F_n\left(x + \frac{a}{x}\right)}{F(x)} \to 1 - e^{-a}. \tag{6.9}$$

Иными словами, условная вероятность неравенств $x < S_n^* < x + a/x$ при условии, что $S_n^* > x$, стремится к $1 - e^{-a}$.

[Указание. Воспользоваться (5.2).]

14. Вероятности больших отклонений. Исходя из (2.4), доказать следующую теорему. Если $n \to \infty$, а k меняется так, что $(k - np)/n \to 0$, то

$$b(k;\, n,\, p) \sim \frac{h}{(2\pi)^{1/2}}\, e^{-x^2/2 - f(x)}, \tag{6.10}$$

где $x = (k - np)\, h$ и

$$f(x) = \sum_{\nu=3}^{\infty} \frac{p^{\nu-1} - (-q)^{\nu-1}}{\nu(\nu-1)}\, h^{\nu-2} x^{\nu}. \tag{6.11}$$

Замечание. Если $x^3 h \to 0$, то $f(x) \to 0$ и (6.10) сводится к (2.11). Если x имеет порядок $h^{-1/3}$ (и мало по сравнению с $h^{-1/2}$), то

$$f(x) \approx \frac{p-q}{6}\, x^3 h. \tag{6.12}$$

Если x имеет порядок $h^{-1/2}$, то

$$f(x) \approx \frac{p-q}{6}\, x^3 h + \frac{p^3 + q^3}{12}\, x^4 h^2 \tag{6.13}$$

и т. д.

15. Продолжение. Доказать, что при $x \to \infty$, $xh \to 0$,

$$f\left(x + \frac{a}{x}\right) - f(x) \to 0, \tag{6.14}$$

откуда

$$F_n(x) \sim e^{f(x)}\, [1 - \Phi(x)]. \tag{6.15}$$

16. Вывести (6.9) из (6.15), принимая лишь, что $xh \to 0$.

17. Если $p > q$, то для больших x

$$P\{S_n > x\} < P\{S_n < -x\}.$$

(Указание. Использовать задачу 14.)

---

[1]) Н. В. Смирнов, О вероятностях больших отклонений, *Математический сборник* М., 40 (1933), 443—454.

[2]) А. Хинчин, Über einen neuen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Math. Annalen*, 101 (1929), 745—752. См. также задачу 16.

18. Новый вывод закона больших чисел. Показать, что закон больших чисел является следствием предельной теоремы Муавра — Лапласа.

19. Новый вывод теоремы о нормальном приближении биномиального распределения[1]). Отправляясь от формул (10.11) — (10.19) гл. VI, доказать, что если $n \to \infty$ и $k \to \infty$ таким образом, что $\xi_k^3 n^{-1/2} \to 0$, то

$$b(k;\ n,\ p) \sim b(m;\ n,\ p)\, e^{-1/2\xi_k^2}. \tag{6.17}$$

20. Доказать, что при $np \leqslant m \leqslant (n+1)p$ имеет место неравенство

$$b\left(m;\ n,\ \frac{m+1}{n+1}\right) \leqslant b(m;\ n,\ p) \leqslant b\left(m;\ n,\ \frac{m}{n}\right). \tag{6.18}$$

Если $(n+1)p - 1 < m \leqslant np$, то верно то же неравенство с заменой $\frac{m+1}{n+1}$ в левой части на $\frac{m}{n+1}$.

21. Доказать, что $b(m;\ n,\ p) \sim \{2\pi(n+1)pq\}^{-1/2}$, и дать верхнюю и нижнюю оценки.

---

[1]) Задачи 19 и 20 вместе показывают, что $b(k;\ n,\ p) \sim \{(n+1)pq\}^{-1/2}\varphi(\xi_k)$. Этот результат совпадает с приближенной формулой (2.11), если заменить x_k на ξ_k, а $h = \{npq\}^{1/2}$ на $h' = \{(n+1)pq\}^{-1/2}$. Так как $x_k \sim \xi_k$ и $h \sim h'$, то обе формулы асимптотически эквивалентны. В действительности же новая формула более точна [при ее выводе удается избежать ошибки, связанной с переходом от (2.7) к (2.8)]. Следует отметить, что вычисления, проведенные в задачах 19 и 20, проще и нагляднее, чем те, которые были использованы в тексте. По существу, они опираются только на стандартную приближенную формулу для логарифма, которая уже была использована в § 8 гл. II и § 3 гл. VI. Короче говоря, новая формула и ее вывод выглядят более предпочтительными, но они не согласуются с традиционным употреблением np вместо $(n+1)p$.

ГЛАВА VIII[1])

НЕОГРАНИЧЕННЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ ИСПЫТАНИЙ БЕРНУЛЛИ

В этой главе изучаются некоторые свойства случайности и, в частности, закон повторного логарифма для испытаний Бернулли.

Различные аспекты теории случайных колебаний, связанных с испытаниями Бернулли (по крайней мере при $p=1/2$), были рассмотрены в гл. III.

§ 1. Бесконечные последовательности испытаний

В предыдущих главах мы имели дело с вероятностями, связанными с n испытаниями Бернулли, и изучали асимптотическое поведение этих вероятностей при $n \to \infty$. Теперь мы перейдем к более общему типу задач, в которых события не могут быть определены в конечном пространстве элементарных событий.

Пример. *Задача о сериях.* Пусть α и β — два положительных целых числа. Рассмотрим неограниченную последовательность испытаний Бернулли, как, например, последовательность бросаний монеты или кости. Предположим, что Павел и Петр поспорили о том, что произойдет раньше: серия, составленная из α последовательных «успехов», или серия из β последовательных «неудач». Смысл вопроса о вероятности события, состоящего в том, что Павел выиграет, достаточно нагляден, но вспомним, что в математической теории термин «событие» означает «множество элементарных событий» и поэтому не имеет смысла, пока не определено соответствующее пространство элементарных событий. Для наших целей обычная модель конечного числа n испытаний Бернулли оказывается недостаточной, но трудность преодолевается простым переходом к пределу. В течение n испытаний Павел может или выиграть, или проиграть, или же победа останется еще не решенной. Пусть соответствующие вероятности будут x_n, y_n и z_n $(x_n+y_n+z_n=1)$. Когда число испытаний n возрастает, вероятность ничейного результата может только убывать, в то время как x_n и y_n, очевидно, возрастают. Следовательно, пределы $x=\lim x_n$, $y=\lim y_n$ и $z=\lim z_n$

---

[1]) Эта глава не связана с материалом предыдущих глав и может быть опущена при первом чтении.

существуют. Никто не усомнится в том, что эти пределы можно назвать вероятностями того, что в конце концов Петр или выиграет, или проиграет, или что игра закончится вничью. Однако описанные события могут быть определены только в пространстве элементарных событий, соответствующем бесконечной последовательности испытаний, а это пространство не является дискретным.

Изложенный пример был рассмотрен только в качестве иллюстрации, и нахождением численных значений x_n, y_n и z_n мы заниматься сейчас не будем. Мы возвратимся к их вычислению в примере 8, б гл. XIII. Пределы x, y и z могут быть вычислены простым методом, который пригоден и в более общих случаях. Этот метод мы изложим здесь из-за его важности и того интереса, который он представляет сам по себе. Пусть A — событие, состоящее в том, что *серия α последовательных успехов произойдет раньше, чем серия β последовательных неудач.* Тогда A означает выигрыш Петра, и $x = P\{A\}$. Если u и v — условные вероятности события A при условиях, что соответственно результатом первого испытания является успех или неудача, то $x = pu + qv$ [гл. V, формула (1.9)]. Предположим сначала, что результатом первого испытания является успех. В этом случае событие A может произойти только следующими α взаимно исключающимися способами:

1) Ближайшие $\alpha - 1$ испытаний приводят к успеху, вероятность чего равна $p^{\alpha-1}$.

2) Первая неудача произойдет при ν-м испытании, где $2 \leqslant \nu \leqslant \alpha$. Обозначим это событие через H_ν. Тогда $P\{H_\nu\} = p^{\nu-2}q$ и $P\{A \mid H_\nu\} = v$. Следовательно, использовав формулу полной вероятности, получим

$$u = p^{\alpha-1} + qv(1 + p + \ldots + p^{\alpha-2}) = p^{\alpha-1} + v(1 - p^{\alpha-1}). \tag{1.1}$$

В случае, когда результатом первого испытания является неудача, аналогичные соображения приводят нас к формуле

$$v = pu(1 + q + \ldots + q^{\beta-2}) = u(1 - q^{\beta-1}). \tag{1.2}$$

Получив, таким образом, два простых линейных уравнения с двумя неизвестными u и v, мы находим для $x = pu + qv$ выражение

$$x = p^{\alpha-1} \frac{1 - q^\beta}{p^{\alpha-1} + q^{\beta-1} - p^{\alpha-1}q^{\beta-1}}. \tag{1.3}$$

Чтобы получить y, мы должны только переменить местами p и q, α и β. Итак,

$$y = q^{\beta-1} \frac{1 - p^\alpha}{p^{\alpha-1} + q^{\beta-1} - p^{\alpha-1}q^{\beta-1}}. \tag{1.4}$$

Так как $x + y = 1$, мы имеем $z = 0$: *вероятность ничейного результата равна нулю.*

Например, при бросании монеты ($p = 1/2$) вероятность того, что серия из двух гербов произойдет раньше, чем серия из трех решеток, равна 0,7; вероятность того, что два последовательных герба выпадут раньше, чем четыре последовательные решетки, равна $^5/_6$; вероятность того, что три последовательных герба выпадут раньше, чем четыре последовательные решетки, равна $^{15}/_{22}$. При бросании кости вероятность того, что два последовательных выпадения единицы произойдут раньше, чем пять последовательных невыпадений, равна 0,1753 и т. д.

В настоящей книге мы ограничиваемся теорией дискретных пространств элементарных событий, и это в значительной степени уменьшает изящество математических рассмотрений. Общая теория рассматривает n испытаний Бернулли лишь как начало бесконечной последовательности испытаний. Элементарными событиями являются тогда бесконечные последовательности, составленные из букв *У* и *Н*, а пространство элементарных событий состоит из всех таких последовательностей. Под конечной последовательностью, например *УУНУ*, понимается тогда совокупность всех элементарных событий с таким началом, т. е. составное событие, состоящее в том, что в бесконечной последовательности испытаний результатами первых четырех испытаний будут соответственно *У*, *У*, *Н* и *У*. В случае бесконечного пространства элементарных событий игра, рассмотренная в вышеприведенном примере, может быть интерпретирована и без обращения к предельному переходу. Возьмем некоторое элементарное событие, т. е. последовательность *УУНУНН*... В ней может встретиться или не встретиться серия из α подряд идущих *У*. Если эта серия встретится, то ей может предшествовать или не предшествовать серия, состоящая из β последовательных *Н*. Таким образом, мы разделяем все элементарные события на три класса, представляющие собой события «Петр выиграл», «Петр проиграл», «нет никакого результата». Их вероятностями являются числа x, y, z, найденные выше. Единственным затруднением является то, что это пространство элементарных событий не является дискретным, а мы еще не определили вероятности в таких пространствах.

Заметим, что мы занимались скорее вопросами терминологии, чем действительными трудностями. В нашем примере вопрос об определении или интерпретации числа x не вызывает сомнений. Затруднение состоит лишь в том, что, желая быть последовательными, мы должны решить, истолковать ли число x как предел вероятности x_n, или же говорить о событии «Петр выиграл», определенном в недискретном пространстве элементарных событий. Мы будем делать и то и другое. Для упрощения терминологии будем говорить о событиях даже тогда, когда они определены в бесконечных пространствах. Однако в целях точности теоремы будут формулироваться в терминах конечных пространств элементарных событий и предельных переходов. Все события, изучаемые в этой главе, обладают одной важной особенностью, которую мы объясним на нашем примере. Событие «Петр выиграл» хотя и определено в бесконечном пространстве, но является суммой событий «Петр выиграл на n-м испытании» ($n = 1, 2, \ldots$), каждое из которых зависит только от конечного числа испытаний. Искомая вероятность x является пределом монотонной последовательности вероятностей x_n, каждая из которых зависит только от конечного множества испытаний. Нам нигде не потребуется теория, идущая дальше модели n испытаний Бернулли; мы только сохранили за собой свободу упрощать неуклюжие выражения, используя в определенных случаях термин «вероятность» вместо термина «предел вероятностей» [1]).

§ 2. Системы игры

Печальный опыт многих игроков учит, что никакая система игры не может увеличить шансов игрока на выигрыш. Если теория вероятностей правильно отображает жизнь, то этому опыту должны соответствовать какие-то утверждения, доказуемые в рамках теории.

Чтобы разобраться в этом, рассмотрим неограниченную последовательность испытаний Бернулли и предположим, что перед каждым испытанием игрок может выбирать, участвует ли он на этот раз в игре или нет. «Система» и представляет собой фиксированное правило выбора тех испытаний, при которых игрок участвует в игре. Например, игрок может включиться в игру лишь при каждом седьмом испытании или же после каждого вступления в игру ожидать, пока 7 раз не выпадет решетка. Он может вступать в игру только после выпадения серии из 13 гербов или же первый раз включиться в игру после выпадения первого герба, второй раз — после первого выпадения двух последовательных гербов и вообще k-й раз после выпадения k гербов подряд. В последнем случае он будет участвовать в игре все более и более редко. Другой возможной системой является участие в игре только тогда, когда общее число выпавших с начала игры гербов превышает общее число решеток (в этом случае игрок скорее всего ожидает выпадения решетки). Мы не будем касаться выбора размера ставок при каждом индивидуальном испытании: мы хотим показать, что никакая «система» не может изменить положение игрока и что результат будет тот же самый, как если бы он участвовал в игре при каждом испытании. Само собой разумеется, что это утверждение может быть доказано лишь для систем в обычном смысле, когда игрок не знает заранее результата будущих испытаний. Приходится также признать, что правило «уходить домой после трех проигрышей» изменяет положение, но мы исключим такие неинтересные системы.

Мы определяем систему как множество фиксированных правил, которые для каждого испытания единственным образом определяют, будет или не будет игрок участвовать в игре при

[1]) Для читателя, знакомого с общей теорией меры, положение может быть охарактеризовано следующим образом. Мы рассматриваем только те события, которые зависят от конечного числа испытаний или же являются пределами *монотонной* последовательности таких событий. Мы вычисляем пределы соответствующих вероятностей и поэтому не нуждаемся для наших целей в теории меры. Однако общая теория меры показывает, что наши пределы не зависят от способа перехода к пределу и являются вполне аддитивными функциями множеств.

этом испытании. Для k-го испытания это решение может зависеть от исходов первых $k-1$ испытаний, но оно должно не зависеть от исходов k-го, $k+1$-го, $k+2$-го, ... испытаний; наконец, правило должно обеспечивать бесконечное продолжение игры. Последнее условие означает следующее. Так как множество правил фиксировано, событие «при n испытаниях игрок участвовал в игре более чем r раз» вполне определено и можно найти его вероятность. Требуется, чтобы для каждого r при $n \to \infty$ эта вероятность стремилась к единице.

Мы сформулируем теперь нашу основную теорему как утверждение, что *при любой системе игры последовательные испытания, при которых игрок участвует в игре, образуют последовательность испытаний Бернулли с неизменной вероятностью успеха.* При соответствующих изменениях формулировки эта теорема оказывается применимой ко всем видам независимых испытаний. Последовательность игр, в которых участвует игрок, является точной копией первоначальных испытаний, так что никакая система не может повлиять на удачу игрока. Важность этого утверждения была впервые подмечена Мизесом, который вводил невозможность успешной действующей системы игры в качестве одной из аксиом. Приводимые здесь формулировки и доказательство заимствованы у Дуба [1]). Для простоты мы принимаем, что $p = 1/2$.

Пусть A_k означает событие «первое вступление игрока в игру происходит на k-м испытании». Из определения системы вытекает, что при $n \to \infty$ вероятность того, что первое вступление произойдет не позже n-го испытания, стремится к единице. Это означает, что

$$P\{A_1\} + P\{A_2\} + \ldots + P\{A_n\} \to 1$$

или

$$\sum P\{A_k\} = 1. \tag{2.1}$$

Далее, пусть B_k означает событие «при k-м испытании выпал герб». Тогда событие B — «в момент первого вступления в игру результатом испытания оказался герб» — окажется суммой взаимно исключающих друг друга событий A_1B_1, A_2B_2, A_3B_3, ... Так как A_k зависит только исходов первых $k-1$ испытаний, а B_k — только от испытания с номером k, то события A_k и B_k независимы и $P\{A_kB_k\} = P\{A_k\} P\{B_k\} = 1/2\, P\{A_k\}$. Таким образом, $P\{B\} = \sum P\{A_kB_k\} = 1/2 \sum P\{A_k\} = 1/2$. Это значит, что при нашей системе вероятность выпадения герба в момент первого вступления в игру равна $1/2$, и то же самое положение сохранится и для каждого из последующих вступлений в игру.

---

[1]) J. L. Doob, Note of probability, *Annals of Mathematics*, **37** (1936), 363—367.

Остается только показать, что вступления в игру независимы, т. е. что вероятность того, что герб выпадет и в момент первого и в момент второго вступления в игру, равна 1/4 (и аналогично для всех других комбинаций и для последуюших вступлений в игру). Для того чтобы проверить это утверждение, обозначим через A_k^* событие, состоящее в том, что второе вступление в игру произойдет на k-м испытании и через E — событие «гербы выпали в моменты первого и второго вступления в игру». Событие E является суммой всех событий $A_jB_jA_k^*B_k$ (если $j<k$, то A_j и A_k^* несовместны и $A_jA_k^*=0$). Следовательно,

$$P\{E\}=\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}P\{A_jB_jA_k^*B_k\}. \tag{2.2}$$

Как и прежде, мы видим, что при фиксированных j и $k>j$ событие B_k (герб выпадает при k-м испытании) не зависит от события $A_jB_jA_k^*$ (последнее зависит только от исходов первых $k-1$ испытаний). Следовательно,

$$P\{E\}=\frac{1}{2}\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}P\{A_jB_jA_k^*\}=\frac{1}{2}\sum_{j=1}^{\infty}P\{A_jB_j\}\sum_{k=j+1}^{\infty}P\{A_k^*|A_jB_j\} \tag{2.3}$$

[см. гл. V, формула (1.8)]. Каков бы ни был исход первого вступления в игру, игра будет продолжена дальше, т. е. рано или поздно произойдет второе вступление в игру. Это значит, что при данном A_jB_j, таком, что $P\{A_jB_j\}>0$, условные вероятности того, что второе вступление в игру произойдет при испытании с некоторым фиксированным номером, дают в сумме единицу. Сумма по k в (2.3) равна, следовательно, единице, и мы уже видели, что $\sum P\{A_jB_j\}=1/2$. Следовательно, $P(E)=1/4$, как и предполагалось. Подобные соображения применимы и при других комбинациях испытаний.

Заметим, что положение становится другим, если игрок может произвольно изменять размер ставки, на которую он играет. При таких условиях существуют выгодные и невыгодные стратегии, и результаты игры зависят от выбора стратегии. Мы вернемся к этому вопросу в § 2 гл, XIV.

§ 3. Леммы Бореля — Кантелли

Две нижеследующие простые леммы, касающиеся бесконечной последовательности испытаний, используются так часто, что заслуживают специального внимания. Мы сформулируем их только для испытаний Бернулли, но они применимы и в более общем случае.

Мы рассмотрим снова бесконечную последовательность испытаний Бернулли. Пусть A_1, A_2, ... — бесконечная последовательность событий, каждое из которых зависит только от конечного числа испытаний; другими словами, мы предполагаем, что для каждого номера k существует целое число n_k, такое, что A_k является событием из пространства первых n_k испытаний Бернулли. Положим

$$a_k = P\{A_k\}. \tag{3.1}$$

(Например, A_k может состоять в том, что испытанию с номером $2k$ непосредственно предшествует по крайней мере k последовательных успехов. Тогда $n_k = 2k$ и $a_k = p^k$.)

При произвольной фиксированной бесконечной последовательности букв $У$ и $Н$ можно установить, осуществилось ли 0, 1, 2, ... или бесконечно большое число событий $\{A_k\}$. Это значит, что мы можем говорить о событии U_r, состоящем в том, что результаты бесконечной последовательности испытаний благоприятствуют более чем r событиям $\{A_k\}$, а также о событии U_∞, состоящем в том, что произошло бесконечно много событий $\{A_k\}$. Событие U_r определено только в бесконечном пространстве элементарных событий, и его вероятностью является предел $P\{U_{n,r}\}$ вероятности того, что в результате n испытаний произойдет более чем r событий $\{A_k\}$. Наконец, $P\{U_\infty\} = \lim P\{U_r\}$; этот предел существует, так как $P\{U_r\}$ могут только убывать, когда r возрастает.

Лемма 1. *Если ряд $\sum a_k$ сходится, то с вероятностью единица произойдет только конечное число событий A_k.* Более точно утверждается, что если r достаточно велико, то $P\{U_r\} < \varepsilon$: *для любого $\varepsilon > 0$ найдется целое число r, такое, что, каково бы ни было n, вероятность того, что в результате n испытаний произойдет одно или больше событий A_{r+1}, A_{r+2}, ..., меньше, чем ε.*

Доказательство. Возьмем r таким, чтобы $a_{r+1} + a_{r+2} + \ldots < \varepsilon$; это возможно, так как ряд $\sum a_k$ сходится. Без ограничения общности можно предполагать, что A_k упорядочены таким образом, что $n_1 \leqslant n_2 \leqslant n_3 \leqslant \ldots$ Пусть N — наибольший индекс, при котором $n_N \leqslant n$ [1]). Тогда A_1, ..., A_N определены в пространстве n испытаний, и лемма утверждает, что вероятность произойти одному или больше событиям A_{r+1}, A_{r+2}, ..., A_N меньше ε. Это верно, так

---

[1]) Если бы неравенство $n_k \leqslant n$ имело место для бесконечной последовательности значений k_1, k_2, k_3, ... индекса k, то последовательность вероятностей a_{k_1}, a_{k_2}, a_{k_3}, ... содержала бы лишь конечное число различных чисел, и из сходимости ряда Σa_k вытекало бы, что все члены этой последовательности, за исключением конечного числа, равны нулю. В этом случае за N следовало бы принять наибольшее среди значений индекса, которым соответствуют отличные от нуля члены. — *Прим. ред.*

как согласно фундаментальному неравенству (6.6) гл. I, мы имеем для этой вероятности

$$P\{A_{r+1} \cup A_{r+2} \cup \ldots \cup A_N\} \leqslant a_{r+1} + a_{r+2} + \ldots + a_N \leqslant \varepsilon, \quad (3.2)$$

что и требовалось доказать.

Удовлетворительное обращение этой леммы известно лишь для специального случая взаимно независимых событий A_k. Такое положение возникает, когда все испытания разделены на взаимно неперекрывающиеся группы и A_k зависит только от испытаний k-й группы (например, A_k может быть событием, состоящим в том, что в результате k-й тысячи испытаний произойдет более чем 600 успехов).

Лемма 2. *Если события A_k взаимно независимы и если ряд $\sum A_k$ расходится, то с вероятностью единица осуществится бесконечно много событий A_k.* Другими словами, для любого r вероятность того, что в результате n испытаний произойдет более чем r событий A_k, стремится к единице при $n \to \infty$.

Доказательство. Как и при доказательстве леммы 1, пусть через $A_1, A_2, \ldots, A_N$ обозначаются события, определенные в пространстве элементарных событий, соответствующем n испытаниям. Вероятность того, что не произойдет ни одно из этих событий, равна, ввиду предположения о независимости, $(1-a_1)(1-a_2)\ldots(1-a_N)$. Так как $1-x<e^{-x}$ при $0<x<1$, то $(1-a_1)(1-a_2)\ldots$ $\ldots(1-a_N)<e^{-(a_1+a_2+\ldots+a_N)}$. Правая часть этого неравенства при возрастании N стремится к нулю. Мы доказали, таким образом, что с вероятностью единица произойдет по меньшей мере одно из событий $\{A_k\}$.

Далее, разобьем последовательность $\{A_k\}$ на две подпоследовательности $\{A_k^*\}$ и $\{A_k^{**}\}$ так, чтобы оба ряда $\sum P\{A_k^*\}$ и $\sum P\{A_k^{**}\}$ расходились. Применив наш результат к этим подпоследовательностям, мы убеждаемся, что с вероятностью единица произойдут по крайней мере одно из событий A_k^* и одно из событий A_k^{**}. Таким образом, с вероятностью единица произойдут по крайней мере два события из последовательности $\{A_k\}$. Применив это утверждение к последовательностям $\{A_k^*\}$ и $\{A_k^{**}\}$, мы убеждаемся, что должны произойти по крайней мере четыре события из последовательности $\{A_k\}$ и т. д.

Пример. Какова вероятность того, что в последовательности испытаний Бернулли комбинация *УНУ* появится бесконечное число раз? Пусть A_k будет событием, состоящим в том, что испытания с номерами k, $k+1$, $k+2$ дадут в результате *УНУ*. События A_k не являются, конечно, взаимно независимыми, но последовательность $A_1, A_4, A_7, A_{10}, \ldots$ содержит только независимые события (так как никакие два из них не зависят от исхода одного и того же испытания). Так как $a_k = p^2q$ не зависит от k, ряд $a_1 + a_4 + a_7 + \ldots$

расходится, и, следовательно, с вероятностью единица комбинация *УНУ* появится бесконечно много раз. Аналогичное доказательство применимо, очевидно, и к другим комбинациям. (Дальнейшие примеры можно найти в задачах 4 и 5.)

§ 4. Усиленный закон больших чисел

Наглядное представление о вероятности основывается на представлении о правильности следующего утверждения: если число успехов в первых n испытаниях последовательности испытаний Бернулли равно S_n, то

$$\frac{S_n}{n} \to p. \tag{4.1}$$

В абстрактной теории это не может быть верным для *каждой* последовательности испытаний; действительно, в наше пространство элементарных событий входят элементарные события, дающие логическую возможность появления в результате испытания, например, только одних успехов; в этом случае $S_n/n = 1$. Однако можно доказать, что (4.1) выполняется с вероятностью единица; поэтому случаи, когда (4.1) не выполняется, являются исключением, которым можно пренебречь.

Отметим, что мы рассматриваем сейчас утверждение более сильное, чем обычный закон больших чисел [гл. VI, формула (4.2)]. Последний говорит о том, что для любого достаточно большого *фиксированного* числа n весьма правдоподобно, что отношение S_n/n окажется близким к p, но он не утверждает, что S_n/n обязано оставаться близким к p, когда число испытаний n возрастает. Этот закон больших чисел оставляет открытой возможность того, что в последующих испытаниях осуществится по крайней мере одно из событий $S_{n+1}/(n+1) < p-\varepsilon$, или $S_{n+2}/(n+2) < p-\varepsilon$, ..., или $S_{2n}/2n < p-\varepsilon$; здесь мы имеем дело с объединением большого числа событий, о которых мы знаем только, что вероятность каждого из них в отдельности мала. Докажем теперь, что с вероятностью единица разность $S_n/n - p$ становится *и остается* малой при неограниченном возрастании n.

Усиленный закон больших чисел. *Для любого* $\varepsilon > 0$ *вероятность того, что осуществится только конечное число событий*

$$\left|\frac{S_n}{n} - p\right| > \varepsilon, \tag{4.2}$$

равна единице. Отсюда следует, что (4.1) выполняется с вероятностью единица. На языке конечных пространств элементарных событий утверждается, что для любых $\varepsilon > 0$ и $\delta > 0$ найдется r, такое, что при всех ν вероятность совместного осуществления ν неравенств

$$\left| \frac{S_{r+k}}{r+k} - p \right| < \varepsilon, \qquad k = 1, 2, \ldots, \nu \tag{4.3}$$

больше, чем $1 - \delta$.

Доказательство. Мы докажем более сильное утверждение. Обозначим через A_k событие

$$|S_k^*| = \left| \frac{S_k - kp}{(kpq)^{1/2}} \right| \geqslant (2a \log k)^{1/2}. \tag{4.4}$$

Тогда по формуле (5.2) гл. VII получим

$$P\{A_k\} < e^{-a \log k} = \frac{1}{k^a}, \tag{4.5}$$

и, следовательно, ряд $\sum P\{A_k\}$ сходится. Таким образом, по лемме 1 предыдущего раздела с *вероятностью единица выполняется только конечное число неравенств* (4.4). С другой стороны, из (4.2) вытекает, что

$$\left| \frac{S_n - np}{(npq)^{1/2}} \right| > \frac{\varepsilon}{(pq)^{1/2}} \cdot n^{1/2}, \tag{4.6}$$

и при больших n правая сторона больше, чем $(2a \log n)^{1/2}$. Следовательно, выполнение бесконечно большого числа неравенств (4.2) влечет за собой осуществление бесконечного числа событий A_k и имеет поэтому вероятность нуль.

Усиленный закон больших чисел был сформулирован впервые Кантелли (1917); еще до этого Борель и Хаусдорф рассматривали некоторые специальные случаи. Усиленный закон, так же как и обычный закон больших чисел, является лишь весьма частным случаем одной общей теоремы о случайных величинах. Рассмотренный совместно с нашей теоремой о невозможности системы игры закон больших чисел влечет за собой существование предела (4.1) не только для первоначальной последовательности испытаний, но и для любых подпоследовательностей, полученных в соответствии с правилом § 2. *Таким образом, эти две теоремы вместе описывают фундаментальное свойство случайности, которое присуще наглядному представлению о вероятности. Значение этого свойства усиленно подчеркивалось Мизесом.*

§ 5. Закон повторного логарифма

Так же как и в гл. VII, рассмотрим нормированное число успехов в n испытаниях:

$$S_n^* = \frac{S_n - np}{(npq)^{1/2}}. \tag{5.1}$$

Предельная теорема Лапласа утверждает, что $P\{S_n^* > x\} \sim 1 - \Phi(x)$. Таким образом, для каждого частного значения n невероятно, чтобы S_n^* было велико; однако интуитивно ясно, что при дальнейшем продолжении испытаний S_n^* примет рано или поздно сколь угодно большие значения. Умеренные значения S_n^* более вероятны, но максимум будет медленно возрастать. С какой скоростью?

В ходе доказательства усиленного закона больших чисел мы видели, что с вероятностью единица при всех достаточно больших n и $a > 1$ выполнено неравенство $S_n^* < (2a \log n)^{1/2}$. Это дает нам верхнюю оценку для S_n^*, далекую, однако, от окончательной. Для того чтобы убедиться в этом, применим наши рассуждения к подпоследовательности $S_2^*, S_4^*, S_8^*, \ldots, S_{16}^*, \ldots$, т. е. определим событие A_k соотношением $S_{2^k}^* \geqslant (2a \log k)^{1/2}$. Неравенство (4.5) показывает теперь, что $S_{2^k}^* < (2a \log k)^{1/2}$ при $a > 1$ и всех достаточно больших k. Но если $n = 2^k$, то $\log k \sim \log\log n$, и мы получаем, что для любого $a > 1$ и всех n вида $n = 2^k$ неравенство

$$S_n^* < (2a \log\log n)^{1/2} \tag{5.2}$$

выполнено начиная с некоторого k. Весьма заманчиво предположить теперь, что (5.2) выполнено для всех достаточно больших n, и действительно это одно из непосредственных следствий закона повторного логарифма. Эта замечательная теорема[1]) устанавливает, что $(2 \log\log n)^{1/2}$ является точной верхней границей в том смысле, что при $a < 1$ с вероятностью единица для бесконечного числа значений n оказывается выполненным обратное неравенство.

Теорема. *С вероятностью единица выполняется соотношение*

$$\limsup_{n\to\infty} \frac{S_n^*}{(2 \log\log n)^{1/2}} = 1. \tag{5.3}$$

Это означает, что при $\lambda > 1$ с вероятностью единица неравенство

$$S_n > np + \lambda (2npq \log\log n)^{1/2} \tag{5.4}$$

осуществится только конечное число раз, а при $\lambda < 1$ с вероятностью единица неравенство (5.4) *окажется выполненным для бесконечного числа значений n.*

---

[1]) А. Хинчин, Über einen Satz der Wahrscheinlichkeitsrechnung. *Fundamenta Mathematicae*, 6 (1924), 9—20. Этому открытию предшествовали частные результаты, полученные другими авторами. Приводимое здесь доказательство построено таким образом, что оно допускает непосредственное обобщение на случай более общих случайных величин.

По соображениям симметрии из уравнения (5.3) вытекает

$$\liminf_{n\to\infty} \frac{S_n^*}{(2\log\log n)^{1/2}} = -1. \tag{5.3a}$$

Доказательство. Мы начнем с двух предварительных замечаний.

1. Существует постоянная $c > 0$, зависящая от p, но не зависящая от n, такая, что при всех n

$$P\{S_n > np\} > c. \tag{5.5}$$

Действительно, рассмотрение биномиального распределения показывает, что левая часть (5.5) никогда не равна нулю, а по предельной теореме Лапласа она стремится к $^1/_2$ при $n\to\infty$. Поэтому левая часть больше постоянной, большей нуля, как и утверждалось.

2. Нам нужна также следующая лемма. Пусть x фиксировано и через A обозначено событие, состоящее в том, что по крайней мере при одном значении k, удовлетворяющем неравенству $k\leqslant n$, имеет место неравенство

$$S_k - kp > x. \tag{5.6}$$

Тогда

$$P\{A\} \leqslant \frac{1}{c} P\{S_n - np > x\}. \tag{5.7}$$

Для доказательства этой леммы обозначим через A_ν событие, состоящее в том, что (5.6) выполнено при $k=\nu$, но не выполнено при $k=1, 2, \ldots, \nu-1$ (здесь $1\leqslant \nu\leqslant n$). События $A_1, A_2, \ldots, A_n$ попарно несовместны, событие A есть сумма. Следовательно,

$$P\{A\} = P\{A_1\} + \ldots + P\{A_n\}. \tag{5.8}$$

Далее, пусть при $\nu < n$ событие U_ν состоит в том, что общее число успехов, выпавших в результате испытаний с номерами $\nu+1$, $\nu+2, \ldots, n$, превышает $(n-\nu)p$. Если A_ν и U_ν осуществятся совместно, то $S_n > S_\nu + (n-\nu)p > np + x$, и так как события $A_\nu U_\nu$ попарно несовместны, то отсюда следует, что

$$P\{S_n - np > x\} \geqslant$$
$$\geqslant P\{A_1U_1\} + P\{A_2U_2\} + \ldots + P\{A_{n-1}U_{n-1}\} + P\{A_n\}. \tag{5.9}$$

Так как A_ν зависит только от первых ν испытаний, а U_ν — только от следующих $n-\nu$ испытаний, то A_ν и U_ν независимы и $P\{A_\nu U_\nu\} = P\{A_\nu\}P\{U_\nu\}$. Из предварительного замечания (5.5) мы знаем, что $P\{U_\nu\} > c$, и, так как $c < 1$, из (5.9) и (5.8) получаем

$$P\{S_n - np > x\} \geqslant c\sum P\{A_\nu\} = cP\{A\}. \tag{5.10}$$

Это и доказывает (5.7).

3. Докажем теперь утверждение нашей теоремы, относящееся к соотношению (5.4) при $\lambda > 1$. Пусть γ — некоторое число, удовлетворяющее неравенству

$$1 < \gamma < \lambda^{1/2}, \tag{5.11}$$

и пусть n_r — целое число, ближайшее к γ^r ($r = 1, 2, 3, \ldots$). Пусть B_r — событие, состоящее в том, что неравенство

$$S_n - np > \lambda (2n_r pq \log\log n_r)^{1/2} \tag{5.12}$$

выполняется по крайней мере при одном n, таком, что $n_r \leqslant n < n_{r+1}$. Очевидно, что (5.4) может оказаться выполненным для бесконечно многих n только тогда, когда осуществится бесконечно много событий B_r. Согласно первой лемме Бореля — Кантелли, достаточно доказать, что ряд

$$\sum P\{B_r\} \tag{5.13}$$

сходится. По лемме (5.7)

$$\begin{aligned} P\{B_r\} &\leqslant c^{-1} P\{S_{n_{r+1}} - n_{r+1}p > \lambda (2n_r pq \log\log n_r)^{1/2}\} = \\ &= c^{-1} P\left\{ S^*_{n_{r+1}} > \lambda \left(2 \frac{n_r}{n_{r+1}} \log\log n_r\right)^{1/2} \right\}. \end{aligned} \tag{5.14}$$

Так как $n_r/n_{r+1} \sim \gamma^{-1} > \lambda^{-1/2}$, то при достаточно большом r

$$P\{B_r\} \leqslant c^{-1} P\{S^*_{n_{r+1}} > 2\lambda (\log\log n_r)^{1/2}\}. \tag{5.15}$$

Из предельной теоремы Муавра — Лапласа [формула (5.2) гл. VII] следует, что при больших r

$$P\{B_r\} \leqslant c^{-1} e^{-\lambda \log\log n_r} = \frac{1}{c (\log n_r)^{\lambda}} \sim \frac{1}{c (r \log \gamma)^{\lambda}}. \tag{5.16}$$

Так как $\lambda > 1$, утверждение (5.13) доказано.

4. Наконец, докажем утверждение, касающееся (5.4) при $\lambda < 1$. На этот раз выберем в качестве λ настолько большое целое число, чтобы

$$\frac{\gamma - 1}{\gamma} > \eta > \lambda, \tag{5.17}$$

где η — некоторая постоянная, которая будет определена позднее. Положим

$$n_r = \gamma^r. \tag{5.18}$$

Вторая лемма Бореля — Кантелли применима только к независимым событиям. Поэтому рассмотрим величины

$$D_r = S_{n_r} - S_{n_{r-1}}. \tag{5.19}$$

D_r является числом успехов, происшедших в испытаниях, следующих за n_{r-1}-м вплоть до n_r-го включительно. Величина D_r имеет биномиальное распределение $b(k;\ n,\ p)$ с $n = n_r - n_{r-1}$. Пусть A_r — событие, состоящее в том, что

$$D_r - (n_r - n_{r-1})\,p > \eta\,(2pqn_r \log\log n_r)^{1/2}. \tag{5.20}$$

Мы утверждаем, что с вероятностью единица осуществится бесконечно много событий A_r. Так как различные события A_r зависят от неперекрывающихся групп испытаний (именно: A_r зависит от испытаний с номерами $n_{r-1} < n \leqslant n_r$), то эти события взаимно независимы и в соответствии со второй леммой Бореля — Кантелли достаточно доказать, что ряд

$$\sum P\{A_r\} \tag{5.21}$$

расходится. Но

$$P\{A_r\} = P\left\{\frac{D_r - (n_r - n_{r-1})\,p}{\{(n_r - n_{r-1})\,pq\}^{1/2}} > \eta\left(\frac{2n_r}{n_r - n_{r-1}}\log\log n_r\right)^{1/2}\right\}. \tag{5.22}$$

Здесь $n_r/(n_r - n_{r-1}) = \gamma/(\gamma - 1) < \eta^{-1}$ по (5.17). Следовательно,

$$P\{A_r\} \geqslant P\left\{\frac{D_r - (n_r - n_{r-1})\,p}{\{(n_e - n_{r-1})\,pq\}^{1/2}} > (2\eta\log\log n_r)^{1/2}\right\}. \tag{5.23}$$

Используя снова оценку (5.2) гл. VII, убеждаемся, что при больших r

$$P\{A_r\} > \frac{1}{2\eta\log\log n_r}\,e^{-\eta\log\log n_r} = \frac{1}{2\eta\,(\log\log n_r)\,(\log n_r)^{\eta}}. \tag{5.24}$$

Так как $n_r = \gamma^r$ и $\eta < 1$, мы убеждаемся, что при больших r $P\{A_r\} > 1/r$, что доказывает (5.21).

В качестве последнего этапа доказательства покажем, что можно пренебречь величиной $S_{n_{r-1}}$ в (5.19). Из первой, уже доказанной, части теоремы следует, что для любого $\varepsilon > 0$ можно найти N такое, что с вероятностью $1 - \varepsilon$ или большей при всех $r > N$ выполняется неравенство

$$|S_{n_{r-1}} + n_{r-1}p| < 2\,(2pqn_{r-1}\log\log n_{r-1})^{1/2}. \tag{5.25}$$

Теперь предположим, что η выбрано столь близким к единице, что

$$1 - \eta < \left(\frac{\eta - \lambda}{2}\right)^2. \tag{5.26}$$

Тогда из (5.17) следует, что

$$4n_{r-1} = 4\,\frac{n_r}{\gamma} < n_r\,(\eta - \lambda)^2. \tag{5.27}$$

Следовательно, из (5.25) получаем

$$S_{n_{r-1}} - n_{r-1}p > -(\eta - \lambda)\,(2pqn_r\log\log n_r)^{1/2}. \tag{5.28}$$

Прибавив (5.28) к (5.20), мы получаем (5.4) с $n = n_r$. Отсюда следует, что с вероятностью $1 - \varepsilon$ или большей неравенство (5.4) выполняется для бесконечно многих r; этим завершается доказательство.

Закон повторного логарифма для испытаний Бернулли представляет собой частный случай более общей теоремы, впервые сформулированной А. Н. Колмогоровым [1]). В настоящее время можно сформулировать и более сильные теоремы (см. задачи 7 и 8).

§ 6. Интерпретация на языке теории чисел

Пусть x — действительное число из интервала $0 \leqslant x < 1$ и

$$x = 0,\ a_1 a_2 a_3 \ldots \tag{6.1}$$

десятичное разложение этого числа (так что каждое из a_j означает одну из цифр 0, 1, ..., 9). Такое разложение единственно для всех чисел, кроме чисел вида $a/10^n$, где n — целое; последние могут быть записаны как в виде конечного разложения (содержащего бесконечно много нулей), так и в виде разложения, содержащего бесконечно много девяток. Чтобы избежать неопределенности, условимся не пользоваться разложениями второго вида.

Десятичное разложение связано с испытаниями Бернулли с вероятностью $p = 1/10$ появления цифры 0, играющей роль «успеха» (все остальные цифры играют роль «неудач»). Если мы рассмотрим цифры после запятой в (6.1) и заменим все нули буквами $У$, а все другие цифры буквами H, то (6.1) будет представлять возможный исход бесконечной последовательности испытаний Бернулли с $p = 1/10$. Обратно, произвольная последовательность букв $У$ и H может быть получена описанным способом из разложения некоторого числа x. Таким образом, каждое событие в пространстве элементарных событий, связанном с испытаниями Бернулли, изображается некоторой совокупностью чисел x. Например, событию «успех в n-м испытании» соответствует множество всех тех x, у которых n-й десятичный знак является нулем. Это множество представляет собой совокупность 10^{n-1} интервалов длиной 10^{-n} каждый, и общая длина этих интервалов равна $1/10$, т. е. вероятности нашего события. Каждому исходу конечного числа испытаний соответствует определенная совокупность интервалов; например, исход первых трех испытаний $УНУ$ изображается девятью интервалами: $0{,}01 \leqslant x < 0{,}011$, $0{,}02 \leqslant x < 0{,}021$, ..., $0{,}09 \leqslant x < 0{,}091$. Вероятность каждого исхода конечного числа испытаний равна суммарной длине соответствующих интервалов на оси x. Вероятности более сложных событий всегда выражаются через вероятности таких исходов, и вычисление их производится в соответствии с правилом сложения, которое верно

[1]) А. Н. Колмогоров, Das Gesetz des iterierten Logarithmus. *Mathematische Annalen*, **101** (1929), 126—135.

для обычной меры Лебега на оси x. Поэтому наши вероятности будут всегда совпадать с мерами соответствующих множеств точек оси x. Мы получили, таким образом, способ перевода всех предельных теорем об испытаниях Бернулли с $p = 1/10$ в теоремы, относящиеся к разложениям в десятичные дроби. Термин «с вероятностью единица» оказывается эквивалентным термину «для почти всех» или «почти всюду».

Рассмотрим случайную величину S_n, равную числу успехов в n испытаниях. Здесь оказывается более удобным подчеркнуть, что S_n является функцией от элементарных событий, и мы используем поэтому запись $S_n(x)$ для *количества нулей среди первых n десятичных знаков числа x*. Очевидно, что $S_n(x)$ является функцией от x, график которой изображается ступенчатой ломаной линией с разрывами в точках вида $a/10^n$, где a — целое число. Отношение $S_n(x)/n$ называется *частотой нуля* среди первых n десятичных знаков числа x.

На языке обычной теории меры закон больших чисел говорит о том, что $S_n(x)/n \to 1/10$ по мере, в то время как усиленный закон больших чисел утверждает, что $S_n(x)/n \to 1/10$ почти всюду. Закон повторного логарифма Хинчина показывает, что

$$\limsup \frac{S_n(x) - n/10}{(n \log \log n)^{1/2}} = (0,3)\, 2^{1/2} \tag{6.2}$$

для почти всех x. Это дает ответ на проблему, изучавшуюся рядом авторов, начиная с Хаусдорфа [1] (1913) и Харди и Литльвуда [2] (1914). Дальнейшие усовершенствования этого результата см. в задачах 7 и 8.

Вместо цифры нуль можно рассмотреть любую другую цифру и, используя усиленный закон больших чисел, показать, что частота каждой из десяти цифр стремится к 1/10 для почти всех x. Аналогичная теорема остается верной, если перейти от десятичной системы исчисления к любой другой. Этот факт, обнаруженный Борелем (1909), обычно выражается утверждением, что почти все числа «нормальны» [3].

§ 7. Задачи

1. Найти такое целое число β, чтобы при бросании кости вероятность того, что серия из трех последовательных единиц встретится раньше, чем серия из β неединиц, оказалась примерно равной половине.

2. Рассмотрим последовательность независимых испытаний с тремя возможными исходами A, B, C и соответствующими вероятностями p, q

---

[1]) F. Hausdorf, Gründzüge der Mengenlehre, Leipzig, 1913.

[2]) Hardy and Littlewood, Some problems of Diophantine approximation. *Acta Mathematica*, **37** (1914), 155—239.

[3]) Вопросы, затронутые в этом параграфе, были подробно освещены в работе А. Я. Хинчина „Метрические задачи теории иррациональных чисел", (*Успехи матем. наук*, вып. I, 1936, 7—32). — *Прим. ред.*

и r ($p+q+r=1$). Найти вероятность того, что серия A длины α произойдет раньше, чем серия B длины β.

3. Продолжение. Найти вероятность того, что серия A длины α произойдет раньше, чем серия B длины β или серия C длины γ.

4. В последовательности испытаний Бернулли пусть A_n будет событием, состоящим в том, что серия из n последовательных успехов произойдет между 2^n-м и 2^{n+1}-м испытаниями. Если $p \geqslant 1/2$, то с вероятностью единица осуществится бесконечно много событий A_n; если $p < 1/2$, то с вероятностью единица осуществится только конечное число событий A_n.

5. Обозначим через N_n длину серии успехов, начиная с n-го испытания (т. е. $N_n = 0$, если результатом n-го испытания является H, и т. д). Доказать, что с вероятностью единица

$$\limsup \frac{N_n}{\operatorname{Log} n} = 1, \tag{7.1}$$

где Log означает логарифм при основании $\frac{1}{p}$.

Указание. Рассмотреть событие A_n, состоящее в том, что после n-го испытания следует серия не менее чем $a \operatorname{Log} n$ успехов. При $a > 1$ дальнейшие рассуждения довольно просты. При $a < 1$ исследовать подпоследовательность событий с номерами $a_1, a_2, \ldots$, где a_n — целое число, ближайшее к $n \operatorname{Log} n$.

6. Вывести из закона повторного логарифма, что с вероятностью единица существует бесконечное число значений n, таких, что все S_k при $n < k < 17n$ положительны. (Замечание. Значительно более сильное утверждение можно доказать, используя результаты гл. III.)

7. Пусть $\varphi(t)$ — положительная монотонно возрастающая функция и пусть n_r — ближайшее целое число к $e^{r/\log r}$. Если ряд

$$\sum \frac{1}{\varphi(n_r)} e^{-\frac{1}{2}\varphi^2(n_r)} \tag{7.2}$$

сходится, то с вероятностью единица неравенство

$$S_n > np + (npq)^{1/2} \varphi(n) \tag{7.3}$$

имеет место только для конечного числа n (отметим, что без уменьшения общности можно предполагать, что $\varphi(n) < 10 (\log\log n)^{1/2}$; закон повторного логарифма обеспечивает верность теоремы для больших $\varphi(n)$).

8. Доказать[1]), что ряд (7.2) сходится тогда и только тогда, когда сходится ряд

$$\sum \frac{\varphi(n)}{n} e^{-\frac{1}{2}\varphi^2(n)}. \tag{7.4}$$

Указание. Собрав члены с номерами n, удовлетворяющими неравенству $n_{r-1} < n < n_r$, заметим, что $n_r - n_{r-1} \sim n_r \left(1 - \frac{1}{\log r}\right)$; следовательно, (7.4) может сходиться только тогда, когда $\varphi^2(n) > 2 \log\log n$.

[1]) Задачи 7 и 8 показывают, что в случае, когда (7.4) сходится, неравенство (7.3) выполняется с вероятностью единица лишь для конечного числа n. Доказать обратное утверждение более трудно. См. W. Feller, The general form of the so-called law of the iterated logarithm, *Transactions of the American Mathematical Society*, **54** (1943), 373—402, где более общая теорема доказывается для произвольных случайных величин. Частный случай испытаний Бернулли с $p = 1/2$ см. у P. Erdös, On the law of the iterated logarithm, *Annals of Mathematics* (2), **43** (1942), 419—436. Закон повторного логарифма получается как частный случай при $\varphi(t) = \lambda (2 \log\log t)^{1/2}$.

ГЛАВА IX

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ; МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

§ 1. Случайные величины

В соответствии с определением, даваемым в курсах анализа, величина y называется *функцией* действительной переменной x, если каждому значению x соответствует некоторое значение y. Это определение может быть расширено и на те случаи, когда незави симая переменная не является действительным числом. Так, мы называем расстояние функцией пары точек; периметр треугольника является функцией, определенной на множестве треугольников; последовательность a_n есть функция, определенная для всех целых положительных чисел; биномиальный коэффициент $\binom{x}{k}$ является функцией, определенной на парах чисел (x, k), в которых второе число является целым и неотрицательным. В том же смысле мы можем сказать, что число успехов S_n в n испытаниях Бернулли является функцией, определенной на пространстве элементарных событий; каждой из 2^n точек этого пространства соответствует некоторое число S_n.

Функция, определенная на пространстве элементарных событий, называется случайной величиной. В предшествующих главах мы занимались случайными величинами, не употребляя еще этого термина. Типичными примерами случайных величин могут служить: число тузов, полученных некоторым игроком при сдаче колоды карт; число дней рождения в мае для группы из n человек; число серий, образованных «успехами» в n испытаниях Бернулли. В любом из этих случаев существует единое правило, сопоставляющее каждому элементарному событию некоторое число X. Можно сказать, что классическая теория вероятностей занималась в основном изучением выигрыша, который есть также случайная величина; действительно, каждая случайная величина может быть интерпретирована как выигрыш игрока в некоторой игре, построенной по надлежащей схеме. Положение частицы при диффузии, энергия, температура и т. п. физической системы являются случайными величинами, но они определены для недискретных пространств элементарных событий, и поэтому их изучение нам придется отложить. В случае дискретного пространства элементарных событий мы можем задать любую случайную величину X, занумеровав в каком-либо порядке все точки пространства и сопоставив каждой из них соответствующее значение величины X.

Термин «случайная величина»[1]) не совсем удачен, и более подходящим был бы термин «случайная функция» (независимой переменной является точка в пространстве элементарных событий, т. е. исход эксперимента).

Пусть X — случайная величина, и x_1, x_2, x_3 ... суть значения [2]), которые она принимает (в последующем x_j в большинстве случаев будут целыми числами). Одно и то же значение x_j может соответствовать, вообще говоря, нескольким элементарным событиям. Совокупность всех этих элементарных событий образует составное событие, состоящее в том, что $X = x_j$. Вероятность этого события обозначается через $P\{X = x_j\}$. *Система равенств*

$$P\{X = x_j\} = f(x_j) \qquad (j = 1, 2, \ldots) \tag{1.1}$$

определяет распределение [3]) *(вероятностей) случайной величины* X. Ясно, что

$$f(x_j) \geqslant 0; \quad \sum f(x_j) = 1. \tag{1.2}$$

В том же смысле можно сказать, что число успехов S_n в n испытаниях Бернулли есть случайная величина с вероятностным распределением $\{b(k; n, p)\}$, тогда как число испытаний до первого успеха включительно есть случайная величина с распределением $\{q^{k-1}p\}$.

Рассмотрим теперь две случайные величины X и Y, определенные на одном и том же пространстве элементарных событий, и обозначим значения, которые они принимают, соответственно через x_1, x_2, ... и y_1, y_2, ...; пусть соответствующие распределения вероят-

---

[1]) В дословном переводе «случайная переменная». — *Прим. ред.*

[2]) Используя стандартную математическую терминологию, следовало бы назвать множество точек $x_1, x_2, \ldots$ областью изменения величины X. К сожалению, в статистической литературе тот же термин употребляется для обозначения разности между максимальным и минимальным значениями X.

[3]) Для дискретной случайной величины X распределение вероятностей является функцией, определенной на множестве возможных значений x_j величины X. Этот термин нужно отличать от термина «функция распределения», который может быть приложен к любой неубывающей функции $F(x)$, стремящейся к нулю при $x \to -\infty$ и к единице при $x \to +\infty$. Функция распределения $F(x)$ дискретной случайной величины X определяется как

$$F(x) = P\{X \leqslant x\} = \sum_{x_j \leqslant x} f(x_j).$$

Последняя сумма взята по всем x_j, не превышающим x. Таким образом, функция распределения может быть вычислена по распределению вероятностей, и наоборот. В этом томе мы не будем иметь дела с функциями распределения.

ностей будут $\{f(x_j)\}$ и $\{g(y_k)\}$. Совокупность элементарных событий, для которых выполнены условия $X = x_j$ и $Y = y_k$, образует событие, вероятность которого будет обозначаться через $P\{X = x_j, Y = y_k\}$. *Система равенств*

$$P\{X = x_j,\ Y = y_k\} = p(x_j,\ y_k) \qquad (j,\ k = 1,\ 2,\ \ldots) \quad (1.3)$$

задает совместное распределение вероятностей случайных величин X и Y. Такие распределения удобнее всего записывать в виде прямоугольных таблиц, аналогичных табл. 1 и 2. Ясно, что

$$p(x_j,\ y_k) \geqslant 0; \quad \sum_{j,\,k} p(x_j,\ y_k) = 1. \qquad (1.4)$$

Кроме того, при каждом фиксированном j

$$p(x_j,\ y_1) + p(x_j,\ y_2) + p(x_j,\ y_3) + \ldots = P\{X = x_j\} = f(x_j) \quad (1.5)$$

и при каждом фиксированном k

$$p(x_1,\ y_k) + p(x_2,\ y_k) + p(x_3,\ y_k) + \ldots = P\{Y = y_k\} = g(y_k). \quad (1.6)$$

Другими словами, суммируя вероятности, стоящие в отдельных строках и столбцах, мы получим вероятностные распределения величин X и Y. Их можно представить так, как это показано в табл. 1 и 2; и тогда они называются *«крайними распределениями»*. Эпитет «крайние» связан с нашей формой записи совместного распределения X и Y в виде прямоугольной таблицы и используется также для стилистической ясности, когда в одном и том же контексте говорят о совместном распределении двух случайных величин и их частных (крайних) распределениях. Строго говоря, эпитет «крайний» может быть опущен.

Понятие совместного распределения переносится также и на *системы более чем двух случайных величин.*

Примеры. а) *Случайное размещение* 3 *шаров по* 3 *ящикам.* Обратимся к пространству элементарных событий, состоящему из 27 точек, которое формально определяется табл. 1, поясняющей пример 2, а гл. I. Как и в том примере, свяжем с каждой точкой вероятность 1/27. Пусть N означает число занятых ящиков, а при $i = 1,\ 2,\ 3$ пусть X_i означает число шаров в i-м ящике. Это наглядные описания. Говоря формально, N — функция, определенная на пространстве элементарных событий и принимающая значение 1 в точках с номерами 1—3, значение 2 в точках с номерами 4—21 и, наконец, значение 3 в точках с номерами 22—27. Соответственно этому, вероятностное распределение определяется системой равенств $P\{N = 1\} = \frac{1}{9}$, $P\{N = 2\} = \frac{2}{3}$, $P\{N = 3\} = \frac{2}{9}$. Совместные распределения $(N,\ X_1)$ и $(X_1,\ X_2)$ приведены в табл. 1 и 2.

Таблица 1

Совместное распределение (N, X_1) в примере „а"

	X_1: 0	1	2	3	Распределение N
N 1	$2q$	0	0	q	$3q=\frac{1}{9}$
2	$6q$	$6q$	$6q$	0	$18q=\frac{2}{3}$
3	0	$6q$	0	0	$6q=\frac{2}{9}$
Распределение X_1	$8q$	$12q$	$6q$	q	
$E(N)=\frac{19}{9}$	$E(N^2)=\frac{129}{27}$				$D(N)=\frac{26}{81}$
$E(X_1)=1$	$E(X_1^2)=\frac{45}{27}$				$D(X_1)=\frac{2}{3}$
$E(NX_1)=\frac{19}{9}$					$\mathrm{Cov}(N, X_1)=0$

N — число занятых ящиков, X_1 — число шаров в первом ящике, когда три шара случайно размещаются по 3 ящикам. Для краткости положено $q=\frac{1}{27}$.

б) *Кости.* Пусть через X_1, X_2, X_3 соответственно обозначены числа единиц, двоек и троек, выпавших при n бросаниях правильной кости. Вероятность $p(k_1, k_2, k_3)$ того, что n бросаний дали k_1 единиц, k_2 двоек, k_3 троек и $n-k_1-k_2-k_3$ прочих значений, определяется полиномиальным распределением (9.2) гл. VI при $p_1=p_2=p_3=\frac{1}{6}$, $p_4=\frac{1}{2}$, т. е.

$$p(k_1, k_2, k_3)=\frac{n!}{k_1!\,k_2!\,k_3!\,(n-k_1-k_2-k_3)!}\,3^{n-k_1-k_2-k_3}\cdot 6^{-n}. \qquad (1.7)$$

Эта формула задает совместное распределение величин X_1, X_2, X_3. Фиксируя k_1 и k_2 и суммируя по $k_3=0, 1, 2, \ldots, n-k_1-k_2$, мы получим, используя формулу бинома Ньютона,

$$p(k_1, k_2)=\frac{n!}{k_1!\,k_2!\,(n-k_1-k_2)!}\cdot 4^{n-k_1-k_2}6^{-n}. \qquad (1.8)$$

Мы пришли к совместному распределению величин (X_1, X_2), которое является «крайним распределением» для трехмерного распределения величин X_1, X_2 и X_3. Нет нужды говорить о том, что формулу (1.8) можно получить и непосредственно из определения полиномиального

Таблица 2

Совместное распределение (X_1, X_2) в примере „а“

		X_1: 0	1	2	3	Распределение X_1
	0	q	$3q$	$3q$	q	$8q$
	1	$3q$	$6q$	$3q$	0	$12q$
X_2	2	$3q$	$3q$	0	0	$6q$
	3	q	0	0	0	q
Распределение X_2		$8q$	$12q$	$6q$	q	
$E(X_i)=1$		$E\left(X_i^2\right)=\frac{45}{27}$				$D(X_i)=\frac{2}{3}$
$E(X_1, X_2)=\frac{2}{3}$						$\operatorname{Cov}(X_1, X_2)=-\frac{1}{3}$

X_i есть число шаров в i-м ящике при случайном размещении 3 шаров по 3 ящикам. Для краткости положено $q=\frac{1}{27}$.

распределения. Суммируя еще раз по $k_2=0, 1, \ldots, n-k_1$, мы придем к биномиальному распределению $b(k_1; n, 1/6)$ для величины X_1.

в) *Выбор.* Пусть генеральная совокупность n элементов состоит из трех классов с объемами $n_1=np_1$, $n_2=np_2$ и $n_3=np_3$ соответственно ($p_1+p_2+p_3=1$). Из этой совокупности извлекается случайно выборка объема r. Обозначим через X_1 и X_2 числа элементов первого и второго классов в выборке. Если выбор производился с возвращением, то $P\{X_1=k_1, X_2=k_2\}$ определяется полиномиальным распределением

$$f(k_1, k_2)=\frac{r!}{k_1!\,k_2!\,(r-k_1-k_2)!}\,p_1^{k_1}p_2^{k_2}p_3^{r-k_1-k_2}. \tag{1.9}$$

(См. формулу (9.2) гл. VI.) Случайные величины X_i имеют биномиальное распределение $\{b(k; r, p_i)\}$. Если выбор производился *без возвращения*, то вероятность $P\{X_1=k_1, X_2=k_2\}$ определяется двойным гипергеометрическим распределением (6.5) гл. II, а X_1 имеет обычное гипергеометрическое распределение (6.1) гл. II.

г) *Выборка случайного объема.* Рассмотрим еще раз предыдущий пример, но теперь предположим, что объем r выборки не фиксирован заранее, а определяется исходом случайного эксперимента. Говоря точнее, допустим, что объем выборки является случайной величиной, имеющей распределение Пуассона. Вероятность того, что объем выборки равен r, равна $p(r; \lambda)=e^{-\lambda}\frac{\lambda^r}{r!}$, а если объем r

задан, то (условная) вероятность того, что $X_1 = k_1$ и $X_2 = k_2$, определяется формулой (1.9). Для совместного распределения (X_1, X_2) получаем следующее выражение:

$$P\{X_1 = k_1, X_2 = k_2\} = e^{-\lambda} \sum_{r=k_1+k_2}^{\infty} \lambda^r f(k_1, k_2)/r! =$$

$$= e^{-\lambda} \frac{(\lambda p_1)^{k_1} (\lambda p_2)^{k_2}}{k_1!\, k_2!} \sum_{k_3=0}^{\infty} \frac{(\lambda p_3)^{k_3}}{k_3!} = e^{-\lambda(1-p_3)} \frac{(\lambda p_1)^{k_1} (\lambda p_2)^{k_2}}{k_1!\, k_2!} \tag{1.10}$$

или

$$P\{X_1 = k_1, X_2 = k_2\} = p(k_1; \lambda p_1)\, p(k_2; \lambda p_2). \tag{1.11}$$

Суммируя по k_2, находим распределение X_1. Это — распределение Пуассона $p(k; \lambda p_1)$. (Задача 27 гл. VI повторяет это утверждение.) Совместное распределение (X_1, X_2) задается бесконечной таблицей умножения с «крайними распределениями» $\{p(k; \lambda p_1)\}$ и $\{p(k; \lambda p_2)\}$. Случайные величины, обладающие таким свойством, называются *независимыми*.

В обозначениях (1.3) условная вероятность события $Y = y_k$ при условии $X = x_j$ (и $f(x_j) > 0$) задается формулой

$$P\{Y = y_k \mid X = x_j\} = \frac{p(x_j, y_k)}{f(x_j)}. \tag{1.12}$$

Удобно сокращенно обозначать (1.12) через $P\{Y = y_k \mid X\}$; эта система формул задает (*условное*) *распределение* Y *при условии, что известно* X. Достаточно взглянуть на табл. 1 и 2, чтобы убедиться, что условная вероятность (1.12), вообще говоря, отличается от $g(y_k)$. Это означает, что по известному значению X можно сделать некоторые заключения относительно значения Y, и наоборот; две случайные величины оказываются *зависимыми*. Максимальная степень зависимости достигается тогда, когда Y является функцией от X, т. е. когда значение X однозначным образом определяет Y. Например, если монета бросается n раз, а X и Y — числа выпавших гербов и решеток, то $Y = n - X$. Аналогично, когда $Y = X^2$, мы можем по X вычислить Y. Это означает, что в каждом из столбцов таблицы совместного распределения на всех местах, кроме одного, стоят нули. Если, с другой стороны, $p(x_j, y_k) = f(x_j)\, g(y_k)$ для всех комбинаций x_j, y_k, то события $X = x_j$ и $Y = y_k$ независимы; совместное распределение имеет вид таблицы умножения. В этом случае мы говорим о *независимых* случайных величинах.

Они возникают, в частности, в связи с независимыми испытаниями, например когда X и Y являются результатами двух бросаний кости. Иллюстрацию другого рода дает пример (г). Заметим, что совместное распределение X и Y определяет распределения самих величин X и Y, но мы не можем определить совместное распределение величин X и Y по их частным распределениям. Две случайные величины X и Y, имеющие одинаковое распределение, могут оказаться как зависимыми, так и независимыми. Например, две случайные величины X и Y из табл. 2 имеют одинаковые распределения и зависимы. С другой стороны, если X и Y имеют тот же смысл, но относятся к независимым играм, то их частные распределения будут одинаковы, но X и Y окажутся независимыми, и их совместное распределение вероятностей примет вид таблицы умножения.

Все рассмотренные понятия применимы также к случаю более чем двух случайных величин. Мы подводим итог следующим определением:

Определение. *Случайная величина X есть функция, определенная на данном пространстве элементарных событий, т. е. она задается посредством сопоставления каждому элементарному событию некоторого действительного числа. Система уравнений* (1.1) *определяет распределение (вероятностей) случайной величины X. Если две или более случайные величины $X_1, X_2, \ldots, X_n$ определены на одном и том же пространстве элементарных событий, то их совместное распределение задается системой равенств, в которых всем комбинациям $X_1 = x_{j_1}$, $X_2 = x_{j_2}$ и т. д. приписываются определенные вероятности. Случайные величины $X_1, \ldots, X_n$ называются взаимно независимыми, если для любой комбинации значений $x_{j_1}, x_{j_2}, \ldots, x_{j_n}$*

$$P\{X_1 = x_{j_1},\ X_2 = x_{j_2}, \ldots, X_n = x_{j_n}\} = \\ = P\{X_1 = x_{j_1}\} P\{X_2 = x_{j_2}\} \ldots P\{X_n = x_{j_n}\}. \quad (1.13)$$

В § 4 гл. V мы определили пространство элементарных событий, соответствующее n взаимно независимым испытаниям. Сравнивая это определение с (1.14), мы видим, что *если X_k зависит только от исхода k-го испытания, то случайные величины $X_1, X_2, \ldots, X_n$ взаимно независимы.* Вообще, если случайная величина U зависит только от исхода первых k испытаний, а другая случайная величина V зависит только от исхода остальных $n-k$ испытаний, то U и V независимы (ср. с задачей 25).

Мы можем представлять себе случайную величину как набор числовых отметок на точках пространства элементарных событий. Это представление имеет вполне наглядный смысл в случае игральной кости, где грани занумерованы и где мы говорим о числах как о возможных исходах отдельного испытания. На математическом языке можно сказать, что случайная величина является отображением основного пространства элементарных событий в новое пространство, точками которого являются $x_1, x_2, \ldots$.

Если $\{f(x_j)\}$ удовлетворяет очевидным условиям (1.2), *то законно говорить о случайной величине X, принимающей* *значения* $x_1, x_2, \ldots$ *с вероятностями* $f(x_1), f(x_2), \ldots,$ *без дальнейших указаний на первоначальное пространство элементарных событий; точками* $x_1, x_2, \ldots$ *образуется новое пространство элементарных событий. Задание распределения вероятностей эквивалентно заданию некоторого пространства элементарных событий, точками которого являются действительные числа. Говорить о двух независимых случайных величинах* X *и* Y *с распределениями* $\{f(x_j)\}$ *и* $\{g(y_k)\}$ *значит говорить о пространстве элементарных событий, точками которого являются пары чисел* (x_j, y_k), *которым отнесены вероятности по правилу* $P\{(x_j, y_k)\} = f(x_j)g(y_k)$. *Аналогично за пространство элементарных событий, соответствующее совокупности* n *случайных величин* $(X_1, \ldots, X_n)$, *можно взять множество точек* $(x_{j_1}, x_{j_2}, \ldots, x_{j_n})$ *в* n*-мерном пространстве с вероятностями, определяемыми совместным распределением. Случайные величины взаимно независимы, если их совместное распределение задается формулой* (1.13).

Пример. д) *Испытания Бернулли с переменными вероятностями.*

Рассмотрим n независимых испытаний, каждое из которых имеет только два исхода: $У$ и $Н$. Вероятность появления $У$ при k-м испытании равна p_k, и, следовательно, вероятность появления $Н$ равна $q_k = 1 - p_k$. Если $p_k = p$, то описанная схема сводится к схеме испытаний Бернулли. Разумнее всего трактовать эту схему следующим образом. Введем вместо $У$ и $Н$ числа 1 и 0. После этого модель можно полностью определить, сказав, что мы имеем n независимых случайных величин X_k с распределениями $P\{X_k = 1\} = p_k$; $P\{X_k = 0\} = q_k$. Эту схему называют иногда «схемой Пуассона». (См. примеры 5, б и 6, б, гл. XI.)

Ясно, что одно и то же распределение может возникнуть при различных пространствах элементарных событий. Если мы говорим, что случайная величина X принимает значения 0 и 1 с вероятностями 1/2, то мы подразумеваем пространство элементарных событий, состоящее из двух точек 0 и 1. Однако случайная величина X может быть задана тем условием, что она равна 0 или 1 в зависимости от того, выпадает ли при десятом бросании монеты герб или решетка; в этом случае X определено на пространстве последовательностей ($ГГР\ldots$), и это пространство состоит из 2^{10} точек.

В принципе можно ограничить теорию вероятностей изучением пространств элементарных событий, определенных в терминах распределений вероятностей случайных величин. При таком подходе становится излишним обращение к абстрактному [1]) пространству эле-

[1]) То есть к нечисловому.— *Прим. ред.*

ментарных событий, а также использование терминов, подобных «испытаниям» или «исходам эксперимента». Сведение теории вероятностей к случайным величинам является кратчайшим путем для использования анализа и упрощает многие стороны теории. Однако оно имеет и оборотную сторону, состоящую в затушевывании вероятностных оснований. Может создаться неясное представление о случайных величинах, как о «чем-то, принимающем различные значения с различными вероятностями», в то время как случайные величины являются просто обыкновенными функциями (от элементарных событий), и в понятии случайной величины нет ничего специфического, принадлежащего только теории вероятностей.

Пример. е) Пусть X — случайная величина с возможными значениями x_1, x_2, ... и соответствующими вероятностями $f(x_1)$, $f(x_2)$, ... Читатель всегда сумеет, если это поможет его воображению, построить мысленный эксперимент, приводящий к случайной величине X. Для этого достаточно, например, разделить колесо рулетки на дуги l_1, l_2, ..., длины которых относятся, как $f(x_1)$: $f(x_2)$: ..., и вообразить игрока, «получающего» (положительную или отрицательную) сумму x_j, если рулетка остановится против точки, принадлежащей дуге l_j. Тогда X есть выигрыш (или проигрыш) игрока. Выигрыши в n испытаниях соответствуют n независимым случайным величинам с одинаковыми законами распределения $\{f(x_j)\}$. Чтобы получить две случайные величины с заданным совместным распределением $\{p(x_j, y_k)\}$, надо взять дуги, соответствующие каждой комбинации (x_j, y_k), и вообразить двух игроков, получающих суммы x_j и y_k соответственно.

Если X, Y, Z, ... — случайные величины, определенные на одном и том же пространстве элементарных событий, то любая функция $F(X, Y, Z, \ldots)$ также есть случайная величина. Ее распределение может быть легко получено из совместного распределения X, Y, Z, ... собиранием членов, соответствующих комбинациям $(X, Y, Z, \ldots)$, дающим одинаковые значения для $F(X, Y, Z, \ldots)$.

Пример. ж) В примере, иллюстрирующем табл. 2, сумма $X_1 + X_2$ является случайной величиной, принимающей значения 0, 1, 2, 3 с вероятностями q, $6q$, $12q$, $8q$ (где $q = 1/27$). Произведение X_1X_2 принимает значения 0, 1, 2 с вероятностями $15q$, $6q$, $6q$.

§ 2. Математическое ожидание

Чтобы достичь разумной простоты, часто бывает необходимо суммарно описать распределение вероятностей немногими «типичными значениями».

В качестве примера можно привести медиану, которая была использована выше в связи с вопросом о времени ожидания. *Медианой распределения* (1.1) случайной величины X называется такое число x_m, для которого $P\{X \leqslant x_m\} \geqslant \frac{1}{2}$ и $P\{X \geqslant x_m\} \geqslant \frac{1}{2}$. Другими словами, x_m выбрано таким образом, чтобы вероятность того, что X меньше или больше, чем x_m, была возможно более близка к $\frac{1}{2}$.

Однако среди всех «типичных значений», безусловно, самым важным является математическое ожидание или среднее значение. Оно наиболее удобно для аналитических операций и в то же время обладает полезным для статистиков свойством, называемым выборочной устойчивостью. Определение математического ожидания связано с обычным понятием о среднем значении. Если n_k семей из некоторой генеральной совокупности имеют ровно по k детей, то общее число семей равно $n = n_0 + n_1 + n_2 + \ldots$, а общее число детей равно $m = n_1 + 2n_2 + 3n_3 + \ldots$ Среднее число детей, приходящихся на одну семью, равно m/n. Аналогия между вероятностями и частотами подсказывает следующее

Определение. *Пусть X является случайной величиной, принимающей значения x_1, x_2, ... с вероятностями $f(x_1)$, $f(x_2)$, .. Математическое ожидание величины X определяется формулой*

$$E(X) = \sum x_k f(x_k), \tag{2.1}$$

в предположении, что ряд (2.1) *абсолютно сходится. В этом случае мы говорим, что X имеет конечное математическое ожидание. Если ряд $\sum |x_k| f(x_k)$ расходится, то мы говорим, что X не имеет конечного математического ожидания.*

Само собой разумеется, большинство обычно встречающихся случайных величин имеют конечное математическое ожидание; в противном случае это понятие было бы бесплодным. Однако случайные величины, не имеющие математического ожидания, встречаются в связи с важными физическими задачами о возвращении. Термины *математическое ожидание* и *среднее значение* являются *синонимами*. Мы говорим также *о среднем значении распределения*, вместо указания на соответствующую случайную величину. Обозначение $E(X)$ является общепринятым в математике и статистике[1]). В физике обозначение $E(X)$ обычно заменяется на $\overline{X}$, $\langle X \rangle$ или $\langle X \rangle_{Av}$.

Мы желаем научиться вычислять математические ожидания функций, таких, например, как X^2. Эта функция является новой случайной величиной, принимающей значения x_k^2. Вообще говоря, вероятность того, что $X^2 = x_k^2$ равна не $f(x_k)$, а $f(x_k) + f(-x_k)$ и

[1]) В русской литературе по теории вероятностей употребляется также обозначение $M(X)$. — *Прим. ред.*

$E(X^2)$, определяется как сумма $x_k^2\{f(x_k)+f(-x_k)\}$ по всем k, для которых $x_k \geqslant 0$. Очевидно,

$$E(X^2)=\sum x_k^2 f(x_k) \tag{2.2}$$

(в предположении, что ряд сходится). Тем же способом (группировка членов) можно получить и более общую теорему.

Теорема 1. *Для произвольной функции $\varphi(x)$ случайная величина $\varphi(X)$ имеет математическое ожидание*

$$E(\varphi(X))=\sum \varphi(x_k) f(x_k) \tag{2.3}$$

[*при этом ряд абсолютно сходится тогда и только тогда, когда $E(\varphi(X))$ существует*]. *Для любой постоянной a имеем $E(aX)=aE(X)$.*

Если различные случайные величины $X_1, \ldots, X_n$ определены на одном и том же пространстве элементарных событий, то их сумма $X_1+\ldots+X_n$ является новой случайной величиной. Ее возможные значения и соответствующие им вероятности могут быть легко найдены по совместному распределению случайных величин $x_1, \ldots, x_n$, и таким образом может быть вычислено математическое ожидание $E(X_1+\ldots+X_n)$. С помощью простого рассуждения получается следующая важная теорема.

Теорема 2. *Если $X_1, \ldots, X_n$ суть случайные величины с конечными математическими ожиданиями, то математическое ожидание их суммы существует и равно сумме их математических ожиданий:*

$$E(X_1+\ldots+X_n)=E(X_1)+\ldots+E(X_n). \tag{2.4}$$

Доказательство. Достаточно доказать (2.4) для двух случайных величин X и Y. Используя обозначения (1.3), можно написать

$$E(X)+E(Y)=\sum_{j,k} x_j p(x_j, y_k)+\sum_{j,k} y_k p(x_j, y_k), \tag{2.5}$$

где суммирование производится по всем возможным значениям x_j, y_k (которые не обязаны быть все различными). Оба ряда сходятся абсолютно, и поэтому перестановкой членов можно представить их сумму в виде $\sum\limits_{j,k}(x_j+y_k)p(x_j, y_k)$. Однако это и есть определение математического ожидания $X+Y$, чем завершается доказательство.

Ясно, что аналогичная общая теорема о произведении не имеет места. Например, $E(X^2)$, вообще говоря, отличается от $(E(X))^2$. Так, если X — число очков, выпавшее при бросании правильной игральной кости, то $E(X)=7/2$, в то время как $E(X^2)=(1+4+9+$ $+16+25+36)/6=91/6$. Однако для независимых случайных величин действует простое правило умножения

Теорема 3. *Если X и Y — взаимно независимые случайные величины с конечными математическими ожиданиями, то их произведение является случайной величиной с конечным математическим ожиданием и*

$$E(XY) = E(X)E(Y). \tag{2.6}$$

Доказательство. Для вычисления $E(XY)$ умножим каждое возможное значение $x_j y_k$ на соответствующую вероятность. Мы уже отмечали, что значения x_k в формуле (2.1) не обязаны быть различными. Следовательно,

$$E(XY) = \sum_{j,k} x_j y_k f(x_j) g(y_k) = \left\{ \sum_j x_j f(x_j) \right\} \left\{ \sum_k y_k g(y_k) \right\}, \tag{2.7}$$

где перестановка членов законна, так как ряды абсолютно сходятся. Это и доказывает теорему. По индукции это же правило умножения распространяется на любое число *взаимно независимых случайных величин.*

Удобно также иметь обозначение для математического ожидания условного вероятностного распределения. Если X и Y — две случайные величины с совместным распределением (1.3), то *условным математическим ожиданием $E(Y \mid X)$ Y при условии, что известно X*, называется функция

$$\sum_k y_k P\{Y = y_k \mid X = x_j\} = \frac{\sum_k y_k p(x_j, y_k)}{f(x_j)} \tag{2.8}$$

(если только все ряды сходятся абсолютно и $f(x_j) > 0$ при всех j).

§ 3. Примеры и приложения

а) **Биномиальное распределение.** Пусть S_n — число успехов в n испытаниях Бернулли с вероятностью успеха p. Мы знаем, что S_n имеет биномиальное распределение $\{b(k; n, p)\}$ [см. гл. VI, формула (2.1)]. Следовательно,

$$E(S_n) = \sum kb(k; n, p) = np \sum b(k-1; n-1, p).$$

Последняя сумма состоит из всех членов биномиального распределения для $n-1$ испытаний и, следовательно, равна единице. Таким образом, *математическое ожидание для биномиального распределения* равно

$$E(S_n) = np. \tag{3.1}$$

Тот же результат можно получить без вычислений при помощи метода, который часто оказывается полезным. Пусть X_k будет числом успехов, выпавших в k-м испытании. Эта случайная величина принимает значения 0 и 1 с вероятностями p и q. Следовательно, $E(X_k) = 0 \cdot q + 1 \cdot p = p$, а так как

$$S_n = X_1 + X_2 + \ldots + X_n, \tag{3.2}$$

то мы получаем (3.1) непосредственно из теоремы (2.4).

б) **Распределение Пуассона.** Если X имеет распределение Пуассона [см. формулу (5.1) гл. VI], то

$$E(X) = \sum kp(k;\ \lambda) = \lambda \sum p(k-1;\ \lambda).$$

Последний ряд содержит все члены распределения и дает поэтому в сумме единицу. Следовательно, *математическое ожидание для распределения Пуассона равно* λ.

в) **Отрицательное биномиальное распределение.** Рассмотрим случайную величину X с *геометрическим распределением* $P\{X = k\} = q^k p$, где $k = 0, 1, 2, \ldots$. Тогда $E(X) = qp(1 + 2q + 3q^2 + \ldots)$. Справа в скобках стоит производная от ряда, являющегося суммой геометрической прогрессии, так что сумма этого ряда равна $E(X) = qp(1-q)^{-2} = q/p$. В § 8 гл. VI мы установили, что X можно интерпретировать, как число неудач, предшествующих первому успеху в последовательности испытаний Бернулли. Более общей ситуацией является следующая. Рассмотрим пространство элементарных событий, которое соответствует испытаниям Бернулли, продолжающимся до n-го успеха. Пусть X — число неудач между $(k-1)$-м и k-м успехами $(X_1 = X)$. Тогда каждое X_ν имеет геометрическое распределение $\{q^k p\}$ и $E(X_\nu) = q/p$. Сумма $Y_r = X_1 + X_2 + \ldots + X_r$ есть число неудач, предшествующих k-му успеху. Другими словами, Y_r — случайная величина, имеющая отрицательное биномиальное распределение, определяемое одной из двух эквивалентных формул (8.1) или (8.2) гл. VI. Следовательно, *среднее значение отрицательного биномиального распределения равно* rq/p. Это можно проверить и непосредственным подсчетом. Действительно, из формулы (8.2) гл. VI следует, что $kf(k;\ r,\ p) = rp^{-1}qf(k-1;\ r+1,\ p)$ и члены распределения $\{f(k-1;\ r+1,\ p)\}$ дают в сумме единицу. Преимущество этого метода состоит в том, что он применим и в том случае, когда число r *не является целым*. С другой стороны, первоначальные рассуждения позволяют получить нужный результат, не зная точного вида распределения $X_1 + X_2 + \ldots + X_r$.

г) **Время ожидания при выборе.** Из генеральной совокупности, содержащей N различных элементов, производится выбор с возвращением. Вследствие возможности повторений случайная выборка объема r будет, вообще говоря, содержать меньше, чем r различных элементов. При возрастании объема выборки новые элементы будут попадать в выборку все реже и реже. Мы интересуемся объемом выборки S_r, необходимым для получения ровно r элементов. (Как пример рассмотрим генеральную совокупность из 365 возможных дней рождения; здесь S_r представляет собой число людей, вошедших в нашу выборку, прежде чем число различных дней рождения достигнет r. Аналогичная интерпретация возможна и для случайного распределения шаров по ячейкам. Наша задача представляет специфический интерес для коллекционеров, так как пополнение коллекции часто может быть уподоблено случайному выбору [1]).)

Первый элемент входит в выборку при первом извлечении. Вслед за этим потребуется некоторое число извлечений для того, чтобы в выборку впервые вошел элемент, отличный от первого. Это число является случайной величиной X_1. Вообще, пусть X_r будет числом извлечений, следующих за выбором r-го элемента до выбора нового $(r+1)$-го элемента включительно. Тогда $S_r = 1 + X_1 + X_2 + \ldots + X_{r-1}$ является объемом выборки в тот момент, когда в выборку входит r-й новый элемент. Если выборка уже содержит k различных элементов, то вероятность выбора нового элемента равна при каждом извлечении $p = (N-k)/N$.

Распределение случайной величины X_k совпадает с распределением номера первого успеха в последовательности испытаний Бернулли с $p = (N-k)/N$. Следовательно, $E(X_k) = 1 + q/p = N(N-k)$, и по теореме сложения (2.4) получаем

$$E(S_r) = N\left\{\frac{1}{N} + \frac{1}{N-1} + \frac{1}{N-2} + \ldots + \frac{1}{N-r+1}\right\}. \tag{3.3}$$

При $r = N$ получаем математическое ожидание числа извлечений, необходимых для исчерпания всей генеральной совокупности. При $N = 10$ имеем $E(S_{10}) = 29,29\ldots$ и $E(S_5) = 6,46\ldots$ Следовательно, можно ожидать, что первая половина генеральной совокупности будет получена примерно за 6—7 извлечений, в то время как на вторую половину потребуется более 23 извлечений. При больших N неплохим приближением к (3.5) является

$$E(S_r) \approx N \log \frac{N}{N-r+1}. \tag{3.4}$$

В частности, для любой дроби $\alpha < 1$, *математическое ожидание числа извлечений, которое нужно для получения доли α от всей генеральной совокупности, при больших N приблизительно*

---

[1]) G. Polya, Eine Wahrscheinlichkeitsaufgabe zur Kundenwerbung *Zeitschrift für Angewandte Mathematik und Mechanik*, 10 (1930), 96—97. Пойа другими методами разбирает несколько более общую задачу. Существует обширная литература, посвященная различным вариантам задачи о коллекционировании. (См. задачи 24.25 гл. XI, 12—14 настоящей главы и (11.12) гл. II.)

равно $N \log 1/(1-\alpha)$; *математическое ожидание числа извлечений, необходимых для получения всех N элементов, входящих в генеральную совокупность, приблизительно равно* $N \log N$. Отметим, что наш результат получен снова без использования распределения изучаемых случайных величин.

д) **Задача об оценке.** Урна содержит шары, снабженные номерами от 1 до N. *Пусть X — наибольший номер, полученный в результате n извлечений*, если производится случайный выбор с возвращением. Событие $X \leqslant k$ означает, что каждый из извлеченных номеров не превосходит k, так что $P\{X \leqslant k\} = (k/N)^n$. Следовательно, распределение вероятностей для случайной величины X определяется формулой

$$p_k = P\{X = k\} = P\{X \leqslant k\} - P\{X \leqslant k-1\} = \{k^n - (k-1)^n\} N^{-n}.$$

Отсюда следует, что

$$E(X) = \sum_{k=1}^{k=N} k p_k = N^{-n} \sum_{k=1}^{k=N} \{k^{n+1} - (k-1)^{n+1} - (k-1)^n\} =$$

$$= N^{-n} \left\{ N^{n+1} - \sum_{k=1}^{N} (k-1)^n \right\}. \qquad (3.6)$$

При большом N последняя сумма приблизительно равна площади под кривой $y = x^n$, заключенной между прямыми $x = 0$ и $x = N$, т. е. $N^{n+1}/(n+1)$. Отсюда следует, что при больших[1]) N

$$E(X) \approx \frac{n}{n+1} N. \qquad (3.7)$$

Если в городе имеется $N = 1000$ автомобилей и среди них выбирается наудачу $n = 10$ автомобилей, то математическое ожидание максимального наблюденного регистрационного номера равно примерно 910. (Медиана равна 934.) В статистике рассматривается также обратная задача оценки неизвестного истинного числа N по наблюденному максимуму выборки.

Этот метод использовался во время последней войны для оценки объема производства противника (см. задачи 8—11).

е) **Задача Банаха о спичечных коробках.** В § 8 гл. VI мы нашли распределение

$$u_r = \binom{2N-r}{N} 1/2^{2N-r} \qquad (3.8)$$

для числа X спичек, оставшихся в тот момент, когда впервые была вынута пустая коробка. Мы не можем непосредственно вычислить математическое ожидание $E(X) = \mu$; однако следующий косвенный способ применим во многих аналогичных случаях. Используя то обстоятельство, что сумма чисел u_r равна единице (эту сумму было бы нелегко сосчитать непосредственно), получаем

$$N-\mu=\sum_{r=0}^{N-1}(N-r)\,u_r=\sum_{r=0}^{N-1}(N-r)\binom{2N-r}{N-1}\frac{1}{2^{2N-r}}. \tag{3.9}$$

С помощью простых операций над биномиальными коэффициентами последняя сумма может быть приведена к виду

$$\sum_{r=0}^{N-1}(2N-r)\binom{2N-r-1}{N-r-1}\frac{1}{2^{2N-r}}=\frac{2N+1}{2}\sum_{r=0}^{N-1}u_{r+1}-\frac{1}{2}\sum_{r=0}^{N-1}(r+1)\,u_{r+1}. \tag{3.10}$$

Последняя сумма совпадает с суммой, определяющей $\mu=E(X)$. В первую сумму входят все u_r, за исключением u_0, и поэтому она равна $1-u_0$. Таким образом, из (3.9) и (3.10) следует, что

$$N-\mu=\frac{2N+1}{2}(1-u_0)-\frac{\mu}{2} \tag{3.11}$$

или

$$\mu=(2N+1)\,u_0-1=\frac{2N+1}{2^{2N}}\binom{2N}{N}-1. \tag{3.12}$$

Используя формулу Стирлинга, получаем

$$\mu\approx 2\left(\frac{N}{\pi}\right)^{1/2}-1. \tag{3.13}$$

В частности, в табл. 8 гл. VI мы имели $N=50$. Для этого случая $\mu=7{,}04, \ldots,$ а медиана равна 6.

§ 4. Дисперсия

Пусть X — случайная величина с распределением $\{f(x_j)\}$, и пусть r — неотрицательное целое число. *Если математическое ожидание случайной величины* X^r

$$E(X^r)=\sum x_j^r f(x_j) \tag{4.1}$$

существует, то оно называется моментом r-го порядка (r-м моментом) случайной величины X. Если этот ряд не является абсолютно сходящимся, то говорят, что r-й момент не существует. Так как $|X|^{r-1}\leqslant|X|^r+1$, *то из существования r-го момента вытекает существование $(r-1)$-го момента, а следовательно, и всех предшествующих.*

Моменты играют важную роль в общей теории, но в этой книге мы будем пользоваться только вторым моментом. Если он существует, то существует и математическое ожидание

$$\mu = E(X). \tag{4.2}$$

Естественно ввести вместо случайной величины X ее отклонение от среднего значения $X-\mu$. Так как $(x-\mu)^2 \leqslant 2(x^2+\mu^2)$, то мы видим, что второй момент $X-\mu$ существует, если существует $E(X^2)$. Мы находим

$$E((X-\mu)^2) = \sum_j \left(x_j^2 - 2\mu x_j + \mu^2\right) f(x_j). \tag{4.3}$$

Разбивая правую часть этого равенства на три суммы, убеждаемся, что она равняется $E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2$.

Определение. *Пусть X — случайная величина, имеющая второй момент $E(X^2)$, и пусть $\mu = E(X)$ — ее математическое ожидание. Мы определяем дисперсию случайной величины X формулой* [1])

$$D(X) = E(X-\mu)^2 = E(X^2) - \mu^2. \tag{4.4}$$

Положительное значение квадратного корня из этого числа (или нуль) называется стандартным отклонением случайной величины X.

Для простоты мы будем часто говорить о дисперсии некоторого распределения, не указывая соответствующей случайной величины.

Примеры. а) Если X принимает значения $\pm c$, каждое с вероятностью $1/2$, то $D(X) = c^2$.

б) Если X является числом очков, выпадающим при бросании симметричной кости, то $D(X) = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) - (7/2)^2 = 35/12$.

в) *Для распределения Пуассона* $p(k;\lambda)$ математическое ожидание равно λ (см. 3, б) и, следовательно, дисперсия равна

$$\sum k^2 p(k;\lambda) - \lambda^2 = \lambda \sum k p(k-1;\lambda) - \lambda^2 =$$
$$= \lambda \sum (k-1)\, p(k-1;\lambda) + \lambda \sum p(k-1;\lambda) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

В этом случае математическое ожидание и дисперсия оказываются одинаковыми.

г) Для *биномиального распределения* аналогичный расчет показывает, что дисперсия равна

$$\sum k^2 b(k;n,p) - (np)^2 = np \sum kb(k-1;n-1,p) - (np)^2 =$$
$$= np\,\{(n-1)p+1\} - (np)^2 = npq.$$

---

[1]) Автор обозначает дисперсию X через $\mathrm{Var}(X)$. Мы изменили это обозначение на общепринятое в русской литературе обозначение $D(X)$. — *Прим. ред.*

Польза понятия дисперсии выявится только постепенно, в частности, в связи с предельными теоремами (гл. X). Здесь мы отметим лишь, что дисперсия является, хотя и довольно грубой, *мерой разброса возможных значений случайной величины.* Действительно, если $D(X)=\sum (x_j-\mu)^2 f(x_j)$ мала, то малы и все члены этой суммы, и поэтому значения x_j, при которых $|x_j-\mu|$ велико, должны иметь малые вероятности $f(x_j)$. Другими словами, в случае малости дисперсии большие отклонения случайной величины X от ее математического ожидания μ маловероятны, Обратно, большая дисперсия указывает на то, что не все возможные значения X лежат вблизи от среднего значения.

Некоторым читателям может помочь следующая механическая интерпретация. Предположим, что единичная масса распределена по оси x так, что в точке x_j сосредоточена масса $f(x_j)$. Тогда математическое ожидание μ равно абсциссе *центра тяжести,* а дисперсия является *моментом инерции.* Ясно, что различные распределения масс могут иметь одинаковые центр тяжести и момент инерции, но, однако, хорошо известно, что этими двумя величинами могут быть описаны важнейшие механические свойства распределения массы.

Если X представляет собой измеренное значение некоторой физической величины, например, длины или температуры, то численные значения X зависят от начала отсчета и единицы измерения. Изменение последних означает переход от X к новой величине $aX+b$, где a и b — постоянные. Ясно, что $D(X+b)=D(X)$, так что

$$D(aX+b)=a^2D(X). \tag{4.5}$$

Выбор начала отсчета и единицы измерения является в значительной степени произвольным, и часто оказывается удобным взять за начало отсчета среднее значение, а за единицу — стандартное отклонение. Мы поступили так в гл. VII, где ввели нормированное число успехов $S_n^*=(S_n-np)/(npq)^{1/2}$. В общем случае, если X имеет математическое ожидание μ и дисперсию $\sigma^2 (\sigma>0)$, то $X-\mu$ имеет математическое ожидание нуль и дисперсию σ^2 и, следовательно, *случайная величина*

$$X^*=\frac{X-\mu}{\sigma} \tag{4.6}$$

имеет математическое ожидание 0 *и дисперсию* 1. *Она называется нормированной случайной величиной, соответствующей* X. На языке физики переход от X к X^* может быть интерпретирован как введение безразмерной величины.

§ 5. Ковариация. Дисперсия суммы

Пусть X и Y — две случайные величины, определенные на одном и том же пространстве элементарных событий. Тогда $X+Y$ и XY суть новые случайные величины, распределения которых можно найти простой группировкой членов совместного распределения X и Y. Наша цель состоит теперь в вычислении $D(X+Y)$. Для этого введем понятие ковариации, которое будет проанализировано более детально в § 8. Если $\{p(x_j, y_k)\}$ есть совместное распределение величин X и Y, то математическое ожидание XY определяется формулой

$$E(XY)=\sum x_j y_k p(x_j, y_k), \tag{5.1}$$

в предположении, конечно, что ряд абсолютно сходится. Далее, $|x_j y_k| \leqslant (x_j^2+y_k^2)/2$, и поэтому $E(XY)$ обязательно существует, если существуют $E(X^2)$ и $E(Y^2)$. В этом случае существуют также и математические ожидания

$$\mu_x=E(X), \quad \mu_y=E(Y), \tag{5.2}$$

а случайные величины $X-\mu_x$ и $Y-\mu_y$ имеют нулевые математические ожидания. Для их произведения по теореме сложения § 2 получаем

$$\begin{aligned} E((X-\mu_x)(X-\mu_y)) &= E(XY)-\mu_x E(Y)-\mu_y E(X)+\mu_x\mu_y = \\ &= E(XY)-\mu_x\mu_y. \end{aligned} \tag{5.3}$$

Определение. *Ковариацией*[1] *случайных величин X и Y называется*

$$\operatorname{Cov}(X, Y)=E((X-\mu_x)(Y-\mu_y))=E(XY)-\mu_x\mu_y. \tag{5.4}$$

Это определение имеет смысл, если X и Y имеют конечные дисперсии. Из § 2 известно, что для независимых случайных величин $E(XY)=E(X)E(Y)$. Таким образом, из (5.4) следует

Теорема 1. *Если X и Y независимы, то* $\operatorname{Cov}(X, Y)=0$.

Заметим, что *обратное неверно*. Табл. 1 дает пример двух зависимых случайных величин, ковариация которых тем не менее равна нулю. Мы вернемся к этому вопросу в § 8. Следующая теорема очень важна, и формула сложения (5.6) для независимых случайных величин постоянно применяется.

---

[1] В английском оригинале covariance; для этого понятия в русской литературе стандартного термина нет. — *Прим. ред.*

Теорема 2. *Если $X_1, \ldots, X_n$ — случайные величины с конечными дисперсиями $\sigma_1^2, \ldots, \sigma_n^2$ и $S_n = X_1 + \ldots + X_n$, то*

$$D(S_n) = \sum_{k=1}^{n} \sigma_k^2 + 2 \sum_{j,k} \operatorname{Cov}(X_j, X_k), \tag{5.5}$$

где последняя сумма содержит каждую из $\binom{n}{2}$ пар (X_j, X_k) с $j < k$ один и только один раз.

В частности, если X_j взаимно независимы, то верно правило сложения

$$D(S_n) = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2. \tag{5.6}$$

Доказательство. Положим $\mu_k = E(X_k)$ и $m_n = \mu_1 + \ldots + \mu_n = E(S_n)$. Тогда $S_n - m_n = \sum (X_k - \mu_k)$ и

$$(S_n - m_n)^2 = \sum (X_k - \mu_k)^2 + 2 \sum (X_j - \mu_j)(X_k - \mu_k). \tag{5.7}$$

Вычислив математическое ожидание от обеих частей равенства (5.7) и применив теорему сложения, получим (5.5). Равенство (5.6) следует из предыдущей теоремы.

Примеры. а) *Биномиальное распределение* $b(k; n, p)$. В примере 3, а случайные величины X_k взаимно независимы. Имеем $E(X_k^2) = 0^2 \cdot q + 1^2 \cdot p = p$ и $E(X_k) = p$. Следовательно, $\sigma_k^2 = p - p^2 = pq$, и из (5.6) видно, что *дисперсия биномиального распределения* равна npq. Тот же результат был получен в примере 4, г с помощью прямого вычисления.

б) *Испытания Бернулли с переменными вероятностями успеха.* Пусть $X_1, \ldots, X_n$ независимые случайные величины, такие, что X_k принимает значение 1 и 0 с вероятностями p_k и $q_k = 1 - p_k$ соответственно. Тогда $E(X_k) = p_k$ и $D(X_k) = p_k - p_k^2 = p_k q_k$. Полагая опять $S_n = X_1 + \ldots + X_n$, из формулы (5.6) получим

$$D(S_n) = \sum_{k=1}^{n} p_k q_k. \tag{5.8}$$

Как и в примере (1, д), случайную величину S_n можно интерпретировать как полное число успехов в n независимых испытаниях, результатом каждого из которых может быть успех или неудача. Тогда $p = (p_1 + \ldots + p_n)/n$ — средняя вероятность успеха, и вполне естественно сравнивать описанную ситуацию со схемой испытаний Бернулли с постоянной вероятностью успеха p. Это сравнение приводит к поразительному результату. Мы можем переписать соотношение (5.8) в виде $D(S_n) = np - \sum p_k^2$. Кроме того, нетрудно показать (элементарными вычислениями или по индукции), что для всех комбинаций $\{p_k\}$, таких, что $\sum p_k = np$, сумма $\sum p_k^2$ достигает своего наименьшего значения, если все p_k равны между собой. Отсюда следует, что при фиксированной средней вероятности ошибки p дисперсия $D(S_n)$ *достигает максимума при* $p_1 = \ldots = p_n = p$. Таким образом, мы пришли к следующему неожиданному результату: *любые отклонения* p_k *от среднего уровня уменьшают величину случайных колебаний*, которые характеризуются дисперсией[1]). Например, число пожаров в городе за год можно рассматривать как случайную величину; если известно среднее значение, то дисперсия *максимальна* при условии, что вероятность пожара в каждом квартале одна и та же. Если для n машин известно среднее качество p, то *производительность всего комплекса будет менее всего равномерной тогда, когда все машины одинаковы*. (Приложение к современному образованию очевидно, но оно не обнадеживает.)

в) *Совпадение карт*. Колода из n занумерованных карт раскладывается в случайном порядке так, что все $n!$ возможных расположений имеют равные вероятности. Число совпадений, т. е. число карт, попавших на место, соответствующее их номеру, является случайной величиной S_n, которая принимает значения 0, 1, ..., n. Распределение вероятностей для S_n было выведено в § 4 гл. IV. Исходя из этого распределения, можно вычислить и математическое ожидание и дисперсию, но следующий способ проще и поучительней.

Определим случайную величину X_k, принимающую значения 1 и 0 в зависимости от того, оказалась ли карта с номером k на k-м месте или нет. Тогда $S_n = X_1 + \ldots + X_n$. Далее, каждая из карт имеет вероятность $1/n$ появиться на k-м месте. Следовательно, $P\{X_k = 1\} = 1/n$, $P\{X_k = 0\} = (n-1)/n$. Поэтому $E(X) = 1/n$; отсюда следует, что $E(S_n) = 1$: в среднем приходится одно совпадение на всю колоду. Для вычисления дисперсии S_n вычислим сначала дисперсию σ_k^2 случайной величины X_k:

$$\sigma_k^2 = \frac{1}{n} - \left(\frac{1}{n}\right)^2 = \frac{n-1}{n^2}. \tag{5.9}$$

Далее, вычислим $E(X_jX_k)$. Произведение X_jX_k может равняться только 0 и 1, причем последнее происходит, если обе карты (карта с номером j и карта с номером k) займут свои собственные места, вероятность чего равна $\frac{1}{n(n-1)}$. Следовательно,

$$E(X_jX_k) = \frac{1}{n(n-1)}; \tag{5.10}$$

$$\operatorname{Cov}(X_j, X_k) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)}.$$

---

[1]) Более сильные результаты в том же направлении можно найти в работе: W. Hoeffding, On the distribution of the number of successes in independent trials, *Annals of Mathematical Statistics*, 27 (1956), 713—721.

Отсюда, наконец,

$$D(S_n) = n\,\frac{n-1}{n^2} + 2\binom{n}{2}\frac{1}{n^2(n-1)} = 1. \tag{5.11}$$

Итак, для числа совпадений и математическое ожидание и дисперсия равны единице. Этот результат может быть применен к изучавшейся в § 4 гл. IV задаче *угадывания карт*. Там мы рассмотрели три метода угадывания, один из которых приводит к только что рассмотренной схеме. Второй может быть описан как последовательность n испытаний Бернулли с вероятностью $p = 1/n$; в этом случае математическое ожидание числа верных догадок равно $np = 1$, а дисперсия $npq = (n-1)/n$. Математические ожидания в обоих случаях одинаковы; однако бо́льшая величина дисперсии при первом методе указывает на бо́льшие случайные отклонения относительно среднего значения и, таким образом, обещает несколько более острую игру (При колодах более сложного состава различие между двумя дисперсиями будет больше, хотя никогда не станет действительно значительным). При последнем способе угадывания испытуемый все время называет одну и ту же карту; число правильных догадок всегда равно единице, и случайные колебания исключены (дисперсия 0). Мы видим, что стратегия отгадывания не влияет на среднее число верных догадок, хотя и оказывает некоторое влияние на амплитуду случайных отклонений.

г) *Выбор без возвращения.* Предположим, что генеральная совокупность состоит из b черных и g зеленых предметов, и из нее производится случайная выборка объема r (без повторений). Число S_k черных предметов, вошедших в выборку, является случайной величиной с *гипергеометрическим распределением* (гл. II, § 6), определяемым формулой

$$q_k = \frac{\binom{b}{k}\binom{g}{r-k}}{\binom{b+g}{r}}. \tag{5.12}$$

Математическое ожидание и дисперсия могут быть получены прямым вычислением, но предпочтительнее следующий метод. Рассмотрим случайную величину X_k, принимающую значения 1 и 0 в зависимости от того, является ли k-й член выборки черным или зеленым. Из соображений симметрии вероятность того, что $X_k = 1$, равна $b/(b+g)$, и, следовательно,

$$E(X_k) = \frac{b}{b+g}; \qquad D(X_k) = \frac{bg}{(b+g)^2}. \tag{5.13}$$

Далее, при $j \neq k$ произведение $X_jX_k = 1$, если и k-й и j-й члены выборки окажутся черными, а во всех других случаях $X_jX_k = 0$. Вероятность того, что $X_jX_k = 1$, равна

$$b(b-1)/(b+g)(b+g-1),$$

и, следовательно,

$$E(X_jX_k)=\frac{b(b-1)}{(b+g)(b+g-1)};\qquad(5.14)$$

$$\operatorname{Cov}(X_jX_k)=\frac{-bg}{(b+g)^2(b+g-1)}.$$

Таким образом,

$$E(S_r)=\frac{rb}{b+g};\qquad D(S_r)=\frac{rbg}{(b+g)^2}\left\{1-\frac{r-1}{b+g-1}\right\}.\qquad(5.15)$$

При выборе с возвращением мы имели бы такое же математическое ожидание, но несколько бóльшую дисперсию, а именно $rbg/(b+g)^2$.

§ 6. Неравенство Чебышева [1])

Уже отмечалось, что малость дисперсии указывает на то, что большие отклонения от среднего значения маловероятны. Это положение уточняется неравенством Чебышева, представляющим собой чрезвычайно важный и удобный рабочий инструмент теории вероятностей.

Теорема. *Пусть X — случайная величина с математическим ожиданием $\mu=E(X)$ и дисперсией $\sigma^2=D(X)$. Тогда для любого $t>0$*

$$P\{|X-\mu|\geqslant t\}\leqslant\frac{\sigma^2}{t^2}.\qquad(6.1)$$

Доказательство. Дисперсия определяется формулой (4.3) как ряд с положительными слагаемыми. Отбрасывание всех слагаемых, для которых $|x_j-\mu|<t$, может только уменьшить сумму этого ряда, и, следовательно,

$$\sigma^2\geqslant\sum{}^*(x_j-\mu)^2f(x_j),\qquad(6.2)$$

где звездочка означает, что суммирование производится только по тем j, при которых $|x_j-\mu|\geqslant t$. Ясно, что

$$\sum{}^*(x_j-\mu)^2f(x_j)\geqslant t^2\sum{}^*f(x_j)=t^2P\{|X-\mu|\geqslant t\},\qquad(6.3)$$

чем и доказывается теорема.

Неравенство Чебышева нужно рассматривать скорее как теоретическое орудие, чем как практический метод оценки. Важность этого неравенства вытекает из его универсальности, но ни от какого утверждения, имеющего большую общность, нельзя ожидать точных результатов в частных случаях.

Примеры. а) Если X есть число очков, выпавших в результате бросания правильной игральной кости, то (см. пример 4, в) $\mu=7/2$

[1]) П. Л. Чебышев (1821—1894).

и $\sigma^2=35/12$. Максимум отклонения X от μ равен $2{,}5\approx 3\sigma/2$. Вероятность бóльших отклонений равна нулю, в то время как неравенство Чебышева утверждает только, что эта вероятность меньше, чем 0,47.

б) Для биномиального распределения $\{b(k;\ n,\ p)\}$ имеем (см. пример 5, а) $\mu=np$, $\sigma^2=npq$. Мы знаем, что при больших n

$$P\left\{|S_n-np|>x(npq)^{1/2}\right\}\approx 1-\Phi(x)+\Phi(-x). \qquad (6.4)$$

Неравенство Чебышева утверждает только, что левая сторона меньше, чем $1/x^2$; это является, очевидно, оценкой гораздо более грубой, чем (6.4).

§ 7*) Неравенство Колмогорова[1])

В качестве примера более точной оценки докажем следующее утверждение.

Пусть $X_1, \ldots, X_n$ — *взаимно независимые случайные величины с математическими ожиданиями* $\mu_k=E(X_k)$ *и дисперсиями* σ_k^2. *Положим*

$$S_k=X_1+\ldots+X_k, \qquad (7.1)$$

$$m_k=E(S_k)=\mu_1+\ldots+\mu_k,$$

$$s_k^2=D(S_k)=\sigma_1^2+\ldots+\sigma_k^2. \qquad (7.2)$$

При каждом $t>0$ *вероятность совместного осуществления* n *неравенств*

$$|S_k-m_k|<ts_n, \quad k=1,\ 2,\ \ldots,\ n \qquad (7.3)$$

больше, чем $1-t^{-2}$.

При $n=1$ эта теорема сводится к неравенству Чебышева, При $n>1$ неравенство Чебышева дает ту же самую границу для вероятности единственного соотношения $|S_n-m_n|<ts_n$, так что неравенство Колмогорова значительно сильнее.

Доказательство. Оценим вероятность x того, что не выполнено хотя бы одно из неравенств (7.3). Теорема утверждает, что $x\leqslant t^{-2}$.

Определим n случайных величин Y_k следующим образом: $Y_\nu=1$, если выполняется система неравенств

$$|S_\nu-m_\nu|\geqslant ts_n; \qquad (7.4)$$

$$|S_k-m_k|<ts_n \quad \text{при} \quad k=1,\ 2,\ \ldots,\ \nu-1; \qquad (7.5)$$

---

*) См. примечание на стр. 138.

[1]) А. Н. Колмогоров, Über die Summen züfälliger Grössen, *Mathematische Annalen*, 99 (1928), 309—319 и 102 (1929), 484—488.

$Y_\nu = 0$ во всех остальных случаях. Другими словами, Y_ν равно единице для всех тех элементарных событий, для которых первым из невыполняющихся неравенств (7.3) является ν-е. Следовательно, для каждого элементарного события не более чем одна из величин Y_k равна единице, а сумма $Y_1 + Y_2 + \ldots + Y_n$ может принимать только значения 0 или 1; она равна единице тогда и только тогда, когда нарушено одно из неравенств (7.3). Поэтому

$$x = P\{Y_1 + \ldots + Y_n = 1\}. \tag{7.6}$$

Так как $Y_1 + \ldots + Y_n$ равно 0 или 1, имеем $\sum Y_k \leqslant 1$. Умножив эту сумму на $(S_n - m_n)^2$ и вычислив математическое ожидание произведения, получим

$$\sum_{k=1}^{n} E\left(Y_k (S_n - m_n)^2\right) \leqslant s_n^2. \tag{7.7}$$

Для оценки слагаемых в левой части положим

$$U_k = (S_n - m_n) - (S_k - m_k) = \sum_{\nu=k+1}^{n} (X_\nu - \mu_\nu). \tag{7.8}$$

Тогда

$$E(Y_k (S_n - m_n)^2) = \\ = E(Y_k (S_k - m_k)^2) + 2E(Y_k U_k (S_k - m_k)) + E(Y_k U_k^2). \tag{7.9}$$

Однако U_k зависит только от $X_{k+1}, \ldots, X_n$, в то время как Y_k и S_k зависят только от $X_1, \ldots, X_k$. Следовательно, U_k не зависит от $Y_k(S_k - m_k)$, и поэтому $E((Y_k U_k (S_k - m_k)) = E(Y_k (S_k - m_k)) \times \times E(U_k) = 0$, так как $E(U_k) = 0$. Таким образом, из (7.9) получаем

$$E(Y_k (S_n - m_n)^2) \geqslant E(Y_k (S_k - m_k)^2). \tag{7.10}$$

Но $Y_k \neq 0$, только если $|S_k - m_k| \geqslant t s_n$, так что $Y_k (S_k - m_k)^2 \geqslant \geqslant t^2 s_n^2 Y_k$. Следовательно, комбинируя (7.7) и (7.10), получаем

$$s_n^2 \geqslant t^2 s_n^2 E\left(Y_1 + \ldots + Y_n\right). \tag{7.11}$$

Так как $Y_1 + \ldots + Y_n$ может равняться только 0 или 1, математическое ожидание правой части неравенства равно вероятности x, определяемой формулой (7.6) Таким образом, $xt^2 \leqslant 1$, что и требовалось доказать.

§ 8 *). Коэффициент корреляции

Пусть X и Y — случайные величины с математическими ожиданиями μ_x и μ_y и ненулевыми дисперсиями σ_x^2 и σ_y^2. Рассмотрим нормированные случайные величины X^* и Y^*, определяемые фор-

*) См. примечание на стр. 138.

мулой (4.6). Их ковариация называется *коэффициентом корреляции случайных величин X и Y и обозначается* $\rho(X, Y)$. Таким образом, используя (5.4), получаем

$$\rho(X, Y) = \mathrm{Cov}(X^*, Y^*) = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y}. \tag{8.1}$$

Ясно, что коэффициент корреляции не зависит от выбора начала отсчета и единицы измерения, т. е. для любых постоянных a_1, a_2, b_1, b_2, таких, что $a_1 > 0$, $a_2 > 0$, имеем

$$\rho(a_1 X + b_1, a_2 Y + b_2) = \rho(X, Y).$$

Использование коэффициента корреляции равносильно особой форме записи ковариации[1]). К сожалению, термин «корреляция» внушает неправильное представление о свойствах этого коэффициента. Мы знаем из § 5, что $\rho(X, Y) = 0$, если X и Y независимы. Очень важно усвоить, что обратное утверждение неверно. *На самом деле коэффициент корреляции $\rho(X, Y)$ может обращаться в нуль, даже если Y является функцией от X.*

Примеры. а) Пусть X принимает значения ± 1, ± 2, каждое с вероятностью $1/4$. Пусть $X = Y^2$. Совместное распределение задается равенствами $p(-1, 1) = p(1, 1) = p(2, 4) = p(-2, 4) = 1/4$. Из соображений симметрии ясно, что $\rho(X, Y) = 0$, хотя между Y и X имеется функциональная зависимость.

б) Пусть U и V — *независимые* случайные величины с одинаковыми распределениями и пусть $X = U + V$, $Y = U - V$. Тогда $E(XY) = E(U^2) - E(V^2) = 0$ и $E(Y) = 0$. Следовательно, $\mathrm{Cov}(X, Y) = 0$; поэтому также и $\rho(X, Y) = 0$. Например, X и Y могут быть суммой и разностью чисел очков, выпавших на двух костях. Тогда X и Y или оба сразу четны или оба сразу нечетны и, следовательно, зависимы.

Отсюда следует, что коэффициент корреляции никоим образом не является исчерпывающей мерой зависимости между X и Y. Однако $\rho(X, Y)$ связано с *линейной* зависимостью между X и Y.

Теорема. *Всегда $|\rho(X, Y)| \leqslant 1$, причем $\rho(X, Y) = \pm 1$ тогда и только тогда, когда $Y = aX + b$, где a и b — постоянные.*

Доказательство. Пусть X^* и Y^* — нормированные случайные величины; тогда

$$D(X^* \pm Y^*) = D(X^*) \pm 2\,\mathrm{Cov}(X^*, Y^*) + D(Y^*) = 2(1 \pm \rho(X, Y)). \tag{8.2}$$

---

[1]) Физик определил бы коэффициент корреляции как «безразмерную ковариацию».

Левая часть не может быть отрицательной; следовательно, $|\rho(X, Y)| \leqslant 1$. Для $\rho(X, Y) = 1$ необходимо, чтобы $D(X^* - Y^*) = 0$, а это значит, что случайная величина $X^* - Y^*$ принимает только одно значение. В этом случае $X^* - Y^* = \mathrm{const}$ и, следовательно, $Y = aX + \mathrm{const}$, где $a = \sigma_y/\sigma_x$. Аналогичное доказательство применимо и к случаю $\rho(X, Y) = -1$.

§ 9. Задачи

1. Семь шаров случайно распределяются по семи ящикам. Обозначим через X_i число ящиков, содержащих ровно i шаров. Используя вероятности, табулированные в § 5 гл. II, записать совместное распределение X_2 и X_3.

2. Бросаются две правильные кости. Пусть X — число очков на первой кости, а Y — максимальное из двух выпавших чисел. а) Записать совместное распределение X и Y. б) Найти математические ожидания, дисперсии и ковариацию.

3. Пусть X, Y и Z суть соответственно число выпадений герба, число серий гербов и длина максимальной из этих серий в последовательности пяти бросаний монеты. Составьте таблицу из всех 32 элементарных событий вместе с соответствующими им значениями X, Y и Z. Простым подсчетом получите совместное распределение пар (X, Y), (X, Z), (Y, Z) и распределение случайных величин $X + Y$ и XY. Найдите математические ожидания, дисперсии и ковариации.

4. Случайные величины X_1 и X_2 независимы и имеют одно и то же *геометрическое распределение* $\{q^k p\}$, где $k = 0, 1 \ldots$. Пусть Z — максимальная из двух величин $[Z = \max(X_1, X_2)]$. Записать совместное распределение Z и X_1; найти распределение Z.

5. Пусть X_1 и X_2 — независимые случайные величины с распределениями Пуассона $\{p(k; \lambda_1)\}$ и $\{p(k; \lambda_2)\}$. Доказать, что $X_1 + X_2$ имеет распределение Пуассона $\{p(k; \lambda_1 + \lambda_2)\}$.

6. Продолжение. Показать, что *условное распределение X_1, при условии, что известно $X_1 + X_2$, есть биномиальное распределение*, а именно

$$P\{X_1 = k \mid X_1 + X_2 = n\} = b\left(k;\ n,\ \frac{\lambda_1}{\lambda_1 + \lambda_2}\right). \tag{9.1}$$

7. Пусть X_1 и X_2 независимы и имеют одно и то же геометрическое распределение $\{q^k p\}$ (как в задаче 4). Показать без вычислений, что *условное распределение X_1 при условии, что известно $X_1 + X_2$, равномерно*, т. е.

$$P\{X_1 = k \mid X_1 + X_2 = n\} = \frac{1}{n+1}, \quad k = 0, \ldots, n. \tag{9.2}$$

8. Пусть $X_1, \ldots, X_n$ — независимые случайные величины, каждая из которых имеет *равномерное распределение* $P\{X_i = k\} = \frac{1}{N}$ для $k = 1, 2, \ldots, N$. Пусть U_n — минимум среди $X_1, \ldots, X_n$, а V_n — максимум. Найти распределения U_n и V_n. Какова связь этой задачи с примером 3, д?

9. В условиях примера 3, д найти совместное распределение максимального и минимального из наблюдений. Особо разобрать случай $n = 2$. (Указание. Вычислить сначала $P\{X \leqslant r, Y \geqslant s\}$.)

10. Продолжение. Найти условную вероятность того, что первые два наблюдения дали j и k при условии, что $X = r$.

11. Продолжение. Найти $E(X^2)$ и затем получить асимптотическое выражение для дисперсии при $N \to \infty$ (n остается фиксированным).

12. Выборочный контроль. Допустим, что изделия, каждое из которых годно с вероятностью p, подвергаются выборочному контролю, так что для каждого изделия вероятность быть проверенным равна p'. Можно выделить четыре класса изделий, а именно «годные и проверенные», «годные и непроверенные» и т. д., которые имеют своими вероятностями pp', pq', $p'q$, qq', где $q = 1 - p$, $q' = 1 - p'$. Таким образом, мы имеем дело со схемой сложных испытаний Бернулли (см. пример 9.6 гл. VI). Пусть N — число изделий, прошедших через стол контроля (как проверенных, так и непроверенных), прежде чем обнаружено первое бракованное изделие; K — число бракованных (оставшихся необнаруженными) среди этих N изделий. Найти совместное распределение N и K и их частные распределения.

13. Продолжение. Найти $E(K/(N+1))$ и $\mathrm{Cov}(K, N)$). В промышленной практике обнаруженное бракованное изделие заменяется на годное, так что $K/(N+1)$ представляет собой долю бракованных изделий в данной партии и является мерой ее качества. Обратить внимание на то, что $E(K/(N+1)) \neq E(K)/E(N+1)$.

14. В последовательности испытаний Бернулли пусть X — длина серии (все равно из успехов или из неудач), начавшейся при первом испытании. Найти распределение X, $E(X)$, $D(X)$.

15. Продолжение. Пусть Y — длина *второй* серии. Найти распределение для Y, $E(Y)$, $D(Y)$ и совместное распределение X и Y.

16. Если случайные величины X и Y принимают только по два значения каждая и если $\mathrm{Cov}(X, Y) = 0$, то X и Y независимы.

17. Дни рождения. Для группы из n человек найти математическое ожидание числа дней года, на которые приходится ровно по k дней рождения (принимается, что год состоит из 365 дней и все размещения дней рождения (равновероятны).

18. Продолжение. Найти математическое ожидание числа дней года, являющихся общими днями рождения по меньшей мере для двух человек. Как велико должно быть n, чтобы это математическое ожидание превосходило единицу?

19. Некий человек хочет открыть свою дверь и имеет n ключей. По неизвестным нам причинам он испытывает эти ключи независимо и случайно. Найти математическое ожидание и дисперсию числа испытаний: а) если уже испробованный ключ не устраняется из дальнейшего выбора; б) если устраняется (предполагается, что только один ключ подходит к двери. Точное распределение приведено в § 7 гл. II, но при решении настоящей задачи оно не требуется).

20. Пусть (X, Y) — случайные величины, совместное распределение которых является триномиальным [см. формулу (1.9)]. Найти $E(X)$, $D(X)$ и $\mathrm{Cov}(X, Y)$ а) прямым подсчетом, б) представив каждую из случайных величин X и Y в виде суммы n слагаемых и воспользовавшись методом § 5.

21. Найти ковариацию числа выпадений единицы и числа выпадений шестерки при n бросаниях кости.

22. В задаче 24 гл. VI доказать, что математическое ожидание числа животных, пойманных при ν-м отлове, равно $nqp^{\nu-1}$.

23. Показать, что если X имеет *геометрическое* распределение $P\{X = k\} = q^k p$ ($k = 0, 1, 2, \ldots$), то $D(X) = qp^{-2}$. Вывести отсюда, что *отрицательное* биномиальное распределение $\{f(k, r, p)\}$ имеет дисперсию rqp^{-2}, при условии, что r — целое. Доказать прямым подсчетом, что это утверждение сохраняет силу при любом $r > 0$.

24. *В задаче о времени ожидания* (3, г) доказать, что

$$D(S_r) = N\left\{\frac{1}{(N-1)^2} + \frac{2}{(N-2)^2} + \ldots + \frac{r-1}{(N-r+1)^2}\right\}.$$

Указание. Используя то, что $\frac{k}{(N-k)^2}=\frac{N}{(N-k)^2}-\frac{1}{N-k}$, легко проверить, что $\lim N^{-2}D(S_N)=\sum\limits_{1}^{\infty}k^{-2}$. Последняя сумма равна $\frac{\pi^2}{6}$.

25. Продолжение. Пусть Y_r — число извлечений, требуемых для получения r заранее намеченных элементов (а не любых r рлзличных элементов, как в тексте). Найти $E(Y_r)$ и $D(Y_r)$. (Замечение. Точное распределение Y_r было найдено в задаче (11.12) гл. II, но для целей настоящей задачи оно не требуется.)

26. [1]) Большое число N людей подвергается исследованию крови. Это исследование может быт· организовано двумя способами: 1. Кровь каждого человека исследуется отдельно. В этом случае потребуется N анализов. 2. Кровь k людей смешивается, и анализируется полученная смесь. Если результат анализа *отрицателен*, то этого *одного* анализа достаточно для k человек. Если же он *положителен*, то кровь каждого из этих людей приходится исследовать затем отдельно, и в целом на k человек понадобится $k+1$ анализ.

Предполагается, что вероятность положительного результата анализа одна и та же для всех людей и что результаты анализов независимы в теоретико-вероятностном смысле.

а) Чему равна вероятность того, что анализ смешанной крови k людей положителен?

б) Чему равно математическое ожидание числа анализов X, необходимых при втором методе организации исследования?

в) При каком k достигается минимум математического ожидания числа необходимых анализов? Не пытайтесь дать численный ответ, так как задача сводится к довольно громоздкому уравнению для k.

27. Структура выборки. Генеральная совокупность разделена на r классов, объемы которых относятся как $p_1:p_2:\ldots:p_r$. Случайным образом извлекается выборка объема n с возвращением. Найти математическое ожидание числа классов, *не* представленных в выборке.

28. Пусть X — число α серий при случайном расположении r_1, элементов α и r_2 элементов β. Распределение дается в задаче (11.23) гл. II. Найти $E(X)$ и $D(X)$.

29. *В урновой схеме Пойа* (пример 2, гл. V) введем случайную величину X_n, равную 1 или 0 в зависимости от того, черный или красный шар извлечен при n-м испытании. Доказать, что $\rho(X_n, X_m)=\frac{c}{b+r+c}$ при $n\neq m$.

30. Продолжение. Пусть S_n — полное число черных шаров, извлеченных при первых n выниманиях (т. е. $S_n=X_1+X_2+\ldots+X_n$). Найти $E(S_n)$ и $D(S_n)$. (Использовать результаты задачи 19 и 20 гл. V, проверить с помощью рекуррентной формулы задачи 22 гл. V.)

31. Выбор по группам. Город состоит из n кварталов, причем в n_j из них проживает по x_j жителей ($n_1+n_2+\ldots=n$). Пусть $m=\sum n_jx_j/n$ — среднее число жителей, приходящееся на один квартал. Положим также $a^2=\sum n_jx_j^2/n-m^2$. С помощью случайного выбора без

[1]) Эта задача основана на новой методике, которая применялась во второй мировой войне. См. R. Dorfman, The detection of defective members of large populations, *Annals of Mathematical Statistics*, 14 (1943), 436—440. В армейской практике второй метод давал экономию до 80% в числе анализов.

возвращения отобраны r кварталов, и в каждом из них подсчитано число жителей. Пусть $X_1, \ldots, X_r$ — соответствующие числа. Покажите, что

$$E(X_1 + \ldots + X_r) = mr; \quad D(X_1 + \ldots + X_r) = \frac{a^2 r (n-r)}{n-1}.$$

(Заметим, что при выборе с возвращением дисперсия будет больше, а именно $a^2 r$.)

32. Длина случайной цепи [1]). Цепь, лежащая ни плоскости x, y, состоит из n-звеньев; длина каждого звена равна единице. Угол между двумя соседними звеньями равен $\pm\alpha$, где α — положительная постоянная. Каждая из возможностей имеет вероятность $1/2$, и величины углов взаимно независимы. Тогда расстояние между концами цепи (длина цепи) L_n является случайной величиной, и нужно доказать, что

$$E\left(L_n^2\right) = n\,\frac{1+\cos\alpha}{1+\cos\alpha} - 2\cos\alpha\,\frac{1-\cos^n\alpha}{(1-\cos\alpha)^2}. \tag{9.3}$$

Без уменьшения общности можно предположить, что первое звено расположено в положительном направлении оси x. Угол между k-м звеном и положительным направлением оси x является случайной величиной S_{k-1}, причем $S_0 = 0$, $S_k = S_{k-1} + X_k\alpha$, где X_k — взаимно независимые случайные величины, принимающие значения ± 1 с вероятностями $1/2$. Проекция k-го звена на оси X и Y равна соответственно $\cos S_{k-1}$ и $\sin S_{k-1}$. Следовательно, при $n \geqslant 1$

$$L_n^2 = \left(\sum_{k-0}^{n-1}\cos S_k\right)^2 + \left(\sum_{k=0}^{n-1}\sin S_k\right)^2. \tag{9.4}$$

Доказать последовательно по индукции, что при $m < n$:

$$(E\cos S_n) = \cos^n a, \quad E(\sin S_n) = 0; \tag{9.5}$$

$$E(\cos S_m\cdot\cos S_n) = \cos^{n-m}\alpha\cdot E(\cos^2 S_m); \tag{9.6}$$

$$E(\sin S_m\cdot\sin S_n) = \cos^{n-m}\alpha\cdot E(\sin^2 S_m); \tag{9.7}$$

$$E\left(L_n^2\right) - E\left(L_{n-1}^2\right) = 1 + 2\cos\alpha\cdot\frac{1-\cos^{n-1}\alpha}{1-\cos\alpha}. \tag{9.8}$$

(причем $L_0 = 0$), и отсюда получить наконец (9.3).

33. Испытания Бернулли производятся до тех пор, пока число успехов не достигнет r (r — фиксированное целое число). Пусть X — потребовавшееся для этого число испытаний. Найти [2]) $E(r/X)$. (Формула, определяющая математическое ожидание, в данном случае содержит бесконечный ряд, для которого, однако, может быть получено конечное выражение.)

34. При случайном размещении r шаров по n ящикам вероятность того, что найдется в точности m пустых ящиков, удовлетворяет рекуррентной формуле (11.8) гл. II. Пусть m_r — математическое ожидание числа пустых

---

[1]) Это двухмерный аналог задачи о длине длинной полимерной молекулы в химии. Задача иллюстрирует применение к случайным величинам, которые не выражаются в виде суммы простых переменных.

[2]) Если число испытаний фиксировано, то отношение числа успехов к числу испытаний является случайной величиной, математическое ожидание которой равно p. Часто ошибочно предполагают, что то же самое верно и в нашем примере, где число r успехов фиксировано, а число испытаний зависит от случая. Если $p = 1/2$ и $r = 2$, то $E(2/X)$ равно 0,614 вместо 0,5; при $r = 3$ мы находим, что $E(3/X) = 0{,}579$.

ящиков. Доказать, *исходя из рекуррентной формулы*, что

$$m_{r+1} = 1 (1 - n^{-1})\, m_r, \text{ и получить отсюда,}$$

что $m_r = n\left\{1 - \left(1 - \frac{1}{n}\right)^r\right\}$.

35. Пусть S_n — число успехов в n испытаниях Бернулли. Доказать, что

$$E(|S_n - np|) = 2\nu q b(\nu;\ n,\ p),$$

где ν — целое число, удовлетворяющее неравенству $np < \nu \leqslant np + 1$.

36. Пусть $\{X_k\}$ — последовательность взаимно независимых случайных величин с одинаковыми распределениями. Предположим, что X_k принимает только положительные значения и что $E(X_k) = a$ и $E(X_k^{-1}) = b$ существуют. Пусть $S_n = X_1 + \ldots + X_n$. Доказать, что $E(S_n^{-1})$ конечно и $E(X_k/S_n) = 1/n$ при $k = 1, 2, \ldots, n$.

37. Продолжение[1]). Доказать, что

$$E\left(\frac{S_m}{S_n}\right) = \frac{m}{n}, \quad \text{если } m \leqslant n;$$

$$E\left(\frac{S_m}{S_n}\right) = 1 + (m - n)\, aE\left(S_n^{-1}\right), \quad \text{если } m \geqslant n.$$

38. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины с математическим ожиданием m и дисперсией σ^2. Пусть $\overline{X} = (X_1 + \ldots + X_n)/n$. Доказать, что[2])

$$\frac{1}{n-1} E\left(\sum_{k=1}^{n} (X_k - \overline{X})^2\right) = \sigma^2.$$

39. Пусть $X_1, \ldots, X_n$ — взаимно независимые случайные величины. Пусть U — функция от $X_1, \ldots, X_k$, а V — функция от $X_{k+1}, \ldots, X_n$ $(k < n)$. Доказать, что U и V — независимые случайные величины.

40. Обобщенное неравенство Чебышева. Пусть $\varphi(x)$ — монотонно возрастающая положительная четная функция. Предположим, что $E(\varphi(|X|)) = M$ существует. Доказать, что

$$P\{|X| \geqslant t\} \leqslant \frac{M}{\varphi(t)}.$$

41. Неравенство Шварца. Для любых двух случайных величин с конечной дисперсией имеет место неравенство $E^2(XY) \leqslant E(X^2)\, E(Y^2)$. Доказать это, опираясь на факт неотрицательности квадратичной формы $E((tX + Y)^2)$.

---

[1]) K. L. Chung заметил, что результаты задачи 37 можно получить непосредственно из 36.

[2]) Этот результат можно сформулировать по-другому, сказав, что $\sum (X_k - \overline{X})^2/n - 1$ дает несмещенную оценку для σ^2.

ГЛАВА X

ЗАКОНЫ БОЛЬШИХ ЧИСЕЛ

§ 1. Одинаково распределенные случайные величины

Предельные теоремы для испытаний Бернулли, полученные в гл. VII и VIII, являются частными случаями общих предельных теорем, не рассматривающихся в настоящей книге. Однако чтобы установить новую точку зрения на математическое ожидание случайной величины, мы рассмотрим здесь ряд довольно общих формулировок закона больших чисел.

Связь между испытаниями Бернулли и теорией случайных величин станет яснее, если мы рассмотрим зависимость числа успехов S_n от числа испытаний n. При каждом испытании S_n возрастает на 1 или на 0. Это утверждение можно записать в виде

$$S_n = X_1 + \ldots + X_n, \tag{1.1}$$

где случайная величина X_k равна 1 или 0 в зависимости от того, будет ли результатом k-го испытания успех или неудача. Таким образом, S_n является суммой n взаимно независимых случайных величин, каждая из которых принимает значения 1 и 0 с вероятностями p и q. Отсюда остается только один шаг до того, чтобы рассмотреть суммы вида (1.1), где X_k суть независимые случайные величины с произвольными распределениями. Закон больших чисел § 4 гл. VII утверждает, что при больших n весьма правдоподобно, что среднее число успехов S_n/n окажется близким к p. Это утверждение есть частный случай следующей теоремы.

Закон больших чисел. *Пусть* $\{X_k\}$ — *последовательность взаимно независимых случайных величин с одинаковыми распределениями. Если математическое ожидание* $\mu = E(X_k)$ *существует, то для любого* $\varepsilon > 0$ *при* $n \to \infty$

$$P\left\{\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right\} \to 0. \tag{1.2}$$

Иначе говоря, вероятность того, что среднее S_n/n отличается от математического ожидания меньше, чем на произвольно заданное ε, стремится к единице.

В такой общности эта теорема была доказана впервые Хинчиным[1]). Прежние доказательства проводились при излишнем ограничении, состоявшем в требовании конечности дисперсии[2]) $D(X_k)$. Если принять это ограничение, то можно доказать и более точный результат, обобщающий предельную теорему Муавра — Лапласа для испытаний Бернулли.

Центральная предельная теорема. *Пусть* $\{X_k\}$ — *последовательность взаимно независимых случайных величин с одинаковыми распределениями. Предположим, что* $\mu = E(X_k)$ *и* $\sigma^2 = D(X_k)$ *существуют. Пусть* $S_n = X_1 + \ldots + X_n$. *Тогда для любых фиксированных* α, β $(\alpha < \beta)$

$$P\left\{\alpha < \frac{S_n - n\mu}{\sigma n^{1/2}} < \beta\right\} \to \Phi(\beta) - \Phi(\alpha). \tag{1.3}$$

Здесь $\Phi(x)$ — нормальная функция распределения, введенная в § 1 гл. VII. Этой теоремой мы обязаны Линдебергу[3]). Ляпунов и другие авторы доказывали ее раньше при более ограничительных условиях. Необходимо представить себе, что сформулированная выше теорема является только весьма частным случаем гораздо более общей теоремы, которая в свою очередь тесно связана со многими другими предельными теоремами. Доказательство и более общую формулировку этой теоремы мы отложим до второй книги. Здесь лишь отметим, что (1.3) намного сильнее, чем (1,2), так как (1.3) дает оценку для вероятности того, что разность $|1/n \cdot S_n - \mu|$ больше, чем $\sigma/n^{1/2}$. С другой стороны, закон больших чисел (1.2) верен, даже если случайные величины X_k не имеют конечной дисперсии, так что он применим к более общему случаю, чем центральная предельная теорема (1.3). По этим соображениям мы приведем отдельное доказательство закона больших чисел, но сначала проиллюстрируем эти две предельные теоремы примерами.

Примеры. а) Рассмотрим последовательность независимых бросаний симметричной кости. Пусть X_k — число очков, выпавших при k-м бросании. Тогда

$$E(X_k) = (1+2+3+4+5+6)/6 = 3{,}5,$$

а $D(X_k) = (1^2+2^2+3^2+4^2+5^2+6^2)/6 - (3{,}5)^2 = 35/12$, и S_n/n является средним числом очков, выпавших в результате n бросаний.

---

[1]) А. Я. Хинчин, Sur la loi des grands nombres. *Comptes rendus de l'Académie des Sciences*, 189 (1929), 477—479.

[2]) А. А. Марков показал, что достаточно существования $E(|X_k|^{1+a})$ при некотором $a > 0$.

[3]) J. W. Lindeberg, Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15 (1922) 211—225.

Закон больших чисел утверждает: правдоподобно, что при больших n это среднее окажется близким к 3,5. Центральная предельная теорема устанавливает вероятность того, что $|S_n - 3{,}5n| < \alpha(35n/12)^{1/2}$ близка к $\Phi(\alpha) - \Phi(-\alpha)$. При $n = 1000$ и $\alpha = 1$ мы находим, что вероятность неравенства $3450 < S_n < 3550$ равна примерно 0,68. Выбрав для α значение $\alpha_0 = 0{,}6744$, удовлетворяющее соотношению $\Phi(\alpha_0) - \Phi(-\alpha_0) = {}^1/_2$, мы получим, что для S_n шансы находиться внутри или вне интервала 3500 ± 36 примерно одинаковы.

б) *Выборка.* Предположим, что в генеральной совокупности, состоящей из N семей, N_k семей имеют ровно по k детей $\left(k = 0, 1 \ldots; \sum N_k = N\right)$. Если семья выбрана наугад, то число детей в ней является случайной величиной, которая принимает значение ν с вероятностью $p_\nu = N_\nu/N$. При выборе с возвращением можно рассматривать выборку объема n как совокупность n независимых случайных величин или «наблюдений» $X_1, \ldots, X_n$, которые имеют все одно и то же распределение; S_n/n является *средним значением выборки*. Закон больших чисел утверждает, что для достаточно большой случайной выборки ее среднее значение будет, вероятно, близким к $\mu = \sum \nu p_\nu = \sum \nu N_\nu/N$, т. е. к среднему значению генеральной совокупности. Центральная предельная теорема позволяет оценить вероятную величину расхождения между этими средними значениями и определить объем выборки, необходимый для надежной оценки. На практике и μ и σ^2 обычно неизвестны; однако в большинстве случаев удается легко получить предварительную оценку для σ^2 и всегда можно заключить σ^2 в надежные границы. Если мы желаем, чтобы с вероятностью 0,99 или большей среднее значение выборки S_n/n отличалось от неизвестного среднего значения генеральной совокупности менее чем на 1/10, то объем выборки должен быть взят таким, чтобы

$$P\left\{\left|\frac{S_n - n\mu}{n}\right| < 1/10\right\} \geqslant 0{,}99. \tag{1.4}$$

Корень x уравнения $\Phi(x) - \Phi(-x) = 0{,}99$ равен $x = 2{,}57 \ldots$, и, следовательно, n должно быть таким, что $n^{1/2}/10\sigma \geqslant 2{,}57$ или $n > 660\sigma^2$. Осторожная предварительная оценка σ^2 дает возможность найти необходимый объем выборки. Подобная ситуация встречается весьма часто. Так, когда экспериментатор вычисляет среднее из n измерений физической величины, он также опирается на закон больших чисел и использует среднее значение выборки для оценки неизвестного математического ожидания. Надежность этой оценки может быть выражена только через σ^2, и обычно мы вынуждены использовать довольно грубые оценки для σ^2.

в) *Распределение Пуассона.* В § 4 гл. VII мы убедились, что распределение Пуассона $\{p(k; \lambda)\}$ при больших λ может быть аппроксимировано нормальным распределением. Это есть прямое следствие центральной предельной теоремы. Предположим, что случайные величины X_k имеют распределение Пуассона $\{p(k;\ \gamma)\}$. Тогда S_n имеет распределение Пуассона с математическим ожиданием и дисперсией, равными $n\gamma$.

Написав λ вместо $n\gamma$, мы заключаем, что при $n \to \infty$

$$\sum_{k < \lambda + \beta\lambda^{1/2}} e^{-\lambda} \frac{\lambda^k}{k!} \to \Phi(\beta). \tag{1.5}$$

Суммирование производится по всем k от 0 до $\lambda + \beta\lambda^{1/2}$. Очевидно, что (1.5) имеет место и тогда, когда $\lambda \to \infty$ произвольным образом. Эта теорема используется в теории суммирования расходящихся рядов и представляет общий интерес; оценки разности правой и левой части равенства (1.5) могут быть получены из общей теории.

Замечание о случайных величинах, не имеющих математического ожидания. Если математическое ожидание μ не существует, то и закон больших чисел и центральная предельная теорема становятся бессмысленными, но их можно заменить более общими предельными теоремами, которые дают аналогичную информацию. В современной теории случайные величины, не имеющие математического ожидания, играют важную роль, и многие *времена ожидания и возвращения* в физике оказываются величинами именно такого типа. Это верно даже в простом случае игры с бросанием монеты.

Предположим, что n правильных монет бросаются поодиночке. Пусть X_k означает число испытаний, предшествующих моменту, когда впервые общее число гербов будет равно общему числу решеток для k-й монеты. Величины X_k взаимно независимы и одинаково распределены: каждая из них принимает лишь положительные четные значения и $P\{X_k = 2r\} = f_{2r}$ [распределение $\{f_{2r}\}$ определяется формулой (4.2) гл. III]. Согласно теореме 3 § 4 гл. III, распределение суммы $S_n = X_1 + \ldots + X_n$ задается равенствами

$$P\{S_n = 2r\} = f_{2r}^{(n)} \qquad (r = 1,\ 2,\ \ldots), \tag{1.6}$$

где величины $f_{2r}^{(n)}$ определяются формулой (4.11) гл. III. В той же гл. III, § 8 (в), было показано, что при $n \to \infty$

$$P\{S_n < n^2 x\} \to 2\,[1 - \Phi(x^{-1/2})]. \tag{1.7}$$

Итак, мы получили *предельную теорему того же типа, что и центральная предельная теорема, с той заметной разницей, что теперь предельное распределение имеет случайная величина* $\frac{S_n}{n^2}$, *в то время как раньше рассматривались величины* $\frac{S_n}{n}$.

На языке физики величины X_k можно интерпретировать как независимые измерения некоторого параметра, и наша предельная теорема утверждает, что $\frac{S_n}{n}$ возрастает по вероятности линейно вместе с n. Неожиданные следствия этого утверждения обсуждались в гл. III [1]).

§ 2 *). Доказательство закона больших чисел

Мы проведем это доказательство в два этапа. Сначала предположим, что $\sigma^2 = D(X_k)$ существует, и заметим, что в этом случае $D(S_n) = n\sigma^2$ по теореме о дисперсии суммы (формула (5.6) гл. IX). Согласно неравенству Чебышева (§ 6 гл. IX), при любом $t > 0$

$$P\{|S_n - n\mu| > t\} \leqslant \frac{n\sigma^2}{t^2}. \tag{2.1}$$

При $t > \varepsilon n$ левая часть меньше, чем $\sigma^2/\varepsilon^2 n$, а последняя величина стремится к нулю. Это завершает доказательство.

Отбросим теперь ограничительное условие существования $D(X_k)$. Этот случай сводится к предшествующему *методом усечения*, используемым (с различными усовершенствованиями) во многих подобных случаях.

Определим два новых набора случайных величин, зависящих от X_k, следующим образом:

$$\begin{aligned} U_k &= X_k, \; V_k = 0, \quad &\text{если } |X_k| \leqslant \delta n; \\ U_k &= 0, \quad V_k = X_k, \quad &\text{если } |X_k| > \delta n. \end{aligned} \tag{2.2}$$

Здесь $k = 1, \ldots, n$ и $\delta > 0$ фиксировано. Тогда

$$X_k = U_k + V_k \tag{2.3}$$

при всех k.

Пусть $\{f(x_j)\}$ — распределение вероятностей случайных величин X_k (одинаковое для всех x_k). Мы предположили, что $\mu = E(X_k)$ существует, так что сумма

$$\sum |x_j| f(x_j) = A \tag{2.4}$$

конечна. Тогда существует и

$$\mu_n' = E(U_k) = \sum_{|x_j| \leqslant \delta n} x_j f(x_j), \tag{2.5}$$

где суммирование производится по всем тем j, при которых $|x_j| \leqslant \delta n$. Отметим, что хотя μ' и зависит от n, но оно одинаково для

---

[1]) Аналоги закона больших чисел для случая величин, не имеющих математического ожидания, даны в § 4 и задаче 13.

*) См. примечание на стр. 138.

U_1, U_2, ..., U_n. Кроме того, $\mu'_n \to \mu$ при $n \to \infty$, и, следовательно, для произвольного $\frac{\varepsilon}{2} > 0$ и всех достаточно больших n

$$|\mu'_n - \mu| < \frac{\varepsilon}{2}. \tag{2.6}$$

Далее, из (2.5) и (2.4) следует, что

$$D(U_k) \leqslant E(U_k^2) \leqslant \delta n \sum_{|x_j| \leqslant \delta n} |x_j| f(x_j) \leqslant \delta A n. \tag{2.7}$$

U_k ваимно независимы, и с их суммой $U_1 + U_2 + \ldots + U_n$ можно поступить точно так же, как и с X_k в случае конечной дисперсии; применив неравенство Чебышева, мы получим аналогично (2.1)

$$P\left\{\left|\frac{U_1 + \ldots + U_n}{n} - \mu'\right| > \frac{\varepsilon}{2}\right\} < \frac{D(U_k)}{n\left(\frac{\varepsilon}{2}\right)^2} < \frac{4\delta A}{\varepsilon^2}. \tag{2.8}$$

Вследствие (2.6) отсюда вытекает, что

$$P\left\{\left|\frac{U_1 + \ldots + U_n}{n} - \mu\right| > \varepsilon\right\} < \frac{4\delta A}{\varepsilon^2}. \tag{2.9}$$

Далее заметим, что с большой вероятностью $V_k = 0$. Действительно,

$$P\{V_k \neq 0\} = \sum_{|x_j| > \delta n} f(x_j) \leqslant \frac{1}{\delta n} \sum_{|x_j| > \delta n} |x_j| f(x_j). \tag{2.10}$$

Поскольку ряд (2.4) сходится, последняя сумма стремится к нулю при возрастании n. Таким образом, при достаточно большом n

$$P\{V_k \neq 0\} \leqslant \delta/n, \tag{2.11}$$

и следовательно (см. формулу (6.6) гл. I),

$$P\{V_1 + \ldots + V_n \neq 0\} \leqslant \delta. \tag{2.12}$$

Но $S_n = (U_1 + \ldots + U_n) + (V_1 + \ldots + V_n)$, и из (2.9) и (2.12) получаем

$$P\left\{\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right\} \leqslant P\left\{\left|\frac{U_1 + \ldots + U_n}{n} - \mu\right| > \varepsilon\right\} + \\ + P\{V_1 + \ldots + V_n \neq 0\} \leqslant \frac{4\delta A}{\varepsilon^2} + \delta. \tag{2.13}$$

Так как ε и δ произвольны, правая часть может быть сделана сколь угодно малой, что и завершает доказательство.

§ 3. Теория «безобидных» игр

При дальнейшем анализе сущности закона больших чисел будем пользоваться традиционной терминологией игроков, хотя наши рассмотрения допускают в равной степени и более серьезные приложения, а два наших основных предположения более реальны в статистике и физике, чем в азартных играх. Во-первых, предположим, что игрок обладает *неограниченным капиталом*, так что никакой проигрыш не может вызвать окончания игры. (Отбрасывание этого предположения приводит к задаче *о разорении* игрока, которая всегда интригует изучающих теорию вероятностей. Эта задача важна в последовательном анализе Вальда и в теории стохастических процессов, и мы вернемся к ней в гл. XIV.) Во-вторых, предположим, что игрок не имеет права *прервать игру, когда ему заблагорассудится: число n испытаний должно быть фиксировано заранее* и не должно зависеть от хода игры. Иначе игрок, осчастливленный неограниченным капиталом, дождался бы серии удач и в подходящий момент прекратил бы игру. Такого игрока интересует не вероятное колебание в заданный момент, а максимальные колебания в длинной серии партий, которые описываются скорее законом повторного логарифма, чем законом больших чисел (см. гл. VIII, § 5).

Введем случайную величину X_k как (положительный или отрицательный) выигрыш при k-м повторении игры. Тогда сумма $S_n = X_1 + \ldots + X_n$ является суммарным выигрышем при n повторениях игры. Если перед каждым повторением игрок уплачивает за право участия в игре (не обязательно положительный) взнос μ', то $n\mu'$ представляет собой общий уплаченный им взнос, а $S_n - n\mu'$ — *общий чистый выигрыш*. Закон больших чисел применим, если $\mu = E(X_k)$ существует. Грубо говоря, при больших n весьма правдоподобно, что разность $S_n - n\mu$ окажется малой по сравнению с n. Следовательно, если μ' меньше, чем μ, то при больших n игрок будет, вероятно, иметь выигрыш порядка $n(\mu - \mu')$. По тем же соображениям взнос $\mu' > \mu$ практически наверняка приводит к убытку. Короче, случай $\mu' < \mu$ *благоприятен* для игрока, а случай $\mu' > \mu$ *неблагоприятен*.

Заметим, что мы еще ничего не говорили о случае $\mu' = \mu$. В этом случае *единственно* возможным заключением является то, что при достаточно большом *n общий выигрыш или проигрыш $S_n - n\mu$ будет с очень большой вероятностью малым по сравнению с n*. Но при этом неизвестно, окажется ли $S_n - n\mu$ положительным или отрицательным, т. е. будет ли игра выгодной или разорительной. Это не было учтено классической теорией, которая называла $\mu' = \mu$ безобидной ценой, а игру с $\mu' = \mu$ «безобидной». Нужно понимать, что «безобидная» игра может на самом деле быть и явно выгодной и разорительной.

Ясно, что в «нормальном случае» существует не только $E(X_k)$, но и $D(X_k)$. В этом случае закон больших чисел дополняется центральной предельной теоремой, а последняя говорит о том, что весьма правдоподобно, что при «безобидной» игре чистый выигрыш в результате продолжительной игры $S_n - n\mu$ будет иметь величину порядка $n^{1/2}$ и что при достаточно больших n этот выигрыш будет с примерно равными шансами положительным или отрицательным. Таким образом, если применима центральная предельная теорема, то термин «безобидная» игра оказывается оправданным, хотя даже и в этом случае мы имеем дело с предельной теоремой, что подчеркивается словами «в результате продолжительной игры». Тщательный анализ показывает, что сходимость в (1.3) ухудшается при возрастании дисперсии. Если $\sigma^2 = D(X_k)$ велико, то нормальное приближение окажется эффективным только при чрезвычайно больших n.

Для определенности представим машину, при опускании в которую рубля игрок может с вероятностью 10^{-6} выиграть $10^6 - 1$ рублей, а в остальных случаях теряет опущенный рубль. Здесь мы имеем испытания Бернулли, и игра является «безобидной». Проделав миллион испытаний, игрок уплатит за это миллион рублей. За это время он может выиграть 0, 1, 2, ... раз. Согласно приближению Пуассона для биномиального распределения, с точностью до нескольких десятичных знаков вероятность выиграть ровно k раз равна $e^{-1}/k!$. Таким образом, с вероятностью 0,368 ... игрок потеряет миллион, и с той же вероятностью он только окупит свои расходы; он имеет вероятность 0,184 ... приобрести ровно один миллион и т. д. Здесь 10^6 испытаний эквивалентны одному-единственному испытанию при игре с выигрышем, имеющим распределение Пуассона. (Подобная же игра может быть осуществлена сравнением двух больших колод карт. См. гл. IV, § 4.)

Очевидно, бессмысленно применять закон больших чисел в такого рода ситуациях. К этой схеме относится страхование от пожара, автомобильных катастроф и т. п. Риску подвергается большая сумма, но зато соответствующая вероятность очень мала. Однако здесь происходит обычно только одно испытание в год, так что число n испытаний никогда не становится большим. Для застрахованного игра обязательно не является «безобидной», хотя, может быть, экономически вполне выгодной. Закон больших чисел здесь не при чем. Что касается страховой компании, то она имеет дело с большим числом игр, но из-за большой дисперсии все же проявляются случайные колебания. Размер страховых премий должен быть установлен таким, чтобы предотвратить большой убыток в отдельные годы, и, следовательно, компанию интересует скорее задача о разорении, чем закон больших чисел.

Когда дисперсия бесконечна, термин «безобидная» игра становится бессмысленным; нет никаких оснований считать, что общий чистый выигрыш $S_n - n\mu'$ колеблется около нуля. Действительно, существуют примеры «безобидных» игр [1]), в которых вероятность того, что в результате игрок потерпит чистый убыток, стремится к единице. Закон больших чисел утверждает только, что этот убыток будет величиной меньшего порядка, чем n. Однако ничего большего утверждать и нельзя. Если a_n образуют произвольную последовательность, причем $a_n/n \to 0$, то можно устроить «безобидную» игру, в которой вероятность того, что общий чистый убыток в результате n повторений игры превышает a_n, стремится к единице. В задаче 15 содержится пример игры, в которой игрок может быть практически уверен в том, что его убыток превысит $n/\log n$. Эта игра является «безобидной», и взнос за право участия в каждом испытании равен единице. Однако трудно себе представить, чтобы игрок нашел ее «безобидной», если после 1 000 000 игр он практически наверняка проиграет больше, чем 150 000, и очень правдоподобно, что проигрыш будет возрастать и дальше.

§ 4 *). Петербургская игра

В классической теории понятие математического ожидания не было ясно отделено от определения вероятности, и в обращении с ним не было достаточной строгости. Поэтому изучение случайных величин, не имеющих математических ожиданий, сталкивалось с непреодолимыми трудностями, и даже сравнительно недавние дискуссии кажутся странными студенту, изучающему современную теорию вероятностей. Важность случайных величин, не имеющих математических ожиданий, была подчеркнута в заключении § 1, и вполне естественно привести пример аналога закона больших чисел для этих величин. С этой целью мы рассмотрим традиционный парадокс петербургской игры [2]).

Каждое из испытаний петербургской игры состоит в бросании правильной монеты до тех пор, пока не выпадет герб; если это случится на r-м бросании, то игрок получает 2^r рублей. Другими словами, выигрыш при каждом испытании представляет собой случайную величину, принимающую значения $2^1, 2^2, 2^3, \ldots$ с вероятностями 2^{-1}, $2^{-2}, 2^{-3}, \ldots$ соответственно. Математическое ожидание формально определяется суммой $\sum x_r f(x_r)$, в которой $x_r = 2^r$, и $f(x_r) = 2^{-r}$, так что каждое слагаемое равно единице. Таким образом, конечного математического ожидания не существует, и закон больших чисел неприменим. Ясно, что игра станет менее благоприятной для игрока,

[1]) W. Feller, Note on the law of large numbers and „fair“ games, *Annals of Mathematical Statistics*, 16 (1945), 301—304.

*) См. примечание на стр. 138.

[2]) Этот парадокс был рассмотрен Даниилом Бернулли (1700—1782). Отметим, что испытания Бернулли были названы в честь Якова Бернулли.

если изменить правила и установить, что игрок не получает ничего, если при испытании произойдет более чем N бросаний (т. е. если решетка выпадет N раз подряд). В этой измененной игре выигрыш имеет конечное математическое ожидание, равное N, и закон больших чисел применим. Следовательно, если игрок уплачивает перед каждым испытанием $\mu' > 0$ рублей и повторяет игру[1]) n раз, то при достаточно большом n он может быть уверен, что получит чистую прибыль. Это верно для каждого μ', но чем больше μ', тем бóльшим должно быть n, чтобы был вероятен положительный выигрыш. Классическая теория утверждала, что $\mu' = \infty$ является «безобидным» взносом, но современный студент с трудом поймет невнятные рассуждения, приводившие к этому выводу. Оказывается вполне возможным определить взнос за право участия в петербургской игре таким образом, чтобы она имела все свойства «безобидной» игры в классическом смысле, за исключением того, что этот взнос будет зависеть от номера испытания, вместо того чтобы оставаться постоянным. Переменный взнос неудобен в игорном доме; однако петербургская игра и без того неосуществима вследствие ограниченности имеющихся денежных средств. В случае конечности математического ожидания $\mu = E(X_k) > 0$ игра называется безобидной, если при больших n отношение общего выигрыша к общему уплаченному взносу $e_n = n\mu'$ будет, вероятно, близким к единице (т. е. если разность $S_n - e_n$ будет, вероятно, иметь порядок меньший, чем $e_n = n\mu'$). Если $E(X_k)$ не существует, то нельзя полагать $e_n = n\mu'$, и надо определить e_n иначе. Мы будем говорить, что *«игра является без обидной в классическом смысле», если можно определить общий взнос e_n таким образом, чтобы для каждого* $\varepsilon > 0$

$$P\left\{\left|\frac{S_n}{e_n} - 1\right| > \varepsilon\right\} \to 0. \tag{4.1}$$

Это вполне аналогично закону больших чисел, где $e_n = n\mu'$. Закон больших чисел понимается физиком, как обстоятельство, состоящее в том, что среднее из n измерений оказывается близким к μ. В общем случае среднее из n измерений оказывается близким к e_n/n. Наша предельная теорема (4.1) имеет (в тех случаях, когда она применима) то же теоретическое и практическое значение, что и закон больших чисел.

Покажем, что петербургская игра становится безобидной в классическом смысле[2]), если положить $e_n = n \operatorname{Log} n$, где логарифм берется при основании 2, т. е. $2^{\operatorname{Log} n} = n$.

[1]) То есть всю серию бросаний до первого выпадения герба. — *Прим. ред.*

[2]) Это частный случай общего закона больших чисел, из которого легко вывести необходимые и достаточные условия выполнения (4.1); см. W. Feller, *Acta Scientiarum Litterarum Univ. Szeged*, 8 (1937), 191—201.

Доказательство. Используем «метод усечения» из § 2. Вместо (2.2) введем случайные величины U_k и V_k ($k=1, 2, \ldots, n$) следующим образом:

$$\begin{aligned} &U_k = X_k, \quad V_k = 0, \qquad \text{если} \quad X_k \leqslant n \operatorname{Log} n; \\ &U_k = 0, \quad V_k = X_k, \qquad \text{если} \quad X_k > n \operatorname{Log} n. \end{aligned} \tag{4.2}$$

Тогда $X_k = U_k + V_k$ и U_k взаимно независимы. При каждом t имеем [1]) $P\{X_k > t\} \leqslant 2/t$ и, следовательно, $P\{V_k \neq 0\} < 2/(n \operatorname{Log} n)$ или

$$P\{V_1 + V_2 + \ldots + V_n > 0\} < \frac{2}{\operatorname{Log} n} \to 0. \tag{4.3}$$

Поэтому, чтобы проверить (4.1), достаточно доказать, что

$$P\{|U_1 + \ldots + U_n - n \operatorname{Log} n| > \varepsilon n \operatorname{Log} n\} \to 0. \tag{4.4}$$

Положим $\mu = E(U_k)$ и $\sigma^2 = D(U_k)$. Эти величины зависят от n, но одинаковы для $U_1, \ldots, U_n$. Если r наибольшее из целых чисел, удовлетворяющих неравенству $2^r \leqslant n \operatorname{Log} n$, то $\mu_n = r$, и, следовательно, для достаточно больших n

$$\operatorname{Log} n < \mu_n \leqslant \operatorname{Log} n + \operatorname{Log} \operatorname{Log} n. \tag{4.5}$$

Аналогично

$$\sigma_n^2 < E(U_k^2) = 2 + 2^2 + \ldots + 2^r < 2^{r+1} \leqslant 2n \operatorname{Log} n. \tag{4.6}$$

Так как сумма $U_1 + \ldots U_n$ имеет математическое ожидание $n\mu_n$ и дисперсию $n\sigma_n^2$, из неравенства Чебышева получаем

$$P\{|U_1 + \ldots + U_n - n\mu_n| > \varepsilon n\mu_n\} \leqslant \frac{n\sigma_n^2}{\varepsilon^2 n^2 \mu_n^2} < \frac{2}{\varepsilon^2 \operatorname{Log} n} \to 0. \tag{4.7}$$

Но по (4.5) $\mu_n \sim \operatorname{Log} n$ и, следовательно, (4.7) эквивалентно (4.4) [2]), а поэтому и (4.1).

---

[1]) Пусть s — наименьшее целое число, для которого $2^s > t$. Тогда $P(X_k > t) = \sum\limits_{r=s}^{\infty} 2^{-r} = 2/2^s < 2/t$. — *Прим. ред.*

[2]) Этот переход можно обосновать с помощью следующей леммы: *если $\{Y_n\}$ — последовательность случайных величин, для которой при некоторой постоянной a и любом $\varepsilon > 0$ выполняется соотношение $P(|Y_n - a| > \varepsilon) \to 0$, и если $c_n \to 1$, то при любом $\varepsilon > 0$ $P(|c_n Y_n - a| > \varepsilon) \to 0$.* Для доказательства заметим, что

$$P\{|c_n Y_n - a| > \varepsilon\} \leqslant P\left\{|Y_n - a| > \frac{\varepsilon}{2}\right\} + P\left\{|Y_n| > \frac{\varepsilon}{2|c_n - 1|}\right\} \leqslant$$
$$\leqslant P\left\{|Y_n - a| > \frac{\varepsilon}{2}\right\} + P\left\{|Y_n - a| > \frac{\varepsilon}{2|c_n - 1|} - |a|\right\},$$

и правая часть, очевидно, стремится к нулю. — *Прим. ред.*

§ 5. Случайные величины с различными распределениями

До сих пор мы рассматривали только тот случай, когда величины X_k имели одинаковые распределения. При игре это означает, что игрок продолжает все время играть в одну и ту же игру: однако еще интереснее выяснить, что случится, если род игры будет изменяться на каждом шагу. Не обязательно представлять себе игорный дом; статистик, применяющий статистические критерии, занят вполне достойным видом «игры», а в этом случае распределение случайной величины изменяется от одного шага к другому.

Для определенности представим, что задана бесконечная последовательность распределений вероятностей. Тогда при каждом n можно говорить о взаимно независимых случайных величинах $X_1, \ldots, X_n$ с заданными распределениями. Предположим, что математические ожидания и дисперсии существуют, и положим

$$\mu_k = E(X_k), \quad \sigma_k^2 = D(X_k). \tag{5.1}$$

Сумма $S_n = X_1 + \ldots + X_n$ имеет также конечные математическое ожидание и дисперсию

$$m_n = E(S_n); \quad s_n^2 = D(S_n), \tag{5.2}$$

определяемые формулами

$$m_n = \mu_1 + \ldots + \mu_n; \quad s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2 \tag{5.3}$$

[см. гл. IX, формулы (2.4) и (5.6)]. В частном случае одинаковых распределений мы имели $m_n = n\mu$, $s_n^2 = n\sigma^2$.

Говорят, что для последовательности $\{X_k\}$ *выполняется закон больших чисел, если при каждом* $\varepsilon > 0$

$$P\left\{\frac{|S_n - m_n|}{n} > \varepsilon\right\} \to 0. \tag{5.4}$$

О последовательности $\{X_k\}$ *говорят, что она удовлетворяет центральной предельной теореме, если при любых фиксированных* $\alpha < \beta$

$$P\left\{\alpha < \frac{S_n - m_n}{s_n} < \beta\right\} \to \Phi(\beta) - \Phi(\alpha). \tag{5.5}$$

Одной из самых характерных черт теории вероятностей является то, что закон больших чисел и центральная предельная теорема выполняются для поразительно широкого класса последовательностей $\{X_k\}$. В частности, *закон больших чисел выполняется всегда, когда* X_k *равномерно ограничены*, т. е. когда существует такая постоянная A, что при всех k

$$|X_k| < A.$$

Более общее условие, достаточное для того, чтобы выполнялся закон больших чисел, состоит в том, чтобы

$$\frac{s_n}{n} \to 0. \tag{5.6}$$

Это является прямым следствием неравенства Чебышева, и доказательство, приведенное в первом абзаце § 2, здесь полностью применимо. Отметим, однако, что условие (5.6) не является необходимым (см. задачу 14).

Известны различные достаточные условия для центральной предельной теоремы, но все они вытекают из *теоремы Линдеберга* [1]), *по которой центральная предельная теорема выполняется, если при каждом $\varepsilon > 0$ усеченные случайные величины U_k, определенные формулами*

$$\begin{aligned} U_k &= X_k - \mu_k, \quad \text{если} \quad |X_k - \mu_k| \leqslant \varepsilon s_n, \\ U_k &= 0, \qquad\qquad \text{если} \quad |X_k - \mu_k| > \varepsilon s_n, \end{aligned} \tag{5.7}$$

удовлетворяют условиям $s_n \to \infty$ и

$$\frac{1}{s_n^2} \sum_{k=1}^{n} E\left(U_k^2\right) \to 1. \tag{5.8}$$

Если X_k равномерно ограничены, т. е. если $|X_k| < A$, то $U_k = X_k - \mu_k$ при всех n настолько больших, что $s_n > 2A\varepsilon^{-1}$. Левая часть в (5.8) равна тогда 1.

Следовательно, из теоремы Линдеберга вытекает, что *любая равномерно ограниченная последовательность взаимно независимых случайных величин удовлетворяет центральной предельной теореме* в предположении, конечно, что $s_n \to \infty$. (Последнее условие нарушается только в вырожденных случаях.) Позднее было обнаружено, что условие Линдеберга также и необходимо [2]) для того, чтобы (5.5) было выполнено [3]). Доказательство этих теорем откладывается до второй книги, где будут даны также оценки разности между обеими частями соотношения (5.5).

Мы обнаружили, что в случае, когда величины X_k имеют одинаковые распределения, центральная предельная теорема сильнее, чем

[1]) См. примечание 3 на стр. 249.

[2]) При условии, что $s_n \to \infty$, $\sigma_n/s_n \to 0$. — *Прим. ред.*

[3]) W. Feller, Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift*, **40** (1935), 521—559. Получены также обобщения центральной предельной теоремы, применимые к случайным величинам, не имеющим математических ожиданий. Отметим, что мы рассматривали независимые случайные величины. Для зависимых величин условия Линдеберга не являются ни необходимыми, ни достаточными.

закон больших чисел. Однако в общем случае это не так, и мы увидим, что центральная предельная теорема применима иногда к последовательностям, не удовлетворяющим закону больших чисел.

Примеры. а) Пусть $\lambda > 0$ фиксировано и пусть X_k равно $+k^\lambda$ или $-k^\lambda$ с вероятностями $1/2$ (например, бросается монета, причем ставка при k-м бросании равна $\pm k^\lambda$). Здесь $\mu_k=0$, $\sigma_k^2 = k^{2\lambda}$ и

$$s_n^2 = 1^{2\lambda} + 2^{2\lambda} + 3^{2\lambda} + \ldots + n^{2\lambda} \sim \frac{n^{2\lambda+1}}{2\lambda+1}. \tag{5.9}$$

Условие (5.7) *выполняется*, если $\lambda < 1/2$; следовательно, при $\lambda < 1/2$ выполняется закон больших чисел. Скоро мы увидим, что при $\lambda \geqslant 1/2$ закон больших чисел не выполняется. При $k = 1, 2, \ldots, n$ имеем $|X_k| = k^\lambda \leqslant n^\lambda$. Согласно (5.10) для всех n, превосходящих некоторое n_0, $s_n^2 > 1/2 \cdot n^{2\lambda+1}/(2\lambda+1)$. Отсюда легко вывести, что если n больше максимального из чисел $2(2\lambda+1)\varepsilon^{-2}$ и n_0, то усеченные величины U_k совпадают с X_k. Следовательно, при $\lambda > 0$ условия Линдеберга выполняются, и

$$P\left\{\alpha < \left(\frac{2\lambda+1}{n^{2\lambda+1}}\right)^{1/2} S_n < \beta\right\} \to \Phi(\beta) - \Phi(\alpha). \tag{5.10}$$

Следовательно, весьма вероятно, что S_n будет иметь порядок роста $n^{\lambda+1/2}$, так что при $\lambda \geqslant 1/2$ закон больших чисел не выполняется. В этом примере центральная предельная теорема применима при всех $\lambda > 0$, а закон больших чисел — только при $\lambda < 1/2$.

б) Рассмотрим две независимые последовательности по 1000 бросаний монеты. Исследуем *разность* D чисел выпадения герба. Пусть бросания двух последовательностей занумерованы числами от 1 до 1000 и от 1001 до 2000 соответственно. Определим 2000 случайных величин X_k следующим образом: если на k-й монете выпала решетка, то $X_k = 0$, если же выпал герб, то положим $X_k = 1$, если $k \leqslant 1000$, и $X_k = -1$, если $k > 1000$. Тогда $D = X_1 + X_2 + \ldots + X_{2000}$. Далее, $\mu_k = \pm 1/2$ в зависимости от последовательности, к которой принадлежит монета, $\sigma_k^2 = 1/4$, $m_{2000} = 0$, $s_{2000}^2 = 500$. Следовательно, вероятность того, что разность будет иметь значение в пределах $\pm (500)^{1/2}\alpha$ приблизительно равна $\Phi(\alpha) - \Phi(-\alpha)$. Поэтому случайная величина D сравнима с отклонением $S_{2000} - 1000$ числа выпадений герба при 2000 бросаний монеты от среднего числа выпадений 1000.

в) Различные выводы из центральной предельной теоремы можно проиллюстрировать приложениями к *теории наследственности*. В § 5 гл. V мы рассматривали свойства организма, которые зависят существенно лишь от одной пары генов (аллелей). Другие свойства (такие, как рост), согласно нашим представлениям, зависят от совместного действия нескольких пар генов. Для простоты предположим, что для каждой отдельной пары генов существует три генотипа AA, *Aa* и *aa* и что в зависимости от генотипа рост организма увеличивается на x_1, x_2 и x_3 сантиметра соответственно. Генотип индивидуума является случайным событием, и влияние отдельной пары генов на рост определяется случайной величиной X, принимающей значения x_1, x_2 и x_3 с некоторыми вероятностями. Рост зависит от нескольких случайных величин $X_1, X_2, \ldots, X_n$ и так как влияние каждой из них мало, то в первом приближении можно считать, что он просто равен *сумме* $X_1+X_2+\ldots+X_n$. Совершенно очевидно, что не *все* X_k независимы. Однако центральная предельная теорема применима также и к широкому классу зависимых величин и, кроме того, весьма правдоподобно, что большинство X_k можно считать независимыми. Эти рассуждения можно уточнить. Мы привели их только для того, чтобы показать, как центральная предельная теорема объясняет тот факт, что эмпирические распределения, появляющиеся при биометрических исследованиях таких свойств, как рост, близки к нормальному. Более полная теория позволяет выяснить вопрос о наследовании такого рода свойств, например оценить средний рост детей, зная рост их родителей. Соответствующие статистические исследования были начаты еще Галтоном и К. Пирсоном.

§ 6 *). Приложения к комбинаторике

Приведем два примера приложений центральной предельной теоремы к задачам, которые не связаны непосредственно с теорией вероятностей.

В дальнейшем мы рассматриваем пространство, точками которого являются $n!$ перестановок элементов $a_1, a_2, \ldots, a_n$, и приписываем каждой перестановке вероятность $1/n!$.

а) **Инверсии.** Говорят, что элемент a_k образует в данной перестановке r инверсий, если он стоит впереди r элементов с меньшими номерами. Так в перестановке $(a_3a_6a_1a_5a_2a_4)$ элемент a_2 не образует ни одной инверсии, элемент a_3 образует две, a_4 — ни одной, а a_5 — две и a_6 — четыре инверсии. Общее число инверсий в этом случае равно восьми. В перестановке $(a_6a_5a_4a_3a_2a_1)$ имеется 15 инверсий, причем a_k образует $k-1$ инверсий ($k=2$, 3, 4, 5, 6). Пусть X_k — число инверсий, образованных элементом a_k. Тогда $S_n=X_1+ +\ldots+X_n$ равно общему числу инверсий. X_k можно считать случайными величинами, причем X_k принимает значения $0, 1, \ldots, k-1$ с вероятностями $1/k$. Следовательно,

$$\mu_k=\frac{k-1}{2};$$

$$\sigma_k^2=\frac{1+2^2+\ldots+(k-1)^2}{k}-\left(\frac{k-1}{2}\right)^2=\frac{k^2-1}{12}. \qquad (6.1)$$

*) См. примечание на стр. 138.

Число инверсий, образованных элементом a_k, не зависит от относительного расположения элементов $a_1, \ldots, a_{k-1}$; следовательно, X_k взаимно независимы. Из (6.1) получаем

$$m_n = \frac{1+2+\ldots+(n-1)}{2} = \frac{n(n-1)}{4} \sim \frac{n^2}{4}; \tag{6.2}$$

$$s_n^2 = \frac{1}{12}\sum_{k=1}^{n}(k^2-1) = \frac{2n^3+3n^2-5n}{72} \sim \frac{n^3}{36}. \tag{6.3}$$

При больших n имеем $\varepsilon s_n > n \geqslant |X_k - \mu_k|$, и поэтому случайные величины U_k, входящие в условие Линдеберга, совпадают с X_k. Следовательно, применима центральная предельная теорема, и мы заключаем [1]), что *число N_n перестановок, для которых число инверсий лежит внутри границ $n^2/4 \pm (\alpha/6)n^{3/2}$, асимптотически равно* $n!\{\Phi(\alpha) - \Phi(-\alpha)\}$. В частности, примерно для половины всех перестановок число инверсий лежит внутри границ $(n^2/4) \pm (0{,}11)n^{3/2}$.

б) **Циклы.** Каждая перестановка разлагается на циклы. Так, в перестановке $(a_3a_6a_1a_5a_2a_4)$ элементы a_1 и a_3 образуют цикл; четыре оставшихся элемента также образуют цикл; эта перестановка содержит два цикла. Если элемент стоит на своем естественном месте, то он образует цикл, так что тождественная перестановка $(a_1, a_2, \ldots, a_n)$ содержит столько циклов, сколько в ней элементов. С другой стороны, циклические перестановки $(a_2, a_3, \ldots, a_n, a_1)$, $(a_3, a_4, \ldots \ldots, a_n, a_1, a_2)$ и т. д. содержат по одному циклу каждая. При изучении циклов удобно описывать перестановки, указывая с помощью стрелок, на какие места переходят элементы. Например, $1 \to 3 \to 4 \to 1$ указывает на то, что a_1 переходит на третье место, a_3 — на четвертое, a_4 — на первое. Этим завершается первый цикл, и мы продолжаем разложение перестановки на циклы, беря первый (в натуральном порядке) из оставшихся элементов, а именно a_2. Например,

---

[1]) Параметры m_n и s_n, входящие в формулу (5.5), заменены здесь своими асимптотическими выражениями по формулам (6.2) и (6.3). Законность такой замены вытекает из следующей леммы: *Если последовательность $\{Y_n\}$ случайных величин при любых α и β ($\alpha < \beta$) удовлетворяет предельному соотношению $P(\alpha < Y_n < \beta) \to F(\beta) - F(\alpha)$*, где $F(x)$ — *непрерывная функция, то из $\alpha_n \to \alpha$, $\beta_n \to \beta$ вытекает, что*

$$P(\alpha_n < Y_n < \beta_n) \to F(\beta) - F(\alpha).$$

Действительно для всякого $\varepsilon > 0$ существует такой номер $N = N(\varepsilon)$, что при любом $n > N(\varepsilon)$

$$P(\alpha + \varepsilon < Y_n < \beta - \varepsilon) \leqslant P(\alpha_n < Y_n < \beta_n) \leqslant P(\alpha - \varepsilon < Y_n < \beta + \varepsilon);$$

при $n \to \infty$ левая часть этого неравенства стремится к $F(\beta - \varepsilon) - F(\alpha + \varepsilon)$, а правая — к $F(\beta + \alpha) - F(\alpha - \varepsilon)$. Если ε достаточно мало, то оба эти предела сколь угодно мало отличаются от $F(\beta) - F(\alpha)$, что и доказывает лемму. — *Прим. ред.*

$(a_4 a_8 a_1 a_3 a_2 a_5 a_7 a_6)$ может быть записано как $1 \to 3 \to 4 \to 1$, $2 \to 5 \to 6 \to 8 \to 2$, $7 \to 7$. Здесь мы имеем три цикла длины 3, 4 и 1 соответственно. Другими словами, мы строим перестановку $(a_1 \ldots a_n)$ с помощью n последовательных выборов. На первом шагу выбираем место i, занятое a_1, затем место, занятое a_i, и т. д. На первом, втором, ..., n-м шагу имеется n, $n-1$, ..., 1 возможностей и только одна из них завершает цикл.

Для изучения числа циклов рассмотрим случайные величины $X_k (k=1, 2, \ldots, n)$, равные 1 или 0 в зависимости от того, завершается или нет цикл на k-м шагу нашего построения. Таким образом, в предыдущем примере $X_3 = 1$, $X_7 = 1$, $X_8 = 1$, в то время как все остальные X_k равны нулю. Ясно, что $X_1 = 1$ тогда и только тогда, когда a_1 осталось на первом месте. Если a_1 осталось на первом месте, то $X_2 = 1$ тогда и только тогда, когда a_2 остается на втором месте. Из нашего построения следует, что $p\{X_k = 1\} = \frac{1}{n-k+1}$ и $p\{X_k = 0\} = \frac{n-k}{n-k+1}$. Случайные величины X_k взаимно независимы. Математические ожидания и дисперсия определяются формулами

$$\mu_k = \frac{1}{n-k+1}, \quad \sigma_k^2 = \frac{n-k}{(n-k+1)^2}, \tag{6.4}$$

и, следовательно,

$$m_n = 1 + 1/2 + 1/3 + \ldots + 1/n \sim \log n; \tag{6.5}$$

$$s_n^2 = \sum_{k=1}^{n} \frac{n-k}{(n-k+1)^2} \sim \log n. \tag{6.6}$$

$S_n = X_1 + \ldots + X_n$ есть число циклов. *Среднее число циклов равно* m_n; *число перестановок, для которых число циклов заключено между* $\log n + \alpha(\log n)^{1/2}$ *и* $\log n + \beta(\log n)^{1/2}$, *асимптотически равно* $n!\{\Phi(\beta) - \Phi(\alpha)\}$. Усовершенствованные формы центральной предельной теоремы дают более точные оценки [1]).

§ 7 *). Усиленный закон больших чисел

Обычный закон больших чисел (5.4) утверждает, что для каждого отдельного достаточно большого n очень правдоподобно, что отклонение $|S_n - m_n|$ окажется малым по сравнению с n.

---

[1]) Много разнообразных асимптотических оценок, относящихся к комбинаторике, вывел В. Л. Гончаров, Из области комбинаторного анализа, *Изв. АН СССР*, серия математич., 8 (1944), 3—48. Использованный здесь метод проще, но имеет меньшую область применимости. См. W. Feller, The fundamental limit theorems in probability, *Bulletin of the American Mathematical Society*, 51 (1945), 800—832.

*) См. примечание на стр. 138.

Уже указывалось в связи с испытаниями Бернулли (гл. VIII), что отсюда не вытекает, что $|S_n - m_n|/n$ остается малым при всех достаточно больших n. Может случиться, что закон больших чисел применим, и все же $|S_n - m_n|/n$ продолжает колебаться между конечными или бесконечными границами. Закон больших чисел позволяет утверждать лишь, что большие значения $|S_n - m_n|/n$ появляются очень редко.

Говорят, что последовательность X_k удовлетворяет усиленному закону больших чисел, если каждой паре чисел $\varepsilon > 0$, $\delta > 0$ соответствует такое целое число N, что с вероятностью $1-\delta$ или большей для любого $r > 0$ будут выполнены все r неравенств

$$\frac{|S_n - m_n|}{n} < \varepsilon, \quad n = N,\ N+1,\ \ldots,\ N+r. \tag{7.1}$$

Грубо говоря, (7.1) означает, что с вероятностью, близкой к единице, $|S_n - m_n|/n$ остается малым [1]) при всех $n > N$.

Критерий Колмогорова. *Сходимость ряда*

$$\sum \sigma_k^2/k^2 \tag{7.2}$$

является достаточным условием для того, чтобы усиленный закон больших чисел был применим к последовательности взаимно независимых случайных величин X_k.

Доказательство. Пусть событие A_ν состоит в том, что по крайней мере при одном n, удовлетворяющем неравенству $2^{\nu-1} < n \leqslant 2^\nu$, не выполнено неравенство (7.1). Очевидно, достаточно доказать, что каково бы ни было δ, при достаточно большом $\nu\,(\nu > \log N)$ и всех r

$$P\{A_\nu\} + P\{A_{\nu+1}\} + \ldots + P\{A_{\nu+r}\} < \delta.$$

Другими словами, достаточно доказать сходимость ряда $\sum P\{A_\nu\}$. Но если событие A_ν осуществилось, то при некотором n, удовлетворяющем неравенству $2^{\nu-1} < n \leqslant 2^\nu$, выполняется соотношение

$$|S_n - m_n| \geqslant \varepsilon/2 \cdot 2^\nu, \tag{7.3}$$

и по неравенству Колмогорова (гл. IX, § 7) имеем

$$P\{A_\nu\} \leqslant 4\varepsilon^{-2} \cdot s_{2^\nu}^2 \cdot 2^{-2\nu}. \tag{7.4}$$

---

[1]) В общей теории вводится пространство элементарных событий, соответствующее бесконечной последовательности $\{X_k\}$. Усиленный закон больших чисел означает тогда, что с вероятностью единица $|S_n - m_n|/n$ стремится к нулю. В терминологии теории функций действительного переменного усиленный закон говорит о сходимости почти всюду, в то время как обычный закон эквивалентен сходимости по мере.

Следовательно,

$$\sum_{\nu=1}^{\infty} P\{A_\nu\} \leqslant 4\varepsilon^{-2} \sum_{\nu=1}^{\infty} 2^{-2\nu} \sum_{k=1}^{2\nu} \sigma_k^2 = 4\varepsilon^{-2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{2^\nu \geqslant k} 2^{-2\nu} \leqslant 8\varepsilon^{-2} \sum_{k=1}^{\infty} \frac{\sigma_k}{k^2}, \tag{7.5}$$

чем и завершается доказательство.

В качестве приложения критерия Колмогорова нами будет доказана следующая

Теорема. *Для последовательности $\{X_k\}$ взаимно независимых случайных величин, имеющих одинаковые распределения вероятностей $f\{X_j\}$ с конечным математическим ожиданием $\mu = E(X_j)$ усиленный закон больших чисел справедлив.*

Эта теорема, конечно, сильнее, чем обычный закон больших чисел § 1. Обе теоремы изложены отдельно вследствие методологического интереса их доказательств. По поводу обратимости последней теоремы см. задачу 11.

Доказательство. Снова используем метод усечения. Введем две новые последовательности случайных величин U_k и V_k, полагая

$$\begin{aligned} U_k &= X_k, \quad V_k = 0, \quad \text{если } |X_k| < k; \\ U_k &= 0, \quad V_k = X_k, \quad \text{если } |X_k| \geqslant k. \end{aligned} \tag{7.6}$$

Тогда U_k взаимно независимы, и мы покажем, что они удовлетворяют критерию Колмогорова. Имеем

$$\sigma_k^2 \leqslant E(U_k^2) = \sum_{|x_j|<k} x_j^2 f(x_j). \tag{7.7}$$

Для краткости положим

$$a_\nu = \sum_{\nu-1 \leqslant |x_j| \leq \nu} |x_j| f(x_j). \tag{7.8}$$

Тогда ряд $\sum a_\nu$ сходится, так как $E(X_k)$ существует. Кроме того, из (7.7) следует, что

$$\sigma_k^2 \leqslant a_1 + 2a_2 + 3a_3 + \ldots + ka_k, \tag{7.9}$$

и, следовательно,

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \leqslant \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{\nu=1}^{k} \nu a_\nu = \sum_{\nu=1}^{\infty} \nu a_\nu \sum_{k=\nu}^{\infty} \frac{1}{k^2} < 2 \sum_{\nu=1}^{\infty} a_\nu < \infty. \tag{7.10}$$

Математические ожидания величин U_k равны

$$E(U_k) = \mu_k = \sum_{|x_j|<k} x_j f(x_j), \tag{7.11}$$

так что $\mu_k \to \mu$ и, следовательно, $(\mu_1 + \mu_2 + \ldots + \mu_n)/n \to \mu$. Применяя усиленный закон больших чисел к $\{U_k\}$, убеждаемся [1]), что с вероятностью $1 - \delta$ или большей при всех $n \geqslant N$ выполняется неравенство

$$\left| n^{-1} \sum_{k=1}^{n} U_k - \mu \right| < \varepsilon. \tag{7.12}$$

Остается доказать, что величинами V_n можно пренебречь, т. е. что вероятность того, что при $n > N$ хотя бы одно из V_n отлично от нуля, стремится к нулю при $N \to \infty$. Легко видеть, что здесь применима (с очевидными изменениями) первая лемма Бореля — Кантелли (гл. VIII, § 3). В самом деле

$$P\{V_n \neq 0\} = \sum_{|x_j| \geqslant n} f(x_j) \leqslant \frac{a_{n+1}}{n} + \frac{a_{n+2}}{n+1} + \frac{a_{n+3}}{n+2} + \ldots \tag{7.13}$$

и, следовательно,

$$\sum P\{V_n \neq 0\} \leqslant \sum_{n=1}^{\infty} \sum_{\nu=n}^{\infty} \frac{a_{\nu+1}}{\nu} = \sum_{\nu=1}^{\infty} \frac{a_{\nu+1}}{\nu} \sum_{n=1}^{\nu} 1 = \sum_{\nu} a_{\nu+1} < \infty, \tag{7.14}$$

как и утверждалось.

§ 8. Задачи

1. Доказать, что закон больших чисел применим в примере а) § 5 также и при $\lambda \leqslant 0$. Центральная предельная теорема выполняется, если $\lambda \geqslant -1/2$.

2. Установить, будут ли выполнены закон больших чисел и центральная предельная теорема для последовательности взаимно независимых случайных величин X_k с распределениями, задаваемыми следующим образом ($k \geqslant 1$):

а) $P\{X_k = \pm 2^k\} = 1/2$;

б) $P\{X_k = \pm 2^k\} = 2^{-(2k+1)}$, $P\{X_k = 0\} = 1 - 2^{-2k}$;

в) $P\{X_k = \pm k\} = \frac{1}{2} k^{-1/2}$, $P\{X_k = 0\} = 1 - k^{-1/2}$.

---

[1]) Для обоснования этого перехода можно воспользоваться следующей леммой (в которой следует положить $Y_n = \sum_{k=1}^{n} (U_k - \mu_k)/n$; $\delta_n = \mu - \sum_{k=1}^{n} \mu_k/n$): *если для последовательности случайных величин* Y_n *и для любого* $\varepsilon > 0$ $P(|Y_n| > \varepsilon) \to 0$, *то для любой числовой последовательности* $\delta_n \to 0$ *и любого* $\varepsilon > 0$ *выполняется соотношение* $P(|Y_n - \delta_n| > \varepsilon) \to 0$.

Для доказательства заметим, что коль скоро $|\delta_n| < \varepsilon/2$, то неравенство $|Y_n - \delta_n| > \varepsilon$ имеет следствием неравенство $|Y_n| > \varepsilon/2$, и, стало быть, $P(|Y_n - \delta_n| > \varepsilon) \leqslant P(|Y_n| > \varepsilon/2)$. — *Прим. ред.*

3. Условие Ляпунова (1901). Предположим, что $s_n \to \infty$ и что при некотором фиксированном $\delta > 0$ $E(|X_k - \mu_k|^{2+\delta}) = \lambda_k$, причем λ_k / σ_k^2 равномерно ограничены. Показать, что выполняется условие Линдеберга.

4. Найти достаточные условия для $\{L_n\}$, которые обеспечили бы выполнимость закона больших чисел и/ или центральной предельной теоремы для последовательности взаимно независимых величин $\{X_n\}$, где $\{X_k\}$ принимает значения

$$0 \pm \frac{1}{2k+1} L_k, \ \pm \frac{2}{2k+1} L_k, \ldots, \pm \frac{k}{2k+1} L_k$$

с вероятностью $\frac{1}{2k+1}$ каждое.

5. Решить ту же задачу, если X_k принимает значения a_k, $-a_k$ и 0 с вероятностями p_k, p_k, $1-2p_k$ соответственно.

Замечание. *В следующих семи задачах рассматривается закон больших чисел для зависимых случайных величин.*

6. В задаче 13 гл. V пусть $X_k = 1$, если k-е бросание привело к выпадению красной грани, и $X_k = 0$ — в противном случае. Показать, что закон больших чисел неприменим.

7. Пусть $\{X_k\}$ взаимно независимы и имеют одинаковые распределения с математическим ожиданием μ и конечной дисперсией. Пусть $S_n = X_1 + \ldots + X_n$. Доказать, что закон больших чисел не выполняется для последовательности $\{S_n\}$, но выполняется для $a_n S_n$, если $a_n \to 0$.

8. Пусть $\{X_k\}$ — последовательность случайных величин, такая, что X_k может зависеть только от X_{k-1} и X_{k+1}, но не зависит от всех других X_j. Показать, что закон больших чисел выполняется, если X_k имеют конечные дисперсии.

9. Если совместное распределение величин $(X_1, \ldots, X_n)$ определено при каждом n, причем дисперсии ограничены, а ковариации отрицательны, то применим закон больших чисел.

10. Продолжение. Заменим условие $\operatorname{Cov}(X_j, X_k) \leqslant 0$ на предположение, что $\operatorname{Cov}(X_j, X_k) \to 0$ равномерно при $|j-k| \to \infty$. Доказать, что выполнен закон больших чисел.

11. Если $|S_n| < cn$, а $D(S_n) > \alpha n^2$, то закон больших чисел к $\{X_k\}$ не применим.

12. В урновой схеме Пойа (пример 2 в гл. V) пусть X_k равно 0 или 1 в зависимости от того, был ли k-й вытащенный шар черным или красным. Тогда S_n — число черных шаров при n извлечениях.

Доказать, что закон больших чисел к $\{X_k\}$ неприменим. Указание. Использовать задачу 11 этой главы и задачу 30 гл. IX.

13. Взаимно независимые случайные величины X_k принимают значения $r = 2, 3, 4, \ldots$ с вероятностями $p_r = \frac{c}{r^2 \log r}$, где c — постоянная, выбранная так, чтобы $\sum p_r = 1$. Показать, что выполняется обобщенный закон больших чисел (4.1), если положить $e_n = c \cdot n \log \log n$.

14. Пусть $\{X_n\}$ — последовательность взаимно независимых случайных величин, таких, что $X_n = \pm 1$ с вероятностями $(1-2^{-n})/2$ и $X_n = \pm 2^n$ с вероятностями 2^{-n-1}. Доказать, что к $\{X_k\}$ применимы оба закона больших чисел. (Замечание. Это означает, что условия (5.7) не являются необходимыми.)

15. Пример разорительной «безобидной» игры. Пусть возможные значения выигрыша при каждом испытании будут $0, 2, 2^2, 2^3, \ldots$; вероятность того, что выигрыш равен 2^k, равна

$$p_k = \frac{1}{2^k k(k+1)}, \tag{8.1}$$

а вероятность нулевого выигрыша равна $p_0 = 1 - (p_1 + p_2 + \ldots)$. Тогда математическое ожидание величины выигрыша равно

$$\mu = \sum 2^k p_k = (1 - 1/2) + (1/2 - 1/3) + (1/3 - 1/4) + \ldots = 1. \tag{8.2}$$

Предположим, что игрок при каждом испытании уплачивает за право участия в игре рубль, так что после n испытаний его чистый выигрыш (или проигрыш) равен $S_n - n$, где S_n — сумма n независимых случайных величин с заданным выше распределением. Показать, что при каждом $\varepsilon > 0$ *вероятность того, что в n испытаниях игрок проиграет более чем $(1 - \varepsilon)\, n/\mathrm{Log}_2\, n$ рублей, стремится к единице*, причем $\mathrm{Log}_2\, n$ означает логарифм при основании 2. Иначе говоря, надо доказать, что

$$P\left\{ S_n - n < -\frac{(1-\varepsilon)\, n}{\mathrm{Log}_2\, n} \right\} \to 1. \tag{8.3}$$

Указание. Использовать «метод усечения» § 4, но заменить границу $n\,\mathrm{Log}\, n$ в (4.2) на $n/\mathrm{Log}_2\, n$. Показать, что вероятность того, что $U_k = X_k$ при всех $k \leqslant n$, стремится к единице, и доказать, что

$$P\left\{ |U_1 + \ldots + U_n - nE(U_1)| < \frac{\varepsilon n}{\mathrm{Log}_2\, n} \right\} \to 1; \tag{8.4}$$

$$1 - \frac{1}{\mathrm{Log}_2\, n} \geqslant E(U_1) \geqslant 1 - \frac{1+\varepsilon}{\mathrm{Log}_2\, n}. \tag{8.5}$$

Подробности см. в статье, цитированной в § 3.

16. Пусть $\{X_k\}$ — последовательность взаимно независимых случайных величин с одинаковыми распределениями. Предположим, что X_n не имеет конечного математического ожидания, и пусть A — некоторая положительная постоянная. С вероятностью единица осуществится бесконечно много событий $|X_n| > An$.

17. Обращение для усиленного закона больших чисел. В предположениях задачи 16 с вероятностью единица $|S_n| > An$ для бесконечно многих n.

18. Обращение критерия Колмогорова. Если ряд $\sum \sigma_k^2/k^2$ расходится, то существует последовательность $\{X_k\}$ взаимно независимых случайных величин с $D\{X_k\} = \sigma_k^2$, к которой неприменим закон больших чисел. (Указание. Доказать, что сходимость ряда $\sum P\{|X_n| > \varepsilon n\}$ является необходимым условием для выполнения усиленного закона больших чисел.)

ГЛАВА XI

ЦЕЛОЧИСЛЕННЫЕ ВЕЛИЧИНЫ. ПРОИЗВОДЯЩИЕ ФУНКЦИИ

§ 1. Общие положения

Среди дискретных случайных величин особенно важны величины, принимающие только целые значения $k=0, 1, 2, \ldots$ Такие случайные величины называются *целочисленными*. Для изучения таких величин весьма полезен метод производящих функций. Как мы увидим в дальнейшем, этот метод есть частный случай более общего метода характеристических функций, служащего основой для решения многих проблем современной аналитической теории вероятностей. С более общей точки зрения производящие функции относятся к области операционных методов, широко применяемых в теории дифференциальных и интегральных уравнений. В теории вероятностей производящие функции употребляются со времени Муавра и Лапласа, но сила и возможности этого метода редко используются полностью.

Определение. *Пусть $a_0, a_1, a_2, \ldots$ — последовательность действительных чисел. Если ряд*

$$A(s)=a_0+a_1s+a_2s^2+\ldots \tag{1.1}$$

сходится в каком-нибудь интервале $-s_0<s<s_0$, то функция $A(s)$ называется производящей функцией последовательности $\{a_j\}$.

Сама по себе переменная s ничего не обозначает. Если последовательность $\{a_j\}$ ограничена, то сравнение с геометрической прогрессией показывает, что ряд (1.1) сходится по крайней мере при $|s|<1$.

Примеры. Если $a_j=1$ для всех j, то $A(s)=1/(1-s)$. Производящей функцией последовательности $(0, 0, 1, 1, 1, \ldots)$ является $s^2/(1-s)$. Последовательность $a_j=1/j!$ имеет производящую функцию e^s. При фиксированном n последовательность $a_j=\binom{n}{j}$ имеет производящую функцию $(1+s)^n$. Если X — число очков, выпадающих при бросании правильной кости, то распределение вероятностей величины X имеет производящую функцию $(s+s^2+s^3+s^4+s^5+s^6)/6=s(1-s^6)/6(1-s)$.

Рассмотрим теперь целочисленную случайную величину с распределением вероятностей

$$P\{X=j\}=p_j, \qquad j=0, 1, \ldots. \tag{1.2}$$

Удобно иметь специальное обозначение для «хвостов» распределения. Мы будем писать

$$P\{X > j\} = q_j, \tag{1.3}$$

так что

$$q_k = p_{k+1} + p_{k+2} + \ldots, \qquad k \geqslant 0. \tag{1.4}$$

Введем две производящие функции:

$$P(s) = p_0 + p_1 s + p_2 s^2 + \ldots \tag{1.5}$$

и

$$Q(s) = q_0 + q_1 s + q_2 s^2 + \ldots. \tag{1.6}$$

Оба ряда сходятся по крайней мере при $|s| < 1$, так как их коэффициенты ограничены. Более того, (1.2) представляет собой распределение вероятностей, так что $P(1) = 1$, вследствие чего ряд для $P(s)$ *сходится абсолютно* по крайней мере при $-1 \leqslant s \leqslant 1$.

Теорема 1. *При* $-1 < s < 1$ *имеем*

$$Q(s) = \frac{1 - P(s)}{1 - s}. \tag{1.7}$$

Доказательство. Коэффициент при s^n в разложении $(1-s) \cdot Q(s)$ равен $q_n - q_{n-1} = -p_n$, если $n \geqslant 1$, и $q_0 = p_1 + p_2 + \ldots = 1 - p_0$, если $n = 0$. Поэтому $(1-s) \cdot Q(s) = 1 - P(s)$, что и утверждалось.

Рассмотрим теперь производную

$$P'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}. \tag{1.8}$$

Этот ряд сходится по крайней мере при $-1 < s < 1$. При $s = 1$ правая часть формально сводится к $\sum k p_k = E(X)$. Если математическое ожидание существует, то производная $P'(s)$ непрерывна в замкнутом интервале $-1 \leqslant s \leqslant +1$. Если же ряд $\sum k p_k$ расходится, то $P'(s) \to \infty$ при $s \to 1$. В этом случае мы будем говорить, что X имеет *бесконечное математическое ожидание*, и писать $P'(1) = E(X) = \infty$. (Если все рассматриваемые величины положительны, то употребление символа ∞ не приводит к противоречиям.) Применяя теорему о среднем к правой части соотношения (1.7), получаем, что $Q(s) = P'(\sigma)$, где σ — точка, заключенная между s и 1. Функция $Q(s)$ монотонно возрастает при $s \to 1$, поэтому $Q(s) \to E(X)$ (конечному или бесконечному). Итак, доказана

Теорема 2. *Для* $E(X)$ *имеются два выражения*:

$$E(X) = \sum_{j=1}^{\infty} j p_j = \sum_{k=0}^{\infty} q_k \tag{1.9}$$

или в терминах производящих функций

$$E(X) = P'(1) = Q(1). \tag{1.10}$$

Дифференцируя формулу (1.8) и используя соотношение $P'(s) = Q(s) - (1-s)Q'(s)$, находим тем же способом

$$E(X(X-1)) = \sum k(k-1)P_k = P''(1) = 2Q(1). \tag{1.11}$$

Для того чтобы получить дисперсию X, к этому выражению надо прибавить $E(X) - E^2(X)$, что приводит нас к следующей теореме.

Теорема 3. *Дисперсия $D(X)$ может быть вычислена по одной из следующих формул*:

$$D(X) = P''(1) + P'(1) - P'^2(1) = 2Q'(1) + Q(1) - Q^2(1). \tag{1.12}$$

В случае бесконечной дисперсии $P''(s) \to \infty$ при $s \to 1$.

Формулы (1.10) и (1.12) дают зачастую самый простой способ определения $E(X)$ и $D(X)$.

§ 2. Композиция

Пусть X и Y — неотрицательные независимые целочисленные случайные величины с распределениями вероятностей $P\{X = j\} = a_j$ и $P\{Y = j\} = b_j$. Событие $(X = j,\ Y = k)$ имеет вероятность $a_j b_k$. Сумма $S = X + Y$ есть новая случайная величина, и событие $S = r$ есть объединение событий

$$(X = 0,\ Y = r),\ (X = 1,\ Y = r-1),$$
$$(X = 2,\ Y = r-2),\ \ldots,\ (X = r,\ Y = 0).$$

Эти события несовместны, и поэтому распределение $c_r = P\{S = r\}$ задается формулой

$$c_r = a_0 b_r + a_1 b_{r-1} + a_2 b_{r-2} + \ldots + a_{r-1} b_1 + a_r b_0. \tag{2.1}$$

Операция (2.1), сопоставляющая двум последовательностям $\{a_k\}$ и $\{b_k\}$ новую последовательность $\{c_k\}$, встречается так часто, что удобно ввести для нее специальное название и обозначение.

Определение. *Пусть $\{a_k\}$ и $\{b_k\}$ — две произвольные числовые последовательности (не обязательно распределения вероятностей). Новая последовательность $\{c_k\}$, определяемая* *формулой* (2.1) *называется композицией*[1]) *последовательностей* $\{a_k\}$ *и* $\{b_k\}$ *и обозначается*

$$\{c_k\}=\{a_k\}*\{b_k\}. \tag{2.2}$$

Пример. а) Если $a_k=b_k=1$ для всех $k\geqslant 0$, то $c_k=k+1$. Если $a_k=k$, $b_k=1$, то $c_k=1+2+\ldots+k=k(k+1)/2$. Наконец, если $a_0=a_1=1/2$, $a_k=0$ для $k\geqslant 2$, то $c_k=(b_k+b_{k-1})/2$ и т. д.

Пусть последовательности $\{a_k\}$ и $\{b_k\}$ имеют производящие функции $A(s)=\sum a_ks^k$ и $B(s)=\sum b_ks^k$. Произведение $A(s)B(s)$ можно получить почленным перемножением степенных рядов для $A(s)$ и $B(s)$. Собирая члены с одинаковыми степенями s, убеждаемся, что коэффициент c_r при s^r в разложении $A(s)B(s)$ задается формулой (2.1). Таким образом, имеет место следующая

Теорема. *Если* $\{a_k\}$ *и* $\{b_k\}$ — *последовательности с производящими функциями* $A(s)$ *и* $B(s)$ *и* $\{c_k\}$ — *их композиция, то производящая функция* $C(s)=\sum c_ks^k$ *равна произведению*

$$C(s)=A(s)B(s). \tag{2.3}$$

Если X *и* Y — *неотрицательные целочисленные независимые случайные величины с производящими функциями* $A(s)$ *и* $B(s)$, *их сумма* $X+Y$ *имеет производящую функцию* $A(s)B(s)$.

Пусть $\{a_k\}$, $\{b_k\}$, $\{c_k\}$, $\{d_k\}$, ... — какие-либо последовательности. Мы можем образовать композицию $\{a_k\}*\{b_k\}$, затем композицию этой новой последовательности с $\{c_k\}$ и т. д. Производящая функция $\{a_k\}*\{b_k\}*\{c_k\}*\{d_k\}$ равна $A(s)B(s)C(s)D(s)$, и это показывает, что порядок, в котором совершаются композиции, безразличен. Например, $\{a_k\}*\{b_k\}*\{c_k\}=\{c_k\}*\{b_k\}*\{a_k\}$ и т. д. Таким образом, *композиция — ассоциативная и коммутативная операция* (точно так же, как сложение случайных величин).

При изучении сумм независимых случайных величин X_n особый интерес представляет тот частный случай, когда величины X_n имеют одинаковые распределения. Если $\{a_j\}$ — *общее распределение вероятностей величин* X_n, *то распределение величины* $S_n=X_1+\ldots+X_n$ *будем обозначать* $\{a_j\}^{n*}$. Таким образом,

$$\{a_j\}^{2*}=\{a_j\}*\{a_j\},\quad \{a_j\}^{3*}=\{a_j\}^{2*}*\{a_j\},\ \ldots \tag{2.4}$$

[1]) Английский термин — convolution; некоторые американские и английские авторы предпочитают термин faltung, имеющий немецкое происхождение; французский термин — composition.

[Буквальный перевод слов convolution — «свертка». В русской математической литературе употребляются оба термина «свертка» и «композиция», но в литературе по теории вероятностей предпочитается последний. — *Прим. ред.*]

и вообще

$$\{a_j\}^{n*} = \{a_j\}^{(n-1)*} * \{a_j\}. \tag{2.5}$$

Иначе говоря, $\{a_j\}^{n*}$ — *последовательность чисел с производящей функцией* $A^n(s)$. В частности, $\{a_j\}^{1*}$ совпадает с $\{a_j\}$, а $\{a_j\}^{0*}$ — последовательность с производящей функцией $A^0(s) = 1$, т. е. последовательность $(1, 0, 0, 0, \ldots)$.

Примеры. б) *Биномиальное распределение.* Производящая функция биномиального распределения $b(k;\ n,\ p) = \binom{n}{k} p^k q^{n-k}$ равна

$$\sum_{k=0}^{n} \binom{n}{k} (ps)^k q^{n-k} = (q+ps)^n. \tag{2.6}$$

Производящая функция есть n-я степень бинома $q+ps$; это означает, что $\{b(k;\ n,\ p)\}$ есть распределение суммы $S_n = X_1 + \ldots + X_n$ n независимых случайных величин с одинаковыми производящими функциями $q+ps$; каждая величина X_j принимает значение нуль с вероятностью q и значение единица с вероятностью p. Таким образом,

$$\{b(k;\ n,\ p)\} = \{b(k;\ 1,\ p)\}^{n*}. \tag{2.7}$$

Мы несколько раз пользовались представлением величины, имеющей биномиальное распределение, в виде суммы $S_n = X_1 + \ldots + X_n$ (см. примеры 3, а и 5 гл. IX). Предыдущее рассуждение можно обратить и использовать для нового вывода биномиального распределения. Мультипликативное свойство $(q+ps)^m (q+ps)^n = (q+ps)^{m+n}$ показывает также, что

$$\{b(k;\ m,\ p)\} * \{b(k;\ np)\} = \{b(k;\ m+n,\ p)\}. \tag{2.8}$$

Этот результат эквивалентен формуле (10.4) гл. VI. Дифференцируя производящую функцию $(q+ps)^n$, можно совсем просто получить доказательство того, что $E(S_n) = np$ и $D(S_n) = npq$.

в) *Распределение Пуассона.* Производящая функция распределения $p(k;\ \lambda) = e^{-\lambda}\lambda^k/k!$ равна

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{(\lambda s)^k}{k!} = e^{-\lambda+\lambda s}. \tag{2.9}$$

Отсюда следует, что

$$\{p(k;\ \lambda)\} * \{p(k;\ \mu)\} = \{p(k;\ \lambda+\mu)\}, \tag{2.10}$$

что эквивалентно формуле (10.5) гл. VI.

Дифференцированием убеждаемся, что и среднее значение и дисперсия распределения Пуассона равны λ; этот результат был доказан раньше прямым вычислением (см. пример 4 гл. IX).

г) *Геометрическое и отрицательное биномиальное распределения.* Допустим, что случайная величина X имеет геометрическое распределение

$$P\{X=k\}=q^k p, \qquad k=0,\ 1,\ 2,\ \ldots, \tag{2.11}$$

где p и q — положительные постоянные, причем $p+q=1$. Соответствующая производящая функция равна

$$p\sum_{k=0}^{\infty}(qs)^k=\frac{p}{1-qs}. \tag{2.12}$$

Пользуясь результатами § 1, легко убедиться, что $E(X)=\frac{q}{p}$ и $D(X)=\frac{q}{p^2}$. Эти значения совпадают с теми, которые были получены в примере 3,в гл. IX.

Вероятность того, что в последовательности испытаний Бернулли *первый успех* случится после k неудач (т. е. при $(k+1)$-м испытании), равна $q^k p$, так что X можно интерпретировать как *время первого успеха.* Строго говоря, такая интерпретация относится к бесконечному пространству элементарных событий, и преимущество формального определения (2.11) и терминологии случайных величин состоит в том, что мы можем не беспокоиться о структуре исходного пространства элементарных событий. То же самое верно и для *времени ожидания r-го успеха.* Если X_k означает число неудач между $(k-1)$-м и k-м успехами, то $S_r=X_1+X_2+\ldots+X_r$ — полное число неудач, предшествующих r-му успеху (а S_r+r — число испытаний до r-го успеха включительно). То обстоятельство, что мы имеем дело с испытаниями Бернулли, влечет за собой, что величины X_k взаимно независимы и имеют одинаковое распределение (2.11). Это свойство можно принять за *определение* величин X_k. Тогда S_r имеет производящую функцию

$$\left(\frac{p}{1-qs}\right)^r, \tag{2.13}$$

и формула разложения бинома (8.7) гл. II показывает, что коэффициент при s_k равен

$$f(k;\ r,\ p)=\binom{-r}{k}p^r(-q)^k, \qquad k=0,\ 1,\ 2,\ \ldots. \tag{2.14}$$

Отсюда следует, что $P\{S_r=k\}=f(k;\ r,\ p)$ в соответствии с формулой для числа неудач, предшествующих r-му успеху, которая была выведена в § 8 гл. VI. Этот результат можно сформулировать по-другому, сказав, что *распределение $\{f(k;\ r,\ p)\}$ есть r-кратная композиция геометрического распределения,* т. е.

$$\{f(k;\ r,\ p)\}=\{q^k p\}^{r*}. \tag{2.15}$$

До сих пор мы считали r целым. Однако, как это было доказано в § 8 гл. IV, *отрицательное биномиальное распределение* $\{f(k;\ r,\ p)\}$ сохраняет смысл при любом неотрицательном r, не обязательно целом. Производящая функция по-прежнему задается формулой (2.13), поэтому нетрудно показать, что *математическое ожидание и дисперсия этого распределения равны rq/p и rq/p^2 и что*

$$\{f(k;\ r_1,\ p)\} * \{f(k;\ r_2,\ p)\} = \{f(k;\ r_1 + r_2,\ p)\}. \qquad (2.16)$$

§ 3. Приложение к задачам о времени первого достижения и времени первого возвращения в схеме Бернулли

Этот параграф включен главным образом для иллюстрации. Основные результаты будут получены и другими методами (см. пример 3, б гл. XIII и задачу 7 гл. VIII; гл. XIV, § 5 и задачи 11 и 15—17). Для специального случая $p = 1/2$ аналогичные результаты содержатся в гл. III. Однако последующие рассуждения дают прекрасный пример применения метода производящих функций, кроме того, поучительно сравнить различные подходы.

Рассмотрим схему испытаний Бернулли с вероятностью успеха p. Положим $X_k = 1$, если k-е испытание привело к успеху, и $X_k = -1$ в противном случае. Тогда $S_n = X_1 + \ldots + X_n$ означает разность между числом успехов и неудач в n испытаниях. На более образном «игровом» языке S_n представляет чистый выигрыш Петра после n игр. Удобно положить $S_0 = 0$.

а) **Время первого достижения.** Допустим, что Петр решил прекратить игру в тот момент, когда он впервые имеет *положительный* чистый выигрыш (необходимо равный единице). Прямое перечисление всех возможностей показывает, что это может произойти при испытаниях с номерами 1, 3, 5, 7 с вероятностями p, qp^2, $2q^2p^3$, $5q^3p^4$, ..., но общую закономерность уловить трудно. Сумма σ этих вероятностей равна *вероятности того, что чистый выигрыш Петра когда-нибудь будет положительным.* Даже величину σ нельзя получить непосредственно, однако мы покажем, что $\sigma = 1$, если $p \geqslant q$, и $\sigma = p/q$, если $p \leqslant q$. Время ожидания момента, когда чистый выигрыш впервые станет равным x, можно разбить на x частей, отвечающих приращениям чистого выигрыша на единицу. Поэтому *вероятность того, что чистый выигрыш Петра когда-нибудь достигнет уровня x,* равна σ^x. Приступим теперь к вычислению σ и вероятностей $\lambda_n^{(x)}$ того, что чистый выигрыш впервые достигнет уровня x при n-м испытании.

На более формальном языке *мы ищем вероятность λ_n того, что $S_1 \leqslant 0$, $S_2 \leqslant 0$, ..., $S_{n-1} \leqslant 0$, $S_n = 1$.* Условимся говорить, что *первое достижение точки $x > 0$ произошло при n-м испытании, если*

$$S_1 < x,\ S_2 < x,\ \ldots,\ S_{n-1} < x,\ S_n = x. \qquad (3.1)$$

Вероятность этого события будем обозначать через $\lambda_n^{(x)}$, ***положим для краткости*** $\lambda_n^{(1)} = \lambda_n$. В игровых терминах соотношения (3.1) означают, что чистый выигрыш впервые стал равен x при n-м испытании. Термин «первое достижение» подсказан приложениями к теории диффузии.

Допустим теперь, что первое достижение точки $x = 1$ произошло при r-м испытании. Чистый выигрыш в последующих испытаниях не зависит от исхода первых r испытаний и принимает значения $S_1' = X_{r+1}$, $S_2' = X_{r+1} + X_{r+2}, \ldots$. Первое достижение точки $x = 2$ произойдет при n-м испытании тогда и только тогда, когда $S_1' \leqslant 0, \ldots, S_{n-r-1}' \leqslant 0$, $S_{n-r}' = 1$. Вероятность последнего события равна λ_{n-r}. Другими словами, вероятность того, что первые достижения точек $x = 1$ и $x = 2$ произойдут при r-м и n-м испытаниях соответственно равна $\lambda_n \lambda_{n-r}$. Отсюда следует, что полная вероятность первого достижения точки $x = 2$ в момент n равна

$$\lambda_n^{(2)} = \lambda_1\lambda_{n-1} + \lambda_2\lambda_{n-2} + \ldots + \lambda_{n-1}\lambda_1. \tag{3.2}$$

Вспоминая, что $\lambda_0 = 0$, убеждаемся, что распределение $\{\lambda_n^{(2)}\} = \{\lambda_n\} * \{\lambda_n\}$ есть композиция распределения $\{\lambda_n\}$ с самим собой. Вводя производящие функции

$$\lambda(s) = \sum_{n=1}^{\infty} \lambda_n s^n, \qquad \lambda^{(x)}(s) = \sum_{n=1}^{\infty} \lambda_n^{(x)} s^n, \tag{3.3}$$

получаем $\lambda^{(2)}(s) = \lambda^2(s)$. Аналогичные рассуждения, проведенные по индукции, показывают, что вообще

$$\lambda^{(x)}(s) = \lambda^x(s). \tag{3.4}$$

Итак, наша задача сведена к нахождению вероятностей λ_n первого достижения точки $x = 1$. Если $X_1 = 1$, то первое достижение имеет место при первом испытании. Если $X_1 = -1$, то чистый выигрыш $X_2 + X_2 + X_3, \ldots$ в последующих испытаниях должен возрасти на две единицы, и поэтому

$$\lambda_1 = p, \; \lambda_n = q\lambda_{n-1}^{(2)}, \; n > 1. \tag{3.5}$$

Это, очевидно, эквивалентно соотношению

$$\lambda(s) = ps + qs\lambda^2(s), \tag{3.6}$$

которое представляет квадратное уравнение относительно $\lambda(s)$. Один из двух корней не ограничен в окрестности $s = 0$, и единственное ограниченное решение (3.6) имеет вид

$$\lambda(s) = \frac{1 - \{1 - 4pqs^2\}^{1/2}}{2qs}. \tag{3.7}$$

Таким образом, мы нашли производящие функции (3.4) всех времен первого достижения. Биномиальное разложение (8.7) гл. II позволяет нам выписать коэффициенты

$$\lambda_{2m-1} = \frac{1}{2q}\binom{\frac{1}{2}}{m}(4pq)^m(-1)^{m-1}, \quad \lambda_{2m} = 0. \tag{3.8}$$

Впрочем, явные выражения нас не очень интересуют, значительно более поучительно получить нужные выводы непосредственно из производящей функции.

Заметим в первую очередь, что

$$\lambda(1) = \frac{1-|p-q|}{2q}, \tag{3.9}$$

и поэтому $\lambda(1) = 1$ при $p \geqslant q$ и $\lambda(1) = p/q$ при $p < q$. Отсюда следует, что *$\sum \lambda_k$ равна 1 или p/q в зависимости от того, что больше, p или q*. Если $q > p$ (игра разорительная для Петра), то *вероятность того, что сумма S_n никогда не будет положительной, равна $(q-p)/q$*.

В симметрическом случае $p = q = 1/2$ и $\sum \lambda_k = 1$. Если серию бросаний правильной монеты неограничено продолжать, то Петр может быть уверен, что его чистый выигрыш рано или поздно станет положительным. Встает вопрос: сколько времени пройдет до этого момента. Из соотношения $\lambda'(1) = \infty$ следует, что *при бросании монеты число испытаний, предшествующих первому достижению точки $x = 1$, имеет бесконечное математическое ожидание*. Если Петр рассчитывает выиграть единичную сумму в игре с бросанием монеты, прекращая игру в первый подходящий момент, он должен учитывать, что для этого потребуется огромное количество испытаний (и соответственно огромный начальный капитал). Нет нужды говорить о том, что бесконечность математического ожидания времени первого достижения тесно связана с неожиданными свойствами случайных колебаний при игре с бросанием монеты. Последний вопрос подробно обсуждался в гл. III.

Замечание. Сейчас мы располагаем явной формулой для λ_n, но еще остается задача о вычислении вероятностей $\lambda_n^{(x)}$ из соотношений (3.3) или (3.4). Стандартный аналитический метод для решения такого рода задач состоит в применении теории функций комплексного переменного. Поэтому интересно отметить, что принцип отражения позволяет получить явное выражение для $\lambda_n^{(x)}$ по крайней мере в симметрическом случае $p = q = {}^1/_2$. (В обозначениях гл. III имеем $f_{2n}^{(x)} = \lambda_{2n-x}^{(x)}$; см. теорему 2 § 4.) Взглянув теперь на соотношения (3.4) и (3.7), нетрудно прийти к интересному выводу, что при произвольном p вероятность $\lambda_n^{(x)}$ равна соответствующей вероятности для симметрического случая, умноженной на $(4pq)^{1/2 n}\left(\frac{p}{q}\right)^{1/2 x}$. Весьма поучительно разобрать этот случай в деталях и еще раз убедиться в том что *самые элементарные комбинаторные методы позволяют зачастую решить технически трудную задачу, не прибегая к сложному аналитическому аппарату.*

б) **Время возвращения.** Мы будем говорить, что *первое возвращение в нуль произошло при n-м испытании*, если $S_1 \neq 0$, $S_2 \neq 0, \ldots$, $S_{n-1} \neq 0$, $S_n = 0$ (т. е. впервые при n-м испытании общее число успехов стало равно общему числу неудач). Пусть f_n — вероятность этого события. (Ясно, что $f_{2n+1} = 0$ при всех n. Несколько первых вероятностей f_{2n} легко найти прямым пересчетом: $f_2 = 2pq$, $f_4 = 2p^2q^2$, $f_6 = 4p^3q^3$, $f_8 = 10p^4q^4$.)

Пусть $\lambda_n^{(-1)}$ — вероятность того, что первое достижение точки $x = -1$ имеет место при n-м испытании; другими словами, $\lambda_n^{(-1)}$ есть величина, получающаяся из $\lambda_n^{(1)} = \lambda_n$ перестановкой p и q. Как и выше, заметим, что возвращение в нуль при n-м испытании эквивалентно первому достижению либо $+1$, либо -1 в $n-1$ испытаниях, следующих за первым. Отсюда вытекает, что

$$f_n = q\lambda_{n-1} + p\lambda_{n-1}^{(-1)}. \tag{3.10}$$

Умножим эти равенства на s^n и просуммируем по n. Учитывая, что производящие функции последовательностей $\{\lambda_n^{(-1)}\}$ и $\{\lambda_n\}$ получаются одна из другой заменой p на q, получаем

$$F(s) = \sum f_n s^n = 1 - (1 - 4pqs^2)^{1/2}. \tag{3.11}$$

Следовательно, *вероятность $\sum f_n$ того, что полное число успехов когда-нибудь будет равно полному числу неудач, есть $1 - |p - q|$.*

В частности, при $p = q = \frac{1}{2}$ находим, что $\sum f_n = 1$, *но вероятностное распределение $\{f_n\}$ имеет бесконечное математическое ожидание.* Вероятности f_{2n} совершенно другим методом уже были вычислены в § 4 гл. III. Поучительно отметить, что ряд теорем гл. III можно получить непосредственно из производящей функции $F(s)$, не прибегая ни к каким вычислениям и не используя явные выражения для f_{2n}. (См. задачи 6—10.)

Замечание. По существу задача этого раздела аналогична задаче о времени ожидания примера 2, г. В пространстве *бесконечных последовательностей* испытаний Бернулли можно рассмотреть случайную величину N_r, равную числу испытаний от первого достижения $r-1$ и до первого достижения r включительно. $\{N_r\}$ — независимые случайные величины с общей производящей функцией $\lambda(s)$. Сумма $N^{(x)} = N_1 + N_2 + \ldots + N_x$ равна числу испытаний до первого достижения точки x. Соответствующая производящая функция равна $\lambda^x(s)$. Мы формально избежали введения бесконечного пространства элементарных событий, определив случайные величины в терминах их распределений. С аналитической точки зрения развитая теория является строгой и замкнутой, но для вероятностной интерпретации и для интуиции предпочтительно располагать естественным бесконечным пространством элементарных событий.

§ 4. Разложение на простые дроби

Если известна производящая функция $P(s)=\sum p_k s^k$, то коэффициенты p_k можно найти дифференцированием по очевидной формуле $p_k = P^{(k)}(0)/k!$ В практике часто невозможно получить точные выражения для $\{p_k\}$, кроме того, эти точные выражения порой так сложны, что разумное приближение оказывается предпочтительнее. Самым выгодным методом для получения приближенных выражений является, пожалуй, метод разложения на простые дроби. Из теории функций комплексного переменного хорошо известно, что такие разложения допускаются широким классом функций, но мы ограничим свои рассмотрения простым случаем рациональных функций.

Предположим, что производящая функция имеет вид

$$P(s)=\frac{U(s)}{V(s)}, \tag{4.1}$$

где $U(s)$ и $V(s)$ — многочлены, не имеющие общих корней. Допустим для простоты, что степень $U(s)$ меньше степени $V(s)$. Пусть $V(s)$ имеет степень m, и предположим, что все m (действительных или мнимых) корней $s_1, \ldots, s_m$ многочлена $V(s)$ различны. Тогда

$$V(s)=(s-s_1)\,(s-s_2)\,\ldots\,(s-s_m), \tag{4.2}$$

и из алгебры известно, что $P(s)$ может быть разложена на *простые дроби*

$$P(s)=\frac{\rho_1}{s_1-s}+\frac{\rho_2}{s_2-s}+\ldots+\frac{\rho_m}{s_m-s}, \tag{4.3}$$

где $\rho_1, \rho_2, \ldots, \rho_m$ — постоянные. Чтобы найти постоянную ρ_1, умножим (4.3) на $(s-s_1)$; при $s \to s_1$ произведение $(s-s_1)\,P(s)$ стремится к ρ_1. С другой стороны, из (4.1) и (4.2) получаем

$$(s_1-s)\,P(s)=\frac{-U(s)}{(s-s_2)\,(s-s_3)\,\ldots\,(s-s_m)}. \tag{4.4}$$

При $s \to s_1$ числитель в (7.4) стремится к $-U(s_1)$, а знаменатель — к произведению $(s_1-s_2)\ldots(s_1-s_m)$, равному $V'(s_1)$. Итак, $\rho_1=-U(s_1)/V'(s_1)$. Это рассуждение применимо ко всем корням, и поэтому при $k \leqslant m$

$$\rho_k=\frac{-U(s_k)}{V'(s_k)}. \tag{4.5}$$

К сожалению, для приведения (4.1) к виду (4.3) обычно необходимы громоздкие выкладки. Однако, получив разложение (4.3), нетрудно вывести точное выражение для коэффициента при s^n в $P(s)$. Заметим, что

$$\frac{1}{s_k - s} = \frac{1}{s_k} \cdot \frac{1}{1 - s/s_k}. \tag{4.6}$$

При $|s| < |s_k|$ последнюю дробь можно разложить в геометрическую прогрессию

$$\frac{1}{1 - s/s_k} = 1 + \frac{s}{s_k} + \left(\frac{s}{s_k}\right)^2 + \left(\frac{s}{s_k}\right)^3 + \ldots, \tag{4.7}$$

Подставляя выражения (4.7) при $k = 1, 2, \ldots, m$ в (4.3), получим *коэффициент* p_n *при* s^n в виде

$$p_n = \frac{\rho_1}{s_1^{n+1}} + \frac{\rho_2}{s_2^{n+1}} + \ldots + \frac{\rho_m}{s_m^{n+1}}. \tag{4.8}$$

Таким образом, для получения p_n нужно сначала найти корни знаменателя $s_1, s_2, \ldots, s_m$, а затем по (4.5) определить коэффициенты $\rho_1, \ldots, \rho_m$.

Формула (4.8) дает *точное* выражение для вероятности p_n. Работа, требующаяся для вычисления всех m корней, обычно бывает затруднительна, и поэтому формула (4.8) имеет преимущественно теоретический интерес. К счастью, почти всегда для удовлетворительного приближения бывает достаточно одного члена из (4.8). Действительно, пусть s_1 — *наименьший* по абсолютной величине корень. Тогда первый знаменатель в (4.8) является наименьшим. Ясно, что при возрастании n относительная доля остальных членов убывает и первое слагаемое в (4.8) преобладает. Другими словами, *если* s_1 — *наименьший по абсолютной величине корень уравнения* $V(s) = 0$, *то при* $n \to \infty$

$$p_n \sim \frac{\rho_1}{s_1^{n+1}} \tag{4.9}$$

(знак $\sim$ означает, что отношение правой и левой частей стремится к единице). Обычно эта формула дает хорошее приближение даже при сравнительно малых значениях n. Главное преимущество формулы (4.9) состоит в том, что она требует вычисления только одного (минимального по абсолютной величине) корня алгебраического уравнения.

Легко освободиться от тех ограничений, при которых мы вывели асимптотическую формулу (4.9). Чтобы сделать это, допустим сначала, что степень числителя в (4.1) превосходит степень знаменателя m. Пусть $U(s)$ имеет степень $m + r$ $(r \geqslant 0)$; с помощью деления $P(s)$ приводится к сумме многочлена степени r и дроби $U_1(s)/V(s)$, где $U_1(s)$ — многочлен степени, меньшей m. Полученный многочлен влияет только на первые $r + 1$ членов распределения $\{p_n\}$, а дробь

$U_1(s)/V(s)$ можно разложить, как и выше, на простые дроби. Итак, формула (4.9) остается верной. Далее, ограничение, что $V(s)$ имеет только простые корни, также не является необходимым. Из алгебры известно, что любая рациональная функция допускает разложение на простые дроби. Если s_k — двойной корень $V(s)$, то разложение на простые дроби (4.3) будет содержать дополнительный член вида $a/(s-s_k)^2$, что внесет в точное выражение для p_n (4.8) один член вида $a(n+1)s_k^{-(n+2)}$. Но если s_1 — простой корень, то это не повлияет на асимптотическую формулу (4.9). Сформируем этот результат в виде теоремы.

Теорема. *Если наименьший по абсолютной величине корень s_1 знаменателя рациональной функции $P(s)$ есть простой корень, то коэффициент p_n при s^n дается асимптотической формулой $p_n \sim \rho_1 s_1^{-(n+1)}$, где ρ_1 определяется по формуле* (4.5).

Аналогичная асимптотическая формула существует также и в случае, когда s_1 является кратным корнем.

Примеры. а) Пусть a_n — вероятность того, что *число успехов в n* испытаниях Бернулли *четно.* Это событие может осуществиться двумя способами: первое испытание привело к неудаче, остальные четное число раз привели к успеху или же первое испытание привело к успеху, а остальные привели к успеху нечетное число раз. Поэтому

$$a_n = qa_{n-1} + p(1-a_{n-1}), \quad a_0 = 1. \tag{4.10}$$

Умножая на s^n и складывая полученные соотношения, приходим к уравнению $A(s) - 1 = qsA(s) + ps(1-s)^{-1} - psA(s)$ для производящей функции $A(s)$. Следовательно,

$$2A(s) = \{1-s\}^{-1} + \{1-(q-p)s\}^{-1}, \quad 2a_n = 1 + (q-p)^n. \tag{4.11}$$

Заметим, что последняя формула с любой точки зрения предпочтительней очевидного ответа $a_n = b(0;\, n,\, p) + b(2;\, n,\, p) + \ldots$.

б) Пусть q_n — вероятность того, что при n бросаниях монеты не появится ни одна серия из трех последовательных выпадений герба. [Заметим, что $\{q_n\}$ не есть распределение вероятностей; если p_n — вероятность того, что первая серия из трех последовательных гербов окончится на n-м испытании, то $\{p_n\}$ есть распределение вероятностей, а q_n представляет его «хвосты»: $q_n = p_{n+1} + p_{n+2} + \ldots$.]

Легко показать, что q_n удовлетворяет рекуррентной формуле

$$q_n = \frac{1}{2} q_{n-1} + \frac{1}{4} q_{n-2} + \frac{1}{8} q_{n-3}. \tag{4.12}$$

Действительно, событие, состоящее в том, что при n испытаниях ни разу не появится последовательность *ГГГ*, может произойти лишь в том случае, если испытания начинаются с P, $ГР$ или $ГГР$. Вероятности того, что последующие испытания не приведут к серии $ГГГ$, равны соответственно q_{n-1}, q_{n-2} и q_{n-3}. Поэтому правая часть (4.12) содержит вероятности трех взаимно исключающих друг друга возможностей, при которых может осуществиться событие «ни одной серии $ГГГ$».

Очевидно, $q_0 = q_1 = q_2 = 1$, а остальные q_n могут быть последовательно вычислены по формуле (4.12). Чтобы получить производящую функцию $Q(s) = \sum q_n s^n$, умножим обе части равенства (4.12) на s^n и просуммируем по $n = 3, 4, 5, \ldots$. Получим

$$Q(s) - 1 - s - s^2 = \frac{s}{2}\{Q(s) - 1 - s\} + \frac{s^2}{4}\{Q(s) - 1\} + \frac{s^3}{8}Q(s)$$

или

$$Q(s) = \frac{2s^2 + 4s + 8}{8 - 4s - 2s^2 - s^3}. \tag{4.13}$$

Знаменатель имеет действительный корень $s_1 = 1{,}0873778\ldots$ и два комплексных корня.

При $|s| < s_1$ имеем

$$|4s + 2s^2 + s^3| < 4s_1 + 2s_1^2 + s_1^3 = 8,$$

и это же неравенство выполняется при $|s| = s_1$, за исключением случая $s = s_1$. Поэтому другие два корня превосходят s_1 по абсолютной величине. Таким образом, из (4.9) получим

$$q_n \sim \frac{1{,}236840}{(1{,}0873778)^{n+1}}, \tag{4.14}$$

где числитель равен $(2s_1^2 + 4s_1 + 8)/(4 + 4s_1 + 3s_1^3)$. Эта формула дает прекрасное приближение даже при небольших значениях n. Вместо точного значения $q_3 = 0{,}875$ она дает $0{,}8847$ и вместо $q_4 = 0{,}8125$ дает $0{,}81360$. Относительная ошибка быстро убывает, и для $q_{12} = 0{,}41626\ldots$ точное и приближенное значения совпадают в пяти знаках.

§ 5. Двойные производящие функции

Для пары целочисленных случайных величин XY, совместное распределение которых определяется системой равенств

$$P\{X = j,\ Y = k\} = P_{jk}, \qquad j, k = 0, 1, \ldots, \tag{5.1}$$

мы определим производящую функцию от двух переменных s_1 и s_2 формулой

$$P(s_1, s_2) = \sum_{j,k} p_{jk} s_1^j s_2^k. \tag{5.2}$$

Такую производящую функцию для простоты будем называть двойной.

Методы доказательств, которыми мы пользовались в первых двух параграфах, приложимы без существенных изменений, поэтому будет достаточно отметить три свойства, вытекающие из (5.2):

а) *Производящие функции распределений* $P\{X=j\}$ *и* $P\{Y=k\}$ *определяются формулами* $A(s)=P\{s, 1\}$ *и* $B(s)=P(1, s)$.

б) *Производящая функция суммы* $X+Y$ *равна* $P(s, s)$.

в) *Случайные величины независимы тогда и только тогда, когда* $P(s_1, s_2)=A(s_1)\,B(s_2)$ *для всех* s_1, s_2.

Примеры. а) *Двойное распределение Пуассона.* Очевидно, что функция

$$P(s_1, s_2)=e^{-a_1-a_2-b+a_1s_1+a_2s_2+bs_1s_2}, \quad a_i>0, \ b>0, \tag{5.3}$$

разлагается в степенной ряд с положительными коэффициентами, дающими в сумме единицу. Поэтому $P(s_1, s_2)$ представляет собой производящую функцию распределения некоторой пары случайных величин. Каждая из этих случайных величин имеет распределение Пуассона со средним значением, равным соответственно a_1+b и a_2+b, но сумма $X+Y$ имеет производящую функцию $e^{-a_1-a_2-b+(a_1+a_2)s+bs^2}$ и поэтому не распределена по Пуассону. (Это так называемое сложное распределение Пуассона; см. гл. XII, § 2.)

в) *Полиномиальное распределение.* Рассмотрим последовательность n независимых испытаний, каждое из которых приводит к одному из исходов E_0, E_1 или E_2 с вероятностями соответственно p_0, p_1 и p_2. Если X_i — случайная величина, равная числу испытаний, которые привели к E_i, то пара (X_1, X_2) имеет триномиальное распределение с производящей функцией $(p_0+p_1s_1+p_2s_2)^n$.

§ 6*). Теорема непрерывности

В гл. VI было показано, что распределение Пуассона $\{e^{-\lambda}\lambda^k/k!\}$ есть предельная форма биномиального распределения для случая, когда вероятность успеха p зависит от n, причем $np\to\lambda$ при $n\to\infty$. В этом случае $b(k; n, p)\to e^{-\lambda}\lambda^k/k!$. Производящая функция распределения $b(k; n, p)$ равна $(q+ps)^n=\{1-\lambda(1-s)/n\}^n$. Логарифмируя, убеждаемся, что эта производящая функция стремится к функции $e^{-\lambda(1-s)}$, являющейся производящей функцией распределения Пуассона. Покажем, что это свойство производящей функции сохраняется и в общем случае: последовательность распределений вероятностей сходится к предельному распределению тогда и только тогда, когда соответствующие производящие функции сходятся. К сожалению, эта теорема имеет ограниченное применение, так как

*) См. примечание на стр. 138.

наиболее интересными предельными формами дискретных распределений являются непрерывные распределения (например, нормальное распределение как предельная форма биномиального распределения).

Теорема непрерывности. *Предположим, что при любом фиксированном n последовательность $a_{0,n}, a_{1,n}, a_{2,n}, \ldots$ является распределением вероятностей, т. е.*

$$a_{k,n} \geqslant 0, \quad \sum_{k=0}^{\infty} a_{k,n} = 1. \tag{6.1}$$

Для того чтобы при любом фиксированном k выполнялось соотношение

$$a_{k,n} \to a_k \quad \text{при} \quad n \to \infty, \tag{6.2}$$

необходимо и достаточно, чтобы при любом s из интервала $0 \leqslant s < 1$ выполнялось соотношение

$$A_n(s) \to A(s), \tag{6.3}$$

где

$$A_n(s) = \sum_{k=0}^{\infty} a_{k,n} s^k, \quad A(s) = \sum_{k=0}^{\infty} a_k s^k \tag{6.4}$$

— соответствующие производящие функции.

Замечание. Из (6.2) автоматически следует, что $0 \leqslant a_k < 1$ и $\sum a_k \leqslant 1$. Поэтому производящая функция $A(s)$ определена по крайней мере при $|s| \leqslant 1$. Но предельная последовательность $\{a_k\}$ не обязательно является распределением вероятностей; например, если первые n членов распределения $\{a_{k,n}\}$ равны нулю, то предельная последовательность $\{a_k\}$ состоит из одних нулей. Для того чтобы последовательность $\{a_k\}$ была распределением вероятностей, необходимо и достаточно выполнение условия $\sum a_k = 1$ или условия $A(1) = 1$.

Доказательство[1]). Допустим сначала, что (6.2) выполнено. Для фиксированного s $(0 < s < 1)$ и фиксированного ε можно выбрать r так, чтобы выполнялось неравенство $s^r/(1 - s) < \varepsilon$. Тогда

$$|A_n(s) - A(s)| \leqslant \sum_{k=0}^{r} |a_{k,n} - a_k| s^k + 2\varepsilon. \tag{6.5}$$

Сумма в правой части содержит конечное число членов, каждый из которых стремится к нулю. Поэтому $|A_n(s) - A(s)|$ сколь

---

[1]) Эта теорема есть частный случай теоремы непрерывности для преобразований Лапласа — Стильтьеса и доказывается так же, как и эта более общая теорема непрерывности. В литературе теорема непрерывности часто формулируется и доказывается при излишних ограничениях.

угодно мало при достаточно большом n. Теперь допустим, что выполнено (6.3). Воспользуемся тем хорошо известным фактом[1]), что из данной последовательности распределений всегда можно выбрать сходящуюся подпоследовательность $\{a_{k,n}\}$. Если бы (6.2) не было верно, то можно было бы выбрать две подпоследовательности, сходящиеся к различным предельным последовательностям $\{a_k^*\}$ и $\{a_k^{**}\}$. Соответствующие подпоследовательности последовательности $\{A_n(s)\}$ сходились бы соответственно к $A^*(s) = \sum a_k^* s^k$ и $A^{**}(s) = \sum a_k^{**} s^k$. Это, однако, невозможно, так как мы допустили, что (6.3) выполнено. Следовательно, из (6.3) следует, что выполняется (6.2).

Примеры. а) *Отрицательное биномиальное распределение.* Мы видели в примере 2, г, что производящая функция распределения $\{f(k;\ r,\ p)\}$ имеет вид $p^r(1-qs)^{-r}$.

Пусть λ фиксировано и пусть $p \to 1$, $q \to 0$, причем $q = \lambda/r$ $(r \to \infty)$. Тогда

$$\left(\frac{p}{1-qs}\right)^r = \left(\frac{1-\lambda/r}{1-\lambda s/r}\right)^r. \tag{6.6}$$

Логарифмируя, убеждаемся, что правая часть равенства (6.6) стремится к функции $e^{-\lambda+\lambda s}$, являющейся производящей функцией распределения Пуассона $\{e^{-\lambda}\lambda^k/k!\}$. Поэтому, если $r \to \infty$ и $rq \to \lambda$,

$$f(k;\ r,\ p) \to e^{-\lambda}\frac{\lambda^k}{k!}. \tag{6.7}$$

б) *Испытания Бернулли с переменными вероятностями.* Рассмотрим n независимых испытаний, таких, что k-е испытание оканчивается успехом с вероятностью p_k и неудачей с вероятностью $q_k = 1 - p_k$. Число успехов S_n при n испытаниях можно записать в виде суммы $S_n = X_1 + X_2 + \ldots + X_n$ n взаимно независимых случайных величин X_k, имеющих распределение $P\{X_k = 0\} = q_k$; $P\{X_k = 1\} = p_k$. Производящая функция величины X_k равна $q_k + p_k s$, так что производящая функция величины S_n равна

$$P(s) = (q_1 + p_1 s)(q_2 + p_2 s) \ldots (q_n + p_n s). \tag{6.8}$$

В качестве приложения этой схемы допустим, что в k-м доме некоторого города в течение дня может возникнуть пожар с небольшой вероятностью p_k. Тогда сумма $p_1 + \ldots + p_n$, где n — число домов в городе, будет ожидаемым числом пожаров. В гл. VI мы видели, что если все p_k равны между собой и воз-

---

[1]) Это легко доказать с помощью указанного Г. Кантором «диагонального процесса», излагаемого во всех книгах по теории множеств. Указанное утверждение, между прочим, есть частный случай известной теоремы Хелли.

никновения пожара в различных домах суть независимые события, то число пожаров есть случайная величина, имеющая распределение, близкое к распределению Пуассона. Сейчас мы покажем, что это заключение остается в силе также и при более реальном предположении, что вероятности p_k не равны между собой. Этот результат укрепит нашу уверенность в том, что распределение Пуассона позволяет осуществить адекватное описание явлений, состоящих в суммарном действии многих маловероятных событий («успехов»). Примерами служат несчастные случаи и телефонные вызовы.

Воспользуемся уже знакомой нам схемой с n величинами, причем вероятности p_k меняются вместе с n так, что наибольшая из вероятностей p_k стремится к нулю, а сумма $p_1+p_2+\ldots+p_n=\lambda$ остается постоянной. Тогда, согласно (6.8),

$$\log P(s)=\sum_{k=1}^{n}\log\{1-p_k(1-s)\}. \tag{6.9}$$

Так как $p_k\to 0$, то можно воспользоваться тем, что $\log(1-x)=-x-\theta x$, где $\theta\to 0$ при $x\to 0$. Отсюда следует, что

$$\log P(s)=-(1-s)\left\{\sum_{k=1}^{n}(p_k+\theta_k p_k)\right\}\to-\lambda(1-s), \tag{6.10}$$

так что $P(s)$ стремится к производящей функции распределения Пуассона. Следовательно, S_n *имеет в пределе распределение Пуассона*. Мы заключаем, что при большом n и не очень большом $\lambda=p_1+p_2+\ldots+p_n$ распределение величины S_n может быть приближенно представлено распределением Пуассона. (См. также пример 5, б гл. IX.)

§ 7. Задачи

1. Пусть X — случайная величина с производящей функцией $P(s)$. Найти производящие функции величин $X+1$ и $2X$.

2. Продолжение. Найти производящие функции последовательностей: а) $P\{X\leqslant n\}$, б) $P\{X<n\}$, в) $P\{X\geqslant n\}$, г) $P\{X>n+1\}$, д) $P\{X=2n\}$.

3. Рассмотрим последовательность испытаний Бернулли. Пусть u_n — вероятность того, что первая комбинация *УН* появится при $(n-1)$-м и n-м испытаниях. Найти производящую функцию, среднее значение и дисперсию.

4. Выяснить, какие из формул § 12 гл. II представляют композиции и где могли быть использованы производящие функции.

5. Пусть a_n — число способов, которыми можно получить сумму в n очков при произвольном числе бросаний правильной кости. Показать, что производящая функция последовательности $\{a_n\}$ равна $\{1-s-s^2-s^3-s^4-s^5-s^6\}^{-1}$.

Замечание. *В задачах* 6—10, *относящихся к игре с бросанием монеты, используются обычные обозначения. Среди прочих результатов они содержат прямой вывод некоторых соотношений из гл.* III. *Мы пишем* $u_n=P\{S_n=0\}$ *и* $f_n=P\{S_1\neq 0, S_2\neq 0\ \ldots\ S_{n-1}\neq 0,\ S_n=0\}$

(*первое возвращение*); *по определению*, $u_0 = 1$, $f_0 = 0$. *Производящая функция* $F(s) = 1 - \{1 - s^2\}^{1/2}$ *последовательности* $\{f_n\}$ *считается известной* (*из гл.* III). *Все вычисления практически тривиальны*; *точная формула для коэффициентов не нужна.*

6. Производящая функция последовательности $\{u_n\}$ равна $U(s) = \{1 - s^2\}^{-1/2}$.

7. Вероятность того, что до времени $2n$ не появится нуль, совпадает с вероятностью того, что $S_{2n} = 0$.

8. Вероятность того, что $S_{2n} = 0$ и все суммы $S_1, S_2, \ldots, S_{2n} = 0$ неотрицательны, равна $2f_{2n+2}$.

9. Вероятность того, что первому изменению знака предшествует $2n$ испытаний, равна $2f_{2n+2}$.

10. Последовательность вероятностей того, что ровно k из n сумм $S_1, \ldots, S_n$ равны нулю, имеет производящую функцию $F^{(k)}(s)\, U(s)\,(1+s)$.

11. В последовательности испытаний Бернулли с $p > q$ пусть a_n — вероятность того, что существует индекс $j > n$, для которого $S_j = 0$. Показать, что производящая функция последовательности $\{a_n\}$ имеет вид $4pq\,[p - q + (1 - 4pqs^2)^{1/2}]^{-1}\,(1+s)$.

12. Найти производящую функцию величины S_r (r фиксировано) *в задаче о времени ожидания* (3, г) гл. IX. Проверить формулу (3.3) гл. IX, для среднего значения вычислить дисперсию.

13. Продолжение (другой метод решения той же задачи). Пусть $p_n(r) = P\{S_r = n\}$. Доказать рекуррентную формулу

$$p_{n+1}(r) = \frac{r-1}{N} p_n(r) + \frac{N - r + 1}{N} p_n(r-1). \tag{7.1}$$

Вывести производящую функцию прямо из (7.1).

14. Решить две предыдущие задачи для r заранее выбранных элементов (вместо r произвольных).

15 [1]). Назовем *циклом* последовательность испытаний Бернулли до первой неудачи включительно. Найти производящую функцию и распределение вероятностей общего числа S_r успехов в r циклах.

16. Продолжение. Пусть R — число последовательных циклов до ν-го успеха (т. е. ν-й успех принадлежит R-му циклу). Доказать, что $P(R = r) = p^\nu q^{r-1} \binom{r + \nu - 2}{\nu - 1}$. Найти $E(R)$ и $D(R)$.

17. Продолжение. Рассмотрим *две* последовательности испытаний Бернулли с вероятностями соответственно p_1, q_1 и p_2, q_2. Показать, что

---

[1]) Задачи 15—17 имеют отношение к игре в биллиард. Вероятность попадания p служит мерой искусства игрока. Игрок продолжает игру до первого промаха, так что число собранных им шаров есть длина «цикла попаданий». Игра заканчивается, когда один из игроков наберет N шаров. Таким образом, в задаче 15 ищется распределение вероятностей для числа шаров, собираемых игроком в течение r циклов, в задаче 16 — распределение вероятностей для числа циклов, в течение которых игрок собирает ν шаров, и среднее число циклов (средняя продолжительность игры), в задаче 17 — вероятность ничьей. Дальнейшие подробности см. в статье O. Bottema and S. C. Van Veen, Kansberekningen by het biljartspel, *Nieuw Archief voor Wiskunde*, 22 (1943), 16—33 и 123—158.

вероятность совпадения номеров цикла, при которых произойдет N-й успех в каждой последовательности, можно представить в любой из форм

$$(p_1p_2)^N \sum_{\nu=1}^{\infty} \binom{N+\nu-2}{\nu-1}^2 (q_1q_2)^{\nu-1} = (p_1p_2)^N (1-q_1q_2)^{1-2N} \sum_{k=0}^{N-1} \binom{N-1}{k}^2 (q_1q_2)^k.$$

18. Пусть $\{X_k\}$ — взаимно независимые величины, принимающие каждая значения 0, 1, 2, ..., $a-1$ с вероятностями $1/a$. Пусть $S_n = X_1 + \ldots + X_n$. Показать, что производящая функция величины S_n равна

$$P(s) = \left\{ \frac{1-s^a}{a(1-s)} \right\}^n,$$

откуда

$$P\{S_n = j\} = \frac{1}{a^n} \sum_{\nu=0}^{\infty} (-1)^{\nu+j+a\nu} \binom{n}{\nu} \binom{-n}{j-a\nu}.$$

(Лишь конечное число членов суммы отлично от нуля.)

Замечание. При $a=6$ мы получим вероятность выпадения $j+n$ очков при бросании n костей. Решение этой задачи восходит к Муавру.

19. Продолжение. Вероятность $P(S_n \leqslant j)$ имеет производящую функцию $P(s)/(1-s)$, откуда

$$P\{S_n \leqslant j\} = \frac{1}{a^n} \sum_{\nu} (-1)^{\nu} \binom{n}{\nu} \binom{j-a\nu}{n}.$$

20. Продолжение. Предельная форма. Если $a \to \infty$ и $j \to \infty$, причем $j/a \to x$, то

$$P\{S_n \leqslant j\} \to \frac{1}{n!} \sum_{\nu} (-1)^{\nu} \binom{n}{\nu} (x-\nu)^n,$$

где сумма берется по всем ν из интервала $0 \leqslant \nu < x$.

Замечание. Этот результат принадлежит Лагранжу. В теории геометрических вероятностей правая часть представляет функцию распределения суммы n независимых случайных величин с «равномерным» распределением на интервале (0, 1).

21. Пусть u_n — вероятность того, что число успехов в испытаниях Бернулли делится на 3. Найти рекуррентное соотношение для u_n, а из него — производящую функцию.

22. Продолжение. (Другой метод.) Пусть v_n и w_n — вероятности того, что S_n имеет соответственно вид $3v+1$ и $3v+2$ (так что $u_n + v_n + w_n = 1$). Найти три выполняющихся одновременно рекуррентных соотношения, а из них — три уравнения для производящих функций.

23. Пусть X и Y — независимые случайные величины с производящими функциями $U(s)$ и $V(s)$. Показать, что $P\{X - Y = j\}$ равно коэффициенту при s^j в разложении функции $U(s)V(1/s)$ $(j = 0, \pm 1, \pm 2, \ldots)$.

24. Производящие функции моментов. Пусть X — случайная величина с производящей функцией $P(s)$ и пусть $\sum p_n s^n$ сходится при некотором $s_0 > 1$. Тогда все моменты $m_r = E(X^r)$ существуют, и про-

изводящая функция $F(s)$ последовательности $m_r/r!$ сходится по крайней мере при $|s| < \log s_0$. Кроме того,

$$F(s) = \sum_{r=0}^{\infty} \frac{m_r}{r!} s^r = P(e^s).$$

Замечание. $F(s)$ обычно называют *производящей функцией моментов*, хотя на самом деле она производит последовательность $m_r/r!$

25. Предположим, что $A(s) = \sum a_n s^n$ — рациональная функция $U(s)/V(s)$ и что s_1 — наименьший по абсолютной величине корень многочлена $V(s)$. Показать, что если s_1 имеет кратность r, то

$$a_n \sim \frac{\rho_1}{s_1^{n+r}} \binom{n+r-1}{r-1},$$

где

$$\rho_1 = -r!\, U(s_1)/V^{(r)}(s_1).$$

26. *Двойное отрицательное биномиальное распределение.* Показать, что при положительных значениях параметров выражение $p_0^a\{1-p_1s_1-p_2s_2\}^{-a}$ представляет производящую функцию распределения пары случайных величин $(\boldsymbol{X}, \boldsymbol{Y})$, таких, что частные распределения $\boldsymbol{X}$, $\boldsymbol{Y}$ и $\boldsymbol{X}+\boldsymbol{Y}$ являются отрицательными биномиальными распределениями [1]).

[1]) Распределение этого типа было использовано G. E. Bates и J. Neyman при исследованиях частоты катастроф. См. University of California Publications in Statistics, 1, 1952.

ГЛАВА XII [1])

СЛОЖНЫЕ РАСПРЕДЕЛЕНИЯ. ВЕТВЯЩИЕСЯ ПРОЦЕССЫ

§ 1. Суммы случайного числа величин

Пусть $\{X_k\}$ — последовательность взаимно независимых случайных величин с одинаковыми распределениями $P\{X_k = j\} = f_j$. Производящую функцию распределения $\{f_j\}$ обозначим через $f(s) = \sum f_i s^i$. Часто рассматривают суммы $S_N = X_1 + X_2 + \ldots + X_N$, где число членов N — случайная величина, не зависящая от X_j. Пусть $P\{N = n\} = g_n$ — распределение N, а $g(s) = \sum g_n s^n$ — соответствующая производящая функция. Распределение $\{h_j\}$ случайной величины S_N получается с помощью формулы полной вероятности

$$h_j = P\{S_N = j\} = \sum_{n=0}^{\infty} P\{N = n\} P\{X_1 + \ldots + X_n = j\}. \quad (1.1)$$

Если N принимает лишь конечное число значений, то сумма S_N определена в пространстве конечного числа X_k. В противном случае вероятностное определение S_N как суммы требует рассмотрения пространства бесконечных последовательностей $\{X_k\}$. Но мы будем иметь дело только с распределением S_N; поэтому для наших целей достаточно принять формулу (1.1) за определение случайной величины S_N в пространстве, состоящем из точек 0, 1, 2, ...

При фиксированном n распределение $X_1 + X_2 + \ldots + X_n$ есть n-кратная композиция $\{f_j\}$, и поэтому (1.1) можно переписать в следующем компактном виде:

$$\{h_j\} = \sum_{n=0}^{\infty} g_n \{f_j\}^{n*}. \quad (1.2)$$

Эту формулу можно упростить, используя производящие функции. Производящая функция распределения $\{f_j\}^{n*}$ равна $f^n(s)$, и из (1.2) следует, что производящая функция суммы S_N имеет вид

$$h(s) = \sum_{j=0}^{\infty} h_j s^j = \sum_{n=0}^{\infty} g_n f^n(s). \quad (1.3)$$

[1]) Материал этой главы в дальнейшем использоваться не будет.

Правая часть представляет ряд Тейлора для $g(s)$ с заменой s на $f(s)$; поэтому он сводится к $g(f(s))$. Итак, доказана

Теорема. *Производящая функция суммы* $S_N = X_1 + X_2 + \ldots$ $\ldots + X_n$ *есть сложная функция* $g(f(s))$.

Особый интерес представляют два следующих частных случая.

а) Если величины X_i принимают только значения 1 и 0 с вероятностями p и q, то $f(s) = q + ps$ и, значит, $h(s) = g(q + ps)$.

б) Если N имеет распределение Пуассона с параметром t, то

$$h(s) = e^{-t+tf(s)}. \tag{1.4}$$

Распределение с такой производящей функцией называется *сложным распределением Пуассона*.

Если X_i имеют распределение Бернулли, т. е. $P\{X_j = 1\} = p$ и $P\{X_j = 0\} = q$, а N имеет распределение Пуассона, то $h(s) = e^{-tp+tps}$. *Сумма* S_N *имеет распределение Пуассона с параметром* tp.

Примеры. а) Мы видели в примере 7 в гл. VI, что под действием рентгеновских лучей хромосомы в клетках могут расщепляться. Если доза и время облучения фиксированы, то число N расщеплений в заданной группе клеток подчиняется распределению Пуассона. Каждое расщепление может с вероятностью p привести к гибели клетки, а с вероятностью $q = 1 - p$ клетка выживает. Здесь S_N — число *наблюденных* расщеплений[1]). Эта случайная величина имеет распределение Пуассона с параметром tp.

б) В экспериментах[2]) по определению численности вида животных пусть g_n вероятность того, что имеется n животных данного вида. Если каждое животное может быть поймано с фиксированной вероятностью p, то (при условии независимости) число пойманных животных данного вида S_N есть случайная величина с производящей функцией $g(q + ps)$. К данной схеме сводятся много других задач. Например, пусть какое-то насекомое кладет n яиц с вероятностью g_n, и каждое яйцо развивается в личинку с вероятностью p. Тогда S_N — число яиц, развившихся в личинки. Аналогично, если g_n — вероятность того, что в семье имеется n детей, и если вероятность того, что каждый ребенок мальчик равна p, а девочка — q $(p + q = 1)$, то S_N — число мальчиков в семье.

в) Каждое растение имеет большое количество семян, однако для отдельного семени вероятность сохраниться и развиться очень мала, так что разумно предположить, что число потомков отдельного

---

[1]) См. D. G. Catcheside, Genetic effects of radiations, *Advances in Genetics* опубликовано M. Demerec, т. 2. Academic Press, New York, 1948, 271—358, в частности 339.

[2]) D. G. Kendall, On some modes of population growth leading to R. A. Fisher's logarithmic series distribution, *Biometrika*, 35 (1948), 6—15.

растения имеет распределение Пуассона. Если $\{g_n\}$ — распределение числа растений-родителей, то $g(e^{-\lambda+\lambda s})$ — производящая функция числа потомков.

§ 2. Сложное распределение Пуассона

Рассмотрим вначале два типичных примера.

Примеры. а) Предположим, что число ударов молнии за время t имеет распределение Пуассона со средним значением λt. Если $\{f_n\}$ вероятностное распределение ущерба, причиненного фиксированным ударом, то (предполагая независимость) вероятностное распределение полного ущерба за время t является сложным распределением Пуассона $\{h_j\} = e^{-\lambda t} \sum \frac{(\lambda t)^n}{n!} \{f_j\}^{n*}$ с производящей функцией

$$h(s, t) = e^{-\lambda t + \lambda t f(s)}. \tag{2.1}$$

б) В экологии обычно предполагают, что число выводков животных, обитающих на некотором участке земли, распределено по Пуассону со средним значением, пропорциональным площади участка. Если $\{f_k\}$ — распределение числа животных в выводке, то формула (2.1) дает производящую функцию полного числа животных, обитающих на данном участке земли.

В гл. VI было показано, что многие явления, зависящие от интервала времени или области пространства, подчиняются распределению Пуассона, а предшествующие примеры разъясняют, почему сложное распределение Пуассона также часто связано с такого рода явлениями.

Производящая функция (2.1) обладает замечательным свойством

$$h(s;\ t_1 + t_2) = h(s;\ t_1)\, h(s;\ t_2). \tag{2.2}$$

Наглядно это можно пояснить следующим образом. С каждым интервалом времени продолжительности t связана случайная величина с производящей функцией $h(s;\ t)$, которую мы будем называть *вкладом* этого интервала. Вклады двух непересекающихся интервалов независимы, а это означает, что деление интервала времени $t = t_1 + t_2$ на две части приводит к разложению $X(t) = X(t_1) + X(t_2)$ соответствующего вклада на сумму двух независимых случайных величин.

В следующем параграфе будет показано, что (в классе целочисленных случайных величин) *только сложное распределение Пуассона* обладает этим свойством. Прежде чем формулировать эту теорему, приведем два примера.

Примеры. в) *Отрицательное биномиальное распределение* с производящей функцией

$$h(t;\ s) = \left(\frac{p}{1 - qs}\right)^t, \qquad p + q = 1, \tag{2.3}$$

обладает свойством (2.2). Следовательно, *отрицательное биномиальное распределение является сложным распределением Пуассона*. Производящая функция (2.3) принимает форму (2.1), если положить

$$\lambda = \log\frac{1}{p}, \quad f(s) = \frac{1}{\lambda}\log\frac{1}{1-qs}, \quad f_n = \frac{\lambda q^n}{n}. \tag{2.4}$$

Распределение $\{\lambda q^n/n\}$ называется логарифмическим распределением.

г) *Кратное распределение Пуассона*. Допустим, что мы классифицируем автомобильные катастрофы в зависимости от числа пострадавших машин на одиночные, двойные и т. д. Предположим, далее, что число одиночных, двойных и т. д. катастроф имеет распределение Пуассона со средними значениями $\lambda_1 t$, $\lambda_2 t$, ... и что эти случайные величины независимы. Тогда общее число пострадавших машин за время t имеет производящую функцию

$$e^{-\lambda_1 t(1-s)} e^{-\lambda_2 t(1-s^2)} e^{-\lambda_3 t(1-s^3)} \ldots . \tag{2.5}$$

Это опять сложное распределение Пуассона, для которого $\lambda = \sum \lambda_i$ и $f_i = \lambda_i/\lambda$. Обратно, производящую функцию любого сложного распределения Пуассона можно записать в форме (2.5) и истолковать его как общий эффект одиночных, двойных и т. д. аварий.

§ 3. Безгранично делимые законы

Вероятностное распределение $\{h_i\}$, $i=0, 1, 2, \ldots$ *называется безгранично делимым, если для любого n оно может быть представлено как n-кратная композиция вероятностного распределения* $\{\varphi_i\}$ *с самим собой, т. е. если его производящая функция* $h(s)$ *обладает тем свойством, что* $h^{1/n}(s) = \varphi(s)$, *где* $\varphi(s)$ — *производящая функция распределения* $\{\varphi_i\}$.

Заметим, что *если* $h(s; t)$ *удовлетворяет условию* (2.2), *то* $h(s; t) = h^n(s; t/n)$ *и значит соответствующее распределение безгранично делимо*. Утверждение предыдущего параграфа заключается в следующей теореме (которая является частным случаем важной общей теоремы Леви, относящейся к произвольным вероятностным распределениям).

Теорема. *Если* $\{h_i\}$ — *безгранично делимое распределение, то его производящая функция может быть записана в форме* (2.1) (*скажем, при* $t=1$).

(Заметим, что $h^t(s) = h(s; t)$ удовлетворяет (2.2).)

Доказательство. Предположим, что $h^{1/n}(s)$ является производящей функцией некоторого вероятностного распределения при любом n. Это возможно только при условии, что $h(0) = h_0 > 0$.

Значит, функция $h(s)$ должна быть положительна в некотором интервале $|s| \leqslant a \leqslant 1$, и в нем же $0 < 1 - h(s) < 1$. Поэтому $\log h(s) = \log(1 - \{1 - h(s)\})$ разлагается в ряд Тейлора

$$\log h(s) = \sum_{i=0}^{\infty} \chi_i s^i, \qquad -a < s < a. \tag{3.1}$$

Положив $s = 0$, убеждаемся, что $\chi_0 < 0$. Мы хотим показать, что все остальные χ_i неотрицательны. Допустим противное, и пусть $r \geqslant 1$ — *наименьший* индекс, для которого $\chi_r < 0$. Для того чтобы избежать слишком длинных формул, положим

$$A(s) = \sum_{\nu=1}^{r-1} \chi_\nu s^\nu, \quad B(s) = \sum_{\nu=r+1}^{\infty} \chi_\nu s^\nu, \quad \frac{1}{n} = \varepsilon, \tag{3.2}$$

так что

$$h^{1/n}(s) = e^{\varepsilon\chi_0} \cdot e^{\varepsilon A(s)} \cdot e^{\varepsilon\chi_r s^r} \cdot e^{\varepsilon B(s)}. \tag{3.3}$$

По предположению, $h^{1/n}(s) = \sum \varphi_k s^k$, где $\varphi_k \geqslant 0$. Рассмотрим, в частности, коэффициент φ_r при s^r. Степенной ряд $B(s)$ содержит только члены со степенями, большими, чем r, и поэтому не влияет на φ_r. Поэтому φ_r есть коэффициент при s^r в выражении

$$e^{\varepsilon\chi_0}\left(1 + \varepsilon A(s) + \frac{1}{2}\varepsilon^2 A^2(s) + \ldots\right)(1 + \varepsilon\chi_r s^r). \tag{3.4}$$

Так как $A(s)$ есть полином степени не выше, чем $r - 1$, то легко установить, что

$$\varphi_r = e^{\varepsilon\chi_0}[\varepsilon\chi_r + \varepsilon^2 p(\varepsilon)], \tag{3.5}$$

где $p(\varepsilon)$ — полином от ε. Если $\chi_r < 0$, то правая часть соотношения (3.5) будет отрицательна при ε достаточно малом и, значит, $\varphi_r < 0$, что невозможно. Этим доказано, что при $r = 1, 2, \ldots, \chi_r \geqslant 0$. Более того, $h(1) = 1$ и, значит, $\log h(1) = \sum \chi_\nu = 0$, т. е. $-\chi_0 = \chi_1 + \chi_2 + \ldots$. Для того чтобы переписать $h(s)$ в форме (2.1), остается положить $-\chi_0 = \lambda$ и $f_i = \chi_i/\lambda$.

§ 4. Примеры ветвящихся процессов

Рассмотрим теперь один случайный процесс, служащий упрощенной моделью многих реальных процессов и дающий вместе с тем прекрасную иллюстрацию применения производящих функций. Этот процесс описывается следующим образом.

Рассмотрим частицы, способные производить новые такие же частицы. Некоторая частица образует начальное или нулевое поколение. Каждая частица с вероятностью p_k ($k = 0, 1, 2, \ldots$) порождает k новых частиц; непосредственные потомки n-го поколения образуют $(n+1)$-е поколение. Частицы каждого поколения размножаются независимо друг от друга.

Прежде чем дать строгие формулировки в терминах случайных величин, приведем несколько типичных примеров.

Примеры. а) *Ядерные цепные реакции.* Это приложение приобрело известность в связи с осуществлением расщепления ядра[1]). Частицами являются нейтроны, подверженные случайным столкновениям с другими частицами. При столкновении каждая частица порождает фиксированное число m непосредственных потомков. Пусть p — вероятность того, что частица рано или поздно придет в столкновение с другой частицей; тогда $q=1-p$ есть вероятность того, что частица останется бездейственной (удалится или поглотится каким-то другим путем). В этом случае возможны лишь числа потомков 0 и m и соответствующие вероятности равны q и p (т. е. $p_0=q$, $p_m=p$, $p_j=0$ при всех остальных j). В худшем случае первая частица останется бездейственной и процесс никогда не начнется. В лучшем случае будет m частиц первого поколения, m^2 частиц второго и т. д. Если p близко к единице, то весьма вероятно, что число частиц будет очень быстро возрастать. Теоретически это число может расти неограниченно. С физической точки зрения вероятности столкновений не могут оставаться неизменными при очень большом числе частиц и поэтому независимость нарушится. Однако для обычных цепных реакций теоретическое утверждение «неограниченно возрастающее число частиц» может быть переведено как «взрыв».

б) *Выживание фамилии.* Здесь, как часто в жизни, учитываются лишь потомки мужского пола; они играют роль частиц, а p_k есть вероятность того, что новорожденный мальчик произведет в течение своей жизни k потомков мужского пола. Наша схема вводит два искусственных упрощения. Рождаемость подвержена вековым колебаниям, и поэтому в действительности распределение $\{p_k\}$ меняется от поколения к поколению. Кроме того, одинаковая наследственность и одинаковая окружающая среда ведут к сходству братьев, что противоречит предположению о независимости. Наша модель может быть усовершенствована с учетом этих возражений, но существенные черты останутся теми же. Мы найдем вероятность того, что в n-м поколении данную фамилию будут иметь k мужчин; в частности, определим вероятность вымирания фамилии. Выживание фамилии — первая цепная реакция, изученная методами теории вероятностей. Впервые задача была рассмотрена Ф. Гальтоном (1889); за детальным расчетом отсылаем читателя к книге А. Лотка[2]). Лотка показывает, что данным по Америке довольно хорошо соответствует распределе-

---

[1]) Заимствовано из статьи E. Schroedinger, Probability problems in nuclear chemistry, *Proceedings of the Royal Irish Academy*, 51, sect. A, № 1 (Dec. 1945).

[2]) A. Lotka, Théorie annalytique des associations biologiques, т. 2. *Actualités scientifiques et industrielles*, № 780 (1939), 123—136, Paris.

ние $p_0 = 0{,}4825$, $p_k = 0{,}2126 \cdot 0{,}5893^{k-1}$ ($k \geqslant 1$), которое, за исключением первого члена, является геометрическим распределением.

в) *Гены и мутации.* Каждый ген данного организма (см. § 5 гл. V) может появиться у 1, 2, 3, ... прямых потомков. Наша схема описывает этот процесс, пренебрегая, конечно, изменениями внутри популяции и во времени. Эта схема в частности, находит применение при изучении мутации[1]) или изменений вида генов. При самопроизвольной мутации возникает единственный ген нового вида, который играет роль производящей частицы нулевого поколения. Теория позволяет оценивать вероятности выживания и распространения нового гена. Рассмотрим для определенности (следуя Фишеру) растение кукурузы, являющееся отцовским растением для каких-нибудь 100 семян и материнским для такого же числа семян. Если объем популяции остается постоянным, то в среднем два из этих двухсот семян разовьются в новые растения. Каждое из семян с вероятностью $\frac{1}{2}$ получает данный ген. Вероятность того, что мутировавший ген будет представлен ровно в k растениях первого поколения, совпадает поэтому с вероятностью k успехов в последовательности 200 испытаний Бернулли с вероятностью успеха $p = \frac{1}{200}$. Разумно считать, что распределение $\{p_k\}$ есть просто распределение Пуассона с параметром 1. Если гены нового вида обеспечивают определенные биологические преимущества, то приходим к распределению Пуассона с $\lambda > 1$.

г) *Очереди*[2]). Теория ветвящихся процессов полезна при изучении случайных колебаний очередей (на почте, телефонных станциях и т. п.). Покупатель, подошедший к свободному прилавку и не затративший времени на ожидание обслуживания, играет роль «основателя рода»; покупатели, подошедшие пока его обслуживают и образующие очередь, суть его прямые потомки. Процесс продолжается до тех пор, пока не рассосется очередь. Нас интересует развитие процесса до этого момента.

§ 5. Вероятности вырождения в ветвящихся процессах

Для математического описания цепной реакции введем случайную величину X_n, представляющую число частиц в n-м поколении. Тогда $X_0 = 1$, а X_1 имеет заданное распределение $\{p_k\}$. Число непосредственных потомков каждой из X_1 частиц первого поколения является

---

[1]) R. A., Fisher, The genetical theory of natural selection, Oxford, 1930, стр. 73 и далее.

[2]) D. G. Kendall, Stochastic processes and population growth, *Journal Royal Statistics Society*, 11 (1949), 230—265.

случайной величиной с тем же самым распределением. Поэтому $X_2 = U_1 + U_2 + \ldots + U_{X_1}$, где каждое U_j имеет распределение $\{p_k\}$. По предположению, эти величины U_k взаимно независимы. Таким образом,

$$P(s) = \sum_{k=0}^{\infty} p_k s^k$$

есть общая производящая функция величин U_j и X_1. Следовательно, по теореме предыдущего параграфа производящая функция величины X_2 равна $P_2(s) = P(P(s))$. Таким же образом X_3 частиц третьего поколения суть внуки X_1 частиц первого поколения, так что X_3 есть сумма X_1 взаимно независимых величин, имеющих каждая то же распределение, что и X_2. Это значит, что производящая функция величины X_3 равна $P_3(s) = P(P_2(s))$. Аналогичные рассуждения показывают, что вообще *производящая функция $P_{n+1}(s)$ числа частиц X_{n+1} в $(n+1)$-м поколении определяется рекуррентным соотношением*

$$P_1(s) = P(s), \quad P_{n+1}(s) = P(P_n(s)). \tag{5.1}$$

В примере а) $P(s) = q + ps^m$, так что $P_2(s) = q + p(q + ps^m)^m$, $P_3(s) = q + p\{q + p(q + ps^m)^m\}^m$ и т. д. Для распределения Пуассона $P(s) = e^{-\lambda(1-s)}$, $P_2(s) = e^{-\lambda + \lambda e^{-\lambda + \lambda s}}$ и т. д. Эти формулы громоздки, но они позволят нам вывести важные общие заключения.

Прежде всего найдем вероятность x_n того, что процесс оборвется на n-м поколении или ранее. Это значит, что $X_n = 0$, откуда $x_n = P_n(0)$. Из определения ясно, что x_n может только возрастать с ростом n. Действительно, $x_1 = P(0) = p_0$ и $x_{n+1} = P(x_n)$. Далее, $P(s)$—возрастающая функция, следовательно, $x_2 = P(x_1) > P(0) = x_1$; по индукции $x_{n+1} = P(x_n) > P(x_{n-1}) = x_n$. Отсюда следует, что $x_n \to \zeta$, где ζ удовлетворяет уравнению

$$\zeta = P(\zeta). \tag{5.2}$$

Мы утверждаем, что ζ — *наименьший* корень уравнения (5.2). Действительно, если η — какой-нибудь другой корень, то, поскольку $P(s)$ — возрастающая функция, $x_1 = P(0) < P(\eta) = \eta$; по индукции, если $x_n < \eta$, то также $x_{n+1} = P(x_n) < P(\eta) = \eta$, так что $\zeta \leqslant \eta$; этим доказано, что уравнение (5.2) не может иметь корня, меньшего ζ.

График функции $y = P(s)$ — выпуклая вниз кривая; биссектриса $y = s$ пересекает эту кривую не более чем в двух точках. Одной такой точкой всегда является точка $(1, 1)$, и поэтому уравнение (5.3) имеет не более одного корня ζ, удовлетворяющего неравенству $0 < \zeta < 1$. Если такой корень существует, отношение $\frac{1 - P(\zeta)}{1 - \zeta}$ равно единице, и по теореме о среднем существует точка x, заключенная между ζ и 1, такая, что производная $P'(x)=1$. Отсюда вытекает, что уравнение (5.2) имеет корень $\zeta<1$ только тогда, когда $P'(1)>1$. С другой стороны, если $P'(1)\leqslant 1$, то $\frac{1-P(s)}{1-s}<1$ при всех $s<1$. Это означает, что $P(s)>s$; график функции $P(s)$ лежит выше биссектрисы и, следовательно, уравнение (5.2) не имеет корней. Итак, корень $\zeta<1$ уравнения $\zeta=P(\zeta)$ существует тогда и только тогда, когда $P'(1)>1$; этот корень единствен. Заметим теперь, что $P'(1)=\sum kp_k$ есть математическое ожидание числа непосредственных потомков частицы. Мы можем, следовательно, сформулировать основной вывод:

Пусть $\mu=\sum kp_k$ — математическое ожидание числа непосредственных потомков одной частицы. Если $\mu\leqslant 1$, то вероятность того, что процесс оборвется до n-го поколения (т. е. $X_n=0$), стремится к единице с ростом n. Если $\mu>1$, то существует единственный корень $\zeta<1$ уравнения (6.3), *а ζ есть предел вероятности того, что процесс оборвется после конечного числа поколений.*

Разность $1-\zeta$ можно назвать вероятностью бесконечного продолжения процесса. Обычно x_n быстро сходится к ζ, так что если процесс окончился, то весьма вероятно, что это произошло в одном из первых поколений. Поэтому практически ζ является вероятностью быстрого замирания процесса. В примере 4,в мы можем назвать $1-\zeta$ вероятностью выживания нового типа генов. Если вначале вместо одной имеется r частиц, то вероятность вымирания всех r частиц равна ζ^r, а вероятность сохранения потомства хотя бы одной из частиц равна $1-\zeta^r$. Если начальное число r велико, $1-\zeta^r$ близко к единице даже при относительно большом ζ. В ядерных цепных реакциях примера а) это всегда имеет место, и поэтому можно сказать, что при $\mu>1$ вероятность взрыва близка к единице, тогда как при $\mu\leqslant 1$ с вероятностью единица процесс оборвется после конечного числа поколений.

Можно найти также математическое ожидание числа частиц в n-м поколении $E(X_n)=P_n'(1)$. Поскольку $P_n(s)=P(P_{n-1}(s))$, имеем

$$P_n'(1)=P'(P_{n-1}(1))P_{n-1}'(1)=P'(1)P_{n-1}'(1)=\mu E(X_{n-1})$$

и по индукции

$$E(X_n)=\mu^n. \tag{5.3}$$

Поэтому при $\mu>1$ следует ожидать роста числа частиц по показательному закону. Это утверждение можно уточнить. Легко убедиться, что не только $P_n(0)\to\zeta$, но также и $P_n(s)\to\zeta$ при всех $s<1$. Это значит, что коэффициенты при $s, s^2, s^3, \ldots$ стремятся к нулю. *После большого числа поколений вероятность того, что не существует потомков, близка к* ζ, *а вероятность того, что число потомков превосходит любую наперед заданную границу, близка к* $1-\zeta$; чрезвычайно маловероятно иметь ограниченное число потомков[1]).

§ 6. Задачи

1. Распределение (1.1) имеет среднее значение $E(N)\,E(X)$ и дисперсию $E(N)\,D(X)+D(N)\,E^2(X)$. Доказать это, а) используя производящие функции, б) непосредственно из определения условного математического ожидания.

2. Задача об отлове животных (пример 1, б). Найти результирующее распределение, если $\{g_n\}$ — геометрическое распределение. Если $\{g_n\}$ — логарифмическое распределение [см. формулу (2.4)], то результирующим распределением является опять-таки логарифмическое, но с дополнительным членом.

3. В последовательности N испытаний Бернулли, где N — случайная величина с распределением Пуассона, число успехов и неудач — независимые случайные величины. Обобщить этот результат на случай полиномиального распределения а) непосредственно, б) используя производящую функцию полиномиального распределения. [См. пример (1.2) гл. IX.]

4. Пусть N имеет распределение Пуассона с параметром λ, и пусть N шаров случайным образом размещаются по n ящикам. Показать без вычислений, что вероятность найти ровно m пустых ящиков равна $\binom{n}{m} e^{-\lambda m/n}\left[(1-e)^{-\lambda/n}\right]^{n-m}$.

5. Продолжение[2]). Показать, что если по r ящикам случайно размещается фиксированное число n шаров, то вероятность найти ровно m пустых ящиков равна коэффициенту при $e^{-\lambda}\lambda^r/r!$ в разложении приведенного выражения. а) Найти связь с производящей функцией моментов (задача 24 гл. XI). б) Воспользоваться этим результатом для упрощенного вывода формулы (11.7) гл. II.

6. Смесь вероятностных распределений. Пусть $\{f_j\}$ и $\{g_i\}$ — два вероятностных распределения $\alpha>0$, $\beta>0$, $\alpha+\beta=1$. Тогда $\{\alpha f_i+\beta g_i\}$ — тоже вероятностное распределение. Обсудить его смысл и связь с урновой моделью § 2 гл. V. Обобщить на случай более чем двух распределений. Может ли смесь двух сложных распределений Пуассона быть опять сложным распределением Пуассона?

7. Доказать, что в ветвящемся процессе $D(X_{n+1})=\mu D(X_n)+\mu^{2n}\sigma^2$, используя а) производящие функции, б) условные математические ожидания. Вывести отсюда, что $D(X_n)=\lambda^2(\mu^{2n-2}+\mu^{2n-3}+\ldots+\mu^{n-1})$.

8. Продолжение. Показать, что при $n>m$ имеет место формула $E(X_nX_m)=\mu^{n-m}E\left(X_m^2\right)$.

9. Продолжение. Показать, что двойная производящая функция пары X_m, X_n имеет вид $P_m(s_1P_{n-m}(s_2))$. Воспользоваться этим для доказательства утверждения задачи 8.

---

[1]) Относительно поведения X_n см. T. E. Harris, Branching processes, *Ann. Math. Statistics*, 19 (1948), 474—494.

[2]) Этот изящный метод вывода различных комбинаторных формул с помощью введения случайного параметра был предложен C. Domb, On the use of a random parameter in combinatorial problems, *Proceedings Royal Philosophical Society*, Sec. A, 65 (1952), 305—309.

ГЛАВА XIII

РЕКУРРЕНТНЫЕ СОБЫТИЯ. УРАВНЕНИЕ ВОССТАНОВЛЕНИЯ

§ 1. Наглядное введение и примеры

Мы будем иметь дело с некоторыми повторяющимися событиями, связанными с повторными испытаниями. Грубо говоря, событие $\mathscr{E}$ пригодно для последующей теории, если после каждого его осуществления все начинается сначала в том смысле, что испытания, проведенные после наступления $\mathscr{E}$, являются точной копией эксперимента в целом. Времена ожиданий между последовательными появлениями $\mathscr{E}$ являются независимыми одинаково распределенными случайными величинами. Рассмотрим простейший частный случай. Пусть в последовательности испытаний Бернулли $\mathscr{E}$ означает «успех». Время ожидания первого успеха имеет геометрическое распределение; после первого успеха испытания начинаются заново, и число испытаний между r-м и $(r+1)$-м успехами имеет то же самое геометрическое распределение. Время ожидания r-го успеха есть сумма r независимых случайных величин (см. пример 3, в гл. IX). Приведем пример противоположного характера. Пусть люди выбираются по одному и пусть $\mathscr{E}$ означает «два человека в выборке имеют общий день рождения». В этом случае событие $\mathscr{E}$ не возвращает нас к начальному положению. Выбор можно продолжать до тех пор, пока не найдется еще одна пара людей с общим днем рождения, но эта вторая фаза эксперимента не является точной копией начальной фазы. Чем больше выборка, тем больше вероятность совпадения дня рождения, поэтому продолжительное ожидание первого осуществления $\mathscr{E}$ увеличивает шансы на то, что интервал между первым и вторым наступлениями будет мал. Два последовательных времени ожидания не только имеют различные распределения, но и являются зависимыми случайными величинами. Такие времена ожидания в теории рекуррентных событий не рассматриваются.

При рассмотрении таких явлений, как осуществление двух последовательных успехов, приходится учитывать другое обстоятельство. Определение *первой* серии *УУ* не встречает затруднений, однако если $\mathscr{E}$ означает «последовательность ровно двух успехов», то третье испытание может отменить результат первых двух. Например, если четыре испытания привели к последовательности *УУУН*, то событие $\mathscr{E}$ осуществляется при втором испытании, но полная последовательность не содержит $\mathscr{E}$. Нам важно, чтобы событие «$\mathscr{E}$ произошло при n-м испытании» зависело от исхода первых n испытаний и не зависело от будущего.

Приведем несколько типичных задач, к которым применима теория рекуррентных событий.

Примеры. а) *Серии успехов в испытаниях Бернулли.* Термин «серия успехов длины r» был определен различными способами. Вопрос о том, считать, что ряд из трех последовательных успехов содержит 0, 1 или 2 серии длины 2 является в значительной степени вопросом соглашения и удобства, и для разных целей мы принимали разные определения. Однако если мы хотим применять теорию рекуррентных событий, то понятие серии длины r нужно определить так, чтобы после завершения серии процесс каждый раз начинался снова. Это значит, что мы должны принять следующее определение. *Последовательность n букв У и Н содержит столько серий успехов длины r, сколько в ней имеется неперекрывающихся подпоследовательностей, каждая из которых состоит ровно из r стоящих рядом букв У. Если в последовательности испытаний Бернулли в результате n-го испытания прибавляется новая серия, то мы будем говорить, что эта серия появляется при испытании с номером n.* Так, в последовательности $УУУ|УН|УУУ|УУУ$ имеется три серии длины 3, появившиеся при 3, 8 и 11 испытаниях; в той же последовательности содержится пять серий длины 2 и они появляются при 2, 4, 7, 9 и 11 испытаниях. Это определение значительно упрощает теорию, так как серии *фиксированной длины* становятся рекуррентными событиями. Оно равносильно подсчету последовательностей, образованных по меньшей мере r следующими друг за другом успехами, с той оговоркой, что $2r$ последовательных успехов считаются за две серии и т. д. (Мы вернемся к этой теме в § 7 и § 8.)

б) *Задача о счетчиках.* Счетчики того типа, которые используются при изучении космических лучей и α-частиц, могут быть описаны с помощью следующей упрощенной модели[1]). Через равные промежутки времени производятся испытания Бернулли. Счетчик регистрирует успехи, но механизм устроен так, что не реагирует на результаты ровно $(r-1)$ испытаний, следующих за каждой регистрацией. Другими словами, успех регистрируется тогда и только тогда, когда в предыдущих $(r-1)$ испытаниях не было *регистраций.* Показания счетчика представляют исход последовательности *зависимых испытаний*; каждая регистрация приводит к эффекту последействия. Однако если счетчик свободен (способен реагировать на успех), то ситуация полностью эквивалентна началь-

---

[1]) Мы описали дискретный аналог так называемых счетчиков типа I. Счетчики типа II описываются в задаче 10. Подробности см. H. Maier-Leibnitz, Die Koinzidenzmethode und ihre Anwendung auf kernphysikalische Probleme, *Physikalische Zeitschrift*, **43** (1942), 333—362.

ной, и испытания начинаются заново. Допустим, что $\mathscr{E}$ означает «счетчик свободен после испытания», в этом случае $\mathscr{E}$ — типичное рекуррентное событие (см. задачи 9 и 10 настоящей главы и задачу 13 гл. XV).

в) *Возвращение к началу.* Пусть в последовательности испытаний Бернулли с вероятностью успеха p событие $\mathscr{E}$ означает «общее число успехов равно общему числу неудач». Как и раньше, введем для описания схемы Бернулли последовательность независимых случайных величин $\{X_k\}$ с общим распределением $P\{X_k=1\}=p$ и $P\{X_k=-1\}=q$ и положим

$$S_0=0, \quad S_n=X_1+\ldots+X_n. \tag{1.1}$$

Тогда S_n — полная разность между числом успехов и неудач, и наше событие $\mathscr{E}$ *осуществляется при n-м испытании тогда и только тогда, когда $S_n=0$.* Мы будем описывать событие $\mathscr{E}$ как *возвращение в нуль.* Если известно, что $S_n=0$, то подпоследовательность частных сумм

$$S_0'=S_n, \quad S_1'=S_{n+1}, \quad S_2'=S_{n+2}, \ \ldots \tag{1.2}$$

подчиняется тем же вероятностным соотношениям, что и первоначальная последовательность $\{S_k\}$; возвращение в нуль для $\{S_k'\}$ означает возвращение в нуль для $\{S_k\}$, и наоборот. Событие «$\mathscr{E}$ осуществилось впервые при испытании с номером n» или иначе «*первое возвращение в нуль* произошло при n-м испытании» определяется как множество последовательностей $\{X_k\}$, таких, что

$$S_1\neq 0, \quad S_2\neq 0, \ \ldots, \ S_{n-1}\neq 0, \quad S_n=0. \tag{1.3}$$

Если оно осуществилось, то мы говорим, что *время ожидания* $\boldsymbol{T}$ равно n. Для вероятности события (1.3) введем обозначение $f_n=\boldsymbol{P}\{\boldsymbol{T}=n\}$. Несколько первых вероятностей нетрудно найти с помощью прямого подсчета всех допустимых исходов; очевидно, $f_n=0$, когда n нечетно и $f_2=2pq$, $f_4=2p^2q^2$, $f_6=4p^3q^3$, $f_8=10p^4q^4$, $f_{10}=28p^5q^5$. Та же последовательность $\{f_n\}$ представляет вероятностное распределение для времени ожидания между r-м и $(r+1)$-м осуществлениями $\mathscr{E}$, поэтому мы называем $\{f_n\}$ также распределением *времени возвращения.* (Распределение $\{f_n\}$ было найдено в гл. XI с использованием производящих функций. В гл. III разобран частный случай $p=q=\frac{1}{2}$, но полученные там формулы применимы и в общем случае, так как число исходов, удовлетворяющих (1.3), не зависит от p. В настоящей главе мы дадим новый и независимый вывод этих результатов.)

г) *Рекорды в испытаниях Бернулли.* Придерживаясь прежних обозначений, определим новое повторяющееся событие $\mathscr{E}$ следующим образом: «$\mathscr{E}$ *осуществляется при n-м испытании, если* S_n *превосходит все предшествующие частные суммы*», т. е. если

$$S_n > 0, \quad S_n > S_1, \quad S_n > S_2, \ \ldots, \ S_n > S_{n-1}. \tag{1.4}$$

В этом случае мы будем говорить, что n-е испытание (или номер n) приводит к рекорду. В последовательности частных сумм — 1, 0, 1 |2| 1, 2, 3|2, 1, 2, 1, 2, 3, 4, |5| (см. рис. 3 гл. III) рекорды устанавливаются при испытаниях с номерами 3, 4, 7, 14, 15, и времена ожидания между последовательными осуществлениями $\mathscr{E}$ равны соответственно 3, 1, 3, 7, 1. В момент r-го появления $\mathscr{E}$ мы *впервые попадаем в состояние* r; поэтому рекорды можно трактовать как *первые достижения.*

Если $\mathscr{E}$ осуществилось при n-м испытании, то процесс начинается сначала в следующем смысле. Допустим, что выполнены соотношения (1.4), в этом случае *испытание с номером* $n+m$ *приводит к рекорду тогда и только тогда, когда*

$$S_{n+m} > S_n, \ S_{n+m} > S_{n+1}, \ \ldots, \ S_{n+m} > S_{n+m-1}. \tag{1.5}$$

Положим

$$S_k^* = S_{n+k} - S_n = X_{n+1} + \ \ldots \ + X_{n+k}. \tag{1.6}$$

В этом случае $n+m$ является моментом установления рекорда для последовательности $S_1, S_2, \ldots$ тогда и только тогда, когда для $\{S_n^*\}$ устанавливается рекорд в момент m. Ясно, что операция, определяемая формулами (1.6), приводит к независимой точной копии начального пространства элементарных событий и что к событию $\mathscr{E}$ применима теория рекуррентных событий. Заметим, что в этом случае последовательность (1.2) с вероятностной точки зрения отлична от начальной последовательности: после r-го появления $\mathscr{E}$ значения частных сумм S_j колеблются вокруг r, а не вокруг 0. Тем не менее поскольку речь идет о событии $\mathscr{E}$, после каждого осуществления $\mathscr{E}$ испытания начинаются заново.

[Использование понятия рекорда позволяет свести изучение времени первого достижения к рекуррентным событиям, т. е. к суммированию независимых случайных величин. Прямой (эквивалентный) метод дается в § 3 гл. IX. Понятие рекорда полезно и при изучении последовательностей произвольных случайных величин, например в связи с общим законом арксинуса.]

д) Рассмотрим последовательные бросания симметричной кости. Пусть $\mathscr{E}$ означает «единицы, двойки, ..., шестерки появились в равном числе». Здесь рекуррентный характер не требует пояснений.

§ 2. Определения

Рассмотрим последовательность повторных испытаний с возможными исходами E_j $(j = 1, 2, \ldots)$. Не обязательно считать эти испытания независимыми (особенно интересны приложения к цепям Маркова). Как обычно, мы предполагаем, что в принципе можно продолжить испытания неограниченно и вероятности $P\{E_{j_1}, E_{j_2}, \ldots, E_{j_n}\}$ определены согласованным образом для всех конечных наборов. Пусть $\mathcal{E}$ — некоторое свойство конечных последовательностей; другими словами, предположим, что для любой конечной последовательности $(E_{j_1}, E_{j_2}, \ldots, E_{j_n})$ однозначно определено, обладает ли она свойством $\mathcal{E}$ или нет. Условимся считать, что выражение «$\mathcal{E}$ произошло при n-м испытании в (конечной или бесконечной) последовательности $E_{j_1}, E_{j_2}, \ldots, E_{j_n}, \ldots$» эквивалентно выражению «набор $E_{j_1}, E_{j_2}, \ldots, E_{j_n}$ обладает свойством $\mathcal{E}$». Отсюда следует, что наступление $\mathcal{E}$ при n-м испытании зависит от исхода лишь первых n испытаний. Следует также понимать, что *говоря о «рекуррентном событии $\mathcal{E}$» мы на самом деле имеем в виду целый класс событий*, определяемых тем свойством, что $\mathcal{E}$ произошло. Само по себе $\mathcal{E}$, конечно, не является событием. Здесь мы несколько вольно обращаемся с языком, подобно тому, как постоянно употребляются термины, вроде «двумерной задачи», хотя задача сама по себе безразмерна.

Определение 1. *Свойство $\mathcal{E}$ определяет рекуррентное событие, если*

а) *для того, чтобы $\mathcal{E}$ произошло при n-м и при $(n+m)$-м испытаниях в последовательности $(E_{j_1}, E_{j_2}, \ldots, E_{j_{n+m}})$, необходимо и достаточно, чтобы $\mathcal{E}$ наступило при последнем испытании в каждом из двух наборов $(E_{j_1}, E_{j_2}, \ldots, E_{j_n})$ и $(E_{j_{n+1}}, E_{j_{n+2}}, \ldots, E_{j_{n+m}})$*;

б) *в этом случае*

$$P\{E_{j_1}, \ldots, E_{j_{n+m}}\} = P\{E_{j_1}, \ldots, E_{j_n}\}\, P\{E_{j_{n+1}}, \ldots, E_{j_{n+m}}\}.$$

Теперь ясно, какой смысл имеют выражения «в последовательности $(E_{j_1}, E_{j_2}, \ldots)$ событие $\mathcal{E}$ *наступило впервые* на n-м месте» и т. п. Ясно также, что с каждым рекуррентным событием можно связать две последовательности чисел, определенных при $n = 1, 2, \ldots$ следующим образом:

$$\begin{aligned} u_n &= P\{\mathcal{E} \text{ наступило при } n\text{-м испытании}\}, \\ f_n &= P\{\mathcal{E} \text{ впервые наступило при } n\text{-м испытании}\}. \end{aligned} \tag{2.1}$$

Удобно доопределить

$$f_0 = 0, \qquad u_0 = 1 \tag{2.2}$$

и ввести производящие функции

$$F(s) = \sum_{k=1}^{\infty} f_k s^k, \quad U(s) = \sum_{k=0}^{\infty} u_k s^k. \tag{2.3}$$

Заметим, что $\{u_k\}$ не является вероятностным распределением; в действительности в типичных случаях оказывается, что $\sum u_k = \infty$. Однако события «$\mathscr{E}$ наступило впервые при n-м испытании» несовместимы, и поэтому

$$f = \sum_{n=1}^{\infty} f_n \leqslant 1. \tag{2.4}$$

Ясно, что $1 - f$ можно интерпретировать как *вероятность того, что $\mathscr{E}$ ни разу не появится в бесконечно продолжаемой серии испытаний.* Если $f = 1$, то можно ввести случайную величину $\boldsymbol{T}$, имеющую распределение

$$P\{\boldsymbol{T} = n\} = f_n. \tag{2.5}$$

Условимся сохранять обозначение (2.5) даже в том случае, когда $f < 1$. Тогда *$\boldsymbol{T}$ является обобщенной случайной величиной, которая с вероятностью $1 - f$ не принимает числового значения.* (Можно было бы доопределить $\boldsymbol{T}$ с помощью символа ∞, и это не потребовало бы никаких дополнительных объяснений.)

Время ожидания $\mathscr{E}$, т. е. число испытаний до первого появления $\mathscr{E}$ включительно, есть случайная величина, которая имеет распределение (2.5); однако эта случайная величина определена в действительности только в пространстве бесконечных последовательностей испытаний $(E_{j_1}, E_{j_2}, \ldots)$.

По определению рекуррентных событий, вероятность того, что $\mathscr{E}$ произойдет в первый раз при k-м испытании, а во *второй* раз при n-м испытании, равна $f_k f_{n-k}$. Поэтому вероятность $f_n^{(2)}$ того, что $\mathscr{E}$ осуществится во второй раз при испытании с номером n, равна

$$f_n^{(2)} = f_1 f_{n-1} + f_2 f_{n-2} + \ldots + f_{n-1} f_1. \tag{2.6}$$

Справа стоит композиция последовательности $\{f_n\}$ с самой собой, и поэтому $\{f_n^{(2)}\}$ представляет вероятностное распределение суммы двух независимых случайных величин, каждая из которых имеет распределение (2.5). Вообще, если $f_n^{(r)}$ — вероятность того, что r-е наступление $\mathscr{E}$ произойдет при n-м испытании, то

$$f_n^{(r)} = f_1 f_{n-1}^{(r-1)} + f_2 f_{n-2}^{(r-1)} + \ldots + f_{n-1} f_1^{(r-1)}. \tag{2.7}$$

Этот простой факт составляет содержание следующей теоремы.

Теорема. *Пусть $f_n^{(r)}$ — вероятность того, что r-е наступление $\mathcal{E}$ имеет место при n-м испытании. Тогда $\{f_n^{(r)}\}$ есть вероятностное распределение суммы*

$$T^{(r)} = T_1 + T_2 + \ldots + T_r, \tag{2.8}$$

r независимых случайных величин $T_1, T_2, \ldots, T_r$, каждая из которых имеет распределение (2.5). *Другими словами, при фиксированном r последовательность $\{f_n^{(r)}\}$ имеет производящую функцию $F^r(s)$.*

Отсюда, в частности, следует, что

$$\sum_{n=1}^{\infty} f_n^{(r)} = F^r(1) = f^r. \tag{2.9}$$

Другими словами, вероятность того, что $\mathcal{E}$ осуществится не менее, чем r раз, равна f^r (факт, который можно было предвидеть заранее). Введем теперь

Определение 2. *Рекуррентное событие $\mathcal{E}$ называется достоверным, если $f = 1$, и недостоверным, если $f < 1$.*

Для недостоверного события $\mathcal{E}$ вероятность того, что оно произойдет не менее чем r раз, стремится к нулю при $r \to \infty$, в то время как для достоверного события $\mathcal{E}$ эта вероятность остается равной единице. Этот факт можно выразить также, сказав, что *с вероятностью единица достоверное событие $\mathcal{E}$ наступит бесконечно много раз, тогда как недостоверное событие $\mathcal{E}$ произойдет лишь конечное число раз.* (Это утверждение может быть строго доказано в пространстве бесконечных последовательностей испытаний $E_{j_1}, E_{j_2}, \ldots$.)

Нам потребуется еще одно определение. В схеме испытаний Бернулли возвращение к началу [см. пример (1, в)] может произойти лишь при испытании с *четным* номером. В этом случае $f_{2n+1} = u_{2n+1} = 0$ и производящие функции $F(s)$ и $u(s)$ являются степенными скорее относительно s^2, чем относительно s. Аналогично этому, в примере (1, д) событие $\mathcal{E}$ может произойти только при испытаниях с номерами 6, 12, 18, ... Мы будем выражать это, говоря, что событие $\mathcal{E}$ является периодическим. Такие рекуррентные события несколько осложняют теорию. В каждой конкретной задаче ситуация совершенно очевидна, однако во всех общих теоремах приходится учитывать возможность периодического случая.

Определение 3. *Рекуррентное событие $\mathcal{E}$ называется периодическим, если существует такое целое $\lambda > 1$, что $\mathcal{E}$ может* *произойти только при испытаниях с номерами* λ, 2λ, 3λ, ... (*т. е.* $u_n=0$, *если* n *не кратно* λ). *Наибольшее* λ, *обладающее этим свойством, называется периодом* $\mathfrak{E}$.

В заключение заметим, что в пространстве бесконечных последовательностей испытаний E_{j_1}, E_{j_2}, ... число испытаний между $(r-1)$-м и r-м осуществлениями $\mathfrak{E}$ является вполне определенной случайной величиной (возможно обобщенной), распределение которой совпадает с распределением введенной нами ранее величины T_r. Другими словами, наши величины T_r в действительности имеют смысл *времен ожидания между последовательными наступлениями события* $\mathfrak{E}$ (*времен возвращения*). Мы определяем T_r аналитически, для того чтобы избежать ссылок на бесконечные пространства элементарных событий, изучение которых выходит за рамки этой книги, но можно надеяться, что и вероятностные основания интуитивно вполне ясны. Понятие рекуррентного события служит для того, чтобы прийти в некоторых общих случаях к суммам независимых случайных величин. И наоборот, отправляясь от *любого распределения* $\{f_n\}$, $n=1, 2, \ldots$ *можно построить рекуррентное событие.* Докажем это утверждение на примере.

Пример. *Самовосстанавливающиеся совокупности.* Рассмотрим электрическую лампочку, электрический предохранитель или другой прибор, обладающий конечным сроком службы. Как только прибор приходит в негодное состояние, он заменяется новым прибором того же типа, который в свою очередь в необходимый момент заменяется третьим прибором, и т. д. Допустим, что срок службы прибора есть случайная величина, принимающая лишь значения, кратные некоторому единичному интервалу времени (год, день или секунда).

Каждый единичный интервал времени представляет тогда испытание с возможными исходами «замена» и «незамена». Замену можно рассматривать как рекуррентное событие. Если f_n — вероятность того, что новый прибор будет служить ровно n единиц времени, то последовательность $\{f_n\}$ есть распределение вероятностей для времени возвращения. Если достоверно, что время службы конечно, то $\sum f_n = 1$, и рекуррентное событие достоверно. Обычно бывает известно, что время службы не может превосходить некоторого фиксированного числа m; в этом случае производящая функция $F(s)$ есть многочлен, степень которого не превосходит m. В приложениях часто интересуются вероятностями u_n того, что замена произошла в момент времени n. Эти величины можно вычислить исходя из уравнений (3.1). Здесь мы сталкиваемся с классом рекуррентных событий, которые определяются в терминах произвольного распределения $\{f_n\}$. Случай $f<1$ не исключается, $1-f$ означает вероятность того, что прибор имеет бесконечный срок службы.

§ 3. Основные соотношения

Сохраним обозначения (2.2) — (2.4) и постараемся установить связь между распределениями $\{f_n\}$ и $\{u_n\}$. Вероятность того, что $\mathscr{E}$ произошло впервые при испытании с номером ν и снова произошло при последнем n-м испытании ($n > \nu$), равна, по определению, $f_\nu u_{n-\nu}$. Вероятность того, что $\mathscr{E}$ произошло впервые при n-м испытании, равна $f_n = f_n u_0$. Так как эти события несовместны, то

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \ldots + f_n u_0, \quad n \geqslant 1. \tag{3.1}$$

В правой части стоит композиция $\{f_k\} * \{u_k\}$, имеющая производящую функцию $F(s)U(s)$. В левой же части стоят члены последовательности $\{u_n\}$, кроме u_0, так что соответствующая производящая функция равна $U(s) - 1$. Итак, $U(s) - 1 = F(s)U(s)$. Тем самым доказана

Теорема 1. *Производящие функции последовательностей $\{u_n\}$ и $\{f_n\}$ связаны соотношением*

$$U(s) = \frac{1}{1 - F(s)}. \tag{3.2}$$

Замечание. Правую часть соотношения (3.2) можно разложить в ряд $\sum F^r(s)$, сходящийся при $|s| < 1$. Коэффициент $f_n^{(r)}$ при s^n в $F^r(s)$ равен вероятности того, что $\mathscr{E}$ наступает в r-й раз при n-м испытании и, значит, уравнение (3.2) эквивалентно системе соотношений

$$u_n = f_n^{(1)} + f_n^{(2)} + \ldots . \tag{3.3}$$

Эти соотношения выражают тот очевидный факт, что если $\mathscr{E}$ произошло при n-м испытании, то ему предшествовало $0, 1, 2, \ldots, n-1$ наступлений $\mathscr{E}$. (Ясно, что $f_n^{(r)} = 0$ при $r > n$.)

Теорема 2. *Рекуррентное событие $\mathscr{E}$ недостоверно тогда и только тогда, когда ряд*

$$u = \sum_{j=0}^{\infty} u_j \tag{3.4}$$

сходится. В этом случае вероятность f того, что $\mathscr{E}$ когда-нибудь произойдет, равна

$$f = \frac{u-1}{u}. \tag{3.5}$$

Замечание. Можно истолковать u_j как математическое ожидание случайной величины, равной единице, когда $\mathscr{E}$ происходит при j-м испытании, и равной нулю в противном случае. При этом $u_1 + u_2 + \ldots + u_n$ будет ожидаемым числом появлений события $\mathscr{E}$

при n испытаниях, а $u-1$ можно истолковать как ожидаемое число появлений события $\mathscr{E}$ при бесконечном числе испытаний.

Доказательство. Так как коэффициенты u_k неотрицательны, то, очевидно, $U(s)$ монотонно возрастает при $s \to 1$. Поэтому при любом N

$$\sum_{n=0}^{N} u_n \leqslant \lim_{s \to 1} U(s) \leqslant \sum_{n=0}^{\infty} u_n = u.$$

Но $U(s) \to (1-f)^{-1}$, если $f < 1$ и $U(s) \to \infty$, если $f = 1$. Теорема доказана.

Следующая теорема очень важна[1]). Хотя ее тельство имеет элементарный характер, но, поскольку оно не необходимо для понимания вероятностных оснований самого результата, мы отложим его на конец главы. (См. также задачу 1.)

Теорема 3. *Пусть $\mathscr{E}$ — достоверное непериодическое рекуррентное событие. Обозначим через μ среднее значение времени возвращения T_ν, т. е.*

$$\mu = \sum_j f_j = F'(1) \tag{3.6}$$

(*возможно* $\mu = \infty$). *Тогда*

$$u_n \to \mu^{-1} \tag{3.7}$$

при $n \to \infty$ ($u_n \to 0$, *если математическое ожидание времени возвращения бесконечно*).

Теорема 4. *Если $\mathscr{E}$ достоверно и имеет период $\lambda > 0$, то при $n \to \infty$*

$$u_{n\lambda} \to \lambda\mu^{-1} \tag{3.8}$$

и $u_k = 0$ при всех k, которые не делятся на λ.

---

[1]) P. Erdös, W. Feller and H. Pollard, A theorem on power series, *Bulletin of the American Mathematical Society*, 55 (1949), 201—204. Эта теорема была предназначена для того, чтобы сделать более доступными результаты Колмогорова об эргодических свойствах бесконечных марковских цепей. После появления первого издания Чжун заметил, что теорема 3 в действительности эквивалентна эргодической теореме Колмогорова и может быть выведена из нее. Предварительно появилось много работ, посвященных различным частным случаям и вариантам. В дальнейшем, теорему 3 обобщили на случай непрерывных случайных величин и различными способами усовершенствовали Блэкуэлл, Чжун, Эрдеш и Вольфовиц. Блэкуэлл дал изящное доказательство того, что (3.7) выполнено для всех целочисленных случайных величин (не обязательно положительных, как в тексте), если только они имеют положительное математическое ожидание. Его метод основан на использовании моментов рекорда для произвольных величин [ср. пример (1.2)]. См. Blackwell D., Extension of a renewal theorem, *Pacific Journal of Mathematics*, 3 (1953), 315—320.

Доказательство. Так как $\mathscr{E}$ имеет период λ, то в ряд $F(s)=\sum f_n s^n$ входят только степени s^λ, и $F(s^{1/\lambda})=F_1(s)$ также является степенным рядом с положительными коэффициентами. При этом $F_1(1)=1$. В силу теоремы 3 коэффициенты ряда $U_1(s)=\{1-F_1(s)\}^{-1}$ стремятся к μ_1^{-1}, где

$$\mu_1=F_1'(1)=\lambda^{-1}F'(1)=\lambda^{-1}\mu.$$

(Ясно, что μ_1 и μ либо оба конечны, либо оба бесконечны.) Но коэффициент при s^n в $U_1(s)$ равен коэффициенту при $s^{n\lambda}$ в $U(s)$, что и доказывает (3.8).

Примеры. а) Начнем с банального примера. Пусть $\mathscr{E}$ означает «успех» в последовательности испытаний Бернулли. Тогда $u_n=p$, по самому определению. Теорема 2 утверждает, что среднее число испытаний между двумя последовательными успехами равно p^{-1}. В рассматриваемом случае $U(s)=1+ps(1-s)^{-1}=(1-qs)(1-s)^{-1}$ и по теореме 1 $F(s)=ps(1-qs)^{-1}$, откуда следует, что время ожидания между последовательными успехами имеет геометрическое распределение.

б) *Возвращение к началу при испытаниях Бернулли.* [*Пример* (1, в).] Если при k-м испытании общее число успехов равно общему числу неудач, то k должно быть четным числом, $k=2n$, и n испытаний должны привести к успеху, а остальные n — к неудаче. Поэтому вероятность возвращения к равновесию равна

$$u_{2n}=\binom{2n}{n}p^n q^n. \tag{3.9}$$

С помощью нормального приближения для биномиального распределения или с помощью формулы Стирлинга получаем

$$\binom{2n}{n}2^{-2n}\sim\frac{1}{(\pi n)^{1/2}}, \tag{3.10}$$

так что

$$u_{2n}\sim\frac{(4pq)^n}{(\pi n)^{1/2}}, \tag{3.11}$$

где знак $\sim$ означает, что отношение двух величин, соединенных этим знаком, стремится к единице.

Если $p\neq\frac{1}{2}$, то $4pq<1$, и ряд $\sum u_{2n}$ сходится быстрее геометрической прогрессии со знаменателем $4pq$. Если же $p={}^1/_2$, то $u_{2n}\sim(\pi n)^{-1/2}$, и ряд $\sum u_{2n}$ расходится, но $u_{2n}\to 0$. Наши теоремы позволяют заключить, что с вероятностью единица верно следующее утверждение:

Если $p \neq q$, то лишь конечное число частных сумм S_n обращается в нуль. Если $p = q = \frac{1}{2}$, то число возвращений к началу бесконечно, но время возвращения не имеет конечного математического ожидания.

В случае $p \neq q$ это утверждение интуитивно очевидно и следует также из *усиленного* закона больших чисел. На языке азартных игр оно означает, что если игра благоприятна для Петра, то он может быть наверняка уверен в том, что после нескольких случайных колебаний его чистый выигрыш станет положительным и останется таким в дальнейшем. Если $p = q = \frac{1}{2}$, то ситуация интуитивно менее очевидна и служит источником различных парадоксов, связанных со случайными колебаниями. Этот круг вопросов рассмотрен в § 7 гл. III.

Дальнейшие результаты можно получить с помощью производящих функций. Пользуясь легко проверяемой формулой

$$\binom{2n}{n} = \binom{-{}^1/_2}{n} \cdot (-4)^n \tag{3.12}$$

и биномиальным рядом (8.7) гл. II, получим из (3.9)

$$U(s) = \sum_{n=0}^{\infty} u_{2n} s^{2n} = (1 - 4pqs^2)^{-{}^1/_2}. \tag{3.13}$$

Если $p \neq {}^1/_2$, то $u = U(1) = (1 - 4pq)^{-{}^1/_2} = |p - q|^{-1}$. Из (3.5) заключаем, что *вероятность f того, что суммарное число успехов когда-нибудь совпадет с суммарным числом неудач, равна*

$$f = 1 - |p - q|. \tag{3.14}$$

(*Это есть вероятность по крайней мере одного возвращения к началу.*)

Из (3.2) получаем выражение для *производящей функции времени возвращения*

$$F(s) = 1 - (1 - 4pqs^2)^{{}^1/_2}. \tag{3.15}$$

Эта формула особенно интересна в случае $p = q = {}^1/_2$. Тогда

$$F(s) = 1 - (1 - s^2)^{{}^1/_2}, \tag{3.16}$$

и биномиальное разложение показывает, что

$$f_{2n} = (-1)^{n+1} \binom{{}^1/_2}{n} = \frac{1}{n} \binom{2n-2}{n-1} 2^{-2n+1} \tag{3.17}$$

(при n нечетном f_n равно нулю). Равенство (3.17) *дает распределение времени возвращения в игре с бросанием монеты.*

(Эта формула уже была получена другими методами в § 4 гл. III и § 3 гл. XI. Настоящий метод является самым прямым, хотя и не вполне элементарен по своей природе.)

в) *Ничьи в игре с бросанием нескольких монет.* Рассмотрим повторные независимые бросания *двух* монет. Событие $\mathfrak{E}$ происходит, когда суммарное число выпадений герба (а следовательно, и решетки) у обеих монет одно и то же.

Очевидно,

$$u_n = \frac{1}{2^{2n}}\left\{\binom{n}{0}^2 + \binom{n}{1}^2 + \binom{n}{2}^2 + \ldots + \binom{n}{n}^2\right\}. \tag{3.18}$$

Пользуясь формулой (9.8) гл. II, а затем нормальным приближением для биномиального распределения, убеждаемся, что

$$u_n = \binom{2n}{n} 2^{-2n} \sim \frac{1}{(n\pi)^{1/2}}. \tag{3.19}$$

Следовательно, $\sum u_n$ расходится, но $u_n \to 0$. Таким образом, $\mathfrak{E}$ — *достоверное событие, имеющее бесконечное среднее значение времени возвращения.*

Рассмотрим теперь более общий случай одновременного бросания r монет, и пусть $\mathfrak{E}$ — рекуррентное событие, состоящее в том, что *суммарное число выпадений герба для всех монет одно и то же.* Тогда

$$u_n = \frac{1}{2^{rn}}\left\{\binom{n}{0}^r + \binom{n}{1}^r + \ldots + \binom{n}{n}^r\right\}. \tag{3.20}$$

Чтобы оценить u_n, заметим, что максимальный член биномиального распределения $\binom{n}{k} 2^{-n}$ имеет порядок $(2/\pi n)^{1/2}$ и меньше $n^{-1/2}$. Следовательно,

$$u_n < n^{-(r-1)/2} 2^{-n}\left\{\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n}\right\} = n^{-(r-1)/2}. \tag{3.21}$$

Поэтому $\sum u_n$ сходится при $r \geqslant 4$. При $r = 2$, как мы видели, $\sum u_n$ расходится. Остается рассмотреть случай $r = 3$. Воспользуемся нормальным приближением для биномиального распределения. При достаточно большом n и значениях k, лежащих между $^1/_2 n - n^{1/2}$ и $^1/_2 n + n^{1/2}$, выполняется неравенство $\binom{n}{k} 2^{-n} > cn^{-1/2}$, где c — положительная постоянная (скажем, e^{-4}). Поэтому при $r = 3$

$$u_n > 2n^{1/2}(c^3 n^{-3/2}) = 2c^3/n, \tag{3.22}$$

так что $\sum u_n$ расходится. Иными словами, *при двух или трех монетах достоверно, что рано или поздно (и поэтому сколько угодно раз) суммарное число гербов, выпавших при бросании* *этих монет, будет одинаково. При четырех или более монетах это же рекуррентное событие недостоверно.*

г) *Кости.* В примере 1, д мы рассматривали рекуррентное событие $\mathscr{E}$, состоящее в том, что суммарное число выпадений единиц, двоек, троек и т. д. — одно и то же. Очевидно, $\mathscr{E}$ имеет период 6, и $u_{6n}=(6n)!\,(n!)^{-6}6^{-6n}$. Пользуясь формулой Стирлинга, легко убедиться, что u_{6n} имеет порядок $n^{-5/2}$, так что $\sum u_n$ сходится. Таким образом, событие $\mathscr{E}$ недостоверно. Пользуясь (3.7), нетрудно вычислить вероятность возвращения: она равна приблизительно 0,022.

д) Приложения к теории серий см. в § 7 и 8 настоящей главы.

§ 4. Уравнение восстановления

Основное уравнение (3.1) теории рекуррентных событий является частным случаем так называемого уравнения восстановления, которое появляется в целом ряде разнообразных задач. Мы покажем, что теоремы предыдущего параграфа применимы без существенных изменений и к этому более общему уравнению. Здесь мы дадим чисто аналитический вывод, вероятностной интерпретации и примерам будет посвящен следующий параграф.

Пусть $\{a_n\}$ *и* $\{b_n\}$ — *две последовательности, таких, что* $0\leqslant a_n<1$ *и* $b_n>0$ $(n=0,\ 1,\ 2,\ \ldots)$. *Определим третью последовательность* $\{u_n\}$ *рекуррентным соотношением*

$$u_n=b_n+(a_0u_n+a_1u_{n-1}+\ \ldots\ +a_nu_0) \tag{4.1}$$

или

$$\{u_n\}=\{b_n\}+\{a_n\}*\{u_n\}. \tag{4.2}$$

Решая (4.1) последовательно, получим

$$u_0=\frac{b_0}{1-a_0},\qquad u_1=\frac{(b_1+a_1u_0)}{1-a_0},\ \ldots,$$

так что не возникает никаких сомнений в существовании единственного решения $\{u_n\}$. Нас интересует поведение u_n при $n\to\infty$ — вопрос, которому было посвящено большое количество исследований (большей частью дискуссионного характера).

Заметим, что если положить $b_n=0$, $a_n=f_n$ при $n=1,\ 2,\ \ldots$ и $a_0=0$, $b_0=1$, то уравнение (4.1) сведется к уравнению (3.1). Таким образом, уравнение восстановления (4.1) формально содержит основное уравнение теории рекуррентных событий как частный случай. Однако все свойства уравнения восстановления можно вывести из аналогичных свойств уравнения (3.1). Мы снова перейдем к производящим функциям

$$A(s)=\sum a_ns^n,\ \ B(s)=\sum b_ns^n,\ \ U(s)=\sum u_ns^n. \tag{4.3}$$

Поскольку коэффициенты a_n и b_n предполагаются ограниченными, первые два ряда сходятся по крайней мере при $|s| < 1$. Область сходимости третьего ряда вскоре будет выяснена. Уравнение (4.1) можно теперь привести к виду $U(s) = B(s) + A(s)U(s)$, откуда

$$U(s) = \frac{B(s)}{1 - A(s)}. \tag{4.4}$$

При $B(s) \equiv 1$ эта формула сводится к формуле (3.2). Существенная разница состоит в том, что теперь $\{a_n\}$ не обязательно служит распределением вероятностей для времени возвращения, так что $A(s)$ может быть как больше, так и меньше единицы.

Будем говорить, что имеет место *периодический случай, если существует такое целое число* $\lambda > 1$, *что все* a_k, *за исключением, быть может,* $a_\lambda, a_{2\lambda}, a_{3\lambda}, \ldots$ *равны нулю.* Тогда $A(s)$ есть степенной ряд относительно s^λ. Максимальное целое число λ, обладающее указанным свойством, называется *периодом.*

Теорема 1. *Допустим, что последовательность* $\{a_n\}$ *непериодична и что величина* $B(1) = \sum b_n$ *конечна. Тогда:*

а) *Если* $\sum a_n = 1$, *то*

$$u_n \to B(1)\mu^{-1}, \quad \text{где } \mu = \sum n a_n. \tag{4.5}$$

(*В частности,* $u_n \to 0$, *если ряд* $\sum n a_n$ *расходится.*)

б) *Если* $\sum a_n < 1$, *то ряд*

$$\sum u_n = B(1)\{1 - A(1)\}^{-1} \tag{4.6}$$

сходится.

в) *Если* $\sum a_n > 1$ *и также если ряд расходится, то существует единственный положительный корень* $x < 1$ *уравнения* $A(x) = 1$. *В этом случае*

$$u_n \sim \frac{B(x)}{A'(x)} x^{-n}, \tag{4.7}$$

где знак $\sim$ означает, что отношение двух величин, соединенных этим знаком, стремится к единице. [Соотношение (4.7) выражает тот факт, что u_n возрастают по показательному закону. Производная $A'(x)$ конечна, так как $A(s)$ регулярна при $|s| < 1$.]

Доказательство. а) Пусть v_n — коэффициент при s^n в разложении функции $1/\{1 - A(s)\}$. По теореме 3 предыдущего параграфа $v_n \to 1/\mu$. Нам известно, что

$$u_n = v_n b_0 + v_{n-1} b_1 + \ldots + v_0 b_n. \tag{4.8}$$

При любом фиксированном k слагаемое $v_{n-k}\, b_k$ при $n \to \infty$ стремится к b_k/μ. Кроме того, v_n ограничены. Отсюда следует, что u_n

при достаточно большом N сколь угодно мало отличается от величины

$$u_n' = v_n b_0 + v_{n-1} b_1 + \ldots + v_{n-N} b_N \tag{4.9}$$

и что $u_n' \to (b_0 + \ldots + b_N)/\mu$. Последняя величина в свою очередь сколь угодно мало отличается от $B(1)/\mu$.

б) Здесь без существенных изменений применимо доказательство теоремы 2 предыдущего параграфа.

в) Достаточно применить результат пункта (а) к последовательностям $\{a_n x^n\}$, $\{b_n x^n\}$ и $\{u_n x^n\}$, которые имеют производящие функции $A(xs)$, $B(xs)$ и $U(xs)$ и которые, очевидно, связаны тем же соотношением, что и исходные последовательности.

Не рассмотрен еще лишь периодический случай, когда $A(s) = \sum a_{n\lambda} s^{n\lambda}$. В этом случае коэффициенты u_n обладают некоторой периодичностью, и мы разобьем их на группы с одинаковыми фазами: $\{u_0, u_\lambda, u_{2\lambda}, \ldots\}$, $\{u_1, u_{\lambda+1}, u_{2\lambda+1}, \ldots\}$, $\ldots$, $\{u_{\lambda-1}, u_{2\lambda-1}, u_{3\lambda-1}, \ldots\}$.

Из (4.4) видно, что коэффициенты $u_{n\lambda}$ зависят только от b_0, b_λ, $b_{2\lambda}$, $\ldots$ и не зависят от тех b_k, у которых индекс k не делится на λ. Поэтому имеет смысл представить $U(s)$ и $B(s)$ в виде суммы λ степенных рядов относительно s^λ:

$$\begin{aligned} U(s) &= U_0(s) + sU_1(s) + \ldots + s^{\lambda-1} U_{\lambda-1}(s), \\ B(s) &= B_0(s) + sB_1(s) + \ldots s^{\lambda-1} B_{\lambda-1}(s), \end{aligned} \tag{4.10}$$

где

$$U_j(s) = \sum_{n=0}^{\infty} u_{n\lambda+j} s^{n\lambda}, \qquad B_j(s) = \sum_{n=0}^{\infty} b_{n\lambda+j} s^{n\lambda}. \tag{4.11}$$

Тогда из (4.4) следует, что при $j = 0, 1, \ldots, \lambda - 1$

$$U_j(s) = \frac{B_j(s)}{1 - A(s)}. \tag{4.12}$$

Здесь все функции суть степенные ряды относительно s^λ, и поэтому после замены s^λ на t применима предыдущая теорема. Этим доказана

Теорема 2. *В периодическом случае при периоде, равном λ, последовательность $\{u_n\}$ асимптотически периодична; если $A(1) = 1$, то каждая из λ подпоследовательностей $\{u_{n\lambda+j}\}$ имеет предел*

$$\lim_{n\to\infty} u_{n\lambda+j} = \frac{\lambda B_j(1)}{\mu}, \tag{4.13}$$

где $B_j(1) = b_j + b_{\lambda+j} + b_{2\lambda+j} + b_{3\lambda+j} + \ldots$.

Пример. *Повторное осреднение.* Пусть заданы три положительных числа u_1, u_2, u_3, определим бесконечную последовательность $\{u_n\}$ следующим образом:

$$u_4 = \frac{1}{3}(u_1 + u_2 + u_3), \quad u_5 = \frac{1}{3}(u_2 + u_3 + u_4), \; \ldots,$$

$$u_{n+3} = \frac{1}{3}(u_n + u_{n+1} + u_{n+2}). \tag{4.14}$$

Наша цель состоит в исследовании асимптотического поведения $\{u_n\}$ *Мы покажем, что*

$$u_n \to \frac{1}{6}(u_1 + 2u_2 + 3u_3). \tag{4.15}$$

Нет нужды говорить, что аналогичные рассуждения применимы к произвольным средним (см. задачу 5 настоящей главы и задачу 15 гл. XV). Дело в том, что задачи такого рода сводятся к уравнению восстановления (4.1), что позволяет в свою очередь по-новому взглянуть на его природу.

Положим

$$a_0 = 0, \quad a_1 = a_2 = a_3 = \frac{1}{3}, \quad a_n = b_n = 0 \quad \text{при} \quad n \geqslant 4, \tag{4.16}$$

тогда формулы (4.14) и (4.1) совпадают при $n \geqslant 4$. Для того чтобы свести (4.14) к (4.1) при всех n, положим дополнительно $b_0 = u_0 = 0$ и определим b_1, b_2, b_3 следующим образом:

$$b_1 = u_1, \quad b_2 = u_2 - \frac{1}{3}u_1, \quad b_3 = u_3 - \frac{1}{3}(u_1 + u_2). \tag{4.17}$$

Применяя затем теорему 1 (а), мы получаем соотношение (4.15) без всяких вычислений. Так как производящая функция $U(s)$ рациональна, то, разлагая ее на простые дроби, нетрудно убедиться в том, что последовательность $\{u_n\}$ стремится к пределу с показательной скоростью. Нетрудно дать и точную оценку скорости сходимости.

§ 5. Рекуррентные события с запаздыванием

Теперь мы введем некоторое обобщение понятия рекуррентных событий. Это обобщение столь очевидно, что оно и не заслуживало бы особого упоминания, если бы не было удобно иметь специальный термин для обозначения таких событий и запись основного уравнения.

Может быть наилучшее наглядное описание рекуррентных событий с запаздыванием состоит в том, что они связаны с испытаниями, «начало которых пропущено и которые наблюдаются с середины». Время ожидания *первого* наступления $\mathcal{E}$ имеет распределение $\{b_n\}$, отличное от распределения $\{f_n\}$ интервалов времени между последующими осуществлениями $\mathscr{E}$. Теория, развитая в предыдущих параграфах, проходит и в этом случае без существенных изменений. Единственная особенность состоит в том, что теперь испытаниям, следующим за каждым наступлением $\mathscr{E}$, отвечают пространства элементарных событий, являющиеся точными копиями определенного пространства, которое не идентично начальному,

Описанная ситуация столь проста, что вряд ли есть нужда в формальных определениях. Условимся говорить о *рекуррентном событии с запаздыванием, когда определение рекуррентного события применимо, если только отбросить испытания, предшествующие первому появлению $\mathscr{E}$. Подразумевается, что время ожидания первого появления $\mathscr{E}$ есть случайная величина, независимая от последующих времен возвращения, хотя ее распределение $\{b_n\}$ может быть отлично от общего для всех времен возвращения распределения $\{f_n\}$*.

Нетрудно вычислить вероятности u_n того, что событие $\mathscr{E}$ произойдет при n-м испытании непосредственно из данного выше определения, опираясь на результаты § 3. Однако предпочтительнее действовать независимо и написать новое уравнение, типа уравнения восстановления.

Вероятность того, что $\mathscr{E}$ произойдет при испытании с номером $(n-k)$, а *следующий раз* — при n-м испытании, равна $u_{n-k}f_k$. Эти события несовместимы и их объединение при $k=1, 2, \ldots, n-1$ есть событие, состоящее в том, что осуществится при n-м и одном из предшествующих испытаний. Вероятность того, что $\mathscr{E}$ произойдет при n-м испытании впервые, равна b_n, и поэтому при $n \geqslant 1$

$$u_n = b_n + u_{n-1}f_1 + \ldots + u_1 f_{n-1}. \tag{5.1}$$

Для событий с запаздыванием естественно положить

$$u_0 = f_0 = b_0 = 0, \tag{5.2}$$

после чего уравнение (5.1) сводится к уравнению восстановления:

$$\{u_n\} = \{b_n\} + \{u_n\} * \{f_n\}. \tag{5.3}$$

Соответствующие производящие функции связаны следующим соотношением:

$$U(s) = \frac{B(s)}{1-F(s)}. \tag{5.4}$$

Из результатов предыдущего параграфа как специальный частный случай вытекает

Теорема. *Если $\mathscr{E}$ — непериодическое достоверное событие (т. е. $\sum f_n = 1$), то*

$$u_n \to \mu^{-1} \sum b_n \qquad \mu = \sum n f_n. \tag{5.5}$$

Если же $\sum f_n < 1$ (*т. е. событие* $\mathcal{E}$ *недостоверно*), *то*

$$\sum u_n = (1-f)^{-1} \sum b_n. \tag{5.6}$$

В периодическом случае применима теорема 2 § 4.

Примеры. а) Предположим, что в *задаче о счетчиках* (1, б) к началу опыта прибор был уже блокирован две единицы времени (другими словами наблюдения начались через две единицы времени после начала некоторой регистрации). Счетчик будет блокирован еще по меньшей мере $r-2$ единицы времени и станет свободным после $(r-1)$-го испытания, если это испытание приведет к неудаче; в противном случае счетчик сработает (регистрация) и поэтому останется блокированным еще не менее чем при r испытаниях, и т. д. Отсюда следует, что $b_{r-2} = q$, $b_{2r-2} = pq$, $b_{3r-2} = qp^2, \ldots$.

б) *Самовосстанавливающиеся совокупности.* В примере 2 мы рассматривали прибор, срок службы которого является случайной величиной с распределением $\{f_n\}$. Как только прибор выходит из строя, он заменяется новым прибором того же типа, и этот процесс продолжается. Замену можно рассматривать как рекуррентное событие $\mathcal{E}$. В § 2 мы предполагали, что в момент времени 0 установлен новый прибор. Предположим вместо этого, что в момент a возраст прибора равен k. Тогда $\mathcal{E}$ становится рекуррентным событием с задержкой. Вычислим распределение вероятностей для времени ожидания первого наступления $\mathcal{E}$. Ясно, что b_n равно условной вероятности того, что срок службы прибора будет равен $n+k$, при условии, что он уже прожил k единиц времени. Поэтому

$$b_n = \frac{f_{n+k}}{r_k}, \quad r_k = f_k + f_{k+1} + f_{k+2} + \ldots . \tag{5.7}$$

В приложениях естественно рассматривать не один прибор, а целую совокупность одинаковых приборов. Предположим, что *начальная совокупность* (в момент времени 0) *состоит из* N *элементов, среди которых* v_k *имеют возраст* k $\left(\sum v_k = N\right)$. Каждый из этих элементов дает начало ряду заменяющих его элементов, и в любой момент времени существует определенная вероятность того, что в этом ряду потребуется замена. Сумма этих вероятностей для всех N элементов есть *среднее число* u_n *замен в момент* n. Очевидно, u_n удовлетворяют основному уравнению (5.3), если положить

$$b_n = \sum_{k=0}^{\infty} \frac{v_k f_{n+k}}{r_k}, \tag{5.8}$$

и наши теоремы показывают, что последовательность $\{u_n\}$ имеет предел.

Нетрудно вычислить не только предел u_n, но и распределение возрастов в момент n и предельное поведение этого распределения.

Пусть $v_k(n)$ — среднее число элементов возраста k в момент n (так что $v_k(0)=v_k$). Тогда, очевидно,

$$\begin{aligned} v_k(n) &= u_{n-k} r_k, \quad \text{если} \quad k < n, \\ v_k(n) &= \frac{v_{k-n} r_k}{r_{k-n}}, \quad \text{если} \quad k \geqslant n. \end{aligned} \tag{5.9}$$

Мы знаем, что в непериодическом случае $u_n \to B(1)/\mu = N/\mu$ при $n \to \infty$, так что из (5.2) следует, что $v_k(n) \to N r_k/\mu$. Поэтому в непериодическом случае существует устойчивое *предельное распределение возрастов*: в пределе математическое ожидание числа элементов возраста k равно $N r_k/\mu$, где N — (постоянный) объем совокупности, а $\mu = \sum r_k$ — средняя продолжительность службы (если $\mu = \infty$, то возрасты неограниченно растут). Важно, что *предельное распределение возрастов не зависит от начального распределения* а зависит только от распределения срока службы $\{a_n\}$ (см. задачи 17 и 18).

В качестве численной иллюстрации рассмотрим совокупность из 1000 элементов с распределением возрастов $v_0 = 500$, $v_1 = 320$, $v_2 = 74$, $v_3 = 100$, $v_4 = 6$. Распределение времени службы пусть будет $a_1 = 0{,}20$, $a_2 = 0{,}43$, $a_3 = 0{,}17$, $a_4 = 0{,}17$, $a_5 = 0{,}03$ (ни один элемент не может достигнуть возраста, большего 5). Здесь $U(s)$ — рациональная функция,

$$U(s) = s\,\frac{397 + 332s + 159s^2 + 97s^3 + 15s^4}{1 - 0{,}20s - 0{,}43s^2 - 0{,}17s^3 - 0{,}17s^4 - 0{,}03s^5}, \tag{5.10}$$

которую можно разложить на простые дроби

$$U(s) = \frac{1250s}{3(1-s)} - \frac{972s}{61(1+3s/5)} + \frac{38s}{87(1+s/5)} - \frac{78{,}225s^2 + 22{,}125s}{5307(1+s^2/4)}.$$

Распределение возрастов $\{v_k(n)\}$ при $n = 1, 2, 3, \ldots$ можно вычислить непосредственно из уравнения восстановления. Эти распределения вместе с их предельной формой представлены в столбцах табл. 1. Интересно отметить, что сходимость в этом случае не монотонная.

в) *Теория популяций.* Эта теория аналогична теории восстановления. Разница состоит в том, что объем популяции меняется и роль замен играют рождения особей женского пола. Существенно и то, что каждая особь женского пола может произвести нуль, одну или более дочерних особей, так что соответствующая линия может или оборваться, или, наоборот, разветвиться. Обозначим через a_n вероятность того, что только что родившаяся особь женского пола разовьется и в возрасте n произведет новую женскую особь (зависимостью от числа и возраста предыдущих детей пренебрегаем). Тогда

Таблица 1

k	n 0	1	2	3	4	5	6	7	∞
0	500	397	411,4	412	423,8	414,3	417,0	416,0	416,7
1	320	400	317,6	329,1	329,6	339,0	331,5	333,6	333,3
2	74	148	185	146,9	152,2	152,4	156,8	153,3	154,2
3	100	40	80	100	79,4	82,3	82,4	84,8	83,3
4	6	15	6	12	15	11,9	12,3	12,4	12,5

$\sum a_n$ — среднее число дочерних особей, и поэтому могут иметь место все три возможности: $\sum a_n < 1$, $\sum a_n = 1$ и $\sum a_n > 1$. С очевидными изменениями применимы все предыдущие рассуждения.

§ 6. Число осуществлений события $\mathcal{E}$

До сих пор мы рассматривали первое, второе, ..., r-е осуществление рекуррентного события $\mathcal{E}$ и считали число испытаний случайной величиной. Часто более удобной бывает другая точка зрения, именно фиксируется число n испытаний, а *число* N_n *осуществлений* $\mathcal{E}$ рассматривается как случайная величина. Мы исследуем асимптотическое поведение N_n при больших n.

Как и в (2.8), пусть $T^{(r)}$ означает число испытаний до r-го появления $\mathcal{E}$ включительно. Распределения вероятностей величин $T^{(r)}$ и N_n связаны очевидным тождеством:

$$P\{N_n \geqslant r\} = P\{T^{(r)} \leqslant n\}. \tag{6.1}$$

Рассмотрим вначале самый простой случай, когда $\mathcal{E}$ достоверно, а распределение $\{f_n\}$ соответствующего времени возвращения имеет конечное математическое ожидание μ и конечную дисперсию σ^2. Так как $T^{(r)}$ представляет сумму r независимых одинаково распределенных величин, то по центральной предельной теореме (§ 1 гл. X) при любом фиксированном x и $r \to \infty$

$$P\left\{\frac{T^{(r)} - r\mu}{\sigma r^{1/2}} < x\right\} \to \Phi(x), \tag{6.2}$$

где $\Phi(x)$ — нормальная функция распределения. Пусть теперь $n \to \infty$ и $r \to \infty$ таким образом, что

$$\frac{n - r\mu}{\sigma r^{1/2}} \to x; \tag{6.3}$$

тогда из (6.1) и (6.2) немедленно вытекает, что

$$P\{N_n \geqslant r\} \to \Phi(x). \tag{6.4}$$

Чтобы записать это соотношение в более удобной форме, введем *нормированную случайную величину*:

$$N_n^* = \frac{N_n - n\mu^{-1}}{\sigma n^{1/2}\mu^{-3/2}}. \tag{6.5}$$

Тогда неравенство $N_n \geqslant r$ эквивалентно неравенству

$$N_n^* \geqslant \frac{r - n\mu^{-1}}{\sigma n^{1/2}\mu^{-3/2}} = -\frac{n - r\mu}{\sigma r^{1/2}}\left(\frac{r\mu}{n}\right)^{1/2}, \tag{6.6}$$

и соотношение (6.3) показывает, что правая часть асимптотически равна $-x$. Поэтому

$$P\{N_n^* \geqslant -x\} \to \Phi(x) \quad \text{или} \quad P\{N_n^* < -x\} \to 1 - \Phi(x). \tag{6.7}$$

Итак, нами доказана следующая теорема.

Теорема 1. Нормальное приближение. *Если рекуррентное событие $\mathcal{E}$ достоверно и его время возвращения имеет конечное математическое ожидание μ и конечную дисперсию σ^2, то число испытаний $T^{(r)}$ до k-го появления события $\mathcal{E}$ и число N_n появлений события $\mathcal{E}$ в первых n испытаниях распределены асимптотически нормально согласно соотношениям* (6.2) *и* (6.7).

Заметим, что (6.7) есть пример приложения центральной предельной теоремы к последовательности *зависимых* величин. Полученный нами предельный результат делает правдоподобными соотношения

$$E(N_n) \sim \frac{n}{\mu}, \quad D(N_n) \sim \frac{n\sigma^2}{\mu^3}, \tag{6.8}$$

но их точное доказательство требует дополнительных рассуждений.

Полезность теоремы 1 мы проиллюстрируем в следующем параграфе, где эта теорема будет использована в теории серий. Следует, однако, помнить, что наиболее интересные времена возвращения в теории случайных колебаний и различных физических процессах имеют *бесконечное математическое ожидание*. В этом случае теорема 1 должна быть заменена более общими предельными теоремами[1]).

Интуиция подсказывает нам, что $E(N_n)$ должно возрастать линейно вместе с n [как в (6.8)], просто «потому что при удвоении числа испытаний число наступлений $\mathcal{E}$ должно удвоиться». *Однако*

---

[1]) W. Feller, Fluctuation theory of recurrent events, *Trans. Amer. Math. Soc.*, 67 (1949), 98—119.

это не так. Возвращение к равновесию при бросании правильной монеты [см. пример (1, в)] является типичным для процессов диффузии временем возвращения. Эта случайная величина может еще раз служить примером того, какими неожиданными свойствами обладают общие случайные колебания.

Теорема 2. Парадокс возвращения. *Пусть $\mathscr{E}$ — возвращение к началу в последовательности симметрических испытаний Бернулли (бросаний монеты). Среднее число $E(N_{2n})$ осуществлений $\mathscr{E}$ при $2n$ испытаний задается формулой*

$$E(N_{2n}) = (2n+1)\binom{2n}{n}2^{-2n} - 1, \tag{6.9}$$

так что

$$E(N_{2n}) \sim 2(n/\pi)^{1/2}. \tag{6.10}$$

[$E(N_{2n})$ есть величина порядка $n^{1/2}$, а не возрастает линейно вместе с n].

Доказательство. Воспользовавшись формулой (1.8) гл. XI, мы можем вычислить $E(N_n)$ по «хвостам» (6.1). Имеем

$$E(N_n) = \sum_{r=1}^{\infty} P\{N_n \geqslant r\} = \sum_{r=1}^{\infty} P\{T^{(r)} \leqslant n\}. \tag{6.11}$$

Производящая функция величины $T^{(r)}$ равна $F^r(s)$, где $F(s) = 1 - (1-s^2)^{\frac{1}{2}}$ согласно (3.16). По теореме 1 § 1 гл. XI производящая функция вероятностей $P\{T^{(r)} \leqslant n\}$ имеет вид $F^{(r)}(s)(1-s)^{-1}$. Исходя из (6.11), теперь нетрудно установить для производящей функции $\{E(N_n)\}$ следующее выражение:

$$\sum E(N_n)s^n = \frac{F(s)}{(1-s)(1-F(s))} = \frac{1+s}{(1-s^2)^{3/2}} - \frac{1}{1-s}. \tag{6.12}$$

Отсюда вытекает, что

$$E(N_{2n}) = E(N_{2n+1}) = (-1)^n\binom{-3/2}{n} - 1. \tag{6.13}$$

Эту формулу удобно записать в форме (6.9) [используя (12.5) гл. II].

Любопытные следствия из этой теоремы подробно обсуждались в § 6 гл. III. Согласно теореме 2 этого параграфа, случайная величина $N_n \cdot n^{-2}$ имеет асимптотическое распределение, которое определяется положительной частью нормального распределения. Нормировка $N_n \cdot n^{-2}$ резко отличается от нормировки, которой мы воспользовались в теореме 1 настоящего параграфа.

§ 7*) Приложения к теории серий успехов

В дальнейшем r означает фиксированное целое положительное число, а рекуррентное событие $\mathfrak{E}$ означает появление серии успехов длины r в последовательности испытаний Бернулли. Важно, что длина серии понимается в смысле определения, данного в примере 1, а. В противном случае, серии не являются рекуррентными событиями и вычисления становятся более сложными. Как и прежде [см. (2.1) и (2.2)], *u_n обозначает вероятность наступления $\mathfrak{E}$ при n-м испытании, а f_n — вероятность того, что первая серия длины r появится при n-м испытании.*

Вероятность того, что r испытаний с номерами $n, n-1, \ldots, n-r+1$ приведут каждое к успеху, очевидно, равна p^r. В этом случае событие $\mathfrak{E}$ произойдет при одном из этих r испытаний; вероятность того, что $\mathfrak{E}$ произойдет при $(n-k)$-м испытании ($k=0, 1, \ldots, r-1$) и что последующие k испытаний приведут к успехам, равна $u_{n-k}p^k$. Поскольку такие r возможностей взаимно исключают друг друга, мы получаем следующее рекуррентное соотношение [1]):

$$u_n + u_{n-1}p + \ldots + u_{n-r+1}p^{r-1} = p^r. \tag{7.1}$$

Это равенство верно при $n \geqslant r$. Очевидно,

$$u_1 = u_2 = \ldots = u_{r-1} = 0; \quad u_0 = 1. \tag{7.2}$$

Умножим теперь (7.1) на s^n и просуммируем по $n = r, r+1, r+2, \ldots$. В силу равенств (7.2) получим слева

$$\{U(s) - 1\}(1 + ps + p^2s^2 + \ldots + p^{r-1}s^{r-1}), \tag{7.3}$$

а справа $p^r(s^r + s^{r+1} + \ldots)$. Суммируя эти две геометрические прогрессии, получим

$$\{U(s) - 1\}\frac{1-(ps)^r}{1-ps} = \frac{p^rs^r}{1-s} \tag{7.4}$$

или

$$U(s) = \frac{1 - s + qp^rs^{r+1}}{(1-s)(1-p^rs^r)}. \tag{7.5}$$

Применяя формулу (3.2), получим *производящую функцию времени возвращения*:

$$F(s) = \frac{p^rs^r(1-ps)}{1-s+qp^rs^{r+1}} = \frac{p^rs^r}{1-qs(1+ps+\ldots+p^{r-1}s^{r-1})}. \tag{7.6}$$

---

*) § 7 и § 8 посвящены специальным вопросам и могут быть опущены при первом чтении.

[1]) Классический подход заключается в составлении рекуррентной формулы для f_n. Этот метод сложнее и неприменим, например, для произвольных серий или для событий вида *УУННУУ*, для которых наш метод сохраняет свою силу (см. пример 8, в).

Тот факт, что $F(1)=1$, показывает, что в бесконечно продолжительной последовательности испытаний число серий наперед заданной длины рано или поздно превзойдет сколь угодно большую величину. Среднее значение времени возвращения μ можно получить непосредственно из (7.1), так как мы знаем, что $u_n \to \mu^{-1}$. Если же нас интересует и дисперсия, то проще всего вычислить первую и вторую производные $F(s)$. Для этой цели лучше воспользоваться первым вариантом формулы (7.6). Несложные выкладки показывают, что *математическое ожидание и дисперсия времени возвращения для серии успехов длины r равны* соответственно

$$\mu = \frac{1-p^r}{qp^r}, \qquad \sigma^2 = \frac{1}{(qp^r)^2} - \frac{2r+1}{qp^r} - \frac{p}{q^2}. \tag{7.7}$$

Из теоремы 1 предыдущего параграфа следует, что при больших n *число N_n серий длины r, полученных в n испытаниях, имеет приблизительно нормальное распределение*, т. е. что при фиксированных $\alpha < \beta$ вероятность неравенства

$$\frac{n}{\mu} + \frac{\alpha\sigma n^{1/2}}{\mu^{3/2}} < N_n < \frac{n}{\mu} + \frac{\beta\sigma n^{1/2}}{\mu^{3/2}} \tag{7.8}$$

стремится к $\Phi(\beta) - \Phi(\alpha)$. Впервые это было доказано Мизесом, но без теории рекуррентных событий доказательство требует довольно длинных вычислений. В табл. 2 указаны математические ожидания для ряда типичных времен возвращения.

Таблица 2

Среднее время возвращения для серий успехов, когда производится по одному испытанию в секунду

Длина серии, r	$p=0,6$	$p=0,5$ (монета)	$p=1/6$ (кость)
5	30,7 сек.	1 мин.	2,6 часа
10	6,9 мин.	34,1 мин.	28,0 мес.
15	1,5 часа	18,2 часа	18 098 лет
20	19 часов	25,3 дня	140,7 млн. лет

Метод разложения на простые дроби (§ 4 гл. XI) позволяет получить хорошую приближенную формулу для f_n. Из второго варианта формулы (7.6) нетрудно вывести, что знаменатель имеет единственный *положительный корень* $s=x$.

Для любого действительного или мнимого числа s, по модулю не превосходящего x, имеем

$$|qs(1+ps+\ldots+p^{r-1}s^{r-1})| \leqslant qx(1+px+\ldots+p^{r-1}x^{r-1}) = 1, \tag{7.9}$$

причем знак равенства возможен лишь в случае, если все члены левой части имеют один и тот же аргумент, т. е. если $s = x$. Поэтому x — наименьший по модулю корень знаменателя (4.6). Следовательно, можно воспользоваться формулой (4.9) гл. XI, положив в ней $s_1 = x$. Коэффициент ρ_1 легко вычисляется по формуле (4.5) при $U(s) = p^r s^r (1 - ps)$ и $V(s) = 1 - s + qp^r s^{r+1}$.

Пользуясь тем, что $V(s) = 0$, получаем

$$f_n \sim \frac{(x-1)(1-px)}{(r+1-rx)q} \cdot \frac{1}{x^{n+1}}. \tag{7.10}$$

Вероятность того, что в n испытаниях не окажется ни одной серии, равна $q_n = f_{n+1} + f_{n+2} + \ldots$. Мы получим эту вероятность из (7.18), суммируя геометрическую прогрессию [1])

$$q_n \sim \frac{1-px}{(r+1-rx)q} \cdot \frac{1}{x^{n+1}}. \tag{7.11}$$

Итак, *вероятность отсутствия серий успехов длины r в последовательности из n испытаний асимптотически дается формулой* (7.11). Табл. 3 показывает, что эта формула дает поразительно хорошее приближение даже при очень малых n, и что ошибка приближения быстро уменьшается с ростом n. Мы имеем здесь типичный пример силы метода производящих функций в соединении с методом разложения на простые дроби.

Таблица 3

Вероятность не иметь ни одной серии длины $r = 2$ в n испытаниях с $p = 1/2$

n	q_n точное	Приближенная формула (7.11)	Ошибка	Верхняя оценка ошибки по формуле (3.10)
2	0,75	0,76631	0,0163	0,0835
3	0,625	0,61996	0,0080	0,042
4	0,500	0,50156	0,0016	0,021
5	0,40625	0,40577	0,0005	0,010

Численные оценки. Для читателя, интересующегося практической стороной вопроса, пользуемся случаем показать, что численные оценки, связанные с разложением на простые дроби, зачастую менее сложны, чем это может показаться с первого взгляда. Кроме того, на конкретном примере мы продемонстрируем, как можно оценивать ошибку асимптотических формул.

В связи с приближенной формулой (7.11) возникают два вопроса: во-первых, какова ошибка, связанная с тем, что мы пренебрегли влиянием $r - 1$ корней, отличных от x, и, во-вторых, каково (хотя бы приближенно) значение наименьшего по абсолютной величине корня x.

Из первого варианта формулы (7.6) следует, что все корни знаменателя $F(s)$ удовлетворяют уравнению

$$s = 1 + qp^r s^{r+1}, \tag{7.12}$$

хотя уравнение (7.12) имеет дополнительно посторонний корень $s = p^{-1}$. При положительных значениях s график функции $f(s) = 1 + qp^r s^{r+1}$ представляет выпуклую вниз кривую, которая пересекает биссектрису $y = s$ в точках $s = x$ и $s = p^{-1}$. Между этими двумя корнями график проходит под биссектрисой. Далее, $f'(p^{-1}) = (r+1)q$. Если эта величина превосходит единицу, то график $f(s)$ пересекает биссектрису при $s = p$ снизу и, значит, $p^{-1} > x$. Для определенности предположим, что

$$(r+1)q > 1. \tag{7.13}$$

В этом случае $x < p^{-1}$ и $f(s) < s$ при $x < s < p^{-1}$. Отсюда вытекает, что при любом комплексном числе s, таком, что $x < |s| < p^{-1}$ имеет место неравенство $|f(s)| \leqslant f(|s|) < |s|$, так что ни одного корня s_k не лежит в кольце $x < |s| < p^{-1}$. Так как s — наименьший по абсолютной величине корень, то

$$|s_k| > p^{-1} \tag{7.14}$$

для любого $s_k \neq x$. Исследуя производную в (7.12), нетрудно убедиться в том, что все корни простые.

Точное выражение для q_n есть сумма r слагаемых; каждому корню знаменателя в (7.6) отвечает одно слагаемое. Слагаемое A_k, соответствующее корню s_k, получается из правой части (7.11) заменой x на s_k. Таким образом

$$A_k = \frac{ps_k - 1}{rs_k - (r+1)} \cdot \frac{1}{qs_k^{n+1}}. \tag{7.15}$$

Нам нужна верхняя оценка для первой дроби справа. Для этой цели заметим, что при $s > p^{-1} > (r+1)r^{-1}$

$$\left| \frac{pse^{i\theta} - 1}{rse^{i\theta} - (r+1)} \right| \leqslant \frac{ps+1}{rs+r+1}. \tag{7.16}$$

Действительно, легко видеть, что выражение слева принимает наибольшее и наименьшее значение при $\theta = 0$ и $\theta = \pi$, и непосредственная подстановка показывает, что 0 соответствует минимуму, а π — максимуму. Если учесть (7.13) и (7.14), то отсюда вытекает неравенство

$$|A_k| < \frac{2p^{n+1}}{(r+1+rp^{-1})q} < \frac{2p^{n+2}}{rq(1+p)}. \tag{7.17}$$

Следовательно, *ошибка формулы* (7.11), *связанная с тем, что мы пренебрегли влиянием* $(r-1)$ *корней, отличных от* x, *по модулю меньше, чем*

$$\frac{2(r-1)p}{rq(1+p)}. \tag{7.18}$$

Значение корня x проще всего вычислить исходя из уравнения (7.12) с помощью последовательных приближений $x_0 = 1$, $x_{\nu+1} = f(x_\nu)$. Последовательность приближений сходится к x монотонно, и любой ее член является нижней оценкой для x. С другой стороны, любая величина s, такая, что $s > f(s)$, дает верхнюю оценку. Легко видеть, что

$$x = 1 + qp^r + (r+1)(qp^r)^2 + \dots \tag{7.19}$$

§ 8*). Более общие рекуррентные события

Наш метод применим к значительно более интересным задачам, которые можно рассматривать как дальнейшее обобщение теории серий.

Примеры. а) Пусть $\mathscr{E}$ означает событие «*либо серия успехов длины r, либо серия неудач длины* ρ». Мы имеем дело с *двумя* рекуррентными событиями $\mathscr{E}_1$ и $\mathscr{E}_2$, где $\mathscr{E}_1$ означает «серия успехов длины r», а $\mathscr{E}_2$ — «серия неудач длины ρ». Событию $\mathscr{E}_1$ соответствует производящая функция (7.5), которую мы обозначим теперь $U_1(s)$. Производящая функция, соответствующая событию $\mathscr{E}_2$, получается из (7.5), если поменять местами p и q и заменить r на ρ. Вероятность u_n появления события $\mathscr{E}$ при n-м испытании равна сумме соответствующих вероятностей для событий $\mathscr{E}_1$ и $\mathscr{E}_2$. Исключением является лишь случай $n=0$, для которого $u_0=1$. Из сказанного следует, что

$$U(s)=U_1(s)+U_2(s)-1. \tag{8.1}$$

Производящая функция времени возвращения события $\mathscr{E}$ равна $F(s)=1-U^{-1}(s)$, или

$$F(s)=\frac{(1-ps)\,p^rs^r(1-q^\rho s^\rho)+(1-qs)\,q^\rho s^\rho(1-p^rs^r)}{1-s+qp^rs^{r+1}+pq^\rho s^{\rho+1}-p^rq^\rho s^{r+\rho}}. \tag{8.2}$$

Среднее время возвращения получается дифференцированием:

$$\mu=\frac{(1-p^r)(1-q^\rho)}{qp^r+pq^\rho-p^rq^\rho}. \tag{8.3}$$

При $\rho\to\infty$ это выражение стремится к среднему времени возвращения серий успехов, определяемому формулой (7.7).

б) В § 1 гл. VIII мы вычислили вероятность x того, что *серия успехов длины r появится раньше, чем серия неудач длины* ρ. Предположим теперь, что игра продолжается до тех пор, пока не появится впервые серия успехов длины r или серия неуспехов длины ρ и вычислим распределение вероятностей *продолжительности игры*. Более точно: мы определяем два рекуррентных события $\mathscr{E}_1$ и $\mathscr{E}_2$ так же, как и в примере «а». Пусть x_n — вероятность того, что $\mathscr{E}_1$ произойдет впервые при n-м испытании, причем осуществлению события $\mathscr{E}_1$ не предшествует ни одно осуществление события $\mathscr{E}_2$; f_n — вероятность того, что $\mathscr{E}_1$ произойдет впервые при n-м испытании (без каких-либо условий относительно осуществления $\mathscr{E}_2$). Пусть y_n и g_n — аналогичные вероятности для $\mathscr{E}_2$.

*) См. примечание на стр. 138.

Производящая функция для f_n определяется формулой (7.6), а $G(s)$ получается из нее заменой p на q и r на ρ. Для x_n и y_n имеем очевидные рекуррентные соотношения:

$$\begin{aligned} x_n &= f_n - (y_1 f_{n-1} + y_2 f_{n-2} + \dots + y_{n-1} f_1); \\ y_n &= g_n - (x_1 g_{n-1} + x_2 g_{n-2} + \dots + x_{n-1} g_1). \end{aligned} \tag{8.4}$$

Это — равенства типа композиций, и поэтому для соответствующих производящих функций получаем

$$\begin{aligned} X(s) &= F(s) - Y(s) F(s); \\ Y(s) &= G(s) - X(s) G(s). \end{aligned} \tag{8.5}$$

Равенства (8.5) суть два линейных уравнения с неизвестными $X(s)$ и $Y(s)$. Из них получаем

$$X(s) = \frac{F(s)\{1 - G(s)\}}{1 - F(s) G(s)}; \quad Y(s) = \frac{G(s)\{1 - F(s)\}}{1 - F(s) G(s)}. \tag{8.6}$$

Выражения для x_n и y_n могут быть получены с помощью метода разложения на простые дроби. При $s = 1$ получаем $X(1) = \sum x_n = x$ (где x — вероятность появления события $\mathscr{E}_1$ до появления события $\mathscr{E}_2$). Числитель и знаменатель в формуле (2.6) для $X(s)$ обращаются в 0 при $s = 1$, и поэтому $X(1)$ вычисляется по правилу Лопиталя дифференцированием числителя и знаменателя:

$$X(1) = G'(1) / \{F'(1) + G'(1)\}.$$

Пользуясь значениями $F'(1) = (1 - p^r)/qp^r$ и $G'(1) = (1 - q^\rho)/pq^\rho$, [см. (7.7)], мы получим значение $X(1)$, которое было приведено в формуле (1.3) гл. VIII.

в) Рассмотрим рекуррентное событие, определяемое последовательностью *УУННУУ*. Повторяя рассуждения предыдущего параграфа, убеждаемся, что

$$p^4 q^2 = u_n + p^2 q^2 u_{n-4} + p^3 q^2 u_{n-5}. \tag{8.7}$$

Так как мы знаем, что $u_n \to \mu^{-1}$, то нетрудно найти среднее время возвращения $\mu = p^{-4} q^{-2} + p^{-2} + p^{-1}$. При $p = q = \frac{1}{2}$ получаем $\mu = 70$, в то время как среднее время возвращения для серий успехов длины 6 равно 126. Это показывает, что при бросании правильной монеты *имеется существенная разница между сериями гербов и другими сериями той же длины.*

§ 9. Особенность времен ожидания с геометрическим распределением

Геометрическое распределение времени ожидания обладает интересным и важным свойством, которое отсутствует у всех остальных распределений. Рассмотрим последовательность испытаний Бернулли, и пусть T — число испытаний до первого успеха включительно. Тогда $P\{T > k\} = q^k$. Предположим, что в данной последовательности испытаний первые m испытаний не дали ни одного успеха. Время ожидания T, начиная с m-й неудачи и до ближайшего успеха, не зависит от числа предшествовавших неудач и имеет то же распределение $\{q^k\}$. Другими словами, вероятность того, что время ожидания продлится дополнительно k всегда равна начальной вероятности того, что полное время ожидания превысит k. Если время жизни атома или срок службы прибора имеют геометрическое распределение, то это значит, что *возраст не играет никакой роли*; как бы долго ни жил атом, вероятность того, что он распадется в следующий момент, остается постоянной. Атомы радиоактивных элементов действительно обладают этим свойством (с той оговоркой, что в случае непрерывного времени роль геометрического распределения играет показательное распределение). Наоборот, если некоторое явление характеризуется полным отсутствием последействия, или «старения», то распределение вероятностей его продолжительности должно быть геометрическим или показательным. Типичным в этом отношении является хорошо известный вид телефонных разговоров, на который часто ссылаются как на пример процесса, целиком зависящего от минутных побуждений. Окончание является случайным мгновенным действием, никак не связанным с продолжительностью предшествующей болтовни. Напротив, знание того, что в течение 5 минут не прошел ни один трамвай, обычно увеличивает наши надежды на то, что он скоро появится. В игре с бросанием монеты вероятность того, что полное число гербов равно полному числу решеток при 2-м испытании, равна $^1/_2$. Однако вероятность того, что то же событие произойдет после двух дополнительных испытаний (при 4-м бросании), равна уже $^1/_4$. Это пример «последействия».

Для того чтобы перейти к строгим формулировкам, предположим, что время ожидания T принимает значения $0, 1, 2, \ldots$ с вероятностями $p_0, p_1, p_2, \ldots$.

Пусть распределение величины T обладает следующим свойством: *условная вероятность того, что время ожидания окончится на k-м испытании, при условии, что оно не кончилось раньше, равна p_0 (вероятности окончания при первом испытании). Мы утверждаем, что $p_k = (1 - p_0)^k p_0$, так что T имеет геометрическое распределение.*

Для доказательства введем опять «хвосты» распределения

$$q_k = p_{k+1} + p_{k+2} + p_{k+3} + \ldots = P\{T > k\}.$$

Наша гипотеза записывается в виде неравенства $T > k - 1$, и ее вероятность равна q_{k-1}. Условная вероятность события $T = k$ равна поэтому p_k/q_{k-1}. Наше предположение состоит в том, что при всех $k \geqslant 1$

$$\frac{p_k}{q_{k-1}} = p_0. \tag{9.1}$$

Далее, $p_k = q_{k-1} - q_k$, откуда

$$\frac{q_k}{q_{k-1}} = 1 - p_0. \tag{9.2}$$

Так как $q_0 = p_1 + p_2 + \ldots = 1 - p_0$, то $q_k = (1-p_0)^{k+1}$, откуда $p_k = q_{k-1} - q_k = (1-p_0)^k p_0$, что и требовалось доказать.

В общей теории стохастических процессов описанное свойство отсутствия «последействия» связано с *марковским свойством*. Мы вернемся к этой теме в § 10 гл. XV.

§ 10 *) Доказательство теоремы 3 § 3

В § 3 мы пропустили доказательство теоремы 3. Эту теорему можно сформулировать либо как «тауберову» теорему о степенных рядах, либо следующим элементарным образом. *Пусть дана последовательность* $\{f_n\}$, *такая, что* $f_0 = 0$, $f_n \geqslant 0$, $\sum f_n = 1$ *и наибольший общий делитель тех* n, *для которых* $f_n > 0$, *равен единице. Положим* $u_0 = 1$ *и определим* u_n *при* $n \geqslant 1$ *формулой*

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \ldots + f_n u_0. \tag{10.1}$$

Тогда $u_n \to 1/\mu$, *где* $\mu = \sum n f_n$ (*и* $u_n \to 0$, *если ряд* $\sum n f_n$ *расходится*).

Для доказательства положим

$$r_n = f_{n+1} + f_{n+2} + \ldots, \tag{10.2}$$

так что [по формуле (1.8) гл. XI] имеем

$$\mu = \sum r_n. \tag{10.3}$$

Из (10.2) получаем $r_0 = 1$, $f_1 = r_0 - r_1$, $f_2 = r_1 - r_2$ и т. д. Подставляя эти значения в (10.1), убеждаемся, что $r_0 u_n + r_1 u_{n-1} + \ldots$ $\ldots + r_n u_0 = r_0 u_{n-1} + r_1 u_{n-2} + \ldots + r_{n-1} u_0$. Если мы обозначим левую часть через A_n, то правая часть окажется равной A_{n-1}. Наше равенство означает, что все A_n равны между собой. Далее, $A_0 = r_0 u_0 = 1$, так что $A_n = 1$ при всех n. Итак, при любом n

$$r_0 u_n + r_1 u_{n-1} + \ldots + r_n u_0 = 1. \tag{10.4}$$

Из (10.1) по индукции следует, что $u_n \leqslant 1$. Поэтому существует такое число $\lambda = \limsup u_n$, что для любого $\varepsilon > 0$ и всех достаточно

*) Этот параграф может быть опущен при первом чтении.

больших n величина $u_n < \lambda+\varepsilon$, и в то же время существует такая последовательность $n_1, n_2, \ldots$, что $u_{n_\nu} \to \lambda$. Выберем такое целое число $j > 0$, что $f_j > 0$. Мы утверждаем, что $u_{n_\nu - j} \to \lambda$. Если бы это было не так, то можно было бы найти сколь угодно большие индексы n, для которых одновременно выполнялись бы неравенства

$$u_n > \lambda - \varepsilon, \quad u_{n-j} < \lambda' < \lambda. \tag{10.5}$$

Выберем теперь N настолько большим, что $r_N < \varepsilon$. Поскольку $u_k \leqslant 1$, согласно (10.1), при $n > N$ имеем

$$u_n \leqslant f_0 u_n + f_1 u_{n-1} + \ldots + f_N u_{n-N} + \varepsilon. \tag{10.6}$$

При достаточно большом n каждое u_k в правой части будет меньше $\lambda+\varepsilon$, а $u_{n-j} < \lambda'$. Отсюда

$$u_n < (f_0 + f_1 + \ldots + f_{j-1} + f_{j+1} + \ldots + f_N)(\lambda+\varepsilon) + f_j\lambda' + \varepsilon \leqslant$$
$$\leqslant (1 - f_j)(\lambda+\varepsilon) + f_j\lambda' + \varepsilon < \lambda + 2\varepsilon - f_j(\lambda - \lambda'). \tag{10.7}$$

Если выбрать ε настолько малым, чтобы $f_j(\lambda-\lambda') > 3\varepsilon$, то последнее неравенство будет противоречить первому из неравенств (10.5), и поэтому допущение $\lambda' < \lambda$ невозможно. Этим доказано, что если $u_{n_\nu} \to \lambda$, то и $u_{n_\nu - j} \to \lambda$. Повторно применяя этот вывод, убеждаемся, что *если* $f_j > 0$ *и* $u_{n_\nu} \to \lambda = \limsup u_n$, *то также*

$$u_{n_\nu - j} \to \lambda, \quad u_{n_\nu - 2j} \to \lambda, \quad u_{n_\nu - 3j} \to \lambda \text{ и т. д.}$$

Для простоты рассмотрим сначала случай, когда $f_1 > 0$. Тогда можно положить $j = 1$, откуда заключаем, что $u_{n_\nu - k} \to \lambda$ при любом фиксированном k. Полагая в (10.4) $n = n_\nu$, получим

$$1 \geqslant r_0 u_{n_\nu} + r_1 u_{n_\nu - 1} + \ldots + r_N u_{n_\nu - N}. \tag{10.8}$$

При фиксированном N каждое $u_{n_\nu - k} \to \lambda$, так что $1 \geqslant \lambda(r_0 + r_1 + \ldots + r_N)$. Поскольку N произвольно, заключаем, что $1 \geqslant \lambda\mu$ или $\lambda \leqslant 1/\mu$.

Положим теперь $\gamma = \liminf u_n$. Аналогичное рассуждение показывает, что для любой последовательности n_ν, для которой $u_{n_\nu} \to \gamma$, величина $u_{n_\nu - k}$ также стремится к γ. Если N настолько велико, что $r_N < \varepsilon$, то, согласно (10.4),

$$1 \leqslant r_0 u_{n_\nu} + \ldots + r_N u_{n_\nu - N} + \varepsilon; \tag{10.9}$$

при этом $u_{n_\nu - k} \to \gamma$, так что $1 \leqslant (r_0 + \ldots + r_N)\gamma + \varepsilon$, откуда $\mu\gamma \geqslant 1$. Но, по определению, $\gamma \leqslant \lambda$. Следовательно, $\gamma = \lambda = 1/\mu$, что и требовалось доказать.

Остается случай, когда $f_1 = 0$. Рассмотрим множество всех j, для которых $f_j > 0$. Среди них можно найти конечную совокупность чисел $a, b, c, \ldots, m$, общим наибольшим делителем которых является единица. Мы знаем, что если $u_{n_\nu} \to \lambda$, то при любых фиксированных $x > 0$, $y > 0$, ..., $w > 0$ и $u_{n_\nu - xa} \to \lambda$, $u_{n_\nu - yb} \to \lambda$ и т. д. Отсюда следует, что также $u_{n_\nu - xa - yb - \ldots - wm} \to \lambda$. Другими словами, если целое число k имеет вид $k = xa + yb + \ldots + wm$, где x, y, ..., w — целые положительные числа, то $u_{n_\nu - k} \to \lambda$. Из элементарной теории чисел известно, что *всякое целое* число k, превосходящее произведение $abc \ldots m$, может быть записано в виде $k = xa + yb + \ldots$ $\ldots + wm$, где x, y, ..., w — целые положительные числа. Это значит, что при $k > abc \ldots m$ мы имеем $u_{n_\nu - k} \to \lambda$. Чтобы получить (10.8), достаточно теперь применить (10.4) к $n = n_\nu + ab \ldots m$. Остальная часть доказательства проходит без изменений.

§ 11. Задачи

1. Пусть $F(s)$ — многочлен. Доказать для этого случая все теоремы § 3, пользуясь методом разложения на простые дроби § 7 гл. XI.

2. Пусть многократно бросаются r монет, и пусть $\mathcal{E}$ — рекуррентное событие, состоящее в том, что для каждой из r монет число выпадений герба равно числу выпадений решетки. Является ли $\mathcal{E}$ достоверным или недостоверным событием? Оценить вероятность появления события $\mathcal{E}$ для наименьшего r, при котором оно недостоверно.

3. Пусть $\{X_k\}$ — последовательность взаимно независимых случайных величин, имеющих одно и то же распределение $P\{X_k = a\} = \dfrac{b}{a+b}$, $P\{X_k = -b\} = \dfrac{a}{a+b}$, где a и b — положительные целые числа. Пусть $\mathcal{E}$ означает событие $S_n = 0$.

Доказать, что $\mathcal{E}$ достоверно.

4. Пусть $\{X_k\}$ — произвольная последовательность взаимно независимых одинаково распределенных случайных величин, и пусть $\mathcal{E}$ означает $S_n = 0$, $S_1 \leqslant 0$, $S_2 \leqslant 0$, ..., $S_{n-1} \leqslant 0$. Тогда $\mathcal{E}$ — недостоверное рекуррентное событие, исключая единственный случай, когда $P\{X_k = 0\} = 1$.

5. Обобщить пример § 4 на случай произвольных взвешенных средних. Найти предел.

Замечание. *Задачи* 6—8 *относятся к испытаниям Бернулли с* $p = q = 1/2$ (*бросания монеты*). *Производящая функция* $F(s) = 1 -$ $- (1 - s^2)^{1/2}$ *времени возвращения к равновесию, считается известной.*

6. Пусть $\mathcal{E}_1$ — рекуррентное событие $S_n = 0$, $S_{n-1} < 0$. Найти производящую функцию $F_1(s)$ времени возвращения.

7. Продолжение. Найти производящую функцию времени возвращения для рекордов [пример (1.2)]. (Заметить, что она совпадает с производящей функцией *времени первого достижения* 1, которая была найдена в § 3 гл. XI.)

8. Продолжение. Доказать *теорему*: *вероятность того, что при* $2n$-*м или* $(2n+1)$-*м испытаниях сумма* S_k *принимает значение, которого не было* (*т. е. имеет место первое достижение*), *равна вероятности того, что* $S_{2n} = 0$.

9. В *задаче о подсчете частиц* (1.6): а) найти производящую функцию времени возвращения. (Каков его физический смысл?)

б) Пусть Z_n означает число регистраций при первых испытаниях. Найти $E(Z_n)$ и $D(Z_n)$.

10. Счетчики типа II отличаются от счетчиков, описанных в примере 1, б, тем, что каждый успех блокирует их на r единиц времени (в том числе в продолжении $(r-1)$ испытаний, следующих за успехом, однако успехи, которые имеют место в течение этого периода, увеличивают его продолжительность. Решить задачу 9 для этого случая.

11. Приближенно определить вероятность того, что при 10 000 бросаний монеты число серий успехов длины 3 будет заключено между 700 и 730.

12. Пусть при последовательных бросаниях монеты $\mathscr{E}$ означает рекуррентное событие *ГРГ*. Пусть r_n — вероятность того, что $\mathscr{E}$ ни разу не произойдет при n испытаниях. Найти соответствующую производящую функцию и воспользоваться методом разложения на простые дроби для получения асимптотического выражения.

13. Показать, что в примере 8, б ожидаемая продолжительность игры равна $\mu_1\mu_2/(\mu_1+\mu_2)$, где μ_1 и μ_2 соответственно средние времена возвращения для серии успехов длины r и серии неудач длины ρ.

14. Возможными исходами каждого испытания являются A, B и C; соответствующие вероятности равны α, β и γ $(\alpha+\beta+\gamma=1)$. Найти производящую функцию вероятностей того, что в n испытаниях не появится ни одна серия длины r: а) исходов A, б) исходов B или исходов A, в) любых однотипных исходов.

15. Продолжение. Найти вероятность того, что первая A-серия длины r предшествует первой B-серии длины ρ и оканчивается при n-м испытании. [Заметим, что эта задача *не* сводится к задаче примера (8.б) с $p=\alpha/(\alpha+\beta)$, $q=\beta/(\alpha+\beta)$.]

Замечание. *Следующие задачи относятся к теории восстановления и, в частности, к примеру* 5, б.

16. Постоянство совокупности. Доказать по индукции, что для величин (5.9) $\sum\limits_k v_k(n)=N$ при любом n.

17. Пусть дано распределение срока службы $p_k=q^{k-1}p$ $(p+q=1)$. Найти u_n и предельное распределение в предположении, что начальная совокупность состоит из N элементов нулевого возраста.

18. Распределение возрастов называется *устойчивым*, если $v_k(n)$ не зависит от n. Показать, что распределение устойчиво тогда и только тогда, когда $v_k=Cr_k$, где C — постоянная.

19. Пусть $\mathscr{E}$ — достоверное непериодическое рекуррентное событие. Допустим, что время возвращения имеет конечное математическое ожидание μ и дисперсию σ^2. Положим $q_n=f_{n+1}+f_{n+2}+\ldots$ и $r_n=q_{n+1}+q_{n+2}+\ldots$. Показать, что ряды, определяющие производящие функции $Q(s)$ и $R(s)$ сходятся при $s=1$. Доказать, что

$$\sum\left(u_n-\frac{1}{\mu}\right)s^n=\frac{R(s)}{\mu Q(s)};\tag{11.1}$$

вывести отсюда, что

$$\sum\left(u_n-\frac{1}{\mu}\right)=\frac{\sigma^2-\mu+\mu^2}{2\mu^2}.\tag{11.2}$$

20. Пусть $\mathscr{E}$ — достоверное рекуррентное событие, и N_r — число появлений события $\mathscr{E}$ при r испытаниях. Доказать, что $E(N_r)=u_1+\ldots+u_r$, и, следовательно,

$$E(N_r)\sim\frac{r}{\mu}.\tag{11.3}$$

21. Продолжение. Доказать, что

$$E(N_r^2)=u_1+\ldots+u_r+2\sum_{j=1}^{r-1}u_j(u_1+\ldots+u_{r-j})$$

и, следовательно, $E(N_r^2)$ есть коэффициент при s^r в

$$\frac{F^2(s)+F(s)}{(1-s)\{1-F(s)\}^2}. \tag{11.4}$$

(Заметим, что этот результат можно сформулировать значительно изящнее, если воспользоваться двойными производящими функциями.)

22. Положим $q_{k,n}=P\{N_k=n\}$. Показать, что $q_{k,n}$ есть коэффициент при s^k в разложении функции

$$F^n(s)\frac{\{1-F(s)\}}{1-s}. \tag{11.5}$$

Показать, что $E(N_r)$ и $E(N_r^2)$ суть соответственно коэффициенты при s^r в выражениях (11.4) и

$$\frac{F(s)}{(1-s)\{1-F(s)\}}. \tag{11.6}$$

23. Пользуясь обозначениями задачи 19, показать, что

$$\frac{F(s)}{(1-s)\{1-F(s)\}}=-\frac{1}{1-s}+\frac{1}{\mu(1-s)^2}+\frac{R(s)}{\mu\{1-F(s)\}}. \tag{11.7}$$

Пользуясь задачей 22, вывести отсюда, что

$$E(N_r)=\frac{r+1}{\mu}+\frac{\sigma^2-\mu-\mu^2}{2\mu^2}+\varepsilon_r, \tag{11.8}$$

где $\varepsilon_r\to 0$.

24. Продолжение. Применяя аналогичные рассуждения, показать, что

$$E(N_r^2)=\frac{(r+2)(r+1)}{\mu^2}+\frac{2\sigma^2-2\mu-\mu^2}{\mu^3}r+\alpha_r, \tag{11.9}$$

где α_r ограничено. Отсюда

$$D(N_r)\sim\frac{\sigma^2}{\mu^3}r. \tag{11.10}$$

25. Пусть в последовательности испытаний Бернулли $q_{k,n}$ обозначает вероятность появления n серий успехов длины r при k испытаниях. Пользуясь задачей 22, показать, что производящая функция $Q_k(x)=\sum q_{k,n}x^n$ есть коэффициент при s^k в выражении

$$\frac{1-p^rs^r}{1-s+qp\,s^{r+1}-(1-ps)\,p\,s\,x}.$$

Показать, далее, что наименьший по абсолютной величине корень знаменателя равен $s_1\approx 1+qp^r(1-x)$.

26. Продолжение. *Распределение Пуассона для длинных серий*[1]). Если число испытаний k и длина серии r стремятся к бесконечности так, что $kqp^r\to\lambda$, то вероятность наличия n серий длины r стремится к $e^{-\lambda}\cdot\lambda^n/n!$

Указание. Пользуясь предыдущей задачей, показать, что производящая функция асимптотически равна $\{1+qp^r(1-x)\}^{-k}\sim e^{-\lambda(1-x)}$. Воспользоваться *теоремой непрерывности* § 6 гл. XI.

---

[1]) Сформулированная теорема доказана Мизесом, но метод, которым пользуемся мы, более элементарен.

ГЛАВА XIV

СЛУЧАЙНЫЕ БЛУЖДАНИЯ И ЗАДАЧИ О РАЗОРЕНИИ

§ 1. Общие понятия

Основная часть этой главы посвящена ряду задач, связанных с испытаниями Бернулли, в которых вероятности успеха и неудачи равны соответственно p и q. Для простоты и ясности изложения мы будем формулировать задачи и теоремы в терминах двух наглядных схем.

Во-первых, будем рассматривать игрока, выигрывающего рубль при каждом успехе и проигрывающего рубль при каждой неудаче. Будем предполагать, что игрок и его противник имеют в общей сложности a рублей, причем в начале игры первый из игроков имеет z, а второй $a-z$ рублей. Игра продолжается до тех пор, пока капитал первого игрока либо сократится до нуля, либо возрастет до a рублей, т. е. пока один из двух игроков не разорится. Нас интересует вероятность разорения игрока и распределение вероятностей продолжительности игры. Это — *классическая задача о разорении.*

Физические приложения и аналогии подсказывают другую наглядную схему. Представим себе, что испытания совершаются в моменты времени $t=1, 2, 3, \ldots$, и будем интерпретировать их исходы, как перемещения точки или *частицы* по оси x. В момент $t=0$ частица находится в положении $x=z$, а в моменты $t=1, 2, 3, \ldots$ она перемещается на один шаг вправо или влево в зависимости от успеха или неудачи соответствующего испытания. Таким образом, положение частицы в момент n изображает капитал игрока после окончания n-го испытания. Испытания кончаются, когда частица в первый раз достигает либо точки $x=0$, либо точки $x=a$. Мы будем говорить, что частица совершает *случайное блуждание.* Граничные положения $x=a$ и $x=0$ называются *поглощающими экранами*; разорение игрока интерпретируется, как *поглощение в точке* $x=0$. Будем говорить, что случайное блуждание *ограничено* возможными положениями $x=0, 1, \ldots, a$; при отсутствии поглощающих экранов случайное блуждание называется *неограниченным.* Если $p=q={}^1/_2$, то случайное блуждание называется симметричным. Физики пользуются схемой случайного блуждания как грубым описанием одномерных процессов диффузии и броуновского движения, где материальная частица подвержена действию большого числа столкновений или ударов со стороны молекул, сообщающих ей случайное движение. Если существует *течение* вправо, то удары слева более вероятны, и $p>q$.

В предельном случае при $a \to \infty$ мы получаем случайное блуждание на полупрямой. Частица начинает движение из некоторой точки $z > 0$ и продолжает его до тех пор, пока впервые не достигнет нуля. В этой формулировке мы узнаем *задачу о времени первого достижения*, которая была решена элементарными методами в гл. III (по крайней мере для симметричного случая) и с использованием производящих функций в § 3 гл. IX (см. также задачу 7 гл. XIII). Мы столкнемся с формулами, которые уже были нами получены, они будут выведены по-новому[1]).

В этой главе мы будем пользоваться методом *уравнений в конечных разностях*, которые служат введением в дифференциальные уравнения теории диффузии. Эта аналогия приводит естественным образом к различным видоизменениям и обобщениям классической задачи о разорении. Так, вместо поглощающих экранов можно рассмотреть другие граничные условия. Например, можно представить себе в точке $x = {}^1/_2$ отражающую стенку, обладающую тем свойством, что если частица выходит из $x = 1$ и движется влево, то она отражается в точке $x = {}^1/_2$ и возвращается в точку $x = 1$, вместо того чтобы достигнуть точки $x = 0$. Иными словами, когда частица находится в точке $x = 1$, она может с вероятностью p сдвинуться на один шаг вправо и с вероятностью q остаться на месте. Мы будем говорить, что в этом случае в точке $x = {}^1/_2$ находится *отражающий экран*. В терминах игры это соответствует соглашению о том, что, когда игрок проиграет свой последний рубль, этот рубль будет ему великодушно возвращен его противником, так что игра сможет продолжаться,

Как поглощающий, так и отражающий экраны являются частными случаями так называемого упругого экрана. Мы определим *упругий экран в нуле условием, что из точки* 1 *частица попадает с вероятностью p в точку* 2, *с вероятностью* δq *остается в точке* 1 *и с вероятностью* $(1 - \delta)\, q$ *движется налево и поглощается в нуле* (т. е. процесс прекращается). При $\delta = 0$ эта схема сводится к классической задаче о разорении или к поглощающему экрану, при $\delta = 1$ имеем отражающий экран. Если δ меняется от 0 до 1, то получается семейство промежуточных случаев. Чем больше δ, тем меньше вероятность того, что процесс прекратится; при двух отражающих экранах случайное блуждание никогда не кончится.

Параграфы 2 и 3 посвящены элементарному обсуждению классической задачи о разорении и некоторых связанных с ней вопросов. Следующие три параграфа технически более сложны (и могут быть опущены); в § 4 и 5 выводятся производящие функции, а из них

[1]) Обратно, некоторые новые результаты можно вывести, используя метод гл. III. Решение задачи о разорении методом отражения см. в задачах 7—9.

точные выражения для распределения продолжительности игры и т. п. В § 6 в общих чертах описан предельный переход к уравнению диффузии (формальные решения которого дают предельные распределения для случайного блуждания).

В § 7 рассуждения опять становятся элементарными и посвящены *случайным блужданиям на плоскости и в пространстве*, при которых возникают новые явления. Другому обобщению классической схемы посвящен § 8, где исследуется случайное блуждание на прямой, при котором частица уже не обязана перемещаться единичными шагами, а может менять свое положение произвольными скачками, кратными единице. Такие обобщенные блуждания привлекли широкий интерес в связи с теорией *последовательного анализа*, разработанной Вальдом.

В заключение следует подчеркнуть, что случайные блуждания можно рассматривать как цепи Маркова частного вида. Настоящая глава служит в какой-то мере введением к следующей, где ряд задач о случайных блужданиях (в частности, упругие экраны) будут рассмотрены с другой точки зрения.

Параграф, посвященный задачам, содержит существенные дополнения к тексту и наброски других возможных подходов. Можно надеяться, что сравнение разных методов окажется весьма поучительным. (Если читатель желает обратиться к тексту и графикам гл. III, он должен иметь в виду, что горизонтальную ось следует считать осью времени, а вертикальную — осью x.)

§ 2. Задача о разорении игрока

Рассмотрим задачу, сформулированную в начале настоящей главы. Пусть q_z — вероятность того, что игрок, в конце концов[1]), разорится, а p_z — вероятность того, что он выиграет. В терминах случайных блужданий q_z и p_z, соответственно — вероятности того, что частица, исходящая из точки z, будет поглощена экраном $x=0$ или экраном $x=a$. Мы покажем, что $p_z+q_z=1$, так что нет надобности рассматривать возможность бесконечной игры.

После первого испытания капитал игрока равен либо $z-1$, либо $z+1$, и поэтому при $1<z<a-1$

$$q_z = pq_{z+1} + qq_{z-1}. \tag{2.1}$$

[1]) Строго говоря, вероятность разорения определена лишь в пространстве элементарных исходов бесконечно продолжающейся игры. Однако можно ограничиться рассмотрением пространства исходов n испытаний. Вероятность разориться после числа испытаний, меньшего n, убывает с ростом n и поэтому имеет предел. Этот *предел* мы и называем «вероятностью разорения». Все вероятности, вычисляемые в настоящей главе, могут быть истолкованы аналогичным образом, без ссылок на бесконечные пространства элементарных исходов (см. введение к гл. VIII).

При $z=1$ первое испытание может привести к разорению, и (2.1) нужно заменить равенством $q_1=pq_2+q$. Аналогично при $z=a-1$ первое испытание может окончиться разорением противника, и поэтому $q_{a-1}=qq_{a-2}$. Чтобы придать всем этим равенствам одинаковый вид, положим

$$q_0=1, \quad q_a=0. \tag{2.2}$$

При таком соглашении вероятность разорения q_z удовлетворяет уравнению (2.1) при $z=1, 2, \ldots, a-1$.

Уравнение (2.1) *есть уравнение в конечных разностях, а условия* (2.2) *служат граничными условиями для* q_z. Мы выведем явное выражение для q_z с помощью *метода частных решений*, который будет применяться и в более общих случаях.

Предположим сначала, что $p\neq q$. Легко проверить, что разностное уравнение (2.1) имеет два частных решения $q_z=1$ и $q_z=(q/p)^z$. Отсюда следует, что при произвольных постоянных A и B последовательность

$$q_z=A+B\left(\frac{q}{p}\right)^z \tag{2.3}$$

есть решение уравнения (2.1). Подберем постоянные A и B так, чтобы выполнялись граничные условия. Это значит, что A и B должны удовлетворять двум линейным уравнениям $A+B=1$ и $A+B(q/p)^a=0$. Отсюда следует, что

$$q_z=\frac{(q/p)^a-(q/p)^z}{(q/p)^a-1} \tag{2.4}$$

есть решение разностного уравнения (2.1), удовлетворяющее граничным условиям (2.2). Чтобы доказать, что искомая вероятность разорения равна (2.4), остается показать, что решение (2.4) единственно. Иначе говоря, нужно доказать, что *все* решения уравнения (2.1) могут быть записаны в виде (2.3). Пусть дано произвольное решение уравнения (2.1). Мы можем выбрать постоянные A и B так, чтобы (2.3) совпадало с этим решением при $z=0$ и $z=1$. Но по этим двум значениям можно найти все другие значения с помощью последовательной подстановки в (2.1) $z=1, 2, 3, \ldots$. Это значит, что два решения, совпадающие при $z=0$ и $z=1$, совпадают тождественно, и поэтому любое решение имеет вид (2.3).

Наше рассуждение непригодно при $p=q=1/2$, так как (2.4) не имеет тогда смысла. Это происходит потому, что в случае $p=q=1/2$ два частных решения $q_z=1$ и $q_z=(q/p)^z$ совпадают. Однако в этом случае мы имеем второе частное решение $q_z=z$, и поэтому $q_z=A+Bz$ есть решение уравнения (2.1), зависящее от двух постоянных. Чтобы удовлетворить граничным условиям, нужно положить $A=1$ и $A+Ba=0$. Отсюда

$$q_z=1-\frac{z}{a}. \tag{2.5}$$

[Это же самое численное значение можно получить формально из (2.4), переходя к пределу при $p \to 1/2$ по правилу Лопиталя.]

Итак, мы доказали, что искомая *вероятность разорения игрока дается формулой* (2.4) *при* $p \neq q$ *и формулой* (2.5) *при* $p = q = 1/2$. Вероятность того, что игрок выиграет игру, равна вероятности разорения его противника и получается поэтому из наших формул соответственной заменой p, q и z на q, p и $a - z$. Непосредственно видно, что $p_z + q_z = 1$, как и утверждалось выше.

Наш результат можно сформулировать следующим образом: *пусть в игре с бесконечно богатым и всегда согласным на продолжение игры противником игрок имеет право прекратить игру в любой момент, когда ему это выгодно. Тактика игрока состоит в том, что он играет, пока либо проиграет все свои деньги, либо наберет капитал в* a *рублей* ($a - z$ *рублей чистого выигрыша*). *Тогда* q_z — *вероятность его проигрыша, а* $1 - q_z$ — *вероятность его выигрыша.*

При такой тактике конечный выигрыш или проигрыш есть случайная величина G, принимающая значения $a - z$ и $-z$ с вероятностями соответственно $1 - q_z$ и q_z. Математическое ожидание выигрыша равно

$$E(G) = a(1 - q_z) - z. \tag{2.6}$$

Подставляя в (2.6) значение q_z из формулы (2.5), убеждаемся, что при $q = p = 1/2$ математическое ожидание $E(G) = 0$. Обратно, вследствие (2.6) из равенства $E(G) = 0$ получается формула (2.5), так что $p = q$. Это значит, что при описанной тактике «безобидная» игра остается безобидной и что никакая «небезобидная» игра не может быть превращена с помощью этой тактики в безобидную.

Из (2.5) мы видим, что в случае $p = q = 1/2$ и $a = 1000$ игрок с исходным капиталом $z = 999$ выигрывает рубль с вероятностью 999/1000. При $q = 0{,}6$, $p = 0{,}4$ игра действительно невыгодна, но все же вероятность (2.4) выиграть рубль равна примерно $2/3$. Вообще игрок с относительно большим начальным капиталом z имеет значительные шансы выиграть малую сумму $a - z$, прежде чем разориться[1]).

Рассмотрим теперь, как влияет *изменение ставки*. Если начальные капиталы игрока и его противника равны соответственно z и $a - z$, то вероятность разорения игрока дается формулой (2.4). Допустим теперь, что единицей измерения вместо рубля становится полтинник.

---

[1]) Некий игрок, ежегодно посещая Монте-Карло, всегда успешно покрывал свои расходы и твердо верил в свою магическую власть над случайностью. Но этот факт ничего удивительного не содержит. Если игрок начинал решающую игру, имея десятикратный перевес, то вероятность успеха была близка к $9/10$. Вероятность выпадения десяти успехов подряд равна $(1 - 1/10)^{10} \approx e^{-1} \approx 0{,}37$. *Единичную* неудачу игрок всегда мог бы объяснить своей невнимательностью или недомоганием.

Это означает, что в (2.4) нужно заменить z на $2z$ и a на $2a$. Поэтому новая вероятность разорения равна

$$q_z^* = \frac{(q/p)^{2a} - (q/p)^{2z}}{(q/p)^{2a} - 1} = q_z \cdot \frac{(q/p)^a + (q/p)^z}{(q/p)^a + 1}. \tag{2.7}$$

Если $q > p$, то последняя дробь больше единицы, и $q_z^* > q_z$. *Поэтому, если ставка удваивается, а начальные капиталы остаются неизмененными, то вероятность разорения игрока, вероятность успеха которого $p < 1/2$, уменьшается, а вероятность разорения его противника (которому игра выгодна, поскольку $q > p$) увеличивается.* Аналогичное утверждение верно не только при удвоении, но и при любом увеличении ставки.

Незнакомые с этими классическими результатами люди могут думать, что участвовать в любой «невыгодной» игре неразумно. Принятие такого мнения означало бы отказ от всякого страхования. В действительности теоремы о вероятностях лишь подтверждают разумность действий осторожного человека, страхующего свое имущество от пожара, хотя он «играет» при этом в «невыгодную» игру.

Допустим, например, что Петр имеет 90 рублей, а Павел — 10, и пусть $p = 0{,}45$. Эта игра «невыгодна» для Петра. Если при каждом испытании ставка равна одному рублю, то, согласно табл. 1, вероятность разорения Петра равна приближенно 0,866. Если в той же игре ставка равна 10 рублям, то вероятность разорения Петра уменьшается более чем в четыре раза и составляет около 0,210. Эффект от увеличения ставки гораздо значительней, чем можно было предполагать. Вообще, если при каждом испытании ставится k рублей, то вероятность разорения находится по формуле (2.4) с заменой z на $\frac{z}{k}$ и a на $\frac{a}{k}$, и эта вероятность монотонно убывает с ростом k. Поэтому в игре с постоянной ставкой вероятность разорения игрока будет минимальной, если он выберет максимальную ставку, совместимую с той суммой, которую он хочет выиграть.

§ 3. Средняя продолжительность игры

Распределение вероятностей продолжительности игры будет выведено в следующих параграфах. Однако среднюю продолжительность игры можно вычислить значительно более простым методом, имеющим весьма широкое применение.

Как и выше, рассмотрим классическую задачу о разорении, сформулированную в начале настоящей главы. Предположим, что продолжительность игры имеет конечное математическое ожидание D_z. Строгое доказательство этого факта будет дано в следующем параграфе.

Рассуждая так же, как и при выводе разностного уравнения (2.1) и граничных условий (2.2), убеждаемся, что средняя продожительность игры D_z удовлетворяет разностному уравнению

$$D_z = pD_{z+1} + qD_{z-1} + 1, \qquad 0 < z < a \tag{3.1}$$

и граничным условиям

$$D_0 = 0, \quad D_a = 0. \tag{3.2}$$

Появление слагаемого 1 делает разностное уравнение (3.1) неоднородным. Если $p \neq q$, то одно из частных решений уравнения (3.1) имеет вид $D_z = z/(q-p)$. Легко видеть, что разность Δ_z каких-либо двух решений уравнения (3.1) удовлетворяет однородному уравнению $\Delta_z = p\Delta_{z+1} + q\Delta_{z-1}$. Мы уже знаем, что все решения этого последнего уравнения имеют вид: $A + B(q/p)^z$. Отсюда следует, что при $p \neq q$ все решения уравнения (3.1) имеют вид:

$$D_z = \frac{z}{q-p} + A + B\left(\frac{q}{p}\right)^z. \tag{3.3}$$

Значения постоянных A и B опять получаются из граничных условий (3.2), согласно которым $A + B = 0$ и $A + B(q/p)^a = -a/(q-p)$. Определив A и B, получаем

$$D_z = \frac{z}{q-p} - \frac{a}{q-p} \cdot \frac{1-(q/p)^z}{1-(q/p)^a}. \tag{3.4}$$

Эта формула также теряет смысл при $p = q = {}^1/_2$. В этом случае частным решением уравнения (3.1) вместо $z/(q-p)$ является $-z^2$. Отсюда следует, что при $p = q = {}^1/_2$ все решения уравнения (3.1) имеют вид $D_z = -z^2 + A + Bz$. Искомое решение D_z, удовлетворяющее граничным условиям (3.2), равно поэтому

$$D_z = z(a-z). \tag{3.5}$$

Математическое ожидание продолжительности игры в классической задаче о разорении дается формулой (3.4) *при* $p \neq q$ *и формулой* (3.5) *при* $p = q = {}^1/_2$.

Следует заметить, что продолжительность игры значительно больше, чем обычно полагают. Если два игрока, имеющие вначале по 500 рублей, бросают монету до тех пор, пока один из них не разорится, то средняя продолжительность игры составляет 250 000 испытаний. Если первый игрок имеет только 1 рубль, а его противник — 1000 рублей, то средняя продолжительность игры равна 1000 испытаний. Другие примеры доставляет табл. 1.

Таблица 1

Классическая задача о разорении игрока

p	q	z	a	Вероятность		Математическое ожидание	
				разорения	выигрыша	выигрыша	продолжительности игры
0,5	0,5	9	10	0,1	0,9	0	9
0,5	0,5	90	100	0,1	0,9	0	900
0,5	0,5	900	1000	0,1	0,9	0	90 000
0,5	0,5	950	1000	0,05	0,95	0	47 500
0,5	0,5	8000	10000	0,2	0,8	0	16 000 000
0,45	0,55	9	10	0,210	0,790	— 1,1	11
0,45	0,55	90	100	0,866	0,134	—76,6	765,6
0,45	0,55	99	100	0,182	0,818	—17,2	171,8
0,4	0,6	90	100	0,983	0,017	—88,3	441,3
0,4	0,6	99	100	0,333	0,667	—32,3	161,7

Начальный капитал равен z. Игра оканчивается разорением (проигрыш z рублей) или приобретением капитала a (выигрыш $a-z$ рублей).

Предельный переход при $a\to\infty$. Если в формулах (2.4) и (2.5) перейти к пределу при $a\to\infty$, то получим

$$q_z \to \begin{cases} 1, & \text{если } p \geqslant p \\ \left(\dfrac{q}{p}\right)^z, & \text{если } q < p. \end{cases} \qquad (3.6)$$

Вряд ли кто будет сомневаться в том, что этот предел можно рассматривать как вероятность *разорения в игре с бесконечно богатым противником*, однако, строго говоря, случайное блуждание на полупрямой $(0, \infty)$ нужно исследовать отдельно. Сказать, что при случайном блуждании на полупрямой частица, исходя из $z>0$, достигает 0, в действительности означает то же самое, что при неограниченном случайном блуждании, частица когда-нибудь сдвинется на z единиц влево от исходной точки. Но вероятность последнего события уже была вычислена в § 3 гл. XI и согласуется с формулой (3.6). При игре с бесконечно богатым противником вероятность разорения равна единице, если $q \geqslant p$, и равна $\left(\dfrac{q}{p}\right)^z$, если $q<p$. Во втором случае не имеет смысла говорить о средней продолжительности игры, так как с положительной вероятностью игра может вообще не окончится. При $q>p$ предельный переход дает для средней продолжительности игры значение $z(q-p)^{-1}$, если же $q=p$, то предел бесконечен. Это согласуется с известным нам фактом, что при симметрическом случайном блуждании все времена первого достижения имеют бесконечное математическое ожидание. (Другой вывод этих результатов можно найти в следующем параграфе.)

§ 4*) Производящие функции продолжительности игры и времени первого достижения

Применим метод производящих функций к изучению продолжительности игры в классической задаче о разорении игрока (или об ограниченном случайном блуждании с поглощающими экранами при $x=a$ и $x=0$). Исходным положением блуждающей частицы является $z\,(0<z<a)$. Пусть $u_{z,n}$ — вероятность того, что процесс кончится на n-м шаге у экрана $x=0$ (разорение игрока при n-м испытании). После первого шага частица попадает или в точку $z+1$ или в точку $z-1$, и мы заключаем, что при $1<z<a-1$ и $n\geqslant 1$

$$u_{z,n+1}=pu_{z+1,n}+qu_{z-1,n}. \tag{4.1}$$

Это — разностное уравнение, аналогичное уравнению (2.1), но зависящее от двух переменных z и n. Подобно тому, как мы поступали в § 2, нам следует определить граничные значения $u_{0,n}$, $u_{a,n}$ и $u_{z,0}$ так, чтобы (4.1) выполнялось также при $z=1$, $z=a-1$ и $n=0$. Для этой цели положим

$$u_{0,n}=u_{a,n}=0, \quad \text{когда} \quad n\geqslant 1, \tag{4.2}$$

и

$$u_{0,0}=1, \; u_{z,0}=0, \quad \text{когда} \quad z>0. \tag{4.3}$$

Тогда (4.1) выполняется при всех z, из интервала $0<z<a$, и при всех $n\geqslant 0$.

Введем теперь производящую функцию

$$U_z(s)=\sum_{n=0}^{\infty}u_{z,n}s^n. \tag{4.4}$$

Умножая (4.1) на s^{n+1} и суммируя по $n=0, 1, 2, \ldots$, убеждаемся, что при $0<z<a$

$$U_z(s)=psU_{z+1}(s)+qsU_{z-1}(s). \tag{4.5}$$

Кроме того, равенства (4.2) и (4.3) приводят к граничным условиям:

$$U_0(s)=1; \quad U_a(s)=0. \tag{4.6}$$

Уравнение (4.5) — разностное уравнение, аналогичное уравнению (2.1), а граничные условия (4.6) соответствуют условиям (2.2). Отличие заключается в том, что коэффициенты и неизвестное $U_z(s)$ зависят теперь от переменной s. Но в разностном уравнении s есть просто параметр. Мы сможем применить метод § 2, если только сумеем найти два частных решения уравнения (4.5). Естественно выяснить, не существует ли двух решений $U_z(s)$ вида $U_z(s)=\lambda^z(s)$.

*) Этот параграф вместе с близким по содержанию § 5 может быть опущен при первом чтении.

Подставляя это выражение в (4.5), убеждаемся, что $\lambda(s)$ должно удовлетворять квадратному уравнению

$$\lambda(s) = ps\lambda^2(s) + qs, \tag{4.7}$$

имеющему два корня:

$$\lambda_1(s) = \frac{1+(1-4pqs^2)^{1/2}}{2ps}, \quad \lambda_2(s) = \frac{1-(1-4pqs^2)^{1/2}}{2ps} \tag{4.8}$$

(у нас $0 < s < 1$, и квадратный корень положителен).

Таким образом, мы нашли два частных решения уравнения (4.5). Как и в § 2, заключаем отсюда, что при произвольных функциях $A(s)$ и $B(s)$

$$U_z(s) = A(s)\lambda_1^z(s) + B(s)\lambda_2^z(s) \tag{4.9}$$

есть решение уравнения (4.5). Так как это решение должно удовлетворять граничным условиям, то мы должны иметь $A(s)+B(s)=1$ и $A(s)\lambda_1^a(s)+B(s)\lambda_2^a(s)=0$. Отсюда получаем

$$U_z(s) = \frac{\lambda_1^a(s)\lambda_2^z(s) - \lambda_1^z(s)\lambda_2^a(s)}{\lambda_1^a(s) - \lambda_2^a(s)}. \tag{4.10}$$

Пользуясь очевидным соотношением $\lambda_1(s)\lambda_2(s) = q/p$, можно привести последнюю формулу к более простому виду:

$$U_z(s) = \left(\frac{q}{p}\right)^z \frac{\lambda_1^{a-z}(s) - \lambda_2^{a-z}(s)}{\lambda_1^a(s) - \lambda_2^a(s)}. \tag{4.11}$$

Это — *искомая производящая функция вероятностей разорения при n-м испытании (поглощения в точке $x=0$)*. Соответствующая производящая функция для вероятностей поглощения в точке $x=a$ получается заменой p, q, z соответственно на q, p и $a-z$. Производящей функцией *продолжительности* игры является, конечно, сумма этих двух функций.

Наш метод применим также при $a=\infty$, что соответствует случайному блужданию с одним поглощающим экраном в точке $x=0$ (или игре с бесконечно богатым противником). Мы имеем тогда единственное граничное условие $U_0(s)=1$. Все решения уравнения (4.5) имеют вид (4.9), но, поскольку $\lambda_1(s) > 1$ и $\lambda_2(s) < 1$ при $0 < s < 1$, $U_z(s)$ неограниченно возрастает, если только $A(s)$ не равно нулю. Поэтому искомым решением является

$$U_z(s) = \lambda_2^z(s). \tag{4.12}$$

Это — производящая функция вероятности того, что *частица, начинающая движение из точки $z > 0$, будет поглощена в точке $x=0$ при n-м испытании.*

Другими словами, *при неограниченном случайном блуждании производящая функция времени первого достижения точки,* *лежащей на z единиц влево от начала, имеет вид* (4.12). Для того чтобы получить аналогичную формулу для симметричной точки $x=z$, достаточно заменить p на q. Воспользовавшись этим соображением, из формулы (4.8) сразу получаем *производящую функцию времени первого достижения точки $x=z$ при неограниченном случайном блуждании*

$$\lambda^{z}(s)=\left(\frac{p}{q}\lambda_2(s)\right)^{z}=\lambda_1^{-z}(s). \tag{4.13}$$

В частности, полагая $z=1$, убеждаемся, что $\lambda(s)$ есть производящая функция времени первого достижения соседнего правого положения. Время первого достижения точки $z>1$ при условии, что частица выходит из нуля, равно времени первого достижения 1 плюс время первого достижения 2 при условии, что частица выходит из точки 1 плюс и т. д., и является поэтому суммой z взаимно независимых случайных величин, имеющих каждая производящую функцию $\lambda(s)$. Этим объясняется, почему выражение (4.13) представляет z-ю степень производящей функции $\lambda(s)$.

Положив $s=1$ в соотношении (4.13), мы получим вероятность разорения при игре с бесконечно богатым противником. Она равна $\left(\frac{q}{p}\right)^{z}$ и 1 в зависимости от того $q\leqslant p$ или $q\geqslant p$.

§ 5*). Явные выражения

Выведем теперь с помощью разложения $U_z(s)$ на простые дроби явную формулу для $u_{z,n}$. Формально выражение (4.11) для $U_z(s)$ содержит квадратный корень, но на самом деле $U_z(s)$ — рациональная функция. Действительно, используя (4.8) и формулу бинома, мы видим, что разность $\lambda_1^k(s)-\lambda_2^k(s)$ равна произведению рациональной функции от s на $(1-4pqs^2)^{1/2}$; этот корень появляется как в числителе, так и в знаменателе выражения (4.11), и поэтому $U_z(s)$ равно отношению двух многочленов. Степень знаменателя равна $a-1$ или $a-2$ в зависимости от того, нечетно или четно число a; степень числителя равна $a-1$ или $a-2$ в зависимости от того, нечетно или четно число $a-z$. Ни в одном из случаев степень числителя не превосходит степени знаменателя больше, чем на единицу. Поэтому при $n>1$ можно вычислить $u_{z,n}$ по формуле (7.8) гл. XI при единственном условии, что все корни знаменателя различны.

Можно было бы вычислить корни знаменателя и соответствующие коэффициенты ρ_ν непосредственно, но выкладки упростятся, если ввести новое независимое переменное φ, полагая

$$\frac{1}{\cos\varphi}=2(pq)^{1/2}s. \tag{5.1}$$

*) См. примечание на стр. 344.

Из (4.8) получаем

$$\lambda_{1,2}(s)=\left(\frac{q}{p}\right)^{1/2}(\cos\varphi\pm i\sin\varphi)=\left(\frac{q}{p}\right)^{1/2}e^{\pm i\varphi}, \tag{5.2}$$

откуда по (4.11)

$$U_z(s)=\left(\frac{q}{p}\right)^{z/2}\frac{\sin(a-z)\varphi}{\sin a\varphi}. \tag{5.3}$$

Корни знаменателя, очевидно, равны $\varphi=0,\ \pi/a,\ 2\pi/a,\ \ldots$. Соответствующие значения s равны

$$s_\nu=\frac{1}{2(pq)^{1/2}\cos\nu\pi/a}. \tag{5.4}$$

Здесь, вообще говоря, $\nu=0,\ 1,\ \ldots,\ a$. Но значениям $\nu=0$ и $\nu=a$ отвечают посторонние корни $\varphi=0,\ \pi$, являющиеся также корнями числителя (5.3), а при четном a значению $\nu=a/2$ не соответствует никакое число s_ν. Поэтому в случае нечетного a все $a-1$ корней s_ν получаются при $\nu=1,\ 2,\ \ldots,\ a-1$; в случае четного a значение $\nu=a/2$ должно быть пропущено.

Мы знаем, что

$$\left(\frac{q}{p}\right)^{z/2}\frac{\sin(a-z)\varphi}{\sin a\varphi}=As+B\frac{\rho_1}{s_1-s}+\ldots+\frac{\rho_{a-1}}{s_{a-1}-s}. \tag{5.5}$$

Чтобы найти ρ_ν, умножим обе части на $s_\nu-s$ и перейдем к пределу при $s\to s_\nu$. Получим (полагая $\varphi_\nu=\pi\nu/a$), как и в формуле (7.5) гл. XI:

$$\rho_\nu=-\left(\frac{q}{p}\right)^{z/2}\frac{\sin(a-z)-\pi\nu/a}{a\cdot\cos\nu\pi\cdot(d\varphi/ds)_{s=s_\nu}}=\left(\frac{q}{p}\right)^{z/2}\frac{\sin z\pi\nu/a\cdot\sin\pi\nu/a}{2a(pq)^{1/2}\cos^2\pi\nu/a}. \tag{5.6}$$

Отсюда и из (5.5) получаем, наконец, коэффициент $u_{z,n}$ при s^n:

$$u_{z,n}=a^{-1}2^np^{(n-z)/2}q^{(n+z)/2}\sum_{\nu=1}^{a-1}\cos^{n-1}\frac{\pi\nu}{a}\cdot\sin\frac{\pi\nu}{a}\cdot\sin\frac{\pi z\nu}{a}. \tag{5.7}$$

(Строго говоря, слагаемое $\nu=a/2$ при четном a должно быть пропущено, но оно всегда равно нулю.)

При $n>1$ формула (5.7) дает *вероятность разорения* (*поглощения*) *при n-м испытании*. Эта формула восходит к Лагранжу; известно много различных способов ее вывода [1]). Несмотря на почтенную историю и наличие во многих учебниках, эта формула неоднократно открывалась заново.

---

[1]) Простой вывод, использующий тригонометрическую интерполяцию, приведен Эллисом в *Cambridge Mathematical Journal*, v. 4 (1844). См. также The Mathematical and Other Writings of R. E. Ellis, Cambridge and London, 1863.

Другое явное выражение см. в задаче 13, предельные формы см. в § 6 и в задаче 14. (Аналогичные формулы для блуждания с отражающими барьерами выведены в § 3 гл. VI.)

Если устремить a к бесконечности, то сумму (5.7) можно будет истолковать как риманову интегральную сумму, в пределе превращающуюся в интеграл. Поэтому мы убеждаемся, что при игре *с бесконечно богатым противником (единственный поглощающий экран в точке $x=0$) вероятность $w_{z,n}$ того, что игрок с исходным капиталом $z>0$ разорится на n-м испытании, равна*

$$w_{z,n}=2^n p^{(n-z)/2} q^{(n+z)/2} \int_0^1 \cos^{n-1}\pi x \sin\pi x \sin\pi xz\cdot dx. \tag{5.8}$$

Элементарным путем[1]) этот интеграл можно привести к виду

$$w_{z,n}=\frac{z}{n}\binom{a}{\frac{1}{2}(n-z)} p^{(n-z)/2} q^{(n+z)/2}; \tag{5.9}$$

(биномиальный коэффициент по-прежнему считается равным нулю, если $(n-z)/2$ не является целым числом из интервала $[0, n]$). Соответствующую *производящую* функцию мы нашли в конце § 4; она оказалась равной $\lambda_2^z(s)$.

§ 6. Переход к пределу; процессы диффузии

Мы уже указывали, что наша схема случайного блуждания может служить первым приближением в теории диффузии и броуновского движения, где малые частицы подвергаются большому числу ударов со стороны молекул. Каждый удар оказывает незначительное действие, но наложение многих малых действий вызывает заметное движение. В соответствии с этим займемся теперь изучением случайных блужданий, в которых отдельные шаги крайне малы, но очень быстро следуют друг за другом. В пределе такой процесс будет казаться непрерывным движением. Самое интересное заключается в том, что при таком переходе к пределу наши формулы сохраняют смысл и согласуются с физически осмысленными формулами теории диффузии, выводимыми при много более общих условиях и более

---

[1]) При $p=q=\frac{1}{2}$ формула (5.9) сводится к формуле (4.11) гл. III для распределения *времени первого достижения*. Проверка того, что из (5.8) следует (5.9), весьма затруднительна. Может быть, самый простой способ состоит в доказательстве того, что обе формулы представляют решения разностного уравнения (4.1) с граничными условиями (4.2)—(4.3) в точке $z=0$.

тонкими методами [1]). Этим частично объясняется, почему схема случайного блуждания, несмотря на свою грубость, довольно хорошо описывает процесс диффузии; физический смысл имеет только предельный случай, но различные дискретные схемы приводят к одним и тем же предельным формулам. Положение во многом похоже на то, что мы имеем в центральной предельной теореме, где, как мы видели, при очень общих условиях совместное действие многих случайных компонент практически не зависит от природы отдельных компонент.

Начнем с *неограниченного случайного блуждания, начинающегося в нуле*. Пусть $v_{x,n}$ — вероятность того, что n-й шаг приведет частицу в точку x. Если r из n шагов направлены вправо, то $n-r$ шагов направлены влево, и суммарное перемещение равно $r-(n-r)=2r-n$ единиц. Поскольку это перемещение должно равняться x, имеем $2r-n=x$. Это возможно, лишь если n и x имеют одинаковую четность (последнее означает, что после четного числа шагов абсцисса x является четным числом). Из n шагов r могут выбраны $\binom{n}{r}$ способами, и поэтому

$$v_{x,n}=\binom{n}{\frac{1}{2}(n+x)}p^{(n+x)/2}q^{(n-x)/2}; \tag{6.1}$$

биномиальный коэффициент считается равным нулю, когда $(n+x)/2$ не есть целое число из интервала $[0, n]$.

Существует другой способ вывода формулы (6.1), опирающийся на рассуждение, которое привело нас к разностному уравнению (4.1) и граничным условиям (4.2) и (4.3). Можно проверить, что $v_{x,n}$ должно удовлетворять уравнению в конечных разностях

$$v_{x,n+1}=pv_{x-1,n}+qv_{x+1,n} \tag{6.2}$$

с граничными условиями:

$$v_{0,0}=1, \quad v_{x,0}=0 \qquad \text{при} \quad x\neq 0. \tag{6.3}$$

Зная значения (5.3) и полагая в (6.2) последовательно $n=1, 2, \ldots$, мы получим сначала все значения $v_{x,1}$, а затем последовательно $v_{x,2}, v_{x,3}, \ldots$. Это показывает, что условия (6.2) и (6.3) одно-

---

[1]) Предельные формулы, выведенные в настоящем параграфе, совпадают с формулами классической теории диффузии Эйнштейна — Винера. Более тонкие современные теории Уленбека и Орнштейна здесь не рассматриваются. Открытие связи между случайными блужданиями и диффузией принадлежит в основном Л. Башелье. Работы этого автора имеют скорее эвристический характер, но им получено много новых результатов. Развитая А. Н. Колмогоровым теория стохастических процессов марковского типа в значительной степени примыкает к идеям Башелье. См., в частности, L. Bachelier, Calcul des probabilités, Paris, 1912.

значно определяют $v_{x,n}$. С другой стороны, легко проверить, что (6.1) есть решение уравнения (6.2) при условиях (6.3).

Выберем теперь единицу длины так, чтобы *каждый шаг имел длину* Δx, *и предположим, что промежуток времени между двумя последовательными шагами равен* Δt. За время t частица совершит $t/\Delta t$ скачков, и смещение x эквивалентно теперь $x/\Delta x$ единицам. При этом имеют смысл лишь значения, кратные Δx и Δt, но в пределе, когда $\Delta x \to 0$, $\Delta t \to 0$, становятся возможными все смещения и все интервалы времени.

Не следует ожидать разумных результатов, если мы устремим Δx и Δt к нулю произвольным образом. Достаточно заметить, что максимальное возможное перемещение частицы за время t равно $t\,\Delta x/\Delta t$, так что при $\Delta x/\Delta t \to 0$ в пределе не будет никакого движения. С физической точки зрения масштабы смещения и времени должны находиться в соответствующем соотношении, иначе в пределе процесс будет вырождаться: скорость стремится или к нулю или к бесконечности. Чтобы найти правильное соотношение, заметим, что суммарное перемещение частицы за время t есть сумма примерно $t/\Delta t$ взаимно независимых случайных величин, имеющих каждая среднее значение $(p-q)\,\Delta x$ и дисперсию $\{1-(p-q)^2\}\,(\Delta x)^2 = 4pq\,(\Delta x)^2$. Поэтому среднее значение и дисперсия всего перемещения за время t приблизительно равны $t(p-q)\,\Delta x/\Delta t$ и $4pqt\,(\Delta x)^2/\Delta t$ соответственно. Чтобы получить разумные результаты, нужно устремить Δx и Δt к нулю так, чтобы среднее значение и дисперсия суммарного перемещения частицы оставались конечными при всех t. Конечность дисперсии требует ограниченности величины $(\Delta x)^2/\Delta t$; из конечности математического ожидания следует, что $p-q$ должно быть величиной порядка Δx. Поэтому целесообразно положить

$$\frac{(\Delta x)^2}{\Delta t} = 2D, \quad p = \frac{1}{2} + \frac{c}{2D}\,\Delta x, \quad q = \frac{1}{2} - \frac{c}{2D}\,\Delta x, \tag{6.4}$$

где D и c — постоянные. Численное значение D сказывается только на масштабах; для математической простоты лучше всего было бы положить $D = 1$, но мы сохраним произвольное D, чтобы облегчить сопоставление с физическими теориями. Постоянные D и c являются соответственно *коэффициентом диффузии и скоростью течения*. Если $c = 0$, то случайное блуждание симметрично. В общем случае знак c определяет направление течения. В пределе p и q стремятся к $^1/_2$; при любой другой нормировке частица уносилась бы так быстро, что вероятность конечных смещений стремилась бы к нулю.

Итак, пусть $\Delta x \to 0$, $\Delta t \to 0$, причем Δx и Δt нормированы согласно (6.4). Суммарное перемещение частицы за время $t \approx n\Delta t$ определяется результатами n испытаний Бернулли, и поэтому предельный вид $v_{x,n}$, как известно из гл. VII, будет определяться нормальным распределением. Необходимые вычисления были проделаны в гл. VII. При фиксированном Δx суммарное перемещение равно сумме конечного числа независимых случайных величин и имеет математическое ожидание $t(p-q)\Delta x/\Delta t = 2ct$ и дисперсию $4pqt(\Delta x)^2/\Delta t = 2Dt$. Поэтому *вероятность того, что в момент t перемещение частицы заключено между x_0 и x_1 $(x_0 < x_1)$, стремится к*

$$(2\pi)^{-1/2}\int_{y_0}^{y_1} e^{-\lambda^2/2}d\lambda, \tag{6.5}$$

где

$$y_1 = (x_1 - 2ct)/(2Dt)^{1/2} \quad \text{и} \quad y_0 = (x_0 - 2ct)/(2Dt)^{1/2}.$$

(Согласно центральной предельной теореме, то же заключение справедливо и для более общих случайных блужданий.)

Что касается уравнения (6.2), то мы перейдем к обычным функциональным обозначениям и запишем это уравнение в виде $v(x, t+\Delta t) = p\cdot v(x-\Delta x, t) + q\cdot v(x+\Delta x, t)$. Разлагая функции от $t+\Delta t$ и $x+\Delta x$ в ряды Тейлора и сохраняя в этих разложениях члены не выше второго порядка, формально получаем

$$\Delta t\frac{\partial v(x, t)}{\partial t} = (q-p)\Delta x\cdot\frac{\partial v(x, t)}{\partial x} + \frac{(\Delta x)^2}{2}\frac{\partial^2 v(x, t)}{\partial x^2} + \dots \tag{6.6}$$

В силу (6.4) в пределе получим

$$\frac{\partial v(x, t)}{\partial t} = -2c\cdot\frac{\partial v(x, t)}{\partial x} + D\cdot\frac{\partial^2 v(x, t)}{\partial x^2}. \tag{6.7}$$

Это — хорошо известное *уравнение Фоккера—Планка* для диффузии при наличии течения. Оно может быть выведено из более общих и более убедительных предположений. В обычной теории выражение (6.5) находится, как решение уравнения (6.7); мы же получили оба результата с помощью одного и того же предельного процесса. Мы действовали чисто эвристически, но аналогичные рассуждения можно провести вполне строго. Все формулы для дискретного случайного блуждания допускают простой переход к пределу.

В качестве следующего примера рассмотрим предельное поведение вероятности *первого* достижения. Для простоты рассмотрим сначала формулу (5.9), соответствующую одному экрану. Одна из величин $w_{z,n}$ и $w_{z,n+1}$ обязательно равна нулю. Сумма $w_{z,n} + w_{z,n+1}$ асимптотически представляет вероятность поглощения в течение промежутка времени $(t, t+2\Delta t)$. Мы покажем, что $w_{z,n} + w_{z,n+1} \sim f(z, t)(2\Delta t)$, где $f(z, t)$ — непрерывная функция. Тогда предельная вероятность поглощения в течение интервала времени (t_1, t_2) будет равна интегралу от $f(z, t)$ по интервалу (t_1, t_2). Допустим, что $n-z$ — четное. Тогда $w_{z,n+1} = 0$, и, чтобы найти $f(z, t)$, нужно в (5.9) заменить z на $z/\Delta x$ и n на $t/\Delta t$ и применить (6.4). Пользуясь нормальным приближением для биномиального распределения и последним из соотношений (6.9), мы легко получим[1])

$$f(z,\ t) \sim \frac{z}{2(\pi Dt^3)^{1/2}} e^{-(z+2ct)^2/(4Dt)}. \tag{6.8}$$

Это — предельный вид формулы (5.9); результат снова совпадает с соответствующей формулой теории диффузии. Легко проверить, что $f(-x,\ t)$ есть решение уравнения (6.7). (В определении $w_{z,n}$ переменная z играет ту же роль, что $-x$ в $v_{x,n}$.)

Аналогичные рассуждения применимы к формуле (5.7). Исследование этой формулы показывает, что слагаемые с $\nu = k$ и $\nu = a - k$ при нечетном $n - z$ взаимно уничтожаются, а при четном $n - z$ складываются. Поэтому мы получим предельный вид величины $f(z,\ t) \sim (u_{z,n} + u_{z,n+1})/(2\Delta t)$, суммируя в (5.7) дважды по всем ν из интервала $0 \leqslant \nu \leqslant a/2$. Заменим a, z и n соответственно на $a/\Delta x$, $z/\Delta x$ и $t/\Delta t$ и заметим, что при фиксированном ν

$$\begin{gathered}\sin\frac{\pi\nu\,\Delta x}{a} \sim \frac{\pi\nu\,\Delta x}{a};\\ \left(\cos\frac{\pi\nu\,\Delta x}{a}\right)^{t/\Delta t} \sim \left(1 - \frac{D\pi^2\nu^2\Delta t}{a^2}\right)^{z/\Delta t} \sim e^{-D\pi^2\nu^2 t/a^2};\\ (4pq)^{t/2\Delta t}\left(\frac{q}{p}\right)^{z/2\Delta x} \sim e^{-c(ct+z)/D}.\end{gathered} \tag{6.9}$$

В пределе получаем

$$f(z,\ t) \sim 2\pi Da^{-2}e^{-c(ct+z)/D}\sum_{\nu=1}^{\infty} \nu e^{-D\pi^2\nu^2 t/a^2} \sin\frac{\pi z\nu}{a}. \tag{6.10}$$

Формальный переход к пределу законен вследствие равномерной сходимости ряда (5.7): сумма членов с большими ν мала как в (6.10), так и в исходной сумме (5.7) (где мы имеем $\nu < a/2$).

В теории диффузии (6.10) носит название формулы Фюрта для первого достижения и выводится непосредственно из уравнения Фоккера—Планка. При свободной диффузии интеграл от величины (6.10) по промежутку времени $(t_1,\ t_2)$ дает вероятность того, что частица, выходящая из $z > 0$, в течение этого промежутка времени впервые достигнет нуля, не побывав до этого в точке $x = a$.

§ 7*). Случайные блуждания на плоскости и в пространстве

При двумерном случайном блуждании частица перемещается на единичный шаг в одном из четырех направлений, параллельных координатным осям. Если частица выходит из начала координат, то возможными ее положениями являются все точки плоскости с целочис-

[1]) В симметрическом случае $c = 0$ (т. е. $p = q$) формула (6.8) согласуется с предельным распределением для времени первого достижения, которое элементарными методами получено в гл. III (8.6).

*) См. примечание на стр. 331.

ленными координатами. Каждое положение имеет четыре *соседних*. Подобным же образом в случае трех измерений каждое положение имеет шесть соседних. Чтобы определить случайное блуждание, нужно задать соответствующие четыре или шесть вероятностей. Для простоты мы рассмотрим только *симметричный* случай, в котором все направления равновероятны. Многомерные задачи много сложнее, чем одномерные, так как тут области, которыми ограничено движение частицы, могут иметь произвольную форму, и роль экранов вместо точек играют сложные границы.

Начнем с интересной теоремы, принадлежащей Пойа[1]).

Теорема. *При одномерном и двумерном случайном блуждании частица с вероятностью единица рано или поздно (и поэтому бесконечно много раз) возвратится в свое начальное положение. Однако в случае трех измерений эта вероятность равна всего лишь примерно* 0,35 [математическое ожидание числа возвращений равно тогда $0{,}65 \sum k(0{,}35)^k = 0{,}35/0{,}65 \approx 0{,}53$].

Прежде чем доказывать эту теорему, приведем две другие ее формулировки, принадлежащие Пойа. Во-первых, почти очевидно, что в *случае одного или двух измерений частица с вероятностью единица бесконечное число раз побывает в каждом состоянии*; в случае трех измерений это, однако, неверно. Таким образом, утверждение «все дороги ведут в Рим» в некотором смысле справедливо для двух измерений.

С другой стороны, рассмотрим *две* частицы, совершающие независимые случайные блуждания, причем их перемещения происходят одновременно. Встретятся ли они когда-нибудь? Для простоты изложения определим *расстояние* между двумя возможными положениями как наименьшее число шагов, ведущих из одного положения в другое. (Тогда расстояние равно сумме абсолютных величин разностей координат.) Если обе частицы движутся единичными шагами, то при каждом их перемещении расстояние между ними либо не меняется, либо меняется на две единицы. Поэтому расстояние между двумя частицами либо всегда четно, либо всегда нечетно. Во втором случае две частицы никогда не могут занять одно и то же положение. В первом случае, как легко видеть, вероятность встречи частиц на n-м шаге равна вероятности того, что первая частица за $2n$ шагов достигнет начального положения второй частицы. Поэтому наша теорема утверждает, что в случае двух (но не трех) измерений две частицы наверняка будут бесконечное число раз занимать одно и то же положение. Аналогичное рассуждение показывает, что если

---

[1]) G. Polya, Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz, *Mathematische Annalen*, **84** (1921), 149—160. Числовое значение 0,35 было вычислено W. H. McCrea and F. J. W. Whipple, Random paths in two and three dimensions, *Proceedings of the Royal Society of Edinburgh*, **60** (1940), 281—298.

начальное расстояние между частицами нечетно, то они будут бесконечно много раз занимать соседние положения. Если называть это встречей, то из теоремы следует, что в случае *одного и двух измерений две частицы с достоверностью встретятся бесконечное число раз*; *в случае трех измерений существует положительная вероятность того, что они никогда не встретятся.*

Доказательство. Для случая одного измерения теорема была доказана в примере 3, б гл. XIII, с той разницей, что там мы говорили об игре с бросанием монеты, а не о симметричном случайном блуждании. Для случаев двух и трех измерений теорема может быть доказана совершенно аналогично. Пусть u_n — вероятность того, что n-й шаг приведет частицу в начальное положение. Согласно теореме 2 § 3 гл. XII, мы должны доказать, что в случае двух измерений ряд $\sum u_n$ расходится, а в случае трех измерений $\sum u_n \approx 0{,}53$. В случае двух измерений возвращение в начальное положение возможно только тогда, когда число шагов в положительных направлениях осей x и y равно соответственно числу шагов в отрицательных направлениях. Поэтому при нечетном n имеем $u_n = 0$, а при четном n (согласно полиномиальному распределению гл. VI)

$$u_{2n} = \frac{1}{4^{2n}} \sum_{k=0}^{2n} \frac{(2n)!}{k!k!\,(n-k)!\,(n-k)!} = \frac{1}{4^{2n}} \binom{2n}{n} \sum_{k=0}^{n} \binom{n}{k}^2. \tag{7.1}$$

Последнее выражение равно $4^{-2n}\binom{2n}{n}^2$ [см. формулу (12.11) гл. II]. Применение формулы Стирлинга показывает, что u_{2n} имеет порядок $1/n$, так что ряд $\sum u_{2n}$ расходится.

В случае трех измерений аналогично получаем

$$u_{2n} = \frac{1}{6^{2n}} \sum_{j,k} \frac{(2n)!}{j!j!k!k!\,(n-j-k)!\,(n-j-k)!}, \tag{7.2}$$

где сумма берется по всем j и k, удовлетворяющим условию $j+k \leqslant n$. Легко проверить, что

$$u_{2n} = \frac{1}{2^{2n}} \binom{2n}{n} \sum_{j,k} \left\{ \frac{1}{3^n} \frac{n!}{j!k!\,(n-j-k)!} \right\}^2. \tag{7.3}$$

В скобках стоят вероятности триномиального распределения, сумма которых равна единице. Поэтому сумма квадратов этих вероятностей меньше максимальной из них, соответствующей значениям j и k, близким к $n/3$. Применяя формулу Стирлинга, мы убеждаемся, что максимальная вероятность имеет величину порядка n^{-1}, так что u_{2n} — величина порядка $n^{-3/2}$, и ряд Σu_{2n} сходится.

Теорема Пойа аналогична фактам, изложенным в примере 3, в гл. XII в связи с бросанием нескольких монет.

В заключение этого параграфа приведем еще одну задачу, обобщающую понятие *поглощающего экрана*. Для определенности рассмотрим случай двух измерений, когда вместо интервала $0 \leqslant x \leqslant a$ имеется плоская область D, т. е. некоторое множество точек с целочисленными координатами. Каждая точка имеет четыре соседних, но для некоторых точек области D одна или несколько соседних точек лежат вне D. Такие точки образуют границу области D; все остальные точки называются внутренними. В одномерном случае границу образуют два экрана, и задача состоит в нахождении вероятности того, что, выходя из z, частица достигнет граничной точки $x = 0$ ранее, чем точки $x = a$. Будем аналогично искать вероятность того, что частица достигнет определенного участка границы, не побывав до этого ни в одной граничной точке, не принадлежащей этому участку. Это значит, что все граничные точки делятся на две группы B' и B'', и мы ищем вероятность $u(x, y)$ того, что, выходя из внутренней точки (x, y), частица достигнет какой-нибудь точки из B' ранее, чем точки из B''. В частности, если B' состоит из одной-единственной точки, то $u(x, y)$ есть вероятность того, что частица рано или поздно будет поглощена в этой точке.

Пусть (x, y) — внутренняя точка. Первый шаг переводит частицу в одну из четырех соседних точек $(x \pm 1, y)$, $(x, y \pm 1)$. Если все эти четыре точки являются внутренними, то

$$u(x, y) = \frac{1}{4}\{u(x+1, y) + u(x-1, y) + u(x, y+1) + u(x, y-1)\}. \quad (7.4.)$$

Это — уравнение в конечных разностях, играющее ту же роль, что и (2.1) (с $p = q = 1/2$). Если $(x+1, y)$ — граничная точка, то соответствующее ей слагаемое должно быть заменено на 1 или 0, в зависимости от того, принадлежит ли $(x+1, y)$ к B' или к B''. *Поэтому уравнение* (7.4) *будет справедливо для всех внутренних точек, если условиться, что для граничной точки* (ξ, η) *вероятность* $u(\xi, \eta) = 1$, *если* (ξ, η) — *точка из* B', *и* $u(\xi, \eta) = 0$, *если* (ξ, η) — *точка из* B''. *Это соглашение играет роль граничных условий* (2.2).

Мы имеем теперь систему линейных уравнений (7.4) с неизвестными $u(x, y)$; каждой внутренней точке соответствует одно неизвестное и одно уравнение. Эта система неоднородна, так как в ней встречается по крайней мере одна граничная точка (ξ, η) из B', прибавляющая к правой части (7.4) слагаемое $1/4$. Если область D конечна, то число уравнений равно числу неизвестных. В этом случае система имеет единственное решение тогда и только тогда, когда соответствующая однородная система [$u(\xi, \eta) = 0$ для всех граничных точек] не имеет решения, отличного от нуля. Далее, $u(x, y)$ — среднее арифметическое четырех соседних значений $u(x \pm 1, y)$, $u(x, y \pm 1)$ и поэтому не больше максимального из них. Иными словами, $u(x, y)$ не может иметь во внутренних точках ни максимума, ни минимума, и поэтому наибольшее и наименьшее значения достигаются в граничных точках. Следовательно, если все граничные значения — нули, то таковы же значения $u(x, y)$ во всех внутренних точках. Это доказывает существование и единственность решения уравнения (7.4). Поскольку граничные значения равны 0 и 1, все значения $u(x, y)$ лежат между нулем и единицей, как это и должно быть для вероятностей. Как мы увидим из общей теоремы о бесконечных цепях Маркова[1], аналогичные утверждения справедливы также и для бесконечных областей.

§ 8. Обобщенное одномерное случайное блуждание (последовательный анализ)

Мы возвратимся теперь к одному измерению, но рассмотрим общий случай, когда частица переходит не обязательно в соседнюю точку. Возможными положениями по-прежнему являются $x = 0, \pm 1, \pm 2, \ldots$. Однако мы будем считать, что *на каждом шаге частица с вероятностью p_k переходит из x в точку $x + k$*; индекс k принимает положительные и отрицательные значения или нуль (в обычном случайном блуждании $p_1 = p$, $p_{-1} = q$). Рассмотрим следующую *задачу о разорении. Частица выходит из точки z $(0 < z < a)$; мы ищем вероятность u_z того, что частица достигнет какого-либо из положений $x \leqslant 0$ ранее, чем какого-либо из положений $x \geqslant a$.*

Другими словами, в момент времени n частица находится в точке $z + X_1 + \ldots + X_n$ оси x, где $\{X_k\}$ — независимые случайные величины, имеющие одно и то же распределение $\{p_\nu\}$, процесс продолжается до тех пор пока впервые или $X_1 + \ldots + X_n \leqslant 0$ или $X_1 + \ldots + X_n \geqslant a - z$.

Эта задача приобрела особый интерес в связи с последовательным анализом, в котором X_k — некоторая характеристика выборки или наблюдения. Измерения производятся до тех пор, пока сумма $X_1 + \ldots + X_k$ не окажется вне двух заранее назначенных границ (наших — z и $a - z$). В первом случае (используя статистическую терминологию) гипотеза отвергается, во втором принимается. Впервые такой прием был описан В. Бартки[2]. А. Уолд построил

[1]) Точные решения известны лишь для некоторых частных случаев и, как правило, очень сложны. Решения для случая прямоугольных областей, бесконечных полос и т. д. см. в статье Мак-Кри и Уиппла, цитированной в примечании на стр. 353.

[2]) W. Bartky, Multiple sampling with constant probability, *Ann. Math. Statistics*, **14** (1943), 363—377. Описание см. в примере 2 к гл. XV.

общую теорию; ему же принадлежит приведенная выше формулировка [1]).

Не теряя общности, можно предположить, что скачки возможны как в положительном, так и в отрицательном направлениях. В противном случае мы имели бы либо $u_z = 0$, либо $u_z = 1$ для всех z.

Вероятность разорения на *первом* шаге равна, очевидно,

$$r_z = p_{-z} + p_{-z-1} + p_{-z-2} + \dots \tag{8.1}$$

(эта величина может быть равна нулю). После первого шага процесс продолжается лишь в том случае, если частица перешла в положение x, лежащее между 0 и a. Вероятность скачка из z в x равна p_{x-z}, а вероятность последующего разорения равна u_x. Следовательно,

$$u_z = \sum_{x=1}^{a-1} u_x p_{x-z} + r_z. \tag{8.2}$$

Мы получили $a - 1$ линейных уравнений с $a - 1$ неизвестными u_z. Система неоднородна, так как при $z = 1$ вероятность r_1 отлична от нуля (возможность скачков в отрицательном направлении означает, очевидно, что $r_1 > 0$). Докажем, что соответствующая однородная система

$$u_z = \sum_{x=1}^{a-1} u_x p_{x-z} \tag{8.3}$$

имеет единственное решение 0.

Действительно, если бы однородная система имела еще и другое решение, то одно из значений u_z было бы наибольшим по абсолютной величине, скажем, $u_z = M > 0$. Допустим сначала, что $p_{-1} \neq 0$. Поскольку сумма коэффициентов p_{x-z} в (8.3) равна самое большее единице, равенство (8.3) возможно, лишь если все u_x, действительно входящие в правую часть (т. е. имеющие отличные от нуля коэффициенты), равны M и если сумма коэффициентов равна единице. Поэтому $u_{z-1} = M$. Аналогично доказывается, что $u_{z-2} = u_{z-3} = \dots$ $\dots = u_1 = M$. Но при $z = 1$ сумма коэффициентов p_{x-z} в (8.3) меньше единицы, так что M должно быть нулем. Такое же рассуждение, очевидно, применимо и при $p_{-1} = 0$, так как можно заменить p_{-1} каким-либо другим положительным коэффициентом p_k с $k < 0$.

Следовательно, система уравнений (8.2) имеет единственное решение, и, таким образом, наша задача является вполне определенной.

---

[1]) A. Wald. On cumulative sums of random variables, *Ann. Math. Statistics*, 15 (1944), 283—296. Методы, описанные в данной книге, отличаются от методов Уолда. См. также A. Wald, Sequential analysis, John Wiley and Sons., N. — Y., 1947.

Уравнение (8.2) играет роль уравнения в конечных разностях (2.1). Можно упростить запись, введя граничные условия:

$$\begin{aligned} u_x &= 1, \text{ если } x \leqslant 0; \\ u_x &= 0, \text{ если } x \geqslant a. \end{aligned} \tag{8.4}$$

Тогда (8.2) можно переписать в виде

$$u_z = \sum u_x p_{x-z}, \tag{8.5}$$

где предполагается суммирование уже по всем x [слагаемые, получающиеся при $x \geqslant a$, не влияют на сумму благодаря второму из условий (8.4); слагаемые, соответствующие $x \leqslant 0$, дают в сумме r_z благодаря первому из условий (8.4)].

При большом a непосредственное решение $a-1$ линейных уравнений громоздко, и поэтому удобнее пользоваться *методом частных решений*, подобно тому, как мы поступали в § 2. Этот метод применим, когда в распределении вероятностей $\{p_k\}$ лишь немногие из вероятностей p_k отличны от нуля. Предположим, что отличны от нуля лишь те p_k, для которых $-\nu \leqslant k \leqslant \mu$, так что наибольшие возможные скачки в положительном направлении равны μ, а в отрицательном направлении $-\nu$. Рассмотрим *характеристическое уравнение*

$$\sum p_k s^k = 1, \tag{8.6}$$

являющееся алгебраическим уравнением степени $\nu+\mu$. Если s — корень уравнения (8.6), то $u_z = s^z$ при любом z есть частное решение уравнения (8.5), не удовлетворяющее, однако, граничным условиям (8.4). Если (8.6) имеет $\mu+\nu$ различных корней $s_1, s_2, \ldots$, то при любом z линейная комбинация частных решений

$$u_z = \sum A_k s_k^z \tag{8.7}$$

также есть решение уравнения (8.5), и остается подобрать постоянные A_k так, чтобы удовлетворялись граничные условия. Но при $0 < z < a$ в уравнении (8.3) встречаются лишь значения x из промежутка $-\nu+1 \leqslant x \leqslant a+\mu-1$. Поэтому достаточно удовлетворить граничные условия (8.4) при $x = 0, -1, -2, \ldots, -\nu+1$ и $x = a, a+1, \ldots, a+\mu-1$, так что всего имеется $\nu+\mu$ условий. Если s_k — двойной корень уравнения (8.6), то недостает одной произвольной постоянной; но легко проверить, что в этом случае $u_z = zs_k^z$ есть дополнительное частное решение уравнения (8.5). В любом случае $\mu+\nu$ граничных условий определяют значения $\mu+\nu$ произвольных постоянных.

Пример. Предположим, что каждый отдельный шаг переводит частицу в одно из четырех ближайших положений, и пусть $p_{-2} = p_{-1} = p_1 = p_2 = 1/4$. Характеристическое уравнение (8.6) имеет вид

$s^{-2}+s^{-1}+s+s^2=4$. Положив $t=s+s^{-1}$, мы приведем это уравнение к виду $t^2+t=6$, откуда $t=2$, -3. Решая уравнение $t=s+s^{-1}$ относительно s, находим четыре корня:

$$s_1=s_2=1; \quad s_3=\frac{-3+\sqrt{5}}{2}=s_4^{-1}; \quad s_4=\frac{-3-\sqrt{5}}{2}=s_3^{-1}. \tag{8.8}$$

Поскольку s_1 — двойной корень, общее решение уравнения (8.5) имеет вид

$$u_z=A_1+A_2z+A_3s_3^z+A_4s_4^z. \tag{8.9}$$

Граничные условия имеют вид $u_0=u_{-1}=1$ и $u_a=u_{a+1}=0$. Определив с помощью этих условий значения произвольных постоянных A_1, A_2, A_3, A_4, получим окончательное решение в виде

$$u_z=1-\frac{z}{a}+\frac{(2z-a)\left(s_3^a-s_4^a\right)-a\left(s_3^{2z-a}-s_4^{2z-a}\right)}{a\left\{(a+2)\left(s_3^a-s_4^a\right)-a\left(s_3^{a+2}-s_4^{a+2}\right)\right\}}, \tag{8.10}$$

где s_3 и s_4 даются формулой (8.8).

Численные оценки. Если степень $\mu+\nu$ характеристического уравнения (8.6) не очень мала, то находить корни этого уравнения трудно. На практике удовлетворительные приближенные значения u_z могут быть получены неожиданно простым способом. Рассмотрим сначала случай, когда математическое ожидание распределения вероятностей $\{p_k\}$ равно нулю. Тогда характеристическое уравнение (8.6) имеет двойной корень в точке $s=1$, и $A+Bz$ есть частное решение уравнения (8.5). Конечно, двух постоянных A и B недостаточно для удовлетворения $\mu+\nu$ граничных условий (8.4). Но если определить A и B так, чтобы сумма $A+Bz$ обращалась в нуль при $z=a+\mu-1$ и в единицу при $z=0$, то мы будем иметь $A+Bx\geqslant 1$ при $x\leqslant 0$ и $A+Bx\geqslant 0$ при $a\leqslant x<a+\mu$. Тогда частное решение $A+Bz$ удовлетворяет граничным условиям (8.4) с заменой знака равенства на знак «больше или равно». Разность $A+Bz-u_z$ есть поэтому частное решение уравнения (8.5) с неотрицательными граничными значениями, так что $A+Bz-u_z\geqslant 0$. Аналогично можно получить и нижнюю границу для u_z, определяя A и B так, чтобы сумма $A+Bz$ обращалась в 0 при $z=a$ и в 1 при $z=-\nu+1$. Отсюда

$$\frac{a-z}{a+\nu-1}\leqslant u_z\leqslant\frac{a+\mu-z-1}{a+\mu-1}. \tag{8.11}$$

Если a велико по сравнению с $\mu+\nu$, то неравенство (8.11) служит прекрасной оценкой для u_z. [Конечно, оценка $u_z\approx(1-z/a)$ еще лучше, но ошибка такого приближенного выражения неизвестна.]

Рассмотрим теперь общий случай, когда математическое ожидание распределения $\{p_k\}$ не равно нулю. Характеристическое уравнение (8.6) имеет тогда в точке $s=1$ простой корень. Левая часть (8.6) стремится к ∞ при $s\to 0$ и $s\to\infty$, непрерывна при $s>0$, и ее вторая производная положительна; это значит, что при положительных s кривая $y=\sum p_k s^k$ непрерывна и обращена выпуклостью вниз. Поскольку эта кривая пересекает прямую $y=1$ в точке $s=1$, она пересекает ее еще ровно в одной точке. Следовательно, характеристическое уравнение (8.6) имеет только два положительных корня 1 и s_1. Как и раньше, убеждаемся, что $A+Bs_1^z$ есть частное решение уравнения (8.5), и мы можем применить предыдущие рассуждения, заменяя $A+Bz$ на $A+Bs_1^z$. В этом случае

$$\frac{s_1^a - s_1^z}{s_1^a - s_1^{-\nu+1}} \leqslant u_z \leqslant \frac{s_1^{a+\mu-1} - s_1^z}{s_1^{a+\mu-1} - 1}. \tag{8.12}$$

Таким образом, справедлива следующая

Теорема. *Решение u_z задачи о разорении удовлетворяет неравенству* (8.11), *если математическое ожидание распределения вероятностей $\{p_k\}$ равно нулю, и неравенству* (8.12), *если это математическое ожидание не равно нулю. Здесь s_1 — единственный положительный отличный от единицы корень уравнения* (8.6), *а μ и $-\nu$ определяются соответственно, как наибольший и наименьший индексы, при которых $p_k \neq 0$.*

Пусть $m = \sum kp_k$ — *ожидаемый выигрыш* в одном испытании (или ожидаемая длина одного шага). Из (8.6) нетрудно усмотреть, что $s_1 > 1$ при $m < 0$ и $s_1 < 1$ при $m > 0$. Устремляя a к бесконечности, с помощью доказанной теоремы убеждаемся, что *при игре с бесконечно богатым противником вероятность разорения игрока равна единице тогда и только тогда, когда $m \leqslant 0$.*

Вопрос о *продолжительности игры* может быть рассмотрен с помощью аналогичных методов (см. задачу 4).

§ 9. Задачи

1. Рассмотреть задачу о разорении (§ 2 и 3) для случая обобщенного случайного блуждания, при котором частица может перемещаться на один шаг влево, на один шаг вправо или оставаться на месте с вероятностями α, β и γ, соответственно ($\alpha+\beta+\gamma=1$). (В терминологии азартных игр это означает, что результатом отдельной партии может быть и ничья.)

2. Рассмотреть задачу о разорении § 2 и 3 для случая, когда в начале координат расположен *упругий* экран (определение см. в § 1). Разностные уравнения для вероятности разорения (поглощения в начале координат) и для средней продолжительности игры остаются прежними, но зато изменяются граничные условия в начале координат.

3. Частица при каждом шаге перемещается на *две* единицы налево или на *единицу* направо, причем соответствующие вероятности равны p и q ($p+q=1$). Если движение начинается из точки $x=z>0$, то какова вероятность того, что частица когда либо достигнет начала координат. (Это задача о разорении при игре с бесконечно богатым противником.)

Указания. Разностные уравнения, аналогичные (2.1), имеют частное решение $q_z = 1$ и два частных решения вида λ^z, где λ удовлетворяет квадратному уравнению.

4. В обобщенном случайном блуждании § 8 положим [по аналогии с (8.1)] $\rho_z = p_{a-z} + p_{a+1-z} + \ldots$, и пусть $d_{z,n}$ — вероятность того, что игра продлится ровно n шагов. Показать, что при $n \geqslant 1$

$$d_{z,n+1} = \sum_{x=1}^{a-1} d_{x,n} p_{x-z},$$

где $d_{z,1}=r_z+\rho_z$. Исходя из этого, доказать, что производящие функции $d_z(s)=\sum d_{z,n}s^n$ являются решением системы линейных уравнений

$$s^{-1}d_z(s)-\sum_{x=1}^{a-1}d_x(s)\,p_{x-z}=r_z+\rho_z.$$

С помощью дифференцирования показать, что ожидаемая продолжительность игры e_z есть решение системы уравнений

$$e_z-\sum_{x=1}^{a-1}e_x p_{x-z}=1.$$

5. Пусть в одномерном случайном блуждании с *поглощающими* экранами в точках $x=0$ и $x=a$ через z обозначено исходное положение частицы, а $w_{z,n}(x)$ — вероятность того, что n-й шаг приведет частицу в точку x. Найти разностные уравнения и граничные условия, которым удовлетворяет $w_{z,n}(x)$.

6. Продолжение. Записать граничные условия для случая двух *отражающих экранов* (упругих экранов с $\delta=1$).

Замечание: *В последующих задачах $v_{x,n}$ везде означает вероятность* (6.1) *того, что при неограниченном случайном блуждании с исходным положением в начале координат n-й шаг приведет частицу в точку x.*

7. Метод отражения[1]). Пусть $p=q=\frac{1}{2}$. Пусть далее $u_{z,n}(x)$ — вероятность того, что при случайном блуждании на полупрямой $(0,\infty)$ с поглощающим экраном в начале координат и с исходным положением $z>0$ n-й шаг приведет частицу в точку x. Показать, что $u_{z,n}(x)=v_{x-z,n}-v_{x+z,n}$. (Указание. Доказать, что для $u_{z,n}(x)$ выполняется разностное уравнение, аналогичное уравнению (4.1), с подходящими граничными условиями.)

8. Продолжение. Если в начале координат находится *отражающий экран*, то

$$u_{z,n}(x)=v_{x-z,n}+v_{x+z,n}.$$

9. Продолжение. Показать, что если в точках $x=0$ и $x=a$ поставлены *поглощающие экраны*, то

$$u_{z,n}(x)=\sum_k\{v_{x-z-2ka,n}-v_{x+z-2ka,n}\}, \tag{9.1}$$

где суммирование производится по всем положительным и отрицательным k. Лишь конечное число слагаемых отлично от нуля.) Если оба экрана *отражающие*, то уравнение (9.1) сохраняется с заменой знака *минус* на *плюс*.

10. Распределение максимума. Пусть M_n — максимальная абсцисса частицы за время от 0 до n при симметрическом неограниченном

---

[1]) Задачи 7—9 являются примерами на применение *метода отражения*. Слагаемое $v_{x-z,n}$ соответствует свободной частице в неограниченном случайном блуждании, а $v_{x+z,n}$ — «отраженной частице», выходящей из точки $-z$. Аналогично этому, слагаемые в уравнении (9.1) соответствуют различным отраженным частицам, которые начинают свой путь из точек, получающихся повторными отражениями исходной точки z относительно каждого из экранов. В задачах 12 и 13 мы получим общий результат для несимметрического случая, используя производящие функции. В теории дифференциальных равнений метод отражения всегда связывают с именем Кельвина. Эквивалентный принцип был предложен Д. Андре. См. примечание 4 к гл. III.

случайном блуждании с началом в точке $x=0$. Используя формулу задачи 7, показать, что

$$P\{M_n=z\}=v_{z,n}+v_{z+1,n}. \tag{9.2}$$

11. Пусть $V_x(s)=\sum v_{x,n}s^n$ (см. замечание перед задачей 7). Доказать, что $V_x(s)=V_0(x)\lambda_2^{-x}(s)$ при $x\leqslant 0$ и $V_x(s)=V_0(s)\lambda_1^{-x}(s)$ при $x\geqslant 0$, где $\lambda_1(s)$ и $\lambda_2(s)$ определяются формулой (4.8). Кроме того, $V_0(s)=(1-4pqs^2)^{-1/2}$.

Замечание. Эти соотношения *непосредственно* следуют из того факта, что $\lambda_1(s)$ и $\lambda_2(s)$ являются производящими функциями для времен первого достижения, как это было установлено в заключение § 4.

12. Предположим, что существует единственный поглощающий экран в начале координат. Пусть $u_{z,n}(x)$ — вероятность того, что частица, выходящая из точки $z>0$, через n шагов окажется в точке x, и пусть

$$U_z(s;\ x)=\sum_{n=0}^{\infty}u_{z,n}(x)\,s^n. \tag{9.3}$$

Используя результат задачи 11, показать, что $U_z(s;\ x)=V_{x-z}(s)-\lambda_2^z(s)\,V_x(s)$. Вывести отсюда, что

$$u_{z,n}(x)=v_{x-z,n}-(q/p)^z\cdot v_{x+z,n}. \tag{9.4}$$

Сравнить с результатом задачи 7 и вывести (9.4) из последней с помощью комбинаторных методов.

13. Другая формула для вероятности разорения (5.7). Разлагая выражение (4.11) в геометрическую прогрессию, показать, что

$$u_{z,n}=\sum_{k=0}^{\infty}\left(\frac{p}{q}\right)^{ka}w_{z+2ka,\,n}-\sum_{k=1}^{\infty}\left(\frac{p}{q}\right)^{ka-z}w_{2ka-z,\,n},$$

где $w_{z,n}$ определено формулой (5.9).

14. Показать, что если осуществленный в § 6 переход к пределу применить к выражению для $u_{z,n}$, приведенному в предыдущей задаче, то вероятность поглощения в течение короткого промежутка времени Δt окажется асимптотически равной [1])

$$\frac{1}{2}\,\Delta t\,(\pi Dt^3)^{-1/2}\,e^{-c(ct+z)/D}\sum_{k=-\infty}^{+\infty}(z+2ka)\,e^{-(z+2ka)^2/4Dt}.$$

(Указание. Воспользоваться нормальным приближением для биномиального распределения.)

15 [2]). Метод теории восстановления в приложении к задаче о разорении. Пусть при случайном блуждании с двумя поглощающими экранами (§ 4) $u_{z,n}$ и $u^*_{z,n}$ обозначают соответственно вероятности

---

[1]) Совпадение этой формулы с формулой (6.10) есть факт, известный в теории тэта-функций.

[2]) Задачи 15—17 содержат новый вывод основных результатов относительно одномерного случайного блуждания.

поглощения левым и правым экраном. Посредством надлежащей интерпретации доказать справедливость следующих двух равенств:

$$V_{-z}(s) = U_z(s)\, V_0(s) + U_z^*(s)\, V_{-a}(s);$$

$$V_{a-z}(s) = U_z(s)\, V_a(s) + U_z^*(s)\, V_0(s).$$

Решая эту систему относительно $U_z(s)$, вывести формулу (4.11).

16. Пусть $u_{z,n}(x)$ — вероятность того, что частица, выходящая из точки z, достигнет на n-м шаге точки x, не коснувшись до этого поглощающих экранов. Показать, что в обозначениях задачи 15 соответствующая производящая функция $U_z(s;\, x) = \sum u_{z,n}(x)\, s^n$ равна $U_z(s;\, x) = V_{x-z}(s) - U_z(s)\, V_x(s) - U_z^*(s)\, V_{x-a}(s)$ (не требуется никаких вычислений).

17. Продолжение. Производящую функцию $U_z(s;\, x)$ предыдущей задачи можно получить, положив $U_z(s;\, x) = V_{x-z}(s) - A\lambda_1^z(s) - B\lambda_2^z(s)$ и определив постоянные так, чтобы выполнялись граничные условия $U_z(s;\, x) = 0$ при $z = 0$ и $z = a$. Если в точках $^1/_2$ и $a - {}^1/_2$ поставить *отражающие экраны*, то граничными условиями будут $U_0(s;\, x) = U_1(s;\, x)$ и $U_a(s;\, x) = U_{a-1}(s;\, x)$.

18[1]. Симметричное неограниченное блуждание начинается в начале координат. Вероятность r-го возвращения в начало координат на n-м шаге равна вероятности первого прохождения через точку $x = r$ на $(n - r)$-м шаге. (Указание. Сравнить производящие функции.)

19. Доказать формулу

$$v_{x,n} = (2\pi)^{-1}\, 2^n p^{(n+x)/2} q^{(n-x)/2} \int_{-\pi}^{\pi} \cos^n t \cdot \cos tx \cdot dt,$$

показав, что удовлетворяется соответствующее разностное уравнение. Вывести что

$$V_x(s) = (2\pi)^{-1} \left(\frac{p}{q}\right)^{x/2} \int_{-\pi}^{\pi} \frac{\cos tx}{1 - 2(pq)^{1/2} \cdot s \cdot \cos t}\, dt.$$

20. В трехмерном случайном симметричном блуждании частица с вероятностью единица бесконечное число раз пересечет любую фиксированную прямую $x = m$, $y = n$. (Указание. См. задачу 1.)

21. В двумерном симметричном случайном блуждании, начинающемся в начале координат, вероятность того, что n-й шаг приведет частицу в точку (x, y), равна

$$(2\pi)^{-2}\, 2^{-n} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} (\cos\alpha + \cos\beta)^n \cdot \cos x\alpha \cdot \cos y\beta \cdot d\alpha\, d\beta.$$

Проверить эту формулу и найти аналогичную формулу для трех измерений (Указание. Проверить, что это выражение удовлетворяет соответствующему разностному уравнению.)

---

[1]) Это утверждение составляет содержание теоремы 3 § 4 гл. III.

22. Двумерное симметрическое блуждание начинается в начале координат. Пусть $D_n^2 = x^2 + y^2$ — квадрат расстояния частицы от точки (0, 0) в момент времени n. Доказать, что $E\left(D_n^2\right) = n$. (Указание. Вычислить $E\left(D_{n+1}^2 - D_n^2\right)$.)

23. Доказать, что при симметрическом случайном блуждании в d-мерном пространстве частица с вероятностью 1 бесконечно много раз будет возвращаться к точкам, которые уже были заняты. (Другими словами, траектория частицы имеет бесконечно много самопересечений.) (Указание. При каждом шаге вероятность попадания в новое положение не более $(2d-1)/2d$.)

ГЛАВА XV

ЦЕПИ МАРКОВА

§ 1. Определение

До сих пор мы занимались в основном только независимыми испытаниями, которые могут быть описаны следующим образом. Задано множество возможных исходов $E_1, E_2, \ldots$ (в конечном или бесконечном числе), каждому из которых отнесена определенная вероятность p_k; вероятности совместных исходов n испытаний определяются по правилу умножения $P\{(E_{j_0}, \ldots, E_{j_n})\} = p_{j_0} \cdot p_{j_1} \ldots p_{j_n}$. В теории цепей Маркова[1]) мы рассмотрим простейшее обобщение этой схемы, состоящее в том, что допускается зависимость исхода любого испытания от исхода предыдущего испытания (и только от него). При этом с исходом E_k не связана определенная вероятность p_k, но зато каждой паре (E_j, E_k) соответствует фиксированная *условная вероятность* p_{jk}: если в некотором испытании осуществилось E_j, то вероятность осуществления в следующем испытании исхода E_k равна p_{jk}. Кроме вероятностей p_{jk}, нужно задать вероятность a_k исхода E_k *в начальном* испытании. Чтобы p_{jk} имели указанный смысл, вероятности совместных исходов двух, трех или четырех испытаний нужно определить следующим образом:

$$P\{(E_j, E_k)\} = a_j p_{jk}, \quad P\{(E_j, E_k, E_r)\} = a_j p_{jk} p_{kr},$$
$$P\{(E_j, E_k, E_r, E_s)\} = a_j p_{jk} p_{kr} p_{rs}$$

и, вообще,

$$P\{(E_{j_0}, E_{j_1}, \ldots, E_{j_n})\} = a_{j_0} p_{j_0 j_1} p_{j_1 j_2} \ldots p_{j_{n-2} j_{n-1}} p_{j_{n-1} j_n}. \qquad (1.1)$$

Здесь начальное испытание имеет номер нуль, так что испытание номер один является вторым испытанием. (Такие обозначения удобны и уже использовались в предыдущей главе.)

Примеры. а) Каждая цепь Маркова эквивалентна некоторой урновой схеме в следующем смысле. Допустим, что каждому из рассматриваемых индексов поставлена в соответствие урна, и каждая урна содержит шары с метками $E_1, E_2, \ldots$. Состав урн остается неизменным, но меняется от урны к урне так, что вероятность извлечь шар с меткой E_k из j-й урны равна p_{jk}. При начальном

---

[1]) А. А. Марков (1856—1922).

(нулевом) испытании выбирается урна в соответствии с распределением $\{a_i\}$. Из этой урны случайно извлекают шар, и если его метка E_j, то следующее извлечение производят из j-й урны и т. д. Очевидно, при таком эксперименте вероятность последовательности $(E_{j_0}, \ldots, E_{j_n})$ определяется формулой (1.1). Мы видим, что понятие цепи Маркова не является более общим, чем понятие урновой схемы, но новая символика и терминология более практичны и наглядны.

б) *Независимые испытания* являются, конечно, частным случаем нашей схемы. В этом случае $p_{jk} = a_k$ для любого j.

Ясно, что если вероятность E_k в начальном (нулевом) испытании равна a_k, то $a_k \geqslant 0$, и $\sum a_k = 1$. Аналогично, так как за осуществлением E_j должно следовать осуществление некоторого E_k, при всех j и k должно быть

$$p_{j1} + p_{j2} + p_{j3} + \ldots = 1, \quad p_{jk} \geqslant 0. \tag{1.2}$$

Покажем, что для любых чисел a_k и p_{jk}, удовлетворяющих этим условиям, (1.1) есть законное определение вероятности в пространстве элементарных событий, соответствующем $n+1$ испытаниям. Так как числа, определенные в (1.1), очевидно, неотрицательны, нужно лишь доказать, что их сумма равна единице. Для этого фиксируем $j_0, j_1, \ldots, j_{n-1}$ и просуммируем числа (1.1) по всем возможным j_n. Используя (1.2) при $j = j_{n-1}$, убеждаемся, что сумма равна $a_{j_0} p_{j_0 j_1} \ldots p_{j_{n-2} j_{n-1}}$. Таким образом, сумма всех чисел (1.1) не зависит от n, а так как $\sum a_{j_0} = 1$, то эта сумма равна 1 для всех n.

Определение (1.1) формально содержит зависимость от числа испытаний, но те же соображения, что и выше, показывают согласованность определений (1.1) при всех n. Например, чтобы получить вероятность события «результатом первых двух испытаний является (E_j, E_k)», нужно фиксировать $j_0 = j$ и $j_1 = k$ и просуммировать вероятности (1.1) по всем возможным $j_2, j_3, \ldots, j_n$. Мы только что показали, что эта сумма равна $a_j p_{jk}$ и, таким образом, не зависит от n. Это означает, что обычно нет необходимости указывать явно число испытаний: событие $(E_{j_0}, \ldots, E_{j_r})$ имеет одну и ту же вероятность во всех пространствах элементарных событий, соответствующих более чем r испытаниям. В связи с независимыми испытаниями несколько раз уже говорилось, что с математической точки зрения более удовлетворительно ввести только одно пространство элементарных событий, соответствующее бесконечной последовательности испытаний, и рассматривать конечное число испытаний как начало бесконечной последовательности. Это положение верно и для цепей Маркова. К сожалению, пространство элементарных событий бесконечного числа испытаний лежит вне теории дискретных пространств, которой мы ограничиваемся в настоящей книге.

Резюмируя, мы получаем в качестве отправного пункта следующее

Определение. *Последовательность испытаний с возможными исходами E_1, E_2, ... называется цепью Маркова*[1]), *если вероятности совместного исхода нескольких испытаний определяются формулой* (1.1) *через начальное распределение вероятностей $\{a_k\}$ состояний E_k в момент* 0 *и через постоянные условные вероятности p_{jk} события E_k при условии, что E_j осуществилось в предыдущем испытании.*

Несколько изменим теперь нашу терминологию, чтобы приспособить ее к использованию в физических приложениях. Вместо того, чтобы говорить «результатом n-го испытания оказалось E_k», будем говорить, что *«в момент времени n система находится в состоянии E_k»*. Условная вероятность p_{jk} будет называться *вероятностью перехода $E_j \to E_k$* (из состояния E_j в состояние E_k).

Вероятности перехода p_{jk} могут быть расположены в *матрицу вероятностей перехода*:

$$P = \begin{Vmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{Vmatrix}, \qquad (1.3)$$

где первый индекс означает номер строки, а второй индекс — номер столбца. Ясно, что P — квадратная матрица с неотрицательными элементами, причем сумма элементов каждой строки равна единице. Такие матрицы (конечные или бесконечные) называются *стохастическими матрицами. Любая стохастическая матрица может служить матрицей вероятностей перехода; вместе с нашим начальным распределением $\{a_k\}$ она вполне определяет цепь Маркова.*

В некоторых случаях удобно нумеровать состояния, начиная не с 1, а с 0. В таких случаях к матрице P добавляются строка и столбец с нулевыми номерами.

§ 2. Примеры

В этом параграфе рассматривается много примеров, которые должны помочь читателю освоиться с понятием цепи Маркова. Для

[1]) Мы рассматриваем здесь только частный вид цепей Маркова и, строго говоря, здесь и в следующих параграфах термин «цепь Маркова» должен каждый раз уточняться прибавлением слов «с постоянными вероятностями перехода». Однако, общий тип цепей Маркова рассматривается редко. Он будет определен в § 10, где свойства цепей Маркова будут рассмотрены в связи с общими стохастическими процессами. Там же читатель найдет примеры зависимых испытаний, не образующих цепи Маркова.

экономии места мы будет в дальнейшем ссылаться на эти примеры, иллюстрируя ими различные определения и теоремы. Читатель не должен все время помнить эти примеры; достаточно лишь возвращаться к ним при соответствующих ссылках. Приложения к задаче о тасовании колоды карт см. в § 9.

а) Предположим, что имеется только два возможных состояния E_1 и E_2, которые будем также называть «успехом» и «неудачей». Матрица вероятностей перехода имеет вид

$$P=\left\| \begin{matrix} p & q \\ p' & q' \end{matrix} \right\| \qquad p+q=p'+q'=1,$$

где p и p' вероятности того, что успех следует за успехом и успех следует за неудачей соответственно. В качестве конкретного примера рассмотрим шар, движущийся со скоростью ± 1 вдоль оси x. В моменты времени 1, 2, ... шар изменяет направление движения с вероятностью q и сохраняет его с вероятностью p. Если E_1 означает скорость $+1$, а E_2 — скорость -1, то матрица вероятностей перехода имеет описанный выше вид с заменой q' на p и p' на q. (Этот эксперимент можно провести, используя гладкую доску с установленной на ней правильной системой колышков.)

б) **Случайное блуждение с поглощающимися экранами.** Пусть возможными состояниями будут $E_0, E_1, \ldots, E_a$. Рассмотрим матрицу вероятностей перехода

$$P=\left\| \begin{matrix} 1 & 0 & 0 & 0 \ldots 0 & 0 & 0 \\ q & 0 & p & 0 \ldots 0 & 0 & 0 \\ 0 & q & 0 & p \ldots 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \quad \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \quad \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 \ldots q & 0 & p \\ 0 & 0 & 0 & 0 \ldots 0 & 0 & 1 \end{matrix} \right\|.$$

Из каждого «внутреннего» состояния $E_1, \ldots, E_{a-1}$ возможен переход в любое из двух соседних состояний (с вероятностями $p_{i,\,i+1}=p$ и $p_{i,\,i-1}=q$). Однако из E_0 или E_a не возможен переход ни в какое другое состояние; система может переходить из одного состояния в другое до тех пор, пока не достигнет E_0 или E_a, после этого процесс останавливается. Ясно, что эта цепь Маркова лишь терминологией отличается от модели случайного блуждания с поглощающими экранами в точках 0 и a, которая исследовалась в предыдущей главе. Допустим, что случайное блуждание начинается в фиксированной точке z. На языке цепей Маркова это означает, что начальное распределение имеет вид $a_z = 1$ (и, значит, $a_x = 0$ при $x \neq z$). Если мы хотим выбрать начальное состояние случайно, следует положить $a_k = (a+1)^{-1}$ при $k = 0, 1, \ldots, a$.

в) **Упругие экраны.** Рассмотрим теперь матрицу, которая отличается от матрицы предыдущего примера только 1-й и $(a-1)$-й строками. Выберем $0 \leqslant \delta_0 \leqslant 1$ и $0 \leqslant \delta_a \leqslant 1$ и положим

$$P = \left\| \begin{array}{ccccccc} 1 & 0 & 0 & 0 \ldots & 0 & 0 & 0 \\ (1-\delta_0)q & \delta_0 q & p & 0 \ldots & 0 & 0 & 0 \\ 0 & q & 0 & p \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 \ldots & 0 & p & 0 \\ 0 & 0 & 0 & 0 \ldots & q & \delta_a p & (1-\delta_a)p \\ 0 & 0 & 0 & 0 \ldots & 0 & 0 & 1 \end{array} \right\|.$$

Вероятности перехода те же, что и в предыдущем примере, с тем отличием, что теперь переход из E_1 в E_0 происходит с вероятностью только $(1-\delta_0)q$, а с вероятностью $\delta_0 q$ система остается в состоянии E_1; аналогичное утверждение имеет место и для E_{a-1}. При $\delta_0 = \delta_a = 0$ наша матрица совпадает с предыдущей. Если же $\delta_0 = \delta_a = 1$, то переход в состояние E_0 и E_a невозможен. Система с начальным состоянием E_j может переходить из одного внутреннего состояния в другое, но никогда не достигнет состояний E_0 или E_a. В терминах случайных блужданий последняя ситуация соответствует *отражающим* экранам (см. гл. XIV). Состояния системы можно также интерпретировать как капитал одного из игроков в игре, при которой суммарный капитал двух игроков равен a. Всякий раз, когда первый игрок проигрывает свой последний рубль, он получает его от своего противника с вероятностью δ_0, а с вероятностью $1-\delta_0$ игра прекращается. В случае двух отражающих экранов игра никогда не окончится.

г) **Циклическое случайное блуждание.** Пусть возможными состояниями снова будут $E_1, E_2, \ldots, E_a$, но порядок их будет циклическим, так что E_a имеет соседями E_{a-1} и E_1. Если, как и прежде, система всегда переходит или в правое, или в левое соседнее состояние, то строки матрицы P будут те же, что и в предшествующем примере, за исключением того, что первая строка будет иметь вид $(0, p, 0, 0, \ldots, 0, q)$, а последняя — вид $(p, 0, 0, 0, \ldots, 0, q, 0)$.

Пусть теперь возможны переходы из одного любого состояния в любое другое. Пусть $q_0, q_1, \ldots, q_{a-1}$ соответственно вероятности остаться на месте или передвинуться на $1, 2, \ldots, a-1$ единиц вправо (переход на k единиц вправо — то же самое, что и переход на $a-k$ единиц влево). Тогда P будет циклической матрицей

$$P=\left\|\begin{matrix} q_0 & q_1 & q_2 & \cdots & q_{a-2} & q_{a-1} \\ q_{a-1} & q_0 & q_1 & \cdots & q_{a-3} & q_{a-2} \\ q_{a-2} & q_{a-1} & q_0 & \cdots & q_{a-4} & q_{a-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ q_1, & q_2, & q_3 & \cdots & q_{a-1} & q_0 \end{matrix}\right\|.$$

Если $q_1=p$, $q_{a-1}=q$ и $q_k=0$ при $1<k<a-1$, то случайное блуждание сводится к простейшему случаю, рассматривавшемуся в начале этого примера (продолжение см. в примере 2, г гл. XVI).

д) **Неограниченное случайное блуждание.** Неограниченное одномерное случайное блуждание есть цепь Маркова, состояния которой естественно расположить в бесконечную последовательность $(\ldots, E_{-2}, E_{-1}, E_0, E_1, E_2, \ldots)$. Чтобы записать матрицу вероятностей перехода в обычной форме, нужно переставить состояния. Например, можно записать их в порядке $(E_0, E_1, E_{-1}, E_2, E_{-2}, \ldots)$. Первой строкой матрицы P будет тогда $(0, p, q, 0, 0, \ldots)$, второй $(q, 0, 0, p, 0, 0, \ldots)$ и т. д. К сожалению, естественная симметрия формул при этом исчезает. Положение становится еще хуже в случае блуждания в двух измерениях. В таких случаях методы настоящей главы для вывода точных формул неудобны, но общие теоремы сохраняются и дают ценные результаты.

е) **Модель Эренфестов для процесса диффузии.** Мы уже рассматривали однажды цепь с $a+1$ состояниями $E_0, E_1, \ldots, E_a$ и возможными переходами только в состояния, соседние справа и слева; теперь положим $p_{j,\,j+1}=1-j/a$ и $p_{j,\,j-1}=j/a$, так что

$$P=\left\|\begin{matrix} 0 & 1 & 0 & 0 & \ldots\, 0 & 0 \\ a^{-1} & 0 & 1-a^{-1} & 0 & \ldots\, 0 & 0 \\ 0 & 2a^{-1} & 0 & 1-2a^{-1} & \ldots\, 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots\, 0 & a^{-1} \\ 0 & 0 & 0 & 0 & \ldots\, 1 & 0 \end{matrix}\right\|.$$

Эта цепь имеет две интересные физические интерпретации. При изучении различных задач о возвращении в статистической механике

П. и Т. Эренфесты[1]) описали мыслимый эксперимент, при котором a молекул распределены по двум резервуарам A и B. В момент n случайно выбирается и перемещается из своего резервуара в другой одна молекула. Пусть состояние системы определяется числом молекул в A. Предположим, что в некоторый момент в резервуаре A содержатся ровно k молекул. При следующем испытании система переходит в E_{k-1} или E_{k+1} в зависимости от того, будет ли молекула выбрана из A или из B; соответствующие вероятности равны k/a и $(a-k)/a$, и, следовательно, эксперимент Эренфестов описывается цепью Маркова. Кроме того, цепь Маркова может быть также интерпретирована, как *диффузия* при наличии *центральной силы*, т. е. как случайное блуждание, в котором вероятность перемещения изменяется при изменении положения. При $x=j$ переход частицы направо (или налево) будет более вероятен, если $j<a/2$ (соответственно если $j>a/2$). Это означает, что частица имеет тенденцию двигаться по направлению к $x=a/2$, что соответствует силе притяжения, возрастающей пропорционально расстоянию от точки $x=a/2$.

[Модель Эренфестов была уже описана в примере 2, в гл. V; см. также пример 6, а и задачу 12.]

ж) **Задачи в размещении.** В гл. I мы рассмотрели случайное размещение шаров по a ячейкам. Пусть состояние системы определяется числом занятых ячеек. Если j ячеек заняты, вероятность того, что следующий шар попадет в одну из пустых ячеек, равна $(a-j)/a$. Следовательно, эксперимент описывается цепью с вероятностями перехода $p_{jj}=j/a$, $p_{j,\,j+1}=(a-j)/a$ и $p_{j,\,k}=0$ для всех других комбинаций j и k. Начальное распределение (все ячейки пусты) определяется вероятностями $p_0=1$, $p_k=0$ при $1\leqslant k\leqslant a$ (см. пример 2, д гл. XVI).

з) **Серии успехов.** Рассмотрим последовательность испытаний Бернулли и условимся говорить, что в момент n мы наблюдаем состояние E_0, если n-е испытание привело к неудаче, и состояние E_k ($k=1, 2, \ldots, n$), если последняя неудача наблюдалась при испытании с номером $n-k$ (предполагается, что нулевое испытание привело к неудаче). Другими словами, индекс k состояния E_k указывает на длину серии успехов, оканчивающейся на n-м месте. Очевидно, что мы имеем дело с цепью Маркова, для которой возможны только переходы $E_k \to E_0$ и $E_k \to E_{k+1}$, а соответствующая матрица

[1]) P. und T. Ehrenfest, Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem, *Physikalische Zeitschrift*, 8 (1907), 311—314.

Ming Chen Wang and G. E. Uhlenbeck, On the theory of the Brownian motion II, *Reviews of Modern Physics*, 17 (1945), 323—342.

Более полное рассмотрение (методами, по существу эквивалентными методам гл. XIII) см. M. Kac, Random walk and the theory of Brownian motion, *American Mathematical Monthly*, 54 (1947), 369—391. См. также B. Friedman, A simple urn model, *Communications on Pure and Applied Mathematics*, 2 (1949), 59—70.

вероятностей перехода имеет вид

$$P=\left\|\begin{matrix} q & p & 0 & 0 & 0 & \ldots \\ q & 0 & p & 0 & 0 & \ldots \\ q & 0 & 0 & p & 0 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}\right\|.$$

и) **Рекуррентные события.** Приведенный выше пример является частным случаем более интересной цепи Маркова. Пусть $\mathcal{E}$ — произвольное рекуррентное событие, время возвращения которого имеет распределение $\{f_n\}$. Условимся считать, что при нулевом испытании осуществилось $\mathcal{E}$. Мы говорим, что в момент времени n система находится в состоянии E_0, если при n-м испытании наступило $\mathcal{E}$, и в состоянии E_k, если событие $\mathcal{E}$ произошло в последний раз при испытании с номером $n-k$. (Можно сказать, что мы имеем дело с временем ожидания в обратном направлении.) Как и в предыдущем примере, очевидно, что из состояния E_k можно перейти в следующий момент только в состояние E_0 (если наступит $\mathcal{E}$) или в E_{k+1}. Положим

$$s_k=f_1+\ \ldots\ +f_k,\ q_k=\frac{f_{k+1}}{1-s_k},\quad p_k=1-q_k=\frac{1-s_{k+1}}{1-s_k}. \tag{2.1}$$

Если система находится в состоянии E_k, то время ожидания $\mathcal{E}$ превосходит k, и (условная) вероятность наступления $\mathcal{E}$ при следующем испытании равна q_k. В соответствии с этим вероятности переходов $E_k \to E_{k+1}$ и $E_k \to E_0$ равны соответственно p_k и q_k. Рассмотрим типичную цепочку последовательных состояний системы

$$E_0E_0E_1E_2E_3E_0E_1E_0E_1E_2E_0E_0$$

(первое E_0 отвечает нулевому испытанию). В этом случае для последовательных времен ожидания получаем значения 1, 4, 2, 3, 1, а вероятность нашей цепочки равна $f_1f_4f_2f_3f_1$. Но

$$f_1f_4f_2f_3f_1=q_0p_0p_1p_2q_3p_0q_1p_0p_1q_2q_0, \tag{2.2}$$

что согласуется с формулой (1.1) для вероятностей в цепях Маркова. Это рассуждение применимо ко всем последовательностям, поэтому описанный процесс является цепью Маркова с матрицей вероятностей перехода

$$P=\left\|\begin{matrix} q_0 & p_0 & 0 & 0 & 0 & \ldots \\ q_1 & 0 & p_1 & 0 & 0 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ q_2 & 0 & 0 & p_2 & 0 & \ldots \\ q_3 & 0 & 0 & 0 & p_3 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}\right\|.$$

[Продолжение см. в примере 6, в].

к) **Последовательный анализ.** В гл. XIV, § 8, мы рассмотрели следующее обобщение задачи о разорении, связанное с последовательным анализом. Задана последовательность взаимно независимых случайных величин X_ν, которые принимают только целые значения (положительные или отрицательные) и имеют одинаковые распределения $\{p_k\}$ $k=0, \pm 1, \pm 2, \ldots$. Положим $S_n = X_1 + \ldots + X_n$. Тогда существует наименьший индекс n, при котором или $S_n \geqslant b$ или $S_n \leqslant -z$; здесь b и z — фиксированные положительные числа, а n, конечно, случайная величина. Основная задача теории последовательного анализа Вальда состоит в нахождении распределения вероятностей для n, а также вероятностей осуществления каждой из возможностей $S_n \leqslant -z$ и $S_n \geqslant b$. Интерпретируем эту задачу следующим образом. Положим $a = b + z - 1$ и рассмотрим цепь Маркова с возможными состояниями $x = 0, 1, 2, \ldots, a$. Начальным состоянием системы является z. Если значение S_1 таково, что $-z < S_1 < b$, то мы говорим, что на первом шаге система переходит в состояние $x = S_1 + z$. Если $S_1 \leqslant -z$, то первый шаг переводит систему в состояние $x = 0$; если $S_1 \geqslant b$, то совершается переход в $x = a$. Если уже достигнуто одно из граничных состояний, то система остается в нем навсегда (т. е. процесс прекращается). Во всех остальных случаях процесс продолжается уже описанным образом; в момент t система находится в состоянии $x = S_n + z$, если все предыдущие суммы $S_1, S_2, \ldots, S_{n-1}$ лежат в интервале $-z < S_k < b$. В остальных случаях система находится в состоянии $x = 0$ или $x = a$ в соответствии с тем, была ли первая сумма S_k, оказавшаяся вне этого интервала, отрицательна или положительна. Матрица вероятностей перехода имеет тогда вид

$$P = \begin{Vmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ r_1 & p_0 & p_1 & p_2 & \ldots & p_{a-1} & \rho_1 \\ r_2 & p_{-1} & p_0 & p_1 & \ldots & p_{a-2} & \rho_2 \\ r_3 & p_{-2} & p_{-1} & p_0 & \ldots & p_{a-3} & \rho_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_a & p_{-a+1} & p_{-a+2} & p_{-a+3} & \ldots & p_0 & \rho_a \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{Vmatrix},$$

где

$$r_k = p_{-k} + p_{-k-1} + p_{-k-2} + p_{-k-3} + \ldots$$

и

$$\rho_k = p_{a-k+1} + p_{a-k+2} + \ldots .$$

В качестве примера рассмотрим схему выборочного контроля, предложенную Бартки. Для того чтобы проверить партию изделий, последовательно извлекаются выборки объема N и в каждой из них определяется число бракованных изделий. Предположим, что выборки независимые и что число бракованных изделий в каждой имеет одно и то же биномиальное распределение. Допускается одно бракованное изделие на выборку, так что их полное число в k-й выборке удобно обозначить через X_k+1. Тогда при $k \geqslant 0$

$$p_k = \binom{N}{k+1} p^{k+1} q^{N-k-1} \tag{2.3}$$

и $p_{-1} = q^N$, $p_x = 0$ при $x < -1$. Процедура выборочного контроля состоит в следующем: производится предварительный выбор, и вся партия признается пригодной, если выборка не содержит бракованных изделий, и бракуется, если число бракованных изделий превосходит a. В том и в другом случае процесс оканчивается и никакой цепи Маркова не возникает. Если, однако, число бракованных изделий z лежит в границах 0 и a $(1 \leqslant z \leqslant a)$, то выбор продолжается описанным способом до тех пор, пока состояния цепи заключены между 1 и a. Рано или поздно наша система достигнет или 0, и в этом случае партия принимается, или $a+1$, и в этом случае партия бракуется.

л) **Пример из генетики**[1]. Рассмотрим популяцию, которая сохраняет постоянную численность в результате того, что в каждом из последовательных поколений отбирается N индивидуумов. Некоторый ген, который может иметь вид A или a, представлен в каждом поколении $2N$ раз. Если в n-м поколении форма A встречается j раз то форма a встречается $(2N-j)$ раз. Если принять гипотезу о случайном скрещивании, то распределение этих форм в следующем поколении определяется $2N$ испытаниями Бернулли, при которых A-ген имеет вероятность $j/2N$. Поэтому мы имеем цепь Маркова с вероятностями перехода

$$p_{jk} = \binom{2N}{k} \left(\frac{j}{2N}\right)^k \left(1 - \frac{j}{2N}\right)^{2N-k}.$$

[См. пример 8, в.]

м) Скрещиваются две особи, и среди их прямых потомков случайно выбираются две особи разного пола, которые в свою очередь скрещиваются и т. д. Каждый из родителей принадлежит к одному из генотипов AA, Aa и aa. Следует различать шесть возможных комбинаций родителей, которые мы обозначим следующим образом: $E_1 = AA \times AA$, $E_2 = AA \times Aa$, $E_3 = Aa \times Aa$, $E_4 = Aa \times aa$,

---

[1] Эта задача подробно исследовалась Р. А. Фишером и С. Райтом. Формулировка в терминах цепей Маркова принадлежит Г. Малешоту, Sur un problème de probabilités en chaîne que pose la génétique, *Comptes rendus de l'Académie des Sciences*, 219 (1944), 379—381.

$E_5 = aa \times aa$, $E_6 = AA \times aa$. Используя правила гл. V, нетрудно показать, что матрица вероятностей перехода имеет в этом случае вид

$$\left\| \begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{1}{16} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{16} & \frac{1}{8} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{matrix} \right\|.$$

[Обсуждение продолжено в задаче 4; полное исследование дано в примере 4, б гл. XVI.]

§ 3. Вероятности перехода за n шагов

Переход из E_j в E_k ровно за n шагов может быть осуществлен различными путями $E_j \to E_{j_1} \to E_{j_2} \to \ldots \to E_{j_{n-1}} \to E_k$. Условная вероятность того, что переход совершился по некоторому фиксированному пути, при условии, что в какой-то момент система находилась в состоянии E_j, равна $p_{jj_1}p_{j_1j_2} \ldots p_{j_{n-1}k}$. Сумма соответствующих выражений для всех возможных путей равна *вероятности того, что система в момент $r+n$ находится в состоянии E_k, при условии, что в момент r она находилась в состоянии E_j. Мы будем обозначать эту вероятность через $p_{jk}^{(n)}$.*

В частности, $p_{jk}^{(1)} = p_{jk}$ и

$$p_{jk}^{(2)} = \sum_{\nu} p_{j\nu}p_{\nu k}. \tag{3.1}$$

По индукции легко находим *рекуррентную формулу*

$$p_{jk}^{(n+1)} = \sum_{\nu} p_{j\nu}p_{\nu k}^{(n)}; \tag{3.2}$$

дальнейшая индукция по m показывает, что и вообще

$$p_{jk}^{(m+n)} = \sum_{\nu} p_{j\nu}^{(m)}p_{\nu k}^{(n)}. \tag{3.3}$$

Это уравнение выражает тот факт, что за первые m шагов система переходит из E_j в некоторое промежуточное состояние E_ν, а за последующие n шагов — из E_ν в E_k. Выполнение уравнения (3.3) есть характерное свойство цепи Маркова. Для более общих процессов (см. § 10) выполняется аналогичное уравнение, в котором, однако, последний множитель зависит не только от ν и k, но также и от j.

Подобно вероятностям p_{jk}, образующим матрицу P, *мы расположим* $p_{jk}^{(n)}$ *в матрицу, обозначаемую* P^n. Согласно (3.2), чтобы получить элемент $p_{jk}^{(n+1)}$ матрицы P^{n+1}, нужно умножить элемент j-й строки матрицы P на соответствующий элемент k-го столбца матрицы P^n и сложить все такие произведения. Эта операция называется умножением матрицы P на матрицу P^n и выражается символически равенством $P^{n+1}=PP^n$. Определение произведения матриц позволяет назвать P^n n-й степенью P; уравнение (3.3) выражает ассоциативный закон умножения $P^{m+n}=P^mP^n$.

Для того чтобы уравнение (3.3) было верно для всех $n\geqslant 0$, мы положим $p_{jk}^{(0)}=0$ при $j\neq k$ и $p_{jj}^{(0)}=1$.

Примеры. а) В тривиальном случае *независимых испытаний* все степени P равны. Очевидно, что $P^n=P$ при всех n.

б) В примере 2, з, где рассматривались *серии успехов*, вероятности перехода за n шагов могут быть найдены непосредственно. Например, за три шага система может перейти из состояния E_k только в E_{k+3}, E_0, E_1, E_2, и соответствующие вероятности равны, очевидно, p^3, q, qp и qp^2. Поэтому

$$P^2=\left\|\begin{matrix} q & qp & p^2 & 0 & 0 & \dots \\ q & qp & 0 & p^2 & 0 & \dots \\ q & qp & 0 & 0 & p^2 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots \end{matrix}\right\|,\quad P^3=\left\|\begin{matrix} q & qp & qp^2 & p^3 & 0 & 0 & \dots \\ q & qp & qp^2 & 0 & p^3 & 0 & \dots \\ q & qp & qp^2 & 0 & 0 & p^3 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \dots \end{matrix}\right\|.$$

В этом случае ясно, что последовательность матриц P^n сходится к матрице, у которой все элементы в k-м столбце равны qp^k.

Полные вероятности. Если начальная вероятность состояния E_j равна a_j, то (безусловная) вероятность того, что система в момент n находится в состоянии E_k, равна, очевидно,

$$a_k^{(n)}=\sum_j a_j p_{jk}^{(n)}. \tag{3.4}$$

Важнейшие свойства цепей Маркова зависят от асимптотического поведения вероятностей $p_{jk}^{(n)}$ при $n\to\infty$. Интуитивно можно ожидать, что влияние начального положения постепенно ослабевает, так что при больших n вероятность нахождения системы в момент n в состоянии E_k не зависит от состояния в момент 0. Мы подразумеваем под этим, что $p_{jk}^{(n)}$ стремится к пределу u_k, не зависящему от j (свойство, называющееся *эргодичностью*). Ниже будет показано, что это интуитивное предположение, вообще говоря, верно, хотя существуют и исключения.

§ 4. Замкнутые множества состояний

Будем говорить, что состояние E_k *достижимо из состояния* E_j, *если существует такое* $n \geqslant 0$, *что* $p_{jk}^{(n)} > 0$ (т. е. если имеется положительная вероятность попасть из состояния E_j в состояние E_k, включая случай $E_k = E_j$). Например, при неограниченном случайном блуждании каждое состояние может быть достигнуто из любого другого состояния. Если же некоторое состояние является поглощающим экраном, то из него не достижимо никакое другое состояние.

Определение. *Множество состояний C называется замкнутым, если никакое состояние вне C не может быть достигнуто ни из какого состояния, входящего в C. Наименьшее замкнутое множество, содержащее C, называется замыканием C.*

Если некоторое состояние E_k образует замкнутое множество, то оно называется поглощающим состоянием.

Цепь называется неприводимой, если в ней нет никаких замкнутых множеств, кроме множества всех состояний.

Ясно, что множество C замкнуто тогда и только тогда, когда $p_{jk} = 0$ при всех j и k, таких, что E_j входит в C, а E_k не входит. В этом случае из соотношения (3.2) вытекает, что $p_{jk}^{(n)} = 0$ при любом n. Итак, установлена очевидная

Теорема. *Если в матрице P^n вычеркнуть все строчки и столбцы, соответствующие состояниям, не входящим в замкнутое множество C, то останется матрица, для которой вновь выполнены основные уравнения* (3.2) *и* (3.3).

Это значит, что мы имеем цепь Маркова, определенную на C, и эту *цепь можно изучать независимо от прочих состояний.*

Состояние E_k является поглощающим тогда и только тогда, когда $p_{kk} = 1$. В этом случае матрица, о которой идет речь в последней теореме, сводится к единственному элементу. Замыканием состояния E_j *является множество состояний, которые могут быть достигнуты из* E_j (включая само E_j). Это замечание можно сформулировать в виде следующего полезного критерия:

Критерий неприводимости. *Цепь является неприводимой тогда и только тогда, когда любое ее состояние может быть достигнуто из любого другого состояния.*

Пример. Для того, чтобы найти все замкнутые множества, достаточно знать, какие элементы p_{jk} матрицы p равны нулю, а какие — положительны. В соответствии с этим мы будем использовать символ $*$ для обозначения положительных элементов.

Рассмотрим типичную матрицу, скажем,

$$P=\left\|\begin{matrix} 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \\ 0 & * & * & 0 & * & 0 & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 \\ * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \end{matrix}\right\|.$$

В пятой строке символ $*$ имеется только на пятом месте, и поэтому $p_{55}=1$: состояние E_5 является *поглощающим*. Третья и восьмая строки содержат только по одному положительному элементу и ясно, что E_3 и E_8 образуют *замкнутое* множество. Из E_1 переходы возможны только в E_4 и E_9, а из них — только в E_1, E_4 и E_9. В соответствии с этим три состояния E_1, E_4 и E_9 образуют другое *замкнутое* множество.

Теперь ясно, что сложная структура матрицы обусловлена главным образом неудачными обозначениями. Изменим нумерацию состояний следующим образом:

$$E_1'=E_5;\ E_2'=E_3;\ E_3'=E_8;\ E_4'=E_1;\ E_5'=E_9;$$
$$E_6'=E_4;\ E_7'=E_2;\ E_8'=E_7;\ E_9'=E_6.$$

В результате этого элементы матрицы P изменят свой порядок. Новая матрица P примет вид

$$P'=\left\|\begin{matrix} * & 00 & 000 & 000 \\ 0 & 0* & 000 & 000 \\ 0 & *0 & 000 & 000 \\ 0 & 00 & 0** & 000 \\ 0 & 00 & 0** & 000 \\ 0 & 00 & *00 & 000 \\ * & *0 & 0*0 & *00 \\ 0 & 00 & 000 & *** \\ 0 & 00 & 000 & *00 \end{matrix}\right\|.$$

При такой форме записи замкнутые множества (E_1'), (E_2', E_3') и (E_4', E_5', E_6') становятся очевидными. Из E_7' переход возможен в каждое из трех замкнутых множеств, и поэтому замыканием E_7' является множество состояний E_1', E_2', E_3', E_4', E_5', E_6', E_7'.

Из E_8' переход возможен в E_7' и E_9' и, значит, в каждое из замкнутых множеств: замыкания E_8' и E_9' состоят из всех девяти состояний.

Вычеркивая все строки и все столбцы, не отвечающие замкнутому множеству состояний, мы можем выделить из P' три стохастические матрицы

$$\left\| * \right\| \quad \left\| \begin{matrix} 0 & * \\ * & 0 \end{matrix} \right\| \quad \left\| \begin{matrix} 0 & * & * \\ 0 & * & * \\ * & 0 & 0 \end{matrix} \right\|. \tag{4.1}$$

Никаких других стохастических матриц P' не содержит.

Читателю рекомендуется самому найти поглощающие состояния и замкнутые множества в примерах § 2.

§ 5. Классификация состояния

Рассмотрим произвольное, но фиксированное состояние E_j, и предположим, что в момент 0 система находится в E_j. Каждый раз, когда система возвращается в E_j, восстанавливается первоначальное положение и процесс начинается сначала. Следовательно, *возвращение в E_j есть рекуррентное событие* в смысле определения гл. XIII. Если же исходным состоянием системы является некоторое другое состояние E_i, то возвращение в E_j становится *рекуррентным событием с запаздыванием* в смысле определения § 5 гл. XIII. Поэтому должно быть ясно, что цепи Маркова являются лишь частным случаем рекуррентных событий, новое состоит в том, что теперь мы имеем дело со многими рекуррентными событиями одновременно.

Каждое состояние E_j характеризуется своим *распределением времени возвращения* $\{f_j^{(n)}\}$. Здесь $f_j^{(n)}$—вероятность того, что *первое возвращение* в E_j произойдет в момент n. Величины $f_j^{(n)}$ можно вычислить, отправляясь от $p_{jj}^{(n)}$ и используя следующие очевидные соотношения[1]):

$$\begin{aligned} & f_j^{(1)} = p_{jj}, \quad f_j^{(2)} = p_{jj}^{(2)} - f_j^{(1)} p_{jj}, \ \ldots, \\ & f_j^{(n)} = p_{jj}^{(n)} - f_j^{(1)} p_{jj}^{(n-1)} - f_j^{(2)} p_{jj}^{(n-2)} - \ \ldots \ - f_j^{(n-1)} p_{jj}. \end{aligned} \tag{5.1}$$

[1]) Они устанавливают, что вероятность первого возвращения в состояние E_j в момент n равна вероятности возвращения в момент n минус вероятность того, что первое возвращение имеет место в некоторый момент $\nu=1, 2, \ldots, n-1$, а за ним следует повторное возвращение в E_j в момент n. В обозначениях (3.1) гл. XIII имеем $p_{jj}^{(n)} = u_n$ и $f_j^{(n)} = f_n$.

Эти соотношения являются частным случаем основного уравнения теории рекуррентных событий (3.1) гл. XIII.

Сумма

$$f_j = \sum_{n=1}^{\infty} f_j^{(n)} \tag{5.2}$$

есть вероятность того, что исходя из E_j *система когда-нибудь вернется в* E_j. *Состояние* E_j *называется возвратным, если* $f_j = 1$; *в этом случае среднее время возвращения равно*

$$\mu_j = \sum_{n=1}^{\infty} n f_j^{(n)}. \tag{5.3}$$

Мы будем называть E_j *нулевым состоянием*, если $\mu_j = \infty$.

Если в начальный момент система находится в E_i, то *время ожидания первого достижения* E_j *имеет распределение* $f_{ij}^{(n)}$, *где*

$$f_{ij}^{(1)} = p_{ij}, \quad f_{ij}^{(n)} = p_{ij}^{(n)} - \sum_{\nu=1}^{n-1} f_{ij}^{(n-\nu)} p_{ij}^{(\nu)}. \tag{5.4}$$

Это уравнение опять-таки не является чем-то характерным только для цепей Маркова, оно выполняется для произвольных рекуррентных событий с запаздыванием. Конечно, если E_j недостижимо из E_i, то $f_{ij}^{(n)} = 0$ при всех n. Вообще

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \tag{5.5}$$

есть вероятность того, что исходя из E_i *система когда-либо достигнет* E_j.

Основные факты, доказанные в § 3 и 5 гл. III, могут быть суммированы для нашего случая следующим образом:

(I) *Состояние* E_j *невозвратно, если* $f_j < 1$. *Для этого необходимо и достаточно, чтобы* $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$; *в этом случае автоматически*[1] $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ *при любом* i.

(II) *Состояние* E_j *является возвратным нулевым состоянием, если* $f_j = 1$, *но среднее время возвращения* μ_j *бесконечно. Для*

---

[1]) Это тривиально следует из (5.4). Кроме того, можно воспользоваться теоремой § 5 гл. XIII, из которой наше предложение следует как частный случай.

этого необходимо и достаточно, чтобы $\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$, *но* $p_{jj}^{(n)} \to 0$. *В этом случае*

$$p_{ij}^{(n)} \to 0 \quad \textit{при} \quad n \to \infty \tag{5.6}$$

для любого i.

(III) *Состояние* E_j *называется периодическим с периодом* $t > 1$, *если* $p_{jj}^{(n)} = 0$ *для любого* n *не кратного* t *и* t — *наименьшее целое число, обладающее этим свойством* (*т. е. система не может вернуться в* E_j *за время, отличное от* t, $2t$, $3t$, ...).

(IV) *Если состояние* E_j *является возвратным и непериодическим, то*[1])

$$p_{ij}^{(n)} \to \mu_j^{-1} f_{ij} \quad \textit{при} \quad n \to \infty \tag{5.7}$$

и, в частности,

$$p_{jj}^{(n)} \to \mu_j^{-1} \quad \textit{при} \quad n \to \infty. \tag{5.8}$$

(*Если состояние* E_j *нулевое, то следует положить* $\mu_j^{-1} = 0$.)

(V) *Если состояние* E_j *возвратно и имеет период* t, *то соотношение* (5.8) *надо заменить другим соотношением*:

$$p_{jj}^{(nt)} \to t\mu_j^{-1} \quad \textit{при} \quad n \to \infty. \tag{5.9}$$

Возвратное состояние, не являющееся ни нулевым, ни периодическим, будет называться эргодическим[2]).

Примеры. а) Рассмотрим матрицу P' примера § 4 (опуская штрихи). Так как состояние E_1 является поглощающим, то оно возвратно. Из E_2 система необходимо переходит в E_3, а оттуда возвращается назад в E_2. Поэтому E_2 и E_3 — возвратные периодические состояния с периодом 2 и средним временем возвращения 2. Состояния E_4, E_5, E_6 образуют замкнутое подмножество и переходы между ними регулируются последней из матриц (4.1). Очевидно, что эти состояния возвратные и непериодические. (В дальнейшем мы увидим, что в конечной цепи Маркова невозможны возвратные нулевые состояния.)

---

[1]) См. примечание на стр. 380.

[2]) К сожалению, не существует общепринятой терминологии. А. Н. Колмогоров называет невозвратные состояния *несущественными*, однако новейшие исследования показали, что главный интерес, как практический, так и теоретический, представляют именно невозвратные состояния. Термин «эргодическое состояние» как синоним термина «возвратное, ненулевое, непериодическое состояние» является до некоторой степени общепринятым, но и его иногда заменяют термином «*положительное состояние*». Иногда же «эргодическим» называют любое возвратное состояние.

Из E_7 переход возможен только в одно из трех замкнутых множеств, в дальнейшем система навсегда остается в этом множестве. Поэтому состояние E_7 невозвратно. Из E_9 система переходит в E_7, и возвращение в E_9 невозможно, так что состояние E_9 тоже невозвратно. Наконец, исходя из E_8 система рано или поздно попадет в E_7 или E_9, после чего она уже не сможет вернуться в E_8. Итак, доказана невозвратность всех трех состояний E_7, E_8 и E_9.

б) Напомним, что при неограниченном случайном блуждании [пример 2, д] все состояния возвратны, если $p=q$, и невозвратны в противном случае [см. пример 8, г].

Не всегда легко решить, является ли данное состояние возвратным, и критерий, состоящий в том, что ряд $\sum p_{jj}^{(n)}$ должен расходиться, обычно слишком сложен для применения. Более удобный критерий содержится в теореме следующей главы.

Пусть теперь E_j — фиксированное *возвратное* состояние, а E_k — некоторое другое состояние, которое достижимо из E_j. Кроме того, пусть N — длина *кратчайшего* возможного пути от E_j к E_k, так что $p_{jk}^{(N)}=\alpha>0$. Возвращение из E_k в E_j должно иметь положительную вероятность, так как в противном случае вероятность того, что система не возвратится в E_j, была бы равна по меньшей мере α, и неравенство $f_j\leqslant 1-\alpha<1$ противоречило бы предположению о том, что E_j возвратно. Следовательно, существует такое M, что $p_{jk}^{(M)}=\beta>0$. Далее, при любом n

$$p_{jj}^{(n+N+M)}\geqslant p_{jk}^{(N)}p_{kk}^{(n)}p_{kj}^{(M)}=\alpha\beta\cdot p_{kk}^{(n)} \tag{5.10}$$

и

$$p_{kk}^{(n+N+M)}\geqslant p_{kj}^{(M)}p_{jj}^{(n)}p_{jk}^{(N)}=\alpha\beta\cdot p_{jj}^{(n)}. \tag{5.11}$$

Из этих уравнений вытекает, что асимптотическое поведение последовательностей $p_{jj}^{(n)}$ и $p_{kk}^{(n)}$ одинаково; из этого можно вывести важные заключения. Прежде всего предполагалось, что E_j возвратно и, следовательно, ряд $\sum p_{jj}^{(n)}$ расходится. Согласно (5.11), отсюда следует, что расходится также и ряд $\sum p_{kk}^{(n)}$, так что состояние E_k также должно быть возвратным. Если $p_{jj}^{(n)}\to 0$, то также и $p_{kk}^{(n)}\to 0$, и наоборот. Наконец, предположим, что E_j имеет период t. За $N+M$ шагов можно вернуться в E_j, так что $N+M$ кратно t. Согласно (5.10) и (5.11), отсюда следует, что E_j и E_k должны иметь один и тот же период.

Мы видим, таким образом, что из возвратного состояния достижимы только возвратные состояния; все они являются состояниями одного и того же типа: или все они нулевые, или все эргодические, или все периодические ненулевые с одним и тем же периодом. Множество всех состояний, которые достижимы из E_j, очевидно, замкнуто и образует, следовательно, наименьшее замкнутое множество, содержащее E_j. Отсюда следует, что в неприводимой цепи все состояния достижимы из каждого другого состояния, и если одно из состояний невозвратно, то невозвратны и все остальные. Нами доказана, таким образом, важная

Теорема. *В неприводимой цепи Маркова все состояния принадлежат к одному и тому же типу: или они все невозвратные, или все возвратные нулевые, или все возвратные ненулевые. Во всех случаях они имеют одинаковый период и каждое состояние достижимо из каждого другого состояния.*

В каждой цепи все возвратные состояния можно, и притом единственным способом, разделить на замкнутые множества C_1, C_2, ... так, чтобы из любого состояния данного множества C_ν были достижимы все состояния этого множества и не были достижимы никакие другие состояния. Все состояния, принадлежащие к одному и тому же замкнутому множеству C_ν, имеют один и тот же тип.

Кроме замкнутых множеств C_ν, цепь, вообще говоря, содержит невозвратные состояния, из которых достижимы состояния замкнутых множеств C_ν (но не обратно).

Следствие. *Конечная цепь Маркова не содержит нулевых состояний и не может состоять только из невозвратных состояний.*

Доказательство. Достаточно рассмотреть лишь неприводимые цепи. Если все состояния или невозвратные, или нулевые, то для любой фиксированной пары индексов j, k выполняется предельное соотношение $p_{jk}^{(n)} \to 0$ при $n \to \infty$. В этом случае все элементы любой строки P^n стремятся к нулю, в то время как их сумма остается равной единице. Ясно, что это невозможно в случае конечного числа членов, поэтому в неприводимой конечной цепи нет ни нулевых, ни невозвратных состояний.

Легко видеть, что с помощью подходящей перенумерации состояний (так как это сделано в примере § 4) матрица, соответствующая цепи, скажем, с двумя замкнутыми множествами C_1 и C_2 и дополнительными невозвратными состояниями, может быть записана в виде

$$P = \begin{Vmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ A & B & C \end{Vmatrix}, \tag{5.12}$$

где P_1 и P_2 — матрицы вероятностей перехода внутри замкнутых множеств. Матрица P^n является тогда матрицей того же типа, но с заменой P_1, P_2, C на P_1^n, P_2^n, C^n (и с заменой A и B более сложными матрицами, изучаемыми в § 8). P_1, P_2 и C — квадратные матрицы, но A и B могут быть и прямоугольными, как в примере § 4.

§ 6. Эргодическое свойство непериодических цепей. Стационарные распределения

В этом параграфе мы ограничимся рассмотрением *непериодических* цепей; как и вообще для рекуррентных событий, перенесение результатов на периодический случай почти тривиально, но приводит к усложнению формулировок.

Определение. *Распределение вероятностей $\{v_k\}$ называется стационарным, если*

$$v_j = \sum_i v_i p_{ij}. \tag{6.1}$$

Если начальное распределение $\{a_k\}$ стационарно, то абсолютные вероятности $\{a_k^{(n)}\}$ не зависят от времени n, т. е. $a_k^{(n)} = a_k$. Физический смысл стационарности станет понятным, если мы представим большое число процессов, идущих одновременно. Пусть, например, N частиц независимо осуществляют случайное блуждание одного и того же типа. В момент n среднее число частиц в состоянии E_k равно $Na_k^{(n)}$. При стационарном распределении это среднее число остается постоянным, и мы наблюдаем (при условии что N столь велико, что применим закон больших чисел) состояние макроскопического равновесия, поддерживаемое большим числом переходов в противоположных направлениях.

Большинство статистических равновесий в физике именно такого рода; т. е. они связаны с одновременным наблюдением многих независимых частиц. Типичным примером является симметрическое случайное блуждание (или диффузия); если одновременно наблюдается много частиц, то после достаточно длинного промежутка времени приблизительно половина окажется справа от начала координат, другая половина — слева. Тем не менее из закона арксинуса (§ 5 гл. III) мы знаем, что *большинство частиц индивидуально ведет себя очень нерегулярно* и находится непропорционально большую часть времени по одну сторону от исходной точки. Многих пространных обсуждений и неверных заключений можно было бы избежать, осознав, что понятие статистического равновесия (или устойчивого состояния) ничего не говорит о поведении отдельной частицы. Это следует иметь в виду и в связи со следующей теоремой, которую часто истолковывают как утверждение о «тенденции к равновесию».

Теорема. *Неприводимая непериодическая цепь Маркова принадлежит к одному из следующих двух классов:*

а) *Или все состояния невозвратные, или все состояния нулевые; в этом случае $p_{jk}^{(n)} \to 0$ при $n \to \infty$ для любой пары индексов j, k и не существует стационарного распределения.*

б) *Или же все состояния эргодические, т. е.*

$$\lim_{n\to\infty} p_{jk}^{(n)} = u_k > 0, \tag{6.2}$$

где u_k — величина, обратная среднему времени возвращения в E_k. В этом случае $\{u_k\}$ — стационарное распределение и не существует никаких других стационарных распределений.

Приведем слегка отличную формулировку, которая поможет понять смысл теоремы. Если выполняется соотношение (6.2), то для любого начального распределения $\{a_k\}$

$$a_k^{(n)} = \sum_j a_j p_{jk}^{(n)} \to u_k. \tag{6.3}$$

Поэтому, *если существует стационарное распределение, то оно необходимо единственно и распределение в момент n стремится к нему, каково бы ни было начальное распределение. В противном случае $p_{jk}^{(n)} \to 0$ при $n \to \infty$.*

Доказательство. В силу теоремы § 5, соотношение (6.2) выполняется всякий раз, когда состояние эргодическое.

Чтобы доказать утверждение (б), заметим, во-первых, что

$$\sum u_k \leqslant 1. \tag{6.4}$$

Это следует непосредственно из того, что при фиксированных j и n величины $p_{jk}^{(n)}$ $(k = 1, 2, \ldots)$ в сумме равны единице, так что $u_1 + u_2 + \ldots + u_N \leqslant 1$ при любом N. Положим теперь в (3.3) $n = 1$, и пусть $m \to \infty$. Левая часть стремится к u_k, а слагаемые правой части стремятся к $u_\nu p_{\nu k}$. Сложив произвольное конечное число членов, убеждаемся, что

$$u_k \geqslant \sum_\nu u_\nu p_{\nu k}. \tag{6.5}$$

Суммируя эти неравенства по всем k, мы получаем в каждой из частей неравенства конечную величину $\sum u_k$. Это показывает, что знак неравенства в (6.7) невозможен и, таким образом,

$$u_k = \sum u_j p_{jk}. \tag{6.6}$$

Положив $v_k = u_k \cdot \left(\sum u_j\right)^{-1}$ видим, что распределение $\{v_k\}$ стационарно и, значит, по крайней мере одно такое распределение существует.

Пусть $\{v_k\}$ — произвольное распределение, удовлетворяющее уравнениям (6.1). Умножая соотношение (6.2) на $p_{jk}^{(n)}$ и суммируя по всем j, получим по индукции, что при любом n

$$v_r = \sum_\nu v_\nu p_{\nu r}^{(n)}. \tag{6.7}$$

Устремив n к бесконечности, в пределе получим

$$v_r = (v_1 + v_2 + \ldots)\, u_r = u_r. \tag{6.8}$$

Этим завершается доказательство утверждения (б). Если все состояния или невозвратные, или нулевые, а распределение $\{v_k\}$ стационарно, то уравнение (6.7) по-прежнему имеет место. В то же время $p_{vr}^{(n)} \to 0$, что, очевидно, невозможно. Поэтому стационарное распределение может существовать только в эргодическом случае, и теорема доказана полностью.

Примеры. а) *Модель Эренфестов.* В примере 2, е уравнения (6.1) принимают вид

$$u_k = \left(1 - \frac{k-1}{a}\right) u_{k-1} + \frac{k+1}{a} u_{k+1} \qquad (k = 1, \ldots, a-1);$$

$$u_0 = \frac{u_1}{a}, \qquad u_a = \frac{u_{a-1}}{a}. \tag{6.9}$$

Легко проверить, что искомым решением является

$$u_k = \binom{a}{k} 2^{-a}. \tag{6.10}$$

Это — биномиальное распределение, и поэтому результат может быть интерпретирован следующим образом: каково бы ни было начальное число молекул в первом резервуаре, через достаточно долгое время вероятность нахождения в нем ровно k молекул примерно такова, как если бы a-молекул были распределены случайно, причем вероятность для каждой из молекул находиться в первом сосуде равнялась $^1/_2$. Таким образом, наши абстрактные результаты в этом примере приобретают конкретный физический смысл.

Нормальное приближение для биномиального распределения показывает, что при достаточно большом a мы будем практически уверены, что найдем примерно половину молекул в каждом из резервуаров. Действительно, для физика $a = 10^6$ — число небольшое. Но даже при $a = 10^6$ молекул вероятность найти более 505 000 молекул в одном из резервуаров (флуктуация плотности в одну сотую) есть величина порядка 10^{-23}. При $a = 10^8$ флуктуация удельного веса на одну тысячную имеет ту же самую ничтожную вероятность. Конечно, система может иногда перейти в очень невероятные состояния, но время возвращения в такое состояние фантастически велико по сравнению со временем возвращения в состояние, близкое к равновесию. Физическая необратимость проявляется в том, что, когда система находится в состоянии, далеком от равновесия, гораздо более правдоподобно изменение в сторону равновесия, чем в противоположном направлении.

б) *Дважды стохастическая матрица.* Матрица P называется дважды стохастической, если не только суммы строк, но также и суммы столбцов равны единице. Предположим, что соответствующая цепь содержит конечное число a состояний. Система (6.1) имеет, очевидно, решение $v_k = 1/a$. Отсюда следует, что *если конечная неприводимая непериодическая цепь имеет дважды стохастическую матрицу P, то* $v_k = 1/a$ (т. е. в пределе все состояния имеют равные вероятности).

В этом случае не может быть невозвратных состояний. Напротив, в неприводимой *бесконечной цепи с двойной стохастической матрицей все состояния или невозвратные или нулевые*. Чтобы доказать это утверждение, предположим, что выполняется соотношение (6.2). Так как P — двойная стохастическая матрица, то при любом фиксированном k и сколь угодно большом N, мы имеем

$$1 = \sum_{j=1}^{\infty} p_{jk}^{(n)} \geqslant \sum_{j=1}^{N} p_{jk}^{(n)} \to N u_k. \tag{6.11}$$

Отсюда, очевидно, следует, что $u_k = 0$, в противоречии с предположением (это доказательство применимо и в периодическом случае).

в) *Рекуррентные события.* В примере 2, и мы ввели марковскую цепь, связанную с произвольным рекуррентным событием $\mathfrak{E}$. Теперь мы покажем, что (как это можно было ожидать) все состояния цепи принадлежат к тому же типу, что и $\mathfrak{E}$.

Рассмотрим сначала случай, когда событие $\mathfrak{E}$ невозвратно, т. е. предположим, что $f < 1$. Цепочка переходов $E_l \to E_{l+1} \to \ldots \to E_{l+n}$ имеет вероятность

$$\frac{1-s_{l+1}}{1-s_l} \cdot \frac{1-s_{l+2}}{1-s_{l+1}} \cdots \frac{1-s_{l+n}}{1-s_{l+n-1}} = \frac{1-s_{l+n}}{1-s_l} \geqslant \frac{1-f}{1-s_l}. \tag{6.12}$$

Отсюда следует, что вероятность того, что система никогда не вернется в E_0, положительна, и, значит, все состояния невозвратны. С другой стороны, если $f = 1$, то левая часть соотношения (6.12) стремится к нулю; с вероятностью единица система рано или поздно достигнет состояния E_0. Отсюда вытекает, что состояние E_0 возвратно, и так как каждое состояние достижимо из E_0, то цепь неприводима. Итак, мы доказали, что *если событие $\mathfrak{E}$ невозвратно, то все состояния цепи также невозвратны; если $\mathfrak{E}$ возвратно, то цепь неприводима и все ее состояния возвратны.*

Ясно что цепь и $\mathfrak{E}$ имеют один и тот же период, и мы будем предполагать, что $\mathfrak{E}$ — непериодическое возвратное событие. Мы должны выяснить, когда для нашей цепи существует стационарное распределение, т. е. распределение вероятностей $\{v_k\}$, удовлетворяющее соотношениям (6.1). В рассматриваемом случае соотношения (6.1) сводятся к такой системе уравнений:

$$v_0 = \sum_{l=0}^{\infty} \frac{f_{l+1}}{1-s_l} v_l, \qquad v_k = \frac{1-s_k}{1-s_{k-1}} v_{k-1}. \tag{6.13}$$

Существует единственное решение этой системы, а именно

$$v_k = (1 - s_k) v_0 = r_k v_0, \text{ где } r_k = \sum_{n=k+1}^{\infty} f_n. \tag{6.14}$$

Для того, чтобы $\sum v_k < \infty$, необходимо и достаточно, чтобы $\sum r_k < \infty$. Но $\sum r_k = \sum n f_n$ равна среднему времени возвращения $\mathfrak{E}$ [см. формулу (1.8) гл. XI]. Это показывает, что все *состояния марковской цепи являются нулевыми, если среднее время возвращения $\mathfrak{E}$ бесконечно, и эргодическими, если среднее время возвращения $\mathfrak{E}$ конечно.*

Мы вывели асимптотические свойства марковских цепей, отправляясь от аналогичных свойств рекуррентных событий. Теперь же мы показали, что и наоборот, каждое рекуррентное событие может быть описано в терминах специальной марковской цепи. Эти три темы — асимптотическое поведение цепей Маркова, асимптотическое поведение рекуррентных событий (т. е. теория суммирования целочисленных случайных величин) и теория восстановления — являются различными вариантами одной и той же аналитической задачи и, по существу, эквивалентны.

В заключение этого параграфа отметим, что обычно сравнительно просто решить, существует ли стационарное распределение, и, следовательно, является ли данная неприводимая цепь эргодической. В § 8 будет выведен аналогичный критерий, позволяющий различать невозвратные состояния и возвратные нулевые состояния. В принципе этот вопрос можно решить, исследуя сходимость ряда $\sum p_{jk}^{(n)}$, но на практике этот прямой подход мало пригоден.

§ 7*). Периодические цепи

В предыдущем параграфе мы не рассматривали случая периодических цепей; это было сделано только для того, чтобы не усложнять основных формулировок. Характеристика асимптотического поведения $p_{jk}^{(n)}$ в неприводимой периодической цепи может быть легко выведена из теорем предыдущего параграфа. Мы приведем здесь такой вывод, чтобы изложение было исчерпывающим; однако результаты настоящего параграфа в дальнейшем использоваться не будут.

По теореме § 5 все состояния неприводимой цепи имеют один и тот же период t. Рассмотрим теперь любые два состояния E_j и E_k какой-нибудь неприводимой цепи периода t. Так как каждое состояние достижимо из каждого другого, то существуют такие целые

*) Этот параграф рассматривает специальный круг вопросов и может быть опущен при первом чтении.

числа a, b, что $p_{jk}^{(a)} > 0$ и $p_{kj}^{(b)} > 0$. Так как $p_{jj}^{(a+b)} \geqslant p_{jk}^{(a)} p_{kj}^{(b)}$, то возможно возвращение из E_j обратно в E_j за $a+b$ шагов, так что период t обязательно является делителем для $a+b$. Отсюда следует, что если можно попасть из E_j в E_k и за a_1 и за a_2 шагов, то $a_2 - a_1$ делится на t и, следовательно, при делении a_1 и a_2 на t должен получаться один и тот же остаток. Поэтому при фиксированном E_j каждому состоянию E_k соответствует такое число $\nu (\nu = 0, 1, \ldots, t-1)$, что переход из E_j в E_k возможен только за ν, $\nu + t$, $\nu + 2t$, $\nu + 3t$, ... шагов. Положим $j = 1$; тогда мы получим разбиение всех состояний на t классов G_0, G_1, ..., G_{t-1}, определяемое следующим способом. Если $p_{1k}^{(a)} > 0$ и $a = nt + \nu$ (где ν удовлетворяет неравенству $0 \leqslant \nu < t$), то E_k принадлежит к G_ν. Мы будем представлять себе G_ν расположенными циклически, так что G_0 и G_{t-1} — соседние классы.

Переход за один шаг из состояния, входящего в G_ν, переводит систему в состояние, входящее в следующий класс $G_{\nu+1}$ (или в G_0 при $\nu = t - 1$). Переход за два шага переводит систему в состояние, принадлежащее к $G_{\nu+2}$ (из G_{t-2} система за два шага переходит в G_0, из G_{t-1} — в G_1) и т. д. Наконец, переход за t шагов переводит систему в состояние, принадлежащее тому же классу. Это означает, что в цепи Маркова с матрицей вероятностей перехода P^t каждый класс G_ν образует замкнутое множество. Так как первоначальная цепь неприводима, то каждое состояние достижимо из каждого. Отсюда следует, что в цепи с матрицей вероятностей перехода P^t каждый класс G_ν образует неприводимое замкнутое множество состояний. Нами получена, таким образом следующая теорема:

Теорема. *В неприводимой периодической цепи Маркова все состояния могут быть разделены на t классов G_0, ..., G_{t-1}, так что переход за один шаг из состояния, входящего в G_ν, переводит систему в состояние класса $G_{\nu+1}$ (или G_0, если $\nu = t - 1$). Если мы рассмотрим цепь только в моменты t, $2t$, $3t$, ..., то получим новую цепь Маркова с матрицей вероятностей перехода P^t. В ней каждое из G_ν образует неприводимую замкнутую цепь.*

Наша теорема дает полную информацию относительно *асимптотического поведения* $p_{jk}^{(n)}$. Если все состояния являются невозвратными или нулевыми, то $p_{jk}^{(n)} \to 0$ при любых j и k. В остальных случаях среднее время возвращения μ_k в состояние E_k конечно. Предположим, что E_j принадлежит к G_ν. Но G_ν есть неприводимая непериодическая цепь Маркова с вероятностями перехода $p_{jk}^{(t)}$, и, следовательно, существуют пределы

$$\lim_{n\to\infty} p_{jk}^{(nt)} = \begin{cases} u_k, & \text{если } E_k \text{ входит в } G_\nu, \\ 0 & \text{в других случаях.} \end{cases} \tag{7.1}$$

Здесь u_k — величина, обратная среднему времени возвращения в E_k для новой цепи, один шаг которой соответствует t шагам первоначальной цепи. Следовательно,

$$u_k = \frac{t}{\mu_k}. \tag{7.2}$$

Используя (3.2), мы получаем из (7.1)

$$\lim_{n\to\infty} p_{jk}^{(nt+1)} = \begin{cases} u_k, \text{ если } E_k \text{ входит в } C_{\nu+1}, \\ 0 \text{ в других случаях.} \end{cases} \tag{7.3}$$

Аналогично $p_{jk}^{(nt+2)} \to u_k$, если E_k входит в $G_{\nu+2}$, и т. д. Другими словами, *при фиксированных E_j и E_k последовательность $p_{jk}^{(n)}$ асимптотически периодична; ее период состоит из $t-1$ последовательных нулей и одного положительного элемента, который сходится к $u_k = t/\mu_k$.*

По теореме § 6, внутри каждого из классов G_ν сумма всех u_k равна единице. Так как число классов равно t, то из (7.2) следует, что $\{1/\mu_k\}$ представляет собой распределение вероятностей. Рассуждения § 6 непосредственно показывают, что *это распределение стационарно и что никакого другого стационарного распределения не существует.*

§ 8. Невозвратные состояния

Из возвратного состояния E_j система может перейти только в возвратное состояние E_k, принадлежащее замыканию E_j. В этом случае мы полностью описали асимптотическое поведение вероятностей $p_{jk}^{(n)}$.

Если состояние E_j — невозвратное, а состояние E_k — эргодическое, то в силу (5.7)

$$p_{jk}^{(n)} \to \mu_k^{-1} f_{jk}, \tag{8.1}$$

где μ_k — среднее время возвращения в E_k, а f_{jk} — вероятность того, что исходя из E_j система когда-нибудь достигнет состояния E_k. Но E_k принадлежит к неприводимой цепи, определенной на замкнутом подмножестве C, и исходя из E_k с вероятностью единица достигается любым из состояний, входящих в C. Поэтому при любом фиксированном j вероятность f_{jk} одна и та же для всех состояний из C. Другими словами, *если C — неприводимая цепь с эргодическими состояниями, определенная на замкнутом подмножестве, то для любого E_k из C*

$$p_{jk}^{(n)} \to \mu_k^{-1} x_j, \tag{8.2}$$

где x_j — вероятность того, что система, отправляясь из состояния E_j, попадет в конце концов в множество C. Нет нужды говорить, что для нулевых состояний правую часть формулы (8.2) следует заменить на 0 и что в случае периодического состояния E_k требуется лишь стандартная модификация доказательства.

Для того чтобы закончить картину асимптотического поведения марковских цепей, остается решить следующие

Три задачи. а) *Дано невозвратное состояние E_j и замкнутое множество возвратных состояний C, найти вероятность x_j того, что, отправляясь из C, система когда-нибудь попадет в C (т. е. достигнет некоторого состояния из C).*

б) *Найти вероятность y_j того, что система будет все время находиться в множестве невозвратных состояний.*

в) *Дана неприводимая цепь. Решить вопрос о том, являются ли ее состояния возвратными или невозвратными.*

Сейчас мы увидим, что после незначительного изменения формулировки задача (в) становится частным случаем задачи (а).

Пусть T — множество всех невозвратных состояний. Предположим, что *в начальный момент система находится в невозвратном состоянии E_j, и пусть $x_j^{(n)}$ — вероятность того, что в момент n система впервые попадет в замкнутое множество C. Тогда*

$$x_j = \sum_{n=1}^{\infty} x_j^{(n)} \tag{8.3}$$

есть вероятность того, что система рано или поздно достигнет C и останется в C. По аналогии с простейшим случайным блужданием *мы будем называть x_j вероятностью поглощения в множестве C.* Разность $1 - x_j$ соответствует возможности поглощения другими замкнутыми множествами и (в случае некоторых бесконечных цепей) возможности бесконечного пребывания в невозвратных состояниях.

Ясно, что

$$x_j^{(1)} = \sum_{C} p_{jk}, \tag{8.4}$$

где суммирование производится по всем тем k, при которых E_k содержится в C. Если система достигает C на $(n+1)$-м шагу, то первый шаг должен привести систему из E_j в некоторое также невозвратное состояние. Поэтому ясно, что

$$x_j^{(n+1)} = \sum_{T} p_{j\nu} x_\nu^{(n)}, \tag{8.5}$$

где суммирование производится по всем ν, при которых E_ν — невозвратное состояние. *Уравнения* (8.4) *и* (8.5) *представляют собой рекуррентные соотношения, которые единственным образом определяют* $x_j^{(n)}$. Суммируя (8.3) при $n=1, 2, 3, \ldots$, мы убеждаемся, что *вероятности поглощения* x_j *суть решения системы линейных уравнений*

$$x_j - \sum_T p_{j\nu} x_\nu = x_j^{(1)}. \tag{8.6}$$

Таким образом, мы ответили на вопрос задачи (а); вероятности x_j могут быть получены конструктивно, отправляясь от соотношений (8.3) — (8.5), но более предпочтительным методом является решение системы линейных уравнений (8.6). В этой связи возникает вопрос о единственности решения, но, к счастью, он сводится к специальному случаю задачи (б).

Пусть $y_j^{(n)}$ — *вероятность того, что в момент* n *система находится в невозвратном состоянии.* Очевидно, что

$$\begin{aligned} y_j^{(1)} &= \sum_T p_{j\nu}, \\ y_j^{(n+1)} &= \sum_T p_{j\nu} y_\nu^{(n)}, \end{aligned} \tag{8.7}$$

где суммирование производится по всем ν, при которых E_ν невозвратно. Из (8.7) последовательно вытекает, что $y_j^{(1)} \leqslant 1$, $y_j^{(2)} \leqslant y_j^{(1)}$, и, вообще, $y_j^{(n+1)} \leqslant y_j^{(n)}$. Поэтому предел

$$y_j = \lim_{n \to \infty} y_j^{(n)} \tag{8.8}$$

существует; y_j есть вероятность того, что система навсегда останется в множестве невозвратных состояний. Из формулы (8.8) имеем

$$y_j = \sum_T p_{j\nu} y_\nu. \tag{8.9}$$

Таким образом, вероятности y_j удовлетворяют системе линейных уравнений (8.9), но этим еще не решен основной вопрос о том, когда $y_j = 0$ при всех j. Предположим, что система (8.9) имеет ограниченное решение, скажем,

$$z_j = \sum_T p_{j\nu} z_\nu, \qquad |z_\nu| \leqslant 1. \tag{8.10}$$

Сравнение (8.10) и (8.8) показывает, что $|z_j| \leqslant y_j^{(1)}$ и, следовательно, по индукции $|z_j| \leqslant y_j^{(n)}$ при любом n. Отсюда следует, что $y_j = 0$ при всех j тогда и только тогда, когда система (8.10) не имеет ограниченного решения, отличного от нуля. Наконец, если линейные уравнения (8.6) имеют два различных решения, то их разность удовлетворяет системе (8.10). Нами получена

Теорема 1. *Вероятности* x_j *задачи* (а) *представляют решение системы линейных уравнений* (8.6). *Это решение единственно, исключая тот случай, когда существует состояние* E_j, *такое, что исходя из* E_j *система с положительной вероятностью* y_j *навсегда остается в множестве невозвратных состояний. Вероятности* $\{y_j\}$ *удовлетворяют соотношениям* (8.9).

Замечание. Мы убедились в том, что вероятности $\{y_j\}$ представляют *максимальное* решение (8.9), ограниченное единицей; аналогичным свойством обладают и $\{x_j\}$.

Следствие. *В конечной цепи Маркова вероятность того, что система все время находится в множестве невозвратных состояний, равна нулю. Вероятности* x_j, *исходя из невозвратного состояния* E_j, *достигнуть когда-нибудь замкнутого множества, определяются как единственное решение системы линейных уравнений* (8.6).

Доказательство. Мы должны доказать, что система (8.9) не допускает ненулевого решения. Предположим обратное, и пусть M — максимальная из конечного числа вероятностей y_j. Не ограничивая общности, можно считать состояния расположенными так, что значения y_j *убывают*, т. е. $y_1 = y_2 = \ldots = y_a = M > y_{a+1} \geqslant \geqslant y_{a+2} \geqslant \ldots$. Из (8.9) следует, что при $i \leqslant a$

$$M = \sum_T p_{i\nu} y_\nu = \sum_{\nu=1}^{a} p_{i\nu} + \sum_{\nu \geqslant a+1} p_{i\nu} y_\nu, \tag{8.11}$$

и знак равенства возможен только при условии, что $p_{i\nu} = 0$ при любом $\nu > a$. В этом случае состояния $E_1, E_2, \ldots, E_a$ образуют замкнутое множество, а это невозможно, так как в конечной цепи обязательно имеются возвратные состояния (см. следствие § 5).

Теорема 1 используется для вычисления *вероятностей поглощения*, т. е. вероятностей достижения данного поглощающего множества или состояния.

Примеры. а) *Случайное блуждание с поглощающими экранами* (пример 2, б). Примем за C поглощающее состояние E_0. Тогда $x_1^{(1)} = q$ и $x_j^{(1)} = 0$, если $j > 1$. Система (8.4) имеет вид

$$\begin{aligned} &x_1 - px_2 = q; \\ &x_j - qx_{j-1} - px_{j+1} = 0; \quad (j = 2, 3, \ldots, a-2) \\ &x_{a-1} - qx_a = 0. \end{aligned} \tag{8.12}$$

Эта система совпадает с системой (2.1) — (2.2) гл. XIV, решение которой дается формулой (2.4).

б) *Последовательный анализ* (пример 2, к). Пусть снова C будет состоянием E_0. Тогда $x_j^{(1)} = r_j$ и уравнения (8.6) сводятся к уравнениям (8.2) гл. XIV (где u_x заменяет употребляемое здесь x_j; см. также задачу 4 гл. XIV).

в) *Генетика* (пример 2, л). В этом случае каждое из двух состояний E_0 и E_{2N} образует замкнутое множество. Поглощение в E_0 или E_{2N} означает, что популяция в конце концов будет состоять только из особей *aa* или *AA*. Вероятность поглощения в E_0 за один шаг равна $x_j^{(1)} = p_{j0} = (1 - j/2N)^{2N}$, и поэтому система (8.6) принимает вид

$$x_j - \sum_{\nu=1}^{2N-1} \binom{2N}{\nu} \left(\frac{j}{2N}\right)^{\nu} \left(1 - \frac{j}{2N}\right)^{2N-\nu} x_\nu = \left(1 - \frac{j}{2N}\right)^{2N}. \quad (8.13)$$

Вполне правдоподобной выглядит гипотеза о том, что вероятности выживания каждой из форм генов (генов A и a) прямо пропорциональны их начальному количеству. Если эта гипотеза верна, то система (8.13) должна иметь решение $x_j = 1 - j(2N)^{-1}$. Эти x_j действительно удовлетворяют (8.13). Для того чтобы убедиться в этом, достаточно заметить, что слагаемые в левой части (8.13) представляют члены биномиального распределения со средним значением j.

Наконец, мы дадим решение задачи (в).

Теорема 2. *Пусть $E_0, E_1, \ldots$ — состояния неприводимой цепи. Для того чтобы состояния были невозвратными, необходимо и достаточно, чтобы система уравнений*

$$y_i = \sum_{j=1}^{\infty} p_{ij} y_j, \quad i = 1, 2, \ldots \quad (8.14)$$

имела ненулевое ограниченное решение.

Доказательство. Рассмотрим формулы (8.7) и (8.8) для $\{y_i\}$, интерпретируя T как множество состояний $E_1, E_2, \ldots$ (дополнение к E_0). После этого к нашему случаю применимо доказательство теоремы 1. Вероятности все время оставаться в T (никогда не достигнуть E_0) определяются формулами (8.8) и удовлетворяют уравнениям (8.9).

Примеры. г) *Неограниченное случайное блуждание.* Этот пример является почти точной копией примера 2, д. Следует лишь слегка изменить обозначения, так как номера состояний меняются от $-\infty$ до $+\infty$. Ясно, однако, что наш критерий зависит от существования решения системы уравнений

$$y_i = p y_{i+1} + q y_{i-1}, \quad y_0 = 0, \quad i = \pm 1, \pm 2 \ldots . \quad (8.15)$$

Ясно, что все y_i можно вычислить последовательно, отправляясь от y_{-1} и y_{+1}. При $p>q$

$$y_i=\left\{1-\left(\frac{q}{p}\right)^i\right\}y_1, \quad y_{-i}=0, \quad i=1, 2, \ldots \tag{8.16}$$

— единственное ограниченное решение. Если же $p=q$, то решение имеет вид $y_i=iy_1$ и, следовательно, неограничено. Опираясь на теорию марковских цепей, мы получили уже известный нам результат, состоящий в том, что состояния невозвратны, если $p\neq q$, и возвратны, если $p=q$.

д) Рассмотрим матрицу

$$P=\begin{Vmatrix} q_0 & p_0 & 0 & 0 & 0 & \ldots \\ q_1 & 0 & p_1 & 0 & 0 & \ldots \\ 0 & q_2 & 0 & p_2 & 0 & \ldots \\ 0 & 0 & q_3 & 0 & p_3 & \ldots \\ \cdot & \cdot & \cdot & \cdot & & \ldots \\ \cdot & \cdot & \cdot & \cdot & & \ldots \\ \cdot & \cdot & \cdot & \cdot & & \ldots \end{Vmatrix},$$

которая соответствует случайному блужданию на полупрямой $(0, \infty)$ с переменными вероятностями перехода. Эта матрица играет важную роль в теории *процессов гибели и размножения*, которая обсуждается в гл. XVII. Уравнения (8.14) принимают вид

$$y_1=p_1y_2, \quad y_i=q_iy_{i-1}+p_iy_{i+1}, \quad i=2, 3, \ldots . \tag{8.17}$$

Они могут быть решены последовательно, так как

$$\frac{y_{i+1}-y_i}{y_i-y_{i-1}}=\frac{q_i}{p_i} \tag{8.18}$$

и, следовательно,

$$y_{i+1}-y_i=y_1\frac{q_1q_2\cdots q_i}{p_1p_2\cdots p_i}. \tag{8.19}$$

Складывая эти уравнения, убеждаемся в том, что ограниченное решение существует тогда и только тогда, когда $\sum L_i<\infty$, где $L_i=(q_1q_2\ldots q_i)(p_1p_2\ldots p_i)^{-1}$. Поэтому состояния цепи невозвратны, если $\sum L_i<\infty$ и возвратны в противном случае.

§ 9. Задача о тасовании колоды карт

Колода из N карт, занумерованных числами 1, 2, ..., N, может быть перетасована $N!$ различными способами; каждый из них представляет собой возможное «состояние системы». Каждая отдельная перетасовка колоды вызывает переход из имевшегося состояния в некоторое другое. Например, «снимание» изменяет порядок $(1, 2, \ldots, N)$ на один из N циклически эквивалентных порядков $(r, r+1, \ldots, N, 1, 2, \ldots, r-1)$. Та же самая операция, примененная к обратному порядку $(N, N-1, \ldots, 1)$, приводит к состоянию $(N-r+1, N-r, \ldots, 1, N, N-1, \ldots, N-r+2)$. Другими словами, мы представляем себе каждую отдельную перетасовку колоды как переход $E_j \to E_k$. Если повторяется *точно* одна и та же перетасовка, то система будет проходить (начиная из данного состояния E_j) вполне определенную последовательность состояний, и через конечное число шагов восстановится первоначальный порядок карт. После этого та же самая последовательность состояний будет повторяться периодически. Для большинства операций этот период будет довольно мал, и с помощью такой процедуры никогда не могут быть получены все «состояния»[1]). Рассмотрим, например, операцию, изменяющую в колоде из $2m$ карт порядок $(1, \ldots, 2m)$ на $(1, m+1, 2, m+2, \ldots, m, 2m)$. При колоде из 6 карт четырехкратное применение этой операции восстанавливает первоначальный порядок. При 10 картах начальное расположение повторится после шести операций, так что в результате многократного применения такой операции мы получим только шесть из $10! = 3\,628\,800$ возможных расположений карт.

На практике игрок не использует какого-либо закономерного тасования колоды карт, и наверняка изменения будут вызваны игрой случая; даже игрок, старающийся применять один и тот же прием тасования, не всегда этого достигнет. Предположим, что можно учесть и привычки игрока, и влияние случайных изменений, считая, что каждый способ тасования карт имеет определенную вероятность (быть может, нуль). Нам не нужны никакие предположения о численном значении этих вероятностей, но мы допускаем, что играющий действует каждый раз независимо от предыдущего и не знает порядка карт[2]). Отсюда следует, что последовательность операций соответствует независимым испытаниям с фиксированными вероятностями, а расположения колоды карт образуют состояния цепи Маркова.

Покажем, что матрица вероятностей перехода P является *дважды стохастической* (пример 6, б). Действительно, если некоторая перестановка изменяет состояние (порядок карт) E_j на E_k, то существует некоторое другое состояние E_r, которое при этой перетасовке

---

[1]) На языке теории групп это означает, что группа перестановок не является циклической и поэтому не может порождаться никаким одним элементом.

[2]) Это предположение соответствует обычной ситуации при карточной игре. Легко придумать более сложную технику тасования, при которой операции зависят от предшествующих операций, и окончательная схема не является цепью Маркова (см. пример 10, д).

изменяется на E_j. А это означает, что элементы j-й строки матрицы P совпадают с элементами j-го столбца, хотя и расположены в другом порядке. Сумма элементов любого столбца равна, следовательно, единице.

Отсюда следует, что ни одно из состояний не может быть невозвратным. *Если цепь неприводима и непериодична, то в пределе все состояния имеют равные вероятности.* Другими словами, пригоден *любой* способ тасования, лишь бы он индуцировал неприводимую и непериодическую цепь. Можно допустить, что обычно дело обстоит именно так. Предположим, однако, что колода состоит из четного числа карт, и что наш способ состоит в разделении ее на две равные части и в перетасовании их отдельно друг от друга любым методом. Если эти две колоды сложить затем в первоначальном порядке, то цепь Маркова окажется приводимой (так как не каждое состояние достижимо из любого другого состояния). Если же изменить порядок на обратный, то цепь будет иметь период 2. Обе эти возможности теоретически существуют, но едва ли могут встретиться на практике, так как действие случая мешает осуществить их вполне точно.

Таким образом, можно ожидать, что непрерывное тасование приведет к полной «случайности» и уничтожит все следы первоначального расположения. Отметим, однако, что число операций, необходимое для этой цели, чрезвычайно велико[1]).

§ 10. Общий марковский процесс

В приложениях обычно удобно описывать цепь Маркова в терминах случайных величин. Это может быть сделано простой заменой в предыдущих параграфах символа E_k на целое число k. *Состояние системы в момент n окажется* тогда случайной величиной $X^{(n)}$, которая принимает значение k с вероятностью $a_k^{(n)}$; совместное распределение $X^{(n)}$ и $X^{(n+1)}$ определяется формулой $P\{X^{(n)}=j,\quad X^{(n+1)}=k\}=a_j^{(n)}p_{jk}$; совместное распределение $(X^{(1)}, X^{(2)}, \ldots, X^{(n)})$ задается формулой (1.1). Иногда удобнее сопоставлять состоянию E_k численное значение e_k, не равное k. При такой записи цепь Маркова может рассматриваться как неко-

[1]) Анализу невероятно бедных данных по тасованию посвящена работа W. Feller, Statistical aspects of ESP, *Journal of Parapsychology*, 4 (1940), 271—298. В занимательной статье A review of Dr. Feller's critique, *там же*, 299—319 Гринвуд и Стюарт пытаются показать, что эти результаты обусловлены случаем, однако их арифметика и их эксперименты имеют определенный оттенок сверхестественности.

торый специальный стохастический процесс[1]), т. е. последовательность (зависимых) случайных величин[2]) $(X^{(0)}, X^{(1)}, \ldots)$, каждый конечный набор которых имеет определенное совместное распределение. Индекс n играет роль времени. В гл. XVII мы получим понятие о более общих стохастических процессах, в которых временно́й параметр может изменяться непрерывно. Термин «марковские процессы» применим к очень большому и важному классу стохастических процессов (как с дискретным, так и с непрерывным параметром). Даже в дискретном случае существуют марковские процессы, более общие, чем простые цепи, которые мы изучали до сих пор. Поэтому полезно определить важнейшее свойство марковских процессов, указав при этом на специальные условия, характеризующие цепи Маркова, и, наконец, рассмотреть несколько примеров немарковских процессов.

По существу марковский процесс есть вероятностная аналогия процессов классической механики, в которых дальнейшее развитие процесса вполне определяется состоянием в настоящий момент и не зависит от способа, которым это состояние было достигнуто. Эти механические процессы противоположны процессам с последействием, возникающим, например, в теории пластичности, где вся предыдущая история системы влияет на ее будущее. В стохастических процессах будущее никогда не бывает однозначно определенным, но все же имеются вероятностные соотношения, дающие возможность давать предсказания. Для цепей Маркова, изученных в настоящей главе, вероятностные соотношения, определяющие будущее, зависят от состояния в настоящий момент, но не зависят от того, как возникло это состояние в прошлом. Другими словами, если две независимые системы с одинаковыми вероятностями перехода оказались в одном и том же состоянии, то все вероятности, связанные с их будущим развитием, совпадают. Это довольно неопределенное описание уточняется следующим определением.

Определение. *Последовательность дискретных случайных величин образует марковский процесс, если для каждого конечного набора целых чисел* $n_1 < n_2 < \ldots < n_r < n$ *совместное распределение* $(X^{(n_1)}, X^{(n_2)}, \ldots, X^{(n_r)}, X^{(n)})$ *таково, что условная вероятность соотношения* $X^{(n)} = x$ *при условиях*

---

[1]) Термины «стохастический процесс» и «случайный процесс» — синонимы; они охватывают практически всю теорию вероятностей от бросания монеты до спектрального анализа. На практике термин «стохастический процесс» обычно применяется, если параметр n интерпретируется, как время.

[2]) Эта формулировка приобретает совершенно строгий смысл в бесконечном пространстве элементарных событий, но в действительности мы имеем дело только с совместными распределениями конечных наборов случайных величин.

$X^{(n_1)}=x_1, \ldots, X^{(n_r)}=x_r$ совпадает с условной вероятностью $X^{(n)}=x$ при единственном условии $X^{(n_r)}=x_r$. Здесь $x_1, \ldots, x_r$, x — произвольные числа, для которых наши условия имеют положительные вероятности.

Говоря проще, если задано состояние x_r в момент n_r, то никакие дальнейшие сведения о поведении системы в предыдущее время не изменяют (условной) вероятности состояния x в последующий момент n.

Цепи Маркова, изучавшиеся в этой главе, очевидно, суть марковские процессы, но они обладают следующим дополнительным свойством. *У цепей Маркова, изучавшихся в предыдущих параграфах, вероятности перехода* $p_{jk}=P\{X^{(m+1)}=k \mid X^{(m)}=j\}$ *не зависят от* m. Более общие вероятности перехода

$$p_{jk}^{(n-m)}=P\{X^{(n)}=k \mid X^{(m)}=j\} \qquad (m<n) \tag{10.1}$$

зависят тогда только от разности $n-m$. Мы будем называть такие цепи Маркова однородными (по времени) или цепями Маркова с постоянными вероятностями перехода. Для произвольной дискретной цепи Маркова правая часть (10.1) зависит от m и n. Мы будем обозначать ее через $p_{jk}(m, n)$, так что $p_{jk}(n, n+1)$ суть вероятности перехода за один шаг в момент n. Вместо (1.1) мы получим для вероятности заданного пути $(j_0, j_1, \ldots, j_n)$ выражение

$$a_{j_0}^{(0)} p_{j_0 j_1}(0, 1)\, p_{j_1 j_2}(1, 2) \ldots p_{j_{n-1} j_n}(n-1, n). \tag{10.2}$$

Соответствующим обобщением (3.3), очевидно, будет тождество

$$p_{jk}(m, n)=\sum_{\nu} p_{j\nu}(m, r)\, p_{\nu k}(r, n), \tag{10.3}$$

которое верно при всех тех r, для которых $m<r<n$. Это тождество следует непосредственно из определения марковского процесса, а также из (10.2), и называется уравнением *Колмогорова — Чэпмена*.

В настоящей главе мы занимались в основном изучением асимптотического поведения вероятностей перехода. Из установленных нами свойств только немногие распространяются на более общие дискретные цепи Маркова. Поэтому мы не будем останавливаться на общей теории.

Примеры немарковских процессов. а) *Урновая схема Пойа* (гл. V, пример 2, в). Пусть $X^{(n)}$ равно 1 или 0 в зависимости от того, будет ли при n-м извлечении вынут черный или красный шар. Последовательность $\{X^{(n)}\}$ не образует марковского процесса. Например,

$$P\{X^{(3)}=1\,|\,X^{(2)}=1\}=(b+c)/(b+r+c),$$

но

$$P\{X^{(3)}=1\,|\,X^{(2)}=1,\ X^{(1)}=1\}=(b+2c)/(b+r+2c)$$

(см. гл. V, задачи 19—20). С другой стороны, если $Y^{(n)}$ есть число черных шаров в урне в момент n, то $\{Y^{(n)}\}$ есть обычная однородная цепь Маркова.

б) *Суммы высших порядков.* Пусть $Y_0, Y_1, \ldots$ взаимно независимые случайные величины; положим $S_n=Y_0+\ldots+Y_n$. Разность S_n-S_m $(m<n)$ зависит только от $Y_{m+1}, \ldots, Y_n$, и легко видеть, что последовательность $\{S_n\}$ образует марковский процесс. Определим новую последовательность случайных величин $U_n=S_0+$ $+S_1+\ldots+S_n$ (это означает, что

$$U_n=Y_n+2Y_{n-1}+3Y_{n-2}+\ldots).$$

Последовательность $\{U_n\}$ образует стохастический процесс, для которого все вероятностные соотношения могут быть в принципе выражены через распределения величин Y_k. Процесс $\{U_n\}$, вообще говоря, не есть марковский процесс, так как нет никаких оснований для того, чтобы, например, вероятность $P\{U_n=0\,|\,U_{n-1}=a\}$ совпадала с $P\{U_n=0\,|\,U_{n-1}=a,\ U_{n-2}=b\}$. Знание U_{n-1} и U_{n-2} позволяет сделать более точные предсказания, чем знание одного только U_{n-1}.

В случае непрерывного времени предыдущие суммы заменяются интегралами. В теории диффузии Y_n играет роль ускорения, S_n — скорости, а U_n — положения. Если могут быть измерены только положения, то мы принуждены иметь дело с немарковским процессом, хотя этот процесс и определяется косвенно через марковский процесс.

в) *Скользящие средние.* Пусть снова $\{Y_n\}$ — последовательность взаимно независимых случайных величин. Скользящее среднее порядка r определяется, как $X^{(n)}=(Y_n+Y_{n+1}+\ldots+Y_{n+r-1})/r$. Легко видеть, что $\{X^{(n)}\}$ не образует марковского процесса. Процессы этого типа часто встречаются в приложениях (см. задачу 26).

г) *Задача об уличном движении.* В качестве практического примера немарковского процесса приведем проведенные Р. Фюртом [1]) наблюдения над числом пешеходов на некотором участке улицы. Идеализированная математическая схема этого процесса может быть получена следующим образом. Для простоты предположим, что все пешеходы имеют одинаковую скорость v; кроме того, рассмотрим

---

[1]) R. Fürth, Schwankungserscheinungen in der Physik, Sammlung Vieweg, Braunschweig, 1920, S. 17 ff. Первоначальные наблюдения опубликованы в *Physikalische Zeitschrift*, 19 (1918) и 20 (1919).

только пешеходов, движущихся в одном направлении. В момент времени 0 разделим положительную часть оси x на отрезки фиксированной длины δ, которые могут содержать или не содержать пешеходов. Предположим, что распределение пешеходов в наших отрезках определяется последовательностью испытаний Бернулли. Другими словами, мы имеем последовательность независимых случайных величин Y_k, каждая из которых принимает значения 1 и 0 с вероятностями соответственно p и q. Если отрезок $(k-1)\delta \leqslant x < k\delta$ содержит пешехода, то $Y_k = 1$. Пусть теперь вся ось x движется со скоростью v в отрицательном направлении, и пусть наблюдается число пешеходов на фиксированном интервале длины $N\delta$, совпадающем в момент $t = 0$ с интервалом $0 \leqslant x < N\delta$ движущейся оси x. В момент t этот фиксированный интервал совпадает с интервалом $vt \leqslant \leqslant x < vt + N\delta$ оси x. Пусть наблюдения проводятся в моменты $n\delta/v$, и пусть $X^{(n)}$ — число пешеходов на нашем фиксированном интервале в момент n. Тогда $X^{(n)} = Y_n + Y_{n+1} + \ldots + Y_{n+N-1}$, и процесс является с точностью до множителя $1/N$ процессом скользящих средних. Он является поэтому немарковским. (Переходя к пределу при $\delta \to 0$, мы получаем непрерывную схему, в которой распределение Пуассона заменяет собой биномиальное распределение.)

д) *Суперпозиция марковских процессов* (*сложное тасование*). Существует много технических устройств (например, селекторы на телефонных станциях, счетчики, фильтры), действие которых может быть описано, как суперпозиция двух марковских процессов, результатом которой является, однако, немарковский процесс. Основная идея таких механизмов может быть получена из рассмотрения следующего метода тасования карт.

В дополнение к основной колоде из N карт возьмем такую же вспомогательную колоду и применим к ней обычную технику тасования. Если карты вспомогательной колоды расположились в порядке $(a_1, a_2, \ldots, a_N)$, то мы переставим карты основной колоды так, чтобы первая, вторая, ..., N-я карты попали на места с номерами $a_1, \ldots, a_N$. Итак, тасование вспомогательной колоды определяет косвенным образом последовательные упорядочивания основной колоды. Последние и образуют *стохастический процесс, не являющийся процессом марковского типа*. Для доказательства этого достаточно показать, что знание двух последовательных упорядочений основной колоды дает для оценки будущего больше, чем знание одного только последнего упорядочения. Покажем это в простейшем частном случае.

Пусть $N = 4$. Предположим, что вначале вспомогательная колода имеет порядок (2431). Предположим, кроме того, что операция тасования представляет собой всегда «снимание», т. е. порядок (a_1, a_2, a_3, a_4) переходит в один из трех порядков (a_2, a_3, a_4, a_1), (a_3, a_4, a_1, a_2)

и (a_4, a_1, a_2, a_3). Припишем каждой из этих трех возможностей вероятность $^1/_3$. При этих условиях вспомогательная колода будет в любой момент иметь одно из четырех упорядочений: (2431), (4312), (3124), (1243). С другой стороны, уже несколько проб показывает, что основная колода будет постепенно проходить через все 24 возможных упорядочения, и что каждое из них будет появляться в комбинации с каждым из четырех возможных упорядочений вспомогательной колоды. Следовательно, порядок (1234) основной колоды появится бесконечное число раз и всегда будет следовать за одним из четырех порядков: (2431), (4312), (3124) и (1243). Но вспомогательная колода не может иметь 2 раза подряд один и тот же порядок, и, стало быть, основная колода не может дважды подряд подвергнуться одной и той же перестановке. Следовательно, если в моменты $(n-1)$ и n наблюдались порядки соответственно (1234) и (1243), то в момент $n+1$ состояние (1234) невозможно. Таким образом, знание состояний в моменты $(n-1)$ и n дает более полную информацию, чем знание одного только состояния в момент n.

§ 11 *). Различные дополнения

1. Обращенные вероятности. Хотя более естественно исследовать будущее развитие системы, иногда оказывается необходимым изучить ее прошлое. Рассмотрим цепь Маркова с состояниями E_k и постоянными вероятностями перехода p_{jk}, причем полные вероятности в момент n равны $a_k^{(n)} = \sum a_\nu^{(0)} p_{\nu k}^{(n)}$. *Условная вероятность того, что система находилась в момент $m < n$ в состоянии E_j, при условии, что в момент n она находится в состоянии E_k, равна*

$$q_{kj}(n, m) = \frac{a_j^{(m)}}{a_k^{(n)}} p_{jk}^{(n-m)}, \quad m < n. \tag{11.1}$$

Эта формула имеет смысл, только если $a_k^{(n)} > 0$; в противном случае искомые условные вероятности не определены. Если все $a_k^{(n)}$ положительны, то (11.1) определяет систему вероятностей перехода, удовлетворяющих всем свойствам, необходимым для того, чтобы процесс был марковским. В частности, $q_{kj}(m, n)$ удовлетворяют тождеству Колмогорова — Чэпмена с соответственно измененным направлением времени, а именно

$$q_{kj}(n, m) = \sum_\nu q_{k\nu}(n, r)\, q_{\nu j}(r, m) \tag{11.2}$$

*) См. примечание на стр. 388.

($m < r < n$). Вероятности $q_{kj}(n, m)$ называются *обращенными вероятностями* [1]). Рассмотрим, в частности, неприводимую цепь со стационарным распределением вероятностей $\{u_k\}$. Тогда $a_k^{(n)} = u_k$ при всех n, и $u_k > 0$ (см. § 6 и 7). В этом случае вероятности перехода за один шаг $q_{kj}(n, n+1)$ не зависят от n и равны

$$q_{kj} = \frac{u_j}{u_k} p_{jk}. \tag{11.3}$$

Матрица $\{q_{kj}\}$ является стохастической. Таким образом, обращенные вероятности определяют однородную цепь Маркова. Если $q_{kj} = p_{jk}$, то первоначальная цепь называется *обратимой*; ее вероятностные соотношения симметричны во времени.

2. Центральная предельная теорема. Теория рекуррентных событий содержит дальнейшую информацию относительно цепей Маркова. Пусть E_k — фиксированное возвратное состояние, причем время возвращения в E_k имеет конечную дисперсию σ_k^2 (это условие всегда выполнено, если цепь конечна). Пусть N_n — число попаданий системы в состояние E_k вплоть до момента n. Тогда, как мы знаем из § 4 гл. XII, величина N_n имеет асимптотически нормальное распределение. В обозначениях настоящей главы имеем $E(N_n) = 1/\mu_k = u_k$ (способ вычисления дисперсии в случае конечной цепи будет указан в следующей главе). Отсюда, в частности, вытекает, что для каждого $\varepsilon > 0$ вероятность того, что $|N_n/n - u_k| < \varepsilon$, стремится к единице при $n \to \infty$. Это и есть *обычный закон больших чисел* для числа попаданий в E_k. Аналогично выполняются усиленный закон больших чисел и закон повторного логарифма, что не требует специального доказательства. В случае бесконечной цепи время возвращения в состояние E_k не обязательно имеет конечную дисперсию, даже если математическое ожидание времени возвращения конечно. Однако общая предельная теорема для возвратных событий применима и в этом случае.

Случайные величины N_n могут быть определены следующим образом. Определим последовательность случайных величин X_n, равных 0 или 1 в зависимости от того, находится или не находится в момент n система в состоянии E_k. Тогда $N_n = X_1 + \ldots + X_n$. Это подсказывает воможность следующего обобщения. Припишем состоянию E_k некоторое число x_k, и пусть случайная величина X_n равна x_k, если в момент n система находится в состоянии E_k. Как обычно, мы положим $S_n = X_1 + \ldots + X_n$. В случае конечных цепей Дёблин [2]) показал, что, вообще говоря, для S_n выполняются центральная

[1]) А. Н. Колмогоров, Zur Theorie der Markoffschen Ketten, *Mathematische Annalen*, 112 (1935), 155—160.

[2]) W. Doeblin, Sur les propriétés asymptotiques de mouvements régis par certains types de chaines simples. Thesis, Paris, 1937.

предельная теорема и закон повторного логарифма. Исключение возникает лишь тогда, когда числа x_k выбраны так, что для любого кратчайшего пути, ведущего из E_k снова в E_k, сумма x_ν равна постоянной c, не зависящей от пути.

3. Нестохастические матрицы. Теоремы этой главы описывают асимптотическое поведение степеней P^n произвольной стохастической матрицы P, т. е. матрицы, элементы которой удовлетворяют условиям (1.2). Легко обобщить эти теоремы на более общий класс матриц. *Пусть P — произвольная (конечная или бесконечная) матрица с неотрицательными элементами. Обозначим суммы элементов ее строк через S_j, так что $S_j = \sum p_{jk}$. Предположим, что последовательность S_j ограничена*, т. е., что существует такая константа M, что $S_j \leqslant M$. При этих условиях асимптотическое поведение P^n также описывается нашими теоремами, так как изучение P может быть сведено к изучению стохастических матриц.

Для определенности предположим, что строки и столбцы матрицы занумерованы числами начиная с 1, и рассмотрим сначала случай, когда $S_j \leqslant 1$ при всех j. В этом случае мы расширим (окаймим) матрицу прибавлением строки и столбца с номерами нуль и элементами $p_{00} = 1$, $p_{01} = p_{02} = \ldots = 0$ и $p_{j0} = 1 - S_j$ при $j \geqslant 1$. Новая матрица Q является стохастической, и ее асимптотическое поведение определяется нашими теоремами. С другой стороны, P^n есть матрица, дополнительная к угловому элементу $p_{00}^{(n)}$ матрицы Q^n. В общем случае суммы строк S_j могут превышать единицу, но мы заменим матрицу P матрицей P^* с элементами p_{jk}/M. Суммы строк S_j^* матрицы P^* удовлетворяют условиям $S_j^* \leqslant 1$, и мы можем описать асимптотическое поведение степеней P^{*n}. Однако матрицы P^n и P^{*n} отличаются только множителем M^n, так что наши теоремы действительно описывают асимптотическое поведение $p_{jk}^{(n)}$ во всех случаях.

Матрицы подобного типа встречаются в теории обобщенных случайных блужданий с возникновением или уничтожением массы.

4. Литература. Существует обширная литература о *конечных* цепях Маркова. Детальный обзор различных методов исследования и ссылки на более ранние работы можно найти в обширном трактате М. Фреше [1]). Алгебраическая трактовка конечных цепей будет описана в следующей главе. Законченная теория конечных цепей может быть получена из развитой Фробениусом теории матриц с положительными элементами. Этот метод разрабатывался, в частности,

---

[1]) M. Fréchet, Recherches théoriques modernes sur le calcul des probabilités, v. 2 (Théorie des événements en chaine dans le cas d'un nombre fini d'états possibles), Paris, 1938. Другая монография о марковских цепях принадлежит B. Hostinsky, Méthodes générales du calcul des probabilités, вып. 52 Mémorial des sciences mathématiques, Paris, 1931.

В. И. Романовским [1]); к сожалению, он не переносится на более интересный случай бесконечных цепей, впервые рассмотренный А. Н. Колмогоровым [2]). Исследования Колмогорова продолжены Дёблином [3]) и Дубом [4]). Последний вывел эргодические свойства с помощью общей теории групп.

В недавних работах Чжуна [5]) исследуется, в частности, переход из одного состояния в другое, при условии, что некоторые состояния запрещены. Это приводит к более изящным формулам для центральной предельной теоремы.

Дерман [6]) показал, что в неприводимой цепи с нулевыми состояниями уравнения (6.1) для стационарного распределения имеют единственное решение $\{v_k\}$, такое, что $\sum v_k = \infty$. Обратная формула (11.3) имеет смысл и для таких решений, и современная теория характеризуется усиленным вниманием к неограниченным решениям [7]). Несколько очень общих классов немарковских процессов, связанных с цепями Маркова, систематически изучались Т. Е. Харрисом [8]).

§ 12. Задачи

1. Рассмотрим цепь Маркова, связанную с испытаниями Бернулли следующим образом. Будем говорить, что в момент n система находится в состоянии E_1, если $(n-1)$-е и n-е испытания привели к результату *УУ*. Аналогично E_2, E_3, E_4 соответствуют результатам *УН*, *НУ*, *НН*. Найти матрицу P и ее степени. Обобщить схему.

---

[1]) В. И. Романовский, Дискретные цепи Маркова, М. — Л., 1949. Распределение на комплексной плоскости характеристических корней стохастических матриц подробно исследовано в работах: Н. А. Дмитриев и Е. Б. Дынкин, Характеристические корни стохастических матриц, *Изв. АН СССР*, сер. матем., 10 (1946), 167—184, и Ф. И. Карпелевич, О характеристических корнях матриц с неотрицательными элементами. *Изв. АН СССР*, сер. матем., 15 (1951), 361—383. — *Прим. ред.*

[2]) Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen, *Математический сборник*, 1 (1936), 607—610. Эта статья не содержит никаких доказательств. Полное изложение в *Бюллетене Московского университета*, секц. А, 1 (1937), 1—45.

[3]) W. Doeblin, Sur deux problèmes de M. Kolmogoroff concernant les chaines dénombrables, *Bulletin Société Mathématique de France*, 66 (1939), 1—11.

[4]) J. L. Doob. Topics in the Theory of Markoff Chains, а также Markoff Chains — Denumerable Case, *Transactions of the American Mathematical Society*, 52 (1942), 37—64 и 58 (1945), 455—473.

[5]) K. L. Chung, Contributions to the theory of Markov chains I, *Journal of Research*, National Bureau of Standards, 50 (1953), 203—208 и II *Transactions of the American Mathematical Society*, 76 (1954), 397—419.

[6]) C. Derman, A solution to a set of fundamental equations in Markov chains, *Proceedings of the American Mathematical Society*, 5 (1954), 332—334.

[7]) W. Feller, Boundaries induced by positive matrices, *Transactions of the American Mathematical Society*, 83 (1956), 19—54.

[8]) T. E. Harris, On chains of infinite order, *Pacific Journal of Mathematics*, 5 (1955), добавление 1, 707—724.

2. Классифицировать состояния четырех цепей Маркова, матрицы которых имеют заданные ниже строки. Найти в каждом из случаев P^2 и асимптотическое поведение $p_{jk}^{(n)}$.

а) $\left(0, \frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, 0, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}, 0\right)$;

б) $(0, 0, 0, 1), (0, 0, 0, 1), \left(\frac{1}{2}, \frac{1}{2}, 0, 0\right), (0, 0, 1, 0)$;

в) $\left(\frac{1}{2}, 0, \frac{1}{2}, 0, 0\right), \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0, 0\right), \left(\frac{1}{2}, 0, \frac{1}{2}, 0, 0\right),$

$$\left(0, 0, 0, \frac{1}{2}, \frac{1}{2}\right), \left(0, 0, 0, \frac{1}{2}, \frac{1}{2}\right);$$

г) $\left(0, \frac{1}{2}, \frac{1}{2}, 0, 0, 0\right), \left(0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right),$

$$(1, 0, 0, 0, 0, 0), (1, 0, 0, 0, 0, 0), (1, 0, 0, 0, 0, 0).$$

3. Рассмотрим бросания правильной кости и условимся говорить, что в момент n система находится в состоянии E_j, если j есть наибольшее из чисел очков, выпавших при первых n бросаниях. Найти матрицу P^n и проверить, что (3.3) выполнено.

4. В примере 2, м система исходит из состояний E_k ($k = 2, 3, 4, 6$). Найти вероятность поглощения x_k и y_k в состояниях E_1 и E_5 соответственно. (Решить задачи, исходя из основных определений и не обращаясь к формулам § 8.)

5. Исследовать пример 5, б гл. I, опираясь на теорию цепей Маркова. Вычислить вероятности выигрыша для каждого из игроков.

6. Пусть $\{p_1, p_2, \ldots\}$ — первая строка стохастической матрицы P. В следующих строках $p_{j,\,j-1} = 1$, прочие элементы равны нулю. Решить вопрос о характере состояний и найти стационарное распределение, если оно существует.

7. Пусть $\{q_0, q_1, \ldots\}$ — первый столбец матрицы P и $p_{i,\,i+1} = 1 - q_i$ при $i = 0, 1, \ldots$. Доказать, что состояния невозвратны тогда и только тогда, когда $\sum q_j < \infty$. Когда состояния являются нулевыми? Найти стационарное распределение, если оно существует.

8. Отражающий экран. Рассмотрим матрицу случайного блуждания, для которой $p_{k,\,k+1} = p$, $p_{k,\,k-1} = q$ при $k = 2, 3, \ldots$ и $p_{12} = p$, $p_{11} = q$. Доказать, что состояния являются невозвратными, если $p > q$, возвратными нулевыми, если $p = q$, и эргодическими, если $p < q$. Найти в последнем случае стационарное распределение.

9. Два отражающих экрана. Цепь Маркова с состояниями $1, 2, \ldots, a$ имеет матрицу вероятностей перехода, первая и последняя строки которой равны $(q, p, 0, \ldots, 0)$ и $(0, 0, \ldots, 0, q, p)$ соответственно. В оставшихся строках $p_{k,\,k+1} = p$, $p_{k,\,k-1} = q$. Найти стационарное распределение. Может ли цепь быть периодической?

10. N черных и N белых шаров разложены по двум урнам так, что каждая урна содержит N шаров. Число черных шаров в первой урне определяет состояние системы. На каждом шагу случайно выбирается по одному шару из каждой урны, и эти выбранные шары меняются местами. Найти p_{jk}. Показать, что предельная вероятность u_k равна вероятности получить ровно k черных шаров, если N шаров выбираются случайно из совокупности, содержащей N черных и N белых шаров [1]).

---

[1]) Эта задача восходит еще к Лапласу; см. стр. 49 книги Фреше, указанной в примечании на стр. 404.

11. Цепь с состояниями E_0, E_1, ... имеет вероятности перехода

$$p_{jk} = e^{-\lambda} \sum_{\nu=0}^{j} \binom{j}{\nu} p^{\nu} q^{j-\nu} \frac{\lambda^{k-\nu}}{(k-\nu)!},$$

где слагаемые заменяются нулями при $\nu > k$. Показать, что

$$p_{jk}^{(n)} \to e^{-\lambda/q} \frac{(\lambda/q)^k}{k!}.$$

З а м е ч а н и е. Такая цепь встречается в статистической механике [1]) и может быть интерпретирована следующим образом. Состояние системы определяется числом частиц в некоторой области пространства. За каждый промежуток времени единичной длины частица может покинуть эту область с вероятностью q. Кроме того, в этой области пространства могут появиться новые частицы, и вероятность этого задается выражением Пуассона $e^{-\lambda}\lambda^r/r!$. Стационарным распределением оказывается тогда распределение Пуассона с параметром λ/q.

12. Модель Эренфестов. В примере 2, е пусть вначале в первом резервуаре содержится j молекул, и пусть $X^{(n)} = 2k - a$, если в момент n система находится в состоянии k (так что $X^{(n)}$ есть разность чисел молекул в двух резервуарах). Пусть $e_n = E(X^{(n)})$. Доказать, что $e_{n+1} = (a-2)e_n/a$, откуда $e_n = (1 - 2/a)^n (2j - a)$. (Отметим, что $e_n \to 0$ при $n \to \infty$.)

13. Исследовать задачу о подсчете частиц, пример 1, б гл. XIII, в терминах цепей Маркова.

14. С л у ч а й н о е б л у ж д а н и е с о т р а ж а ю щ и м и э к р а н а м и. Рассмотрим *симметричное* случайное блуждание в ограниченной области плоскости. Граница является отражающей в том смысле, что всякий раз, когда при неограниченном случайном блуждании частица должна покинуть область, она вынуждена возвратиться в предыдущее положение. Показать, что если каждая точка области достижима из каждой другой точки, то существует стационарное распределение, и что $u_k = 1/a$, где a — число состояний в области.

15. П о в т о р н ы е с р е д н и е. Пусть $\{x_1, x_2, \ldots\}$ — ограниченная последовательность чисел, а P — матрица вероятностей перехода в эргодической цепи. Доказать, что $\sum_j p_{ij}^{(n)} x_j \to \sum u_j x_j$. Показать, что процедура повторного осреднения (§ 4 гл. XIII) и задача 5 гл. XIII являются частными случаями этой задачи.

16. В теории массового обслуживания появляется цепь Маркова с матрицей вероятностей перехода

$$\left\| \begin{matrix} p_0 & p_1 & p_2 & p_3 \cdots \\ p_0 & p_1 & p_2 & p_3 \cdots \\ 0 & p_0 & p_1 & p_2 \cdots \\ 0 & 0 & p_0 & p_1 \cdots \\ . & . & . & . \;\cdots \end{matrix} \right\|,$$

где $\{p_k\}$ — произвольное распределение. Используя производящие функции выяснить характер состояний цепи. Найти производящую функцию стационарного распределения, если оно существует.

---

[1]) С. Ч а н д р а с е к а р, Стохастические процессы в физике и астрономии, ИЛ, М., 1947, стр. 1—126, в частности стр. 82.

17. Время ожидания поглощения. Для невозвратного состояния E_j пусть Y_j — момент, когда система исходя из E_j *впервые* достигнет возвратного состояния. Предполагая, что вероятность все время оставаться в множестве невозвратных состояний равна нулю, доказать, что $d_j = \mathbf{E}(Y_j)$ однозначно определяются как решение системы линейных уравнений

$$d_j - \sum_T p_{j\nu} d_\nu = 1,$$

где суммирование распространяется по всем ν, таким, что E_ν — невозвратное состояние. Следует отметить, что d_ν не обязательно конечны.

18. Если число состояний $a < \infty$ и если E_k достижимо из E_j, то оно достижимо за a или меньше шагов.

19. Пусть цепь содержит a состояний, и пусть E_j — возвратное состояние. Тогда существует такое число $q < 1$, что при $n \geqslant a$ вероятность того, что время возвращения в E_j превышает n, меньше, чем q^n. (Указание. Использовать задачу 18.)

20. Состояние E_j конечной цепи является невозвратным тогда и только тогда, когда существует такое E_k, что E_k достижимо из E_j, а E_j недостижимо из E_k. (Для бесконечной цепи это неверно, как показывает задача 7.)

21. Неприводимая цепь, для которой *хотя бы один* диагональный элемент p_{jj} положителен, не может быть периодической.

22. Конечная неприводимая цепь является непериодической тогда и только тогда, когда существует такое n, что $p_{jk}^{(n)} > 0$ при всех j и k.

23. Для цепи из a состояний пусть $(x_1, \ldots, x_a)$ есть решение системы линейных уравнений $x_j = \sum p_{j\nu} x_\nu$. Доказать: а) что состояния E_r, для которых $x_r > 0$, образуют замкнутое (не обязательно неприводимое) множество; б) если E_j и E_k принадлежат к одному и тому же неприводимому множеству, то $x_j = x_k$.

24. Продолжение. Если $(x_1, \ldots, x_a)$ — решение уравнений $x_j = s \sum p_{j\nu} x_\nu$, причем $|s| = 1$, но $s \neq 1$, то существует такое целое число $t > 1$, что $s^t = 1$. Если цепь неприводима, то наименьшее целое число такого рода является периодом цепи.

25. Эргодическая теорема для средних[1]). Для произвольной цепи Маркова определим числа $A_{jk}^{(n)}$ равенствами

$$A_{jk}^{(n)} = 1/n \sum_{\nu=1}^{n} p_{jk}^{(\nu)}.$$

Если E_j и E_k принадлежат к одному и тому же неприводимому замкнутому множеству, то $A_{jk}^{(n)}$ стремятся к пределам, которые не зависят от j (и равны стационарным вероятностям u_k, если последние существуют). Если E_j и E_k принадлежат к различным замкнутым множествам, то $A_{jk}^{(n)} = 0$ при всех n. Если E_k невозвратно, то $A_{jk}^{(n)} \to 0$ для всех j.

26. Скользящие средние. Пусть $\{Y_k\}$ — последовательность взаимно независимых случайных величин, каждая из которых принимает зна-

---

[1]) Эта теорема есть простое следствие результатов настоящей главы. Однако она намного слабее и может быть поэтому доказана более простыми методами. См.: K. Yosida and S. Kakutani, Markoff processes with an enumerable infinite number of possible states, *Japanese Journal of Mathematics*, 16 (1939), 47—55.

чения ± 1 с вероятностями $^1/_2$. Положим $X^{(n)} = \{Y_n + Y_{n+1}\}/2$. Найти вероятности перехода

$$p_{jk}(m, n) = P\{X^{(n)} = k \mid X^{(m)} = j\},$$

где $m < n$, и j, $k = -1, 0, 1$. Показать, что $\{X^{(n)}\}$ не образуют марковского процесса и что (10.3) не выполняется.

27. Рассмотрим последовательность испытаний Бернулли. Условимся говорить, что в момент n система находится в состоянии E_1, если n-е и $(n-1)$-е испытания привели к успеху, и в состоянии E_2 — в противном случае. Найти вероятности перехода за n шагов. Выяснить немарковский характер процесса.

Замечание. Этот процесс получается из цепи задачи 1 объединением трех состояний в одно. Такая процедура *группировки* применима к любым цепям Маркова и нарушает марковское свойство. Процессам такого рода посвящена работа Харриса.

28. Пусть даны две цепи Маркова с одним и тем же числом состояний и матрицами переходных вероятностей P_1 и P_2. Новый процесс определяется начальным распределением и матрицей вероятностей перехода за n шагов $\frac{1}{2}P_1^n + \frac{1}{2}P_2^n$. Выяснить немарковский характер процесса и связь с урновыми схемами гл. V.

ГЛАВА XVI *)

АЛГЕБРАИЧЕСКИЙ МЕТОД ИЗУЧЕНИЯ КОНЕЧНЫХ ЦЕПЕЙ МАРКОВА

В настоящей главе мы рассмотрим цепь Маркова с конечным числом состояний $E_1, \ldots, E_a$ и матрицей вероятностей перехода p_{jk}. Нашей главной целью будет получить явные формулы для вероятностей перехода за n шагов $p_{jk}^{(n)}$. Нам не понадобятся результаты предыдущей главы, за исключением общих понятий и обозначений § 3.

Мы используем метод производящих функций и получим желаемые результаты методом разложения на простые дроби, изложенным в § 4 гл. XI. Наши результаты могут быть легко получены из теории приведения матриц к каноническому виду [1]) (которая в свою очередь может быть выведена из наших результатов). Все эргодические свойства *конечных* цепей, доказанные в гл. XV, также следуют из результатов настоящей главы. Однако для простоты мы пойдем на некоторое ограничение общности и не будем интересоваться исключительными случаями, усложняющими общую теорию и не встречающимися в практических примерах.

Основные черты общего метода излагаются в § 1; затем метод иллюстрируется на частных примерах в § 2 и 3. В § 4 особое внимание обращается на невозвратные состояния и вероятности поглощения. В § 5 теория прилагается к нахождению дисперсии времени возвращения в состояние E_j.

§ 1. Общая теория

Для каждых фиксированных j и k определим производящую функцию

$$P_{jk}(s)=\sum_{n=1}^{\infty} p_{jk}^{(n)} s^{n-1}. \tag{1.1}$$

Умножив эти уравнения на $sp_{\nu j}$ и суммируя по всем j, получим

$$s\sum_{j=1}^{a} p_{\nu j} P_{jk}(s)=P_{\nu k}(s)-p_{\nu k}, \tag{1.2}$$

---

*) См. примечание на стр. 291.

[1]) См. книгу Фреше, на которую мы уже ссылались в § 11 гл. XV.

При каждом фиксированном k мы имеем систему a неоднородных линейных уравнений относительно a неизвестных: $P_{1k}(s), \ldots, P_{ak}(s)$. Теоретически эта система может быть разрешена с помощью детерминантов или последовательным исключением неизвестных. Мы используем только тот факт, что детерминант системы $D(s)$ есть полином степени не выше a и что $P_{\nu k}(s)$ суть *рациональные функции от* s с общим знаменателем $D(s)$. Рассмотрим только тот случай, когда уравнение $D(s)=0$ не имеет кратных корней. Это условие, являющееся некоторым ограничением общности, все же выполняется в большинстве случаев, имеющих практический интерес.

Так как $P_{\nu k}(s)$ суть рациональные функции, можно воспользоваться методом разложения на простые дроби § 4 гл. XI. Из этого разложения следует, что существуют коэффициенты $\rho_{\nu k}^{(1)}, \rho_{\nu k}^{(2)}, \ldots, \rho_{\nu k}^{(a)}$, такие, что

$$p_{\nu k}^{(n)}=\frac{\rho_{\nu k}^{(1)}}{s_1^n}+\frac{\rho_{\nu k}^{(2)}}{s_2^n}+\ldots+\frac{\rho_{\nu k}^{(a)}}{s_a^n}, \tag{1.3}$$

где $s_1, s_2, \ldots$ суть корни уравнения $D(s)=0$. Если степень $D(s)$ меньше a, то и (1.3) будет содержать меньше, чем a членов. Возможно также, что при некоторых частных значениях ν и k один или большее число корней s_r являются общими для числителя и знаменателя и, следовательно, сокращаются. Мы учтем такие случаи, полагая соответствующие коэффициенты $\rho_{\nu k}^{(r)}$ равными нулю.

Мы могли бы вычислить корни s_r и коэффициенты $\rho_{\nu k}^{(r)}$ методами гл. XI, но лучше использовать преимущества, создаваемые некоторыми свойствами цепей Маркова. Умножим (1.3) на $p_{j\nu}$ и просуммируем по $\nu=1, 2, \ldots$. В результате получим

$$p_{jk}^{(n+1)}=\sum_{\nu=1}^{a} p_{j\nu}\left\{\frac{\rho_{\nu k}^{(1)}}{s_1^n}+\ldots+\frac{\rho_{\nu k}^{(a)}}{s_a^n}\right\}. \tag{1.4}$$

Если левую часть заменить значением (1.3), то получится тождество, которое может выполняться при всех n лишь тогда, когда коэффициенты при $s_1^{-n}, \ldots, s_a^{-n}$ в обеих частях тождества равны. Стало быть, для каждого фиксированного r (при $1\leqslant r\leqslant a$) имеем

$$\rho_{jk}^{(r)}=s_r\sum_{\nu=1}^{a} p_{j\nu}\rho_{\nu k}^{(r)}, \qquad r=1, \ldots, a. \tag{1.5а}$$

Аналогичным образом, умножая (1.3) на p_{km} и суммируя по всем k, получаем

$$\rho_{\nu m}^{(r)}=s_r\sum_{k=1}^{a}\rho_{\nu k}^{(r)}p_{km}. \tag{1.5б}$$

Уравнения (1.5а) показывают, что при фиксированных k и r a величин $\rho_{1k}^{(r)}, \ldots, \rho_{ak}^{(r)}$ суть решения системы a линейных уравнений

$$x_j^{(r)} = s_r \sum_{\nu=1}^{a} p_{j\nu} x_\nu^{(r)} \qquad (j = 1, \ldots, a). \tag{1.6а}$$

Аналогично из (1.5б) следует, что при фиксированных ν и r коэффициенты $\rho_{\nu 1}^{(r)}, \ldots, \rho_{\nu a}^{(r)}$ удовлетворяют системе a линейных уравнений

$$y_m^{(r)} = s_r \sum_{k=1}^{a} y_k^{(r)} p_{km} \qquad (m = 1, \ldots, a). \tag{1.6б}$$

Заменим s_r на произвольное s и рассмотрим две более общие системы:

$$x_j = s \sum_{\nu=1}^{a} p_{j\nu} x_\nu \qquad (j = 1, \ldots, a) \tag{1.7а}$$

и

$$y_m = s \sum_{k=1}^{n} y_k p_{km} \qquad (m = 1, \ldots, a). \tag{1.7б}$$

Система a однородных уравнений с a неизвестными может иметь нетривиальное [1]) решение лишь тогда, когда ее детерминант равен нулю. Матрицы двух систем (1.7а) и (1.7б) получаются одна из другой заменой строк на столбцы. Детерминант (1.7а), очевидно, равен детерминанту системы (1.2), откуда следует, что детерминанты двух систем (1.7а) и (1.7б) обращаются в нуль при $s = s_1, s_2, \ldots, s_a$.

Мы можем теперь забыть о производящей функции $P_{jk}(s)$ и определить корни $s_1, \ldots, s_a$, как такие числа (действительные или комплексные), при которых системы (1.7а) и (1.7б) имеют нетривиальные решения. Предположение, что s_r есть простой корень, означает, что для каждого фиксированного r решения $\left(x_1^{(r)}, \ldots, x_a^{(r)}\right)$ и $\left(y_1^{(r)}, \ldots, y_a^{(r)}\right)$ определяются однозначно с точностью до численного множителя. Однако нашей отправной точкой было установление того факта, что при фиксированных k и r строчка $\left(\rho_{1k}^{(r)}, \ldots, \rho_{ak}^{(r)}\right)$ есть решение системы (1.7а), а при фиксированных ν и r строчка $\left(\rho_{\nu 1}^{(r)}, \ldots, \rho_{\nu a}^{(r)}\right)$ есть решение системы (1.7б). Так как эти решения определены с точностью до численного множителя, имеем

$$\rho_{jk}^{(r)} = c_r x_j^{(r)} y_k^{(r)}. \tag{1.8}$$

Остается только вычислить постоянные $c_1, \ldots, c_a$.

---

[1]) Как обычно, мы называем тождественно обращающееся в нуль решение тривиальным и пренебрегаем им.

Из (1.8) и (1.3) получаем

$$p_{jk}^{(n)} = \sum_{r=1}^{a} c_r x_j^{(r)} y_k^{(r)} s_r^{-n}. \tag{1.9}$$

Далее,

$$p_{jk}^{(2n)} = \sum_{\nu=1}^{a} p_{j\nu}^{(n)} p_{\nu k}^{(n)}, \tag{1.10}$$

и мы можем выразить все величины в (1.10) с помощью (1.9). Приравнивая коэффициенты при s_r^{-2n} в обеих частях равенства, получаем

$$1 = c_r \sum_{\nu=1}^{a} x_\nu^{(r)} y_\nu^{(r)}; \tag{1.11}$$

откуда определяется c_r. Правда, решения $x_\nu^{(r)}$ и $y_\nu^{(r)}$ определены только с точностью до численного множителя· Однако если мы заменим $x_j^{(r)}$ на $Ax_j^{(r)}$ и $y_k^{(r)}$ на $By_k^{(r)}$, то c_r заменится на c_r/AB, а величины $\rho_{jk}^{(r)}$ в (1.8) останутся неизменными.

В итоге мы получаем следующий метод вычисления $p_{jk}^{(n)}$.

Составим системы линейных уравнений (1.7а) *и* (1.7б). *Они имеют общий детерминант и допускают нетривиальное решение только в том случае, когда этот детерминант обращается в нуль. Мы предположили, что все корни* s_1, s_2, ... (*число которых, самое большее, равно* a) *простые*; *тогда для каждого* r *решения* $\left(x_1^{(r)}, \ldots, x_a^{(r)}\right)$ и $\left(y_1^{(r)}, \ldots, y_a^{(r)}\right)$ *определены с точностью до произвольного множителя. Найдем эти решения и затем постоянные* c_r *из* (1.11). *Образуем величины* $\rho_{jk}^{(r)}$ *в соответствии с* (1.8); *тогда* $p_{jk}^{(n)}$ *задаются формулой* (1.9).

При каждом фиксированном r величины $\rho_{jk}^{(r)}$ образуют матрицу, которая может быть построена следующим способом. Образуем таблицу произведений, в которой на пересечении j-й строки и k-го столбца стоит произведение $x_j^{(r)} y_k^{(r)}$. Умножив все a^2 элементов этой квадратной таблицы на c_r, мы получим матрицу $\rho_{jk}^{(n)}$. Чтобы построить матрицу $\left(p_{jk}^{(n)}\right)$, мы должны разделить все элементы $\rho_{jk}^{(r)}$ на s_r^n и просуммировать полученные таким образом матрицы по $r = 1, 2, \ldots, a$. Отметим, что корни s_r могут быть простыми, даже если имеются менее чем a корней; однако если имеется a различных корней s_1, s_2, ..., s_a, то они все обязательно простые.

Случай кратных корней требует некоторых изменений, но все же может быть изучен аналогичными методами. Наиболее интересные примеры этого рода будут рассмотрены в § 4.

В алгебре обратные величины $\lambda_r = 1/s_r$ называются *характеристическими корнями* (или собственными значениями) матрицы P. Нуль может быть собственным значением, но ему не соответствует никакой корень s_r. Этим объясняется, почему может быть меньше, чем a корней s_r, несмотря на то, что имеется a характеристических корней. При применении метода производящих функций удобнее пользоваться корнями s_r вместо обратных величин λ_r. Кроме того, это соответствует приему, общепринятому в теории интегральных уравнений, и поэтому более естественно для теории вероятностей.

Среди s_r *всегда встречается значение* $s=1$, *и ему соответствует решение* $(1, 1, \ldots, 1)$ *уравнения* (1.7a). Для всех r имеем $|s_r| \geqslant 1$. Действительно[1]), наличие корня s_r с $|s_r| < 1$ привело бы к решению, стремящемуся к бесконечности [см. (1.3)]. Если $s_1 = 1$ есть единственный корень, такой, что $|s_r| = 1$, то $p_{jk}^{(n)} \to c_1 x_j^{(1)} y_k^{(1)}$. Нетрудно показать, что если существуют другие корни с $|s_r| = 1$, то они обязательно суть корни t-й степени из единицы, где t — целое число; в этом случае цепь имеет период t. За деталями читатель отсылается к трактату Фреше, на который мы ссылались в § 11 гл. XV.

Часто нахождение *всех* корней s_r оказывается громоздким или даже невозможным. Однако ясно, что асимптотическое поведение $p_{jk}^{(n)}$ определяется в первом приближении корнями s_r, для которых $|s_r| = 1$, а во втором приближении — корнями s_r со следующим наименьшим модулем.

Окончательную формулу (1.9) можно очень изящно записать в матричных обозначениях. Пусть $X^{(r)}$ — вектор-столбец (или матрица размеров $a \times 1$) с элементами $x_j^{(r)}$, и пусть $Y^{(r)}$ — вектор-строка (или матрица размеров $1 \times a$) с элементами $y_k^{(r)}$. Тогда $X^r Y^{(r)}$ — матрица размеров $a \times a$ с элементами $x_j^{(r)} y_k^{(r)}$ и формула (1.9) принимает вид

$$P^n = \sum_{r=1}^{a} c_r s_r^{-n} X^{(r)} Y^{(r)}, \quad \text{где } c_r^{-1} = (Y^{(r)} X^{(r)}). \tag{1.12}$$

Векторы $X^{(r)}$ и $Y^{(r)}$ называются собственными векторами матрицы P, а c_r^{-1} — их скалярным (внутренним) произведением.

§ 2. Примеры

а) Рассмотрим сначала цепь, состоящую только из двух состояний. Матрица вероятностей перехода имеет тогда простой вид:

$$P = \begin{pmatrix} 1-p, & p \\ \alpha & 1-\alpha \end{pmatrix},$$

где $0 < p < 1$ и $0 < \alpha < 1$. Уравнения (1.7a) сводятся к $s(1-p)x_1 + sp x_2 = x_1$ и $s\alpha x_1 + s(1-\alpha)x_2 = x_2$. Находим из этих равенств два выражения для отношения x_1/x_2 и приравниваем их друг другу. Убеждаемся, что решение существует только при $s=1$ или $s = 1/(1-\alpha-p)$. Решение, соответствующее $s_1 = 1$, есть $(1, 1)$; решение, соответствующее $s_2 = 1/(1-\alpha-p)$, есть $(p, -\alpha)$. Далее, рассмотрим систему (1.7б), которая теперь имеет вид $s(1-p)y_1 +$

[1]) Прямое доказательство проводится следующим образом. Пусть M — максимальное из чисел $|x_1^{(r)}|, \ldots, |x_a^{(r)}|$ (r — фиксировано). Тогда по (1.6a) $M \leqslant |s_r| \sum p_{j\nu} M = |s_r| M$, или $|s_r| \geqslant 1$.

$+s\alpha y_2=y_1$ и $sp y_1+s(1-\alpha)y_2=y_2$. Мы знаем, что эта система имеет решение только при $s=s_1$ или $s=s_2$. Соответствующие решения суть (α, p) и $(1, -1)$. Из (1.11) получаем $c_1=c_2=1/(\alpha+p)$. Уравнения (1.8) и (1.3) дают теперь возможность записать явные формулы для величин $p_{jk}^{(n)}$. Окончательный результат может быть записан в матричной форме:

$$P^{(n)}=\frac{1}{\alpha+p}\begin{pmatrix}\alpha & p\\ \alpha & p\end{pmatrix}+\frac{(1-\alpha-p)^n}{\alpha+p}\begin{pmatrix}p & -p\\ -\alpha & \alpha\end{pmatrix}$$

(множители, общие для всех четырех элементов, вынесены за матрицу). Так как $|1-\alpha-p|<1$, вторая матрица стремится к нулю при $n\to\infty$, и первая матрица представляет собой предельное выражение для P^n.

б) Пусть

$$P=\begin{Vmatrix}0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1\\ 1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1 & 0\end{Vmatrix}. \tag{2.1}$$

(Это — матрица задачи 2, б гл. XV). Система (1.7а) имеет вид

$$x_1=sx_4,\quad x_2=sx_4,\quad x_3=\frac{s(x_1+x_2)}{2},\quad x_4=sx_3. \tag{2.2}$$

Так как мы имеем право произвольно выбрать постоянный множитель, то можно положить $x_4=1$. Тогда $x_1=s$, $x_2=s$, $x_3=s^2$, $x_4=s^3$, и поэтому должно быть $s^3=1$. Положим теперь

$$\theta=e^{2\pi i/3}=\cos\frac{2\pi}{3}+i\sin\frac{2\pi}{3}; \tag{2.3}$$

тогда корнями уравнения $s^3=1$ будут $s_1=1$, $s_2=\theta$, $s_3=\theta^2$. (Заметим, что имеется только три корня, хотя всего есть четыре состояния.) Решения $x_j^{(r)}$, соответствующие этим трем корням, суть $(1, 1, 1, 1)$, $(\theta, \theta, \theta^2, 1)$ и $(\theta^2, \theta^2, \theta, 1)$.

Из (1.7б) получаем $y_1=sy_3/2$, $y_2=sy_3/2$, $y_3=sy_4$, $y_4=s(y_1+y_2)$. Решения, соответствующие корням $s_1=1$, $s_2=\theta$ и $s_3=\theta^2$, суть $(1, 1, 2, 2)$, $(\theta, \theta, 2, 2\theta^2)$ и $(\theta^2, \theta^2, 2, 2\theta)$. Следовательно, из (1.11) получаем $c_1=1/6$, $c_2=1/(6\theta^2)=\theta/6$, $c_3=1/(6\theta)=\theta^2/6$. Мы можем теперь выписать все $p_{jk}^{(n)}$. Например,

$$\begin{aligned}p_{11}^{(n)}&=p_{22}^{(n)}=\frac{1+\theta^n+\theta^{2n}}{6},\\ p_{13}^{(n)}&=\frac{1+\theta^{2n+2}+\theta^{n+1}}{3},\\ p_{14}^{(n)}&=\frac{1+\theta^{2n+1}+\theta^{n+2}}{3}\end{aligned} \tag{2.4}$$

и т. д. Цепь является, очевидно, периодической с периодом 3.

в) Пусть $p+q=1$, и

$$P=\left\|\begin{matrix} 0 & p & 0 & q \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ p & 0 & q & 0 \end{matrix}\right\|. \qquad (2.5)$$

[Эта матрица описывает циклическое случайное блуждание. См. пример 2, г гл. XV.] Уравнения (1.7а) имеют вид

$$x_1=s(px_2+qx_4), \quad x_2=s(qx_1+px_3),$$
$$x_3=s(qx_2+px_4), \quad x_4=s(px_1+qx_3).$$

Предположим, что $p\neq q$. Из первого и третьего уравнений находим $x_1+x_3=s(x_2+x_4)$; из оставшихся уравнений получаем $x_2+x_4=s(x_1+x_3)$. Следовательно, или $s^2=1$, или же $x_1+x_3=x_2+x_4=0$. Первая возможность приводит к двум корням: $s_1=1$ и $s_2=-1$. С другой стороны, подставляя $x_3=-x_1$ и $x_4=-x_2$ в первые два уравнения, получаем $s^2(p-q)^2=-1$, откуда получаются оставшиеся два корня s_3 и s_4. Таким образом,

$$s_1=1; \quad s_2=-1; \quad s_3=\frac{i}{q-p}; \quad s_4=-\frac{i}{q-p}. \qquad (2.6)$$

(Здесь $i^2=-1$.) Соответствующие решения содержат произвольный множитель, и мы можем поэтому положить $x_4^{(r)}=1$. Тогда легко найти четыре решения (1, 1, 1, 1), (—1, 1, —1, 1), (i, —1, —i, 1) и (—i, —1, i, 1). Система (1.7б) в нашем случае имеет вид

$$y_1=s(qy_2+py_4), \quad y_2=s(py_1+qy_3),$$
$$y_3=s(py_2+qy_4), \quad y_4=s(qy_1+py_3).$$

Четырем корням (2.6) соответствуют решения (1, 1, 1, 1), (—1, 1, —1, 1), (—i, —1, i, 1) и (i, —1, —i, 1). Постоянные c_r находим из (1.11). Оказывается, что $c_1=c_2=c_3=c_4=1/4$. Используя (1.3) и (1.8), мы можем написать явные формулы для каждой из последовательностей $p_{jk}^{(n)}$ ($n=1, 2, 3, \ldots$). В данном случае решения $x_j^{(r)}$ и $y_j^{(r)}$ имеют простой вид (α, α^2, α^3, α^4), где α — одно из четырех чисел 1, —1, i, —i. Это позволяет выразить $p_{jk}^{(n)}$ простой формулой:

$$p_{jk}^{(n)}=\frac{1}{4}\{1+(q-p)^n(i)^{j-k-n}\}\{1+(-1)^{k+j-n}\}. \qquad (2.7)$$

Эта формула остается верной и при $p=q=1/2$.

Очевидно, что слагаемое, содержащее $(q-p)^n$, стремится к нулю, и что второй множитель имеет период 2.

г) Общее циклическое блуждание (пример 2, г гл. XV). В предыдущем примере мы смогли выразить $x_j^{(r)}$ и $y_k^{(r)}$, как степени корней четвертой степени из единицы. Это наводит на мысль попытаться получить аналогичные выражения для общей матрицы примера 2, г гл. XV. Удобно занумеровать состояния числами от 0 до $a-1$. Для краткости положим

$$\theta = e^{2\pi i/a}. \tag{2.8}$$

Это — корень a-й степени из единицы, и все a корней этой степени представляются в виде $1, \theta, \theta^2, \ldots, \theta^{a-1}$. Легко видеть, что системам (1.7а) и (1.7б) удовлетворяют a наборов чисел

$$x_j^{(r)} = \theta^{rj}, \quad y_k^{(r)} = \theta^{-rk} \tag{2.9}$$

при $r = 0, 1, 2, \ldots, a-1$; они соответствуют корням

$$s_r = \left\{ \sum_{\nu=0}^{a-1} q_\nu \theta^{\nu r} \right\}^{-1}. \tag{2.10}$$

Из (1.11) и (2.9) убеждаемся, что $c_r = 1/a$ при всех r, и, таким образом,

$$p_{jk}^{(n)} = \frac{1}{a} \sum_{r=0}^{a-1} \theta^{r(j-k)} \left(\sum_{\nu=0}^{a-1} q_\nu \theta^{\nu r} \right)^n. \tag{2.11}$$

Интересно проверить эту формулу для $n = 1$. Множитель при q_ν тогда равен

$$\sum_{r=0}^{a-1} \theta^{r(j-k+\nu)}. \tag{2.12}$$

Эта сумма равна нулю, исключая тот случай, когда $j-k+\nu$ равно нулю или a. В этом последнем случае каждое слагаемое равно единице. Следовательно, $p_{jk}^{(1)}$ равно q_{k-j}, если $k \geqslant j$, и равно q_{a+k-j}, если $k < j$; но это и есть заданная матрица (p_{jk}).

д) **Задача о размещении.** Пример 2, ж гл. XV показывает, что классическая задача о размещении может быть изучена методом цепей Маркова. Система находится в состоянии j, если имеется j занятых и $a-j$ пустых ячеек. Если таково начальное положение и n дополнительных шаров случайно размещаются по ячейкам, то $p_{jk}^{(n)}$ представляет собой вероятность того, что в результате окажется k занятых и $a-k$ пустых ячеек (так что $p_{jk}^{(n)} = 0$, если $k < j$). При $j = 0$ эта вероятность может быть найдена по формуле (11.7) гл. II. Мы выведем теперь формулу для $p_{jk}^{(n)}$, обобщив, таким образом, результаты гл. II.

Так как $p_{jj}=j/a$ и $p_{j,\,j+1}=(a-j)/a$, то, как легко видеть, система уравнений (1.7а) принимает вид

$$(a-sj)\,x_j=s\,(a-j)\,x_{j+1}\qquad (j=0,\ \ldots,\ a). \tag{2.13}$$

При $s=1$ имеем решение $x_j=1$. Ясно, что если $s\neq 1$, то $x_a=0$, так что $s=1$ — единственное значение, при котором все x_j отличны от нуля. Если s имеет любое другое значение, при котором (2.13) имеет решение, то должен существовать такой индекс $r<a$, что $x_{r+1}=0$, но $x_r\neq 0$; по (2.13) отсюда следует, что $sr=a$. Таким образом, корни s_r, при которых (2.13) имеет решения, суть $s_r=a/r$, $r=1,\ 2,\ \ldots,\ a$. Соответствующие решения (2.13) мы получим последовательно, беря $x_0^{(r)}=1$ и $j=0,\ 1,\ \ldots$. Получаем

$$x_j^{(r)}=\binom{r}{j}:\binom{a}{j}, \tag{2.14}$$

так что $x_j^{(r)}=0$, когда $j>r$.

При $s=s_r$ система (1.7б) имеет вид

$$(r-j)\,y_j^{(r)}=(a-j+1)\,y_{j-1}^{(r)} \tag{2.15}$$

и имеет решение

$$y_j^{(r)}=\binom{a-r}{j-r}(-1)^{j-r}, \tag{2.16}$$

где, конечно, $y_j^{(r)}=0$, если $j<r$. Так как $x_j^{(r)}=0$ для $j>r$, и $y_j^{(r)}=0$ для $j<r$, (из 1.11) получаем $c_r=\left(x_r^{(r)}y_r^{(r)}\right)^{-1}=\binom{a}{r}$ и, следовательно,

$$p_{jk}^{(n)}=\sum_{r=j}^{k}\left(\frac{r}{a}\right)^n\binom{a}{r}\binom{r}{j}\binom{a-r}{k-r}(-1)^{k-r}:\binom{a}{j}. \tag{2.17}$$

Если выразить биномиальные коэффициенты через факториалы, то формула упрощается и принимает вид

$$p_{jk}^{(n)}=\binom{a-j}{a-k}\sum_{\nu=0}^{k-j}\left(\frac{\nu+j}{a}\right)^n(-1)^{k-j-\nu}\binom{k-j}{\nu}, \tag{2.18}$$

где $p_{jk}^{(n)}=0$, если $k<j$.

(Дальнейшие примеры содержатся в двух следующих параграфах.)

§ 3. Случайное блуждание с отражающими экранами

Применение цепей[1]) Маркова будет теперь проиллюстрировано полным рассмотрением случайного блуждания с состояниями $1,\ 2,\ \ldots,\ a$

[1]) Часть последующих рассмотрений повторяет теорию гл. XIV. Наше квадратное уравнение встречалось там в (4.7); величины $\lambda_1(s)$ и $\lambda_2(s)$ были даны в (4.8), а общее решение (3.3) появлялось в гл. XIV в (4.9). Оба метода связаны друг с другом, хотя во многих случаях детали вычислений отличаются коренным образом.

и двумя отражающими экранами. Строки с номерами 2, 3, ..., $a-1$ матрицы P определяются соотношениями $P_{k,k+1}=p$ и $P_{k,k-1}=q$; первая и последняя строки равны $(q, p, 0, \ldots, 0)$ и $(0, \ldots, 0, q, p)$. Матрица примера 2, в гл. XV сводится к этой при $\delta=1$. В терминологии случайных блужданий $p_{kj}^{(n)}$ есть вероятность того, что частица, отправившаяся из точки $x=j$, находится в момент n в точке $x=k$.

Уравнения (1.7а) для $j=2, 3, \ldots, a-1$ принимают вид

$$x_1=s(qx_1+px_2); \quad x_j=s(qx_{j-1}+px_{j+1}) \quad x_a=s(qx_{a-1}+px_a). \tag{3.1}$$

Эта система имеет решение $x_j\equiv 1$, соответствующее корню $s=1$. Чтобы найти все другие решения, применим метод частных решений (который мы использовали для аналогичных уравнений в § 4 гл. XIV). Уравнения (3.1) с $j=2, 3, \ldots, a-1$ удовлетворяются при $x_j=\lambda^j$, где λ — корень квадратного уравнения $\lambda=$ $=qs+\lambda^2ps$. Корни этого уравнения имеют вид

$$\lambda_1(s)=\frac{1+(1-4pqs^2)^{1/2}}{2ps}, \quad \lambda_2(s)=\frac{1-(1-4pqs^2)^{1/2}}{2ps}, \tag{3.2}$$

и общее решение уравнений (3.1) с $j=2, 3, \ldots, a-1$ может быть записано в виде

$$x_j=A(s)\lambda_1^j(s)+B(s)\lambda_2^j(s), \tag{3.3}$$

где $A(s)$ и $B(s)$ произвольны. Выражение (3.3) будет решением первого и последнего уравнений (3.1) в том и только том случае, когда $x_0=x_1$, и $x_a=x_{a+1}$. Для этого необходимо, чтобы $A(s)$ и $B(s)$ удовлетворяли уравнениям

$$\begin{aligned} A(s)\{1-\lambda_1(s)\}+B(s)\{1-\lambda_2(s)\}&=0; \\ A(s)\lambda_1^a(s)\{1-\lambda_1(s)\}+B(s)\lambda_2^a(s)\{1-\lambda_2(s)\}&=0. \end{aligned} \tag{3.4}$$

Однако эти два уравнения совместны лишь тогда, когда

$$\lambda_1^a(s)=\lambda_2^a(s), \tag{3.5}$$

и мы должны определить значения s, при которых возможно (3.5).

Из (3.2) имеем $\lambda_1(s)\lambda_2(s)=q/p$, а из (3.5) следует, что $\lambda_1(s)(p/q)^{1/2}$ и $\lambda_2(s)(p/q)^{1/2}$ должны быть корнями степени $2a$ из единицы. Эти корни могут быть записаны в виде

$$e^{i\pi r/a}=\cos(\pi r/a)+i\sin(\pi r/a), \tag{3.6}$$

где $i^2=-1$ и $r=0, 1, 2, \ldots, 2a-1$. Таким образом, все решения (3.5) содержатся среди значений

$$\lambda_1(s)\left(\frac{q}{p}\right)^{1/2}e^{i\pi r/a}; \quad \lambda_2(s)=\left(\frac{q}{p}\right)^{1/2}e^{-i\pi r/a}, \tag{3.7}$$

Значение $r = a$ должно быть отброшено, так как для него $\lambda_1(s) = \lambda_2(s)$, $A(s) = -B(s)$, так что оно приводит только к тривиальному решению $x_j \equiv 0$. Значению $r = 0$ соответствует решение $x_j \equiv 1$, которое мы уже рассмотрели. Значениям $r = 1, 2, \ldots, a-1$ соответствует $a-1$ различных решений; если мы положим $r = a+1$, $a+2, \ldots, 2a-1$, то получим те же решения с переставленными $\lambda_1(s)$ и $\lambda_2(s)$. Таким образом, мы нашли a различных систем решений (3.1), а мы знаем, что больше их и не может быть. Для каждого значения r мы можем найти из (3.7) корень s_r; он равен $s_r = \{2(pq)^{1/2}\cos \pi r/a\}^{-1}$. Для $r = 1, 2, \ldots, a-1$ мы получаем значения $\lambda_1(s)$ и $\lambda_2(s)$ из (3.7), и тогда по (3.4) $2A(s) = 1 - \lambda_2(s)$ и $2B(s) = -\{1 - \lambda_1(s)\}$. (Вспомним, что постоянный множитель остается произвольным.) Подставляя $A(s)$ и $B(s)$ в (3.3), находим $a-1$ систем решений

$$x_j^{(r)} = \left(\frac{q}{p}\right)^{j/2} \sin\frac{\pi r j}{a} - \left(\frac{q}{p}\right)^{(j+1)/2} \sin\frac{\pi r (j-1)}{a} \tag{3.8}$$

($r = 1, 2, \ldots, a-1$). К ним мы прибавим ранее найденное решение

$$x_j^{(0)} = 1. \tag{3.9}$$

Легко проверить, что (3.8) и (3.9) представляют собой решения заданной системы (3.1).

Мы должны теперь найти решения второй системы линейных уравнений. В нашем случае (1.7б) принимает вид

$$\begin{aligned} &y_1 = sq(y_1 + y_2); \\ &y_k = s(py_{k-1} + qy_{k+1}) \qquad (k = 2, \ldots, a-1); \\ &y_a = sp(y_{a-1} + y_a). \end{aligned} \tag{3.10}$$

Средние уравнения таковы же, как и в (3.1), но с переставленными p и q. Поэтому их общее решение получается из (3.3) простой перестановкой p и q. Первое и последнее уравнения удовлетворяются, если выполнено (3.7), и простое вычисление показывает, что при $r = 1, 2, \ldots, a-1$ решения (3.10) имеют вид

$$y_k^{(r)} = \left(\frac{p}{q}\right)^{k/2} \sin\frac{\pi r k}{a} - \left(\frac{p}{q}\right)^{(k-1)/2} \sin\frac{\pi r (k-1)}{a}. \tag{3.11}$$

При $s = 1$ аналогичным образом получаем

$$y_k^{(0)} = \left(\frac{p}{q}\right)^{k}. \tag{3.12}$$

Последний шаг состоит в вычислении коэффициентов c_r по (1.11). Сумма упростится, если $\sin^2 \pi r j/a$ выразить через косинусы двойного угла, а косинусы записать в комплексной форме. Тогда мы получим сумму конечной геометрической прогрессии и находим

$$c_r = \frac{2p}{a}\left\{1 - 2(pq)^{1/2}\cos\frac{\pi r}{a}\right\}^{-1} \quad (r = 1, 2, \ldots, a-1). \quad (3.13)$$

При $r = 0$ получаем

$$c_0 = \frac{q}{p}\frac{(p/q)-1}{(p/q)^a - 1}, \quad (3.14)$$

в предположении, что $p \neq q$. Если $p = q = {}^1/_2$, то (3.13) остается верным, но (3.14) заменяется на $c_0 = 1/a$. Эти формулы приводят нас к *окончательному результату*:

$$p_{jk}^{(n)} = \frac{(p/q)-1}{(p/q)^a - 1}\left(\frac{p}{q}\right)^{k-1} + 2^{n+1}p^{1+(n-j+k)/2}q^{(n+j-k)/2}a^{-1}\sum_{r=1}^{a-1} S_r, \quad (3.15)$$

где S_r имеет вид

$$\frac{\cos^n\frac{\pi r}{a}\left\{\sin\frac{\pi rj}{a} - \left(\frac{q}{p}\right)^{1/2}\sin\frac{\pi r(j-1)}{a}\right\}\left\{\sin\frac{\pi rk}{a} - \left(\frac{q}{p}\right)^{1/2}\sin\frac{\pi r(k-1)}{a}\right\}}{1 - 2(pq)^{1/2}\cos\frac{\pi r}{a}}.$$

При $n \to \infty$ второе слагаемое в (3.15) стремится к нулю, и мы убеждаемся, что $p_{jk}^{(n)}$ стремятся к стационарному распределению, не зависящему от j. (Это предельное распределение было выведено другими методами в задаче 9 гл. XV.) Переходя к пределу при $a \to \infty$, получаем формулу для случайного блуждания с единственным отражающим экраном: в пределе сумма заменяется на интеграл[1]).

§ 4. Невозвратные состояния; вероятности поглощения

Теорема § 1 была выведена в предположении, что корни s_1, s_2, ... различны. Наличие кратных корней не вносит существенных изменений; поэтому мы рассмотрим только один особо важный частный случай. Корень $s_1 = 1$ будет кратным, если цепь содержит две или больше замкнутых подцепи; в задачах, связанных с вероятностями поглощения, эта ситуация обычна. Легко видоизменить метод § 1 так, чтобы он был пригоден и в этом случае. Для краткости

[1]) Аналогичные формулы в случае одного отражающего и одного поглощающего экранов см. M. Кац, Random walk and the theory of Brownian motion, *American Mathematical Monthly*, 54 (1947), 369—391. Определение отражающего экрана изменено там так, что частица может достигнуть $x = 0$; когда это происходит, следующий шаг переводит частицу в $x = 1$. Точные формулы оказываются тогда более сложными. Кац нашел также формулы для $p_k^{(n)}$ в модели Эренфестов (пример 2, е гл. XV).

к ясности мы проведем эти рассуждения на нескольких примерах, в которых обнаружатся также и основные черты общего случая.

Примеры. а) Рассмотрим матрицу вероятностей перехода

$$P = \begin{Vmatrix} 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/5 & 4/5 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/4 & 1/4 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{Vmatrix}. \qquad (4.1)$$

Ясно, что E_1 и E_2 образуют замкнутое множество (т. е. из них невозможны переходы в остальные четыре состояния; см. § 4 гл. XV). Аналогично E_3 и E_4 образуют другое замкнутое множество. Наконец, E_5 и E_6 суть невозвратные состояния. Через конечное число шагов система перейдет в одно из двух замкнутых множеств и там и останется.

Матрица P имеет вид

$$P = \begin{Vmatrix} A & 0 & 0 \\ 0 & B & 0 \\ U & V & T \end{Vmatrix}, \qquad (4.2)$$

где каждая буква заменяет квадратную матрицу второго порядка, а каждый нуль — матрицу из четырех нулей. Например, A имеет строки $(1/3,\ 2/3)$ и $(2/3,\ 1/3)$; A есть матрица вероятностей перехода, соответствующая цепи, образованной двумя состояниями E_1 и E_2. Эта матрица может быть изучена отдельно, и степени A^n могут быть получены из примера 2,а при $p = \alpha = 2/3$. Если вычислить степени P^2, P^3, ..., то окажется, что первые две строки совершенно не зависят от остальных строк. Точнее, P^n имеет вид

$$P^n = \begin{Vmatrix} A^n & 0 & 0 \\ 0 & B^n & 0 \\ U_n & V_n & T^n \end{Vmatrix}, \qquad (4.3)$$

где A^n, B^n и T^n суть n-е степени A, B и T соответственно и могут быть вычислены[1]) методами § 1 (см. пример 2,а, где проделаны все вычисления). Вместо шести уравнений с шестью неизвестными

---

[1]) В матрице T суммы элементов строк не равны единице, так что T не является стохастической матрицей. Однако методы § 1 применимы в этом случае без изменений, за исключением того, что $s = 1$ более не является корнем (так что $T^n \to 0$).

нам предстоит иметь дело только с системами из двух уравнений с двумя неизвестными.

Отметим, что матрицы U_n и V_n в (4.3) не являются степенями U и V и не могут быть получены теми же простыми методами, какими мы находим A^n, B^n и T^n. Однако при вычислении P^2, P^3, ..., третий и четвертый столбцы не влияют на остальные четыре столбца. Другими словами, если в P^n строки и столбцы, соответствующие E_3 и E_4, вычеркнуть, то мы получим матрицу

$$\left\| \begin{matrix} A^n & 0 \\ U_n & T^n \end{matrix} \right\|, \tag{4.4}$$

которая является n-й степенью соответствующей части матрицы P, а именно, n-й степенью матрицы

$$\left\| \begin{matrix} A & 0 \\ U & T \end{matrix} \right\| = \left\| \begin{matrix} 1/3 & 2/3 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 \\ 1/6 & 1/6 & 1/6 & 1/6 \end{matrix} \right\|. \tag{4.5}$$

Следовательно, (4.4) может быть вычислено методами § 1, которые в нашем случае значительно упрощаются. Матрица V_n может быть вычислена аналогичным способом.

Обычно точные формулы для U_n и V_n интересны постольку, поскольку они связаны с *вероятностями поглощения*. Пусть система отправляется, скажем, из E_5. Чему равна тогда *вероятность* λ *того, что система попадет в конечном счете в замкнутое множество, образованное* E_1 и E_2 (а не в другое замкнутое множество)? Чему равна *вероятность* λ_n *того, что это осуществится точно на n-м шаге*? Ясно, что $p_{51}^{(n)} + p_{52}^{(n)}$ есть вероятность того, что рассматриваемое событие осуществится на n-м шаге или раньше, т. е.

$$p_{51}^{(n)} + p_{52}^{(n)} = \lambda_1 + \lambda_2 + \ldots + \lambda_n.$$

Полагая $n \to \infty$, мы получим λ. Наилучший способ вычисления λ_n заключается в следующем. На $(n-1)$-м шаге система должна попасть в состояние, отличное от E_1 и E_2, т. е. в E_5 или в E_6 (так как из E_3 и E_4 переход в E_1 и E_2 невозможен). Но n-й шаг должен перевести систему в E_1 или E_2. Следовательно,

$$\lambda_n = p_{55}^{(n-1)}(p_{51} + p_{52}) + p_{56}^{(n-1)}(p_{61} + p_{62}) = {}^1/_4 p_{55}^{(n-1)} + {}^1/_3 p_{56}^{(n-1)}. \tag{4.6}$$

Отметим, что λ_n полностью определяется элементами T^{n-1}, а эту матрицу легко вычислить. В нашем случае $p_{55}^{(n)} = p_{56}^{(n)} = {}^1/_4({}^5/_{12})^{n-1}$ и, следовательно, $\lambda_n = {}^7/_{48}({}^5/_{12})^{n-2}$.

б) *Скрещивание прямых потомков.* В качестве второго примера дадим исчерпывающий анализ примера 2, м гл. XV.

Взгляд на соответствующую матрицу показывает, что каждое из состояний E_1 и E_5 образует замкнутое множество. В каком бы состоянии ни находилась система в начальный момент, она неизбежно попадет или в E_1, или в E_5 и в дальнейшем там останется. Натуралиста интересуют соответствующие вероятности и ожидаемая продолжительность процесса. Вычеркнув первые и пятые строки и столбцы, мы получим усеченную матрицу:

$$T=\begin{Vmatrix} {}^1/_2 & {}^1/_4 & 0 & 0 \\ {}^1/_4 & {}^1/_4 & {}^1/_4 & {}^1/_8 \\ 0 & {}^1/_4 & {}^1/_2 & 0 \\ 0 & 1 & 0 & 0 \end{Vmatrix}. \qquad (4.7)$$

Степени T^n могут быть вычислены методами § 1. Они представляют собой вероятности переходов между невозвратными состояниями.

Уравнения (1.7а) принимают вид

$$\begin{aligned} x_1 &= \frac{s(2x_1+x_2)}{4}; \qquad x_2=\frac{s(2x_1+2x_2+2x_3+x_4)}{8}; \\ x_3 &= \frac{s(x_2+2x_3)}{4}; \qquad x_4=sx_2. \end{aligned} \qquad (4.8)$$

Система (4.8) имеет решение лишь тогда, когда ее детерминант обращается в нуль, и это условие приводит к уравнению четвертой степени относительно s. Для упрощения записи положим

$$\theta_1=5^{1/2}-1, \qquad \theta_2=5^{1/2}+1. \qquad (4.9)$$

Тогда четыре корня s_r окажутся равными

$$s_1=2, \quad s_2=4, \quad s_3=\theta_1, \quad s_4=-\theta_2, \qquad (4.10)$$

и соответствующие решения $\left(x_1^{(r)}, \ldots, x_4^{(r)}\right)$ уравнений (4.8) имеют вид

$$(1,\ 0,\ -1,\ 0),\ (1,\ -1,\ 1,\ -4),\ \left(1,\ \theta_1,\ 1,\ \theta_1^2\right),\ \left(1,\ -\theta_2,\ 1,\ \theta_2^2\right). \qquad (4.11)$$

Система линейных уравнений для $y_k^{(r)}$ получается специализацией уравнений (1.7б), и четыре решения в соответствующем порядке окажутся равными

$$\begin{aligned} &(1,\ 0,\ -1,\ 0); \quad (1,\ -1,\ 1,\ -{}^1/_2); \\ &\left(1,\ \theta_1,\ 1,\ \theta_1^2/8\right); \quad \left(1,\ -\theta_2,\ 1,\ \theta_2^2/8\right). \end{aligned} \qquad (4.12)$$

Из (1.11) находим четыре постоянные: $c_1={}^1/_2$, $c_2={}^1/_5$, $c_3=\theta_2^2/40$, $c_4=\theta_1^2/40$. Из (1.8) получаем $\rho_{jk}^{(r)}$, и, наконец, (1.3) дает нам $p_{jk}^{(n)}$ для всех невозвратных состояний, т. е. при j, $k=2$, 3, 4, 6. При фиксированных j, k последовательность $p_{jk}^{(n)}$ есть сумма четырех геометрических прогрессий со знаменателями $s_1, \ldots, s_4$.

Поглощение в E_1 точно на n-м шаге возможно, лишь если $(n-1)$-й шаг переводит систему или в E_2, или в E_3, а n-й шаг — в E_1. Вероятность этого равна $\left(p_{j2}^{(n-1)}/4\right)+\left(p_{j3}^{(n-1)}/16\right)$. Аналогично вероятность поглощения в E_5 равна $\left(p_{j3}^{(n-1)}/16\right)+\left(p_{j4}^{(n-1)}/4\right)$. Суммируя по всем n, мы получим вероятности того, что система в конце концов попадет и останется соответственно в E_1 и E_5. Фактическое вычисление этих вероятностей сводится к суммированию четырех геометрических прогрессий.

§ 5. Приложение к времени возвращения

В задаче 19 гл. XIII показано, как можно вычислить математическое ожидание μ и дисперсию σ^2 времени возвращения для рекуррентного события $\mathfrak{E}$ через вероятности u_n того, что $\mathfrak{E}$ осуществится на n-м испытании. Если $\mathfrak{E}$ не является периодическим, то

$$u_n \to \frac{1}{\mu} \quad \text{и} \quad \sum_{n=0}^{\infty}\left(u_n - \frac{1}{\mu}\right) = \frac{\sigma^2 - \mu + \mu^2}{2\mu^2} \tag{5.1}$$

в предположении, что σ^2 конечно.

Если в качестве события $\mathfrak{E}$ рассмотреть попадание в возвратное состояние E_j (предполагая, что система находится в этом состоянии в начальный момент), то $u_n = p_{jj}^{(n)}$ $\left(\text{и } u_0 = 1\right)$. В конечной цепи Маркова время возвращения всегда имеет конечную дисперсию (см. задачу 19 гл. XV), так что (5.1) применимо. Предположим, что состояние E_j непериодическое и что применима формула (1.3). Тогда $s_1 = 1$ и $|s_r| > 1$ для $r = 2, 3, \ldots$, так что $p_{jj}^{(n)} \to \rho_{jj}^{(1)} \to 1/\mu_j$. Слагаемому $u_n - 1/\mu$ равенства (5.1) соответствует

$$p_{jj}^{(n)} - \frac{1}{\mu_j} = \sum_{r=2}^{a} \rho_{jj}^{(r)} s_r^{-n}. \tag{5.2}$$

Эта формула верна при $n \geqslant 1$. Суммируя геометрическую прогрессию со знаменателем s_r^{-1}, находим

$$\sum_{n=1}^{\infty}\left(p_{jj}^{(n)} - \frac{1}{\mu_j}\right) = \sum_{r=2}^{a} \frac{\rho_{jj}^{(r)}}{s_r - 1}. \tag{5.3}$$

Подставив это выражение в (5.1), убеждаемся, что *если E_j непериодическое возвратное состояние, то математическое ожидание* *времени возвращения в это состояние равно $\mu_j = 1/\rho_{jj}^{(1)}$, а дисперсия этого времени возвращения равна*

$$\sigma_j^2 = \mu_j - \mu_j^2 + 2\mu_j^2 \sum_{r=2}^{a} \frac{\rho_{jj}^{(r)}}{s_r - 1} \tag{5.4}$$

в предположении, конечно, что применима формула (1.3) и $s_1 = 1$. Случай периодического возвратного события и появление кратных корней вносит очевидные видоизменения.

ГЛАВА XVII

ПРОСТЕЙШИЕ СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ[1])

§ 1. Общие понятия

Случайные блуждания и цепи Маркова являются примерами стохастических процессов[2]), в которых изменения могут происходить только в фиксированные моменты времени, скажем, в моменты $t=1, 2, 3, \ldots$. С другой стороны, в § 5 и 6 гл. VI мы уже имели дело с такими явлениями, как телефонные вызовы и радиоактивный распад, которые могут произойти в любой момент. Понятно, что полное описание таких процессов выходит из границ области дискретных вероятностей. Для определенности рассмотрим вызовы, поступающие на телефонную станцию (точнее, рассмотрим идеализированную математическую модель такого явления). Каждому моменту t соответствует испытание, и результат эксперимента может быть описан с помощью функции $X(t)$, равной числу вызовов, поступивших к моменту времени t. Если первый вызов произошел в момент t_1, второй — в момент t_2 и т. д., то функция $X(t)$ равна 0 при $0<t<t_1$, 1 при $t_1<t<t_2$, 2 при $t_2<t<t_3$ и т. д. Обратно, каждая неубывающая функция $X(t)$, принимающая только значения $0, 1, 2, \ldots$, изображает возможное течение нашего процесса. Другими словами, полное описание нашего воображаемого эксперимента требует рассмотрения пространства элементарных событий, точками которого являются функции $X(t)$ (а не последовательности, как в случае дискретных испытаний). Далее, мы можем рассмотреть более сложные события, как, например, «семь вызовов за одну минуту в некоторый день»; это событие является, очевидно, совокупностью всех $X(t)$, удовлетворяющих тому условию, что для некоторой точки t, входящей в определенный интервал, мы имеем $X(t+h)-X(t)\geqslant 7$, где h — промежуток времени в одну минуту. Мы не можем здесь оперировать с таким сложным пространством элементарных событий и должны поэтому отложить изучение более тонких сторон теории до второй книги. К счастью, на некоторые интересные вопросы можно ответить и с помощью тех простых средств, которые имеются в нашем распоряжении.

Если ограничиться рассмотрением числа вызовов $X(t)$ за произвольный, но фиксированный промежуток времени t, то $X(t)$ будет

---

[1]) Эта глава почти независима от гл. X—XVI.

[2]) См. примечание 1 на стр. 398.

обычной случайной величиной, принимающей значения 0, 1, 2, ... Пусть $P_n(t)$ — вероятность того, что $X(t)=n$. Правда, распределение $\{P_n(t)\}$ зависит от промежутка времени t, т. е. от непрерывно меняющегося параметра. Однако многие из распределений вероятностей, уже введенных нами раньше, тоже зависели от параметра, и мы не имеем здесь существенно новых трудностей.

Положение лучше всего иллюстрируется распределением Пуассона

$$P_n(t)=e^{-\lambda t}\frac{(\lambda t)^n}{n!}. \tag{1.1}$$

Оно было выведено в § 5 гл. VI как предельная форма биномиального распределения; более удовлетворительный вывод содержится в § 3 гл. XII. Мы не будем пользоваться результатами этой главы, но ситуация, которая там анализируется, столь проста и столь типична, что короткое резюме может служить лучшим введением к настоящей главе.

Рассмотрим стохастический процесс, представленный целочисленной случайной величиной $X(t)\geqslant 0$. Интуитивно можно интерпретировать $X(t)$, скажем, как суммарный убыток от гроз, округленный с точностью до рублей. Мы придем к особенно простой математической модели, если введем два постулата следующим образом. Приращение $X(t+s)-X(0)$ на интервале времени от 0 до $t+s$ является суммой приращений $X(s)-X(0)$ и $X(t+s)-X(s)$, соответствующих интервалам от 0 до s и от s до $t+s$. Мы постулируем, во-первых, что приращения $X(s)-X(0)$ и $X(t+s)-X(s)$ независимы и, во-вторых, что распределение $X(t+s)-X(s)$ зависит только от t (т. е. только от длины интервала, но не от его положения на оси времени (однородность по времени).

Пусть $h_n(t)$ — вероятность того, что $X(t+s)-X(s)$ примет значение n (где $n=0, 1, 2, \ldots$). Аналитически независимость величин $X(t+s)-X(s)$ и $X(s)-X(0)$ выражается системой равенств

$$h_n(t+s)=\sum_{j=0}^{n} h_j(s)\cdot h_{n-j}(t). \tag{1.2}$$

В § 3 гл. XII было показано, что единственным распределением, которое удовлетворяет уравнениям (1.2), является *сложное распределение Пуассона*. Это означает, что $X(t)$ имеет то же распределение, что и случайная величина S_N, где

$$P\{N=n\}=e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \tag{1.3}$$

а $S_n=Y_1+Y_2+\ldots+Y_n$ — сумма n независимых величин, имеющих одно и то же распределение $\{f_i\}$, $i=0, 1, 2, \ldots$. В нашем примере $\{f_i\}$ можно рассматривать как распределение убытка при отдельном грозовом разряде; тогда формула (1.3) устанавливает, что число грозовых разрядов в интервале времени длины t подчиняется распределению Пуассона (1.1) и что отдельные убытки являются независимыми случайными величинами. Величина S_N имеет то же распределение, что и приращение $X(t+s)-X(s)$ на произвольном интервале длины t, и мы видим, что это полное приращение есть сумма случайного числа N отдельных приращений или *скачков*. Число N скачков имеет распределение Пуассона (1.1), а отдельный скачок имеет распределение вероятностей $\{f_i\}$. Распределение Пуассона (1.1) само представляет собой частный случай, когда все скачки имеют величину 1 (т. е. $f_1=1$, $f_0=f_2=f_3=\ldots=0$, переменные Y_n принимают только значения 1).

Заметим, что мы нашли характеристику простого и сложного распределений Пуассона с помощью внутренних вероятностных свойств. Распределение Пуассона теперь проявляется не как приближение и не как предельная форма для других распределений, а выступает совершенно самостоятельно (оно является, так сказать, выражением физического закона). Его вывод носит чисто аналитический характер, понятия случайного процесса и случайных величин $X(t)$ служат только для того, чтобы получить систему правдоподобных постулатов о распределении $\{h_n(t)\}$. Во многих приложениях не требуется ничего, кроме знания вероятностей $\{h_n(t)\}$. С теоретической точки зрения нужно еще показать, что $\{h_n(t)\}$ действительно определяют семейство случайных величин $X(t)$ и вероятности всех связанных с этими величинами событий, скажем, вероятность того, что $X(t)$ когда-нибудь станет больше, чем $at+b$ (эта задача о разорении важна для теории коллективного риска в страховом деле).

Вопросы такого типа выходят за рамки настоящей книги. Мы будем довольствоваться тем, что переведем физические допущения о процессе на язык основных вероятностей $P_n(t)$ и исследуем $\{P_n(t)\}$ как семейство дискретных вероятностных распределений, зависящих от t.

Искусственное условие ограничиваться лишь дискретными распределениями имеет свою оборотную сторону. Рассмотрим, например, нулевой член в (1.1). Мы интерпретировали

$$P_0(t)=e^{-\lambda t} \tag{1.4}$$

как вероятность того, что за рассматриваемый промежуток времени t не будет ни одного вызова. Эта формулировка подсказывает возможность интерпретировать $P_0(t)$ как вероятность того, что время, прошедшее с произвольного фиксированного момента вплоть до появления первого вызова, превышает t. Можно показать, что эта интерпретация вполне корректна; однако следует отметить, что она приводит к распределению вероятностей, заданному на континууме. Физический смысл нашей первой формулировки состоит в следующем: сделаем ряд «идентичных наблюдений» с фиксированным промежутком времени t. В результате каждого испытания может либо не произойти событие «ни одного вызова» (неудача), либо произойти событие «один или больше вызовов» (успех). Следовательно, мы имеем испытания Бернулли с вероятностью успеха $e^{-\lambda t}$. При второй интерпретации мы ожидаем, пока не произойдет вызов. Любое положительное число представляет собой возможное время ожидания, так что пространством элементарных событий, соответствующим каждому испытанию, является полупрямая $t > 0$. Формула (1.2) представляет собой тогда непрерывное распределение вероятностей и с этой точки зрения будет рассмотрена во второй книге (в результате чего возникает новый подход к распределению Пуассона).

§ 2. Распределение Пуассона

Мы приступаем к новому выводу распределения Пуассона. Этот вывод сам по себе нисколько не лучше описанного выше, но он более естественно приводит к различным обобщениям, которые мы собираемся изучать.

Рассмотрим систему, подверженную мгновенным изменениям. Эти изменения могут быть обусловлены случайными событиями, такими, как распад физической частицы, телефонный вызов, изменение хромосом под действием вредного излучения и т. п.

Все изменения подобны друг другу, и мы интересуемся только их общим числом. Каждое изменение отмечается точкой на временнóй оси, так что мы изучаем некоторое случайное распределение точек на действительной прямой.

Физические процессы, которые мы имеем в виду, характеризуются двумя свойствами: во-первых, они стационарны, а во-вторых, будущие изменения не зависят от изменений в прошлом. Под этим мы подразумеваем, что силы и влияния, которые определяют процесс, остаются неизменными, так что вероятность любого события одинакова для всех интервалов длины t, независимо от того, где эти интервалы расположены, и независимо от прошлой истории системы [1].

---

[1]) На телефонную станцию вызовы поступают во время рабочей части дня чаще, чем, скажем, между полуночью и часом ночи; процесс не является поэтому однородным во времени. Однако по понятным основаниям специалисты по телефонии интересуются в основном рабочей частью дня, а за этот период процесс можно рассматривать как однородный. Опыт показывает, что в течение рабочей части дня число поступивших вызовов с поразительной точностью соответствует распределению Пуассона. Аналогичные замечания применимы к автомобильным катастрофам, которые более часты по воскресеньям, и т. п.

Мы переведем теперь это описание на язык математики. Процесс характеризуется вероятностями $P_n(t)$ того, что в течение интервала времени длины t осуществится ровно n изменений [1]. В частности, $P_0(t)$ есть вероятность того, что не произойдет ни одного изменения, а $1 - P_0(t)$ — вероятность одного или большего числа изменений. Мы предположим [2], что при $t \to 0$

$$\frac{1 - P_0(t)}{t} \to \lambda, \tag{2.1}$$

где λ — положительная постоянная (производная величины $-P_0(t)$ при $t = 0$). Тогда для малого интервала длины h вероятность одного или большего числа изменений равна $1 - P_0(h) = \lambda h + o(h)$, где через $o(h)$ обозначена величина, убывающая быстрее, чем h. Мы можем теперь сформулировать следующие постулаты.

Постулаты процесса Пуассона: каково бы ни было число изменений в период времени $(0, t)$, *(условная) вероятность того, что в течение интервала времени* $(t, t + h)$ *произойдет изменение, равна* $\lambda h + o(h)$, *а вероятность того, что произойдет более чем одно изменение, есть* $o(h)$.

Из этих условий легко выводится система дифференциальных уравнений для $P_n(t)$. Рассмотрим два смежных интервала $(0, t)$ и $(t, t + h)$, где h мало. Если $n \geqslant 1$, то в интервале $(0, t + h)$ может произойти ровно n изменений тремя взаимно исключающими друг друга способами: 1) ни одного изменения за время $(t, t + h)$ и n изменений за время $(0, t)$; 2) одно изменение за время $(t, t + h)$ и $n - 1$ изменений за время $(0, t)$; 3) $x \geqslant 2$ изменений за время $(t, t + h)$ и $n - x$ изменений за время $(0, t)$. В соответствии с нашей гипотезой вероятность первой из возможностей равна произведению $P_n(t)$ на вероятность того, что в интервале $(t, t + h)$ не произойдет ни одного изменения; эта последняя вероятность равна $1 - \lambda h - o(h)$. Аналогично вторая возможность имеет вероятность $P_{n-1}(t) \lambda h + o(h)$, в то время как вероятность третьей возможности убывает быстрее, чем h. Это означает, что

$$P_n(t + h) = P_n(t)(1 - \lambda h) + P_{n-1}(t) \lambda h + o(h) \tag{2.2}$$

или

$$\frac{P_n(t + h) - P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(h)}{h}. \tag{2.3}$$

---

[1] Для неоднородных процессов мы должны ввести вероятности $P_n(t_1, t_2)$ того, что ровно n изменений произойдет в интервале времени $t_1 < t < t_2$.

[2] От этого условия мы в дальнейшем освободимся; см. § 6.

При $h \to 0$ последний член стремится к нулю; следовательно, предел[1]) левой части существует и равен

$$P'_n(t) = -\lambda P_n(t) + \lambda P_{n-1}(t) \quad (n \geqslant 1). \tag{2.4}$$

При $n = 0$ вторая и третья из вышеназванных возможностей не возникают, и поэтому (2.2) заменяется более простым уравнением:

$$P_0(t+h) = P_0(t)(1 - \lambda h) + o(h), \tag{2.5}$$

из которого вытекает

$$P'_0(t) = -\lambda P_0(t). \tag{2.6}$$

Из (2.6) и равенства $P_0(0) = 1$ мы получаем $P_0(t) = e^{-\lambda t}$. Подставляя это выражение в (2.4) при $n = 1$, мы получаем обыкновенное дифференциальное уравнение для $P_1(t)$. Так как $P_1(0) = 0$, мы находим легко, что $P_1(t) = \lambda t e^{-\lambda}$, что соответствует распределению Пуассона (1.1). Действуя таким же образом и дальше, мы найдем последовательно все члены (1.1).

§ 3. Процесс чистого размножения

В процессе Пуассона вероятность изменения за время $(t, t+h)$ не зависит от числа изменений за время $(0, t)$. Простейшее обобщение состоит в отказе от этого предположения. Предположим теперь, что если за время $(0, t)$ осуществилось n изменений, то вероятность нового изменения за время $(t, t+h)$ равна $\lambda_n h$ плюс слагаемое более высокого порядка малости по сравнению с h; вместо одной постоянной λ, характеризующей процесс, мы имеем последовательность постоянных $\lambda_0, \lambda_1, \lambda_2, \ldots$.

Удобно ввести более гибкую терминологию. Вместо того чтобы говорить, что n изменений произошли за время $(0, t)$, будем говорить, что система находится в состоянии E_n. Новое изменение вызывает тогда переход $E_n \to E_{n+1}$. *В процессе чистого размножения* переход из E_n возможен только в E_{n+1}. Такой процесс характеризуют следующие постулаты.

Постулаты. *Если в момент t система находится в состоянии E_n $(n = 0, 1, 2, \ldots)$, то вероятность того, что за время $(t, t+h)$ осуществится переход в E_{n+1}, равна $\lambda_n h + o(h)$. Вероятность иных изменений имеет более высокий порядок малости, чем h.*

---

[1]) Так как мы считаем h положительной величиной, то, строго говоря, $P'_n(t)$ в (2.4) следует рассматривать как правую производную. Но в действительности это обычная двусторонняя производная. В самом деле, член $o(h)$ в формуле (2.2) не зависит от t и потому не изменится, если t заменить на $t - h$. Тогда свойство (2.2) выражает непрерывность, а (2.3) дифференцируемость в обычном смысле. Это замечание применимо и в дальнейшем и не будет повторяться.

Отличительной чертой этого предположения является то, что время, которое система проводит в любом индивидуальном состоянии, не играет роли: как бы долго система ни оставалась в одном состоянии, внезапный переход в другое состояние остается одинаково возможным.

Пусть снова $P_n(t)$ — вероятность того, что в момент t система находится в состоянии E_n. Функции $P_n(t)$ удовлетворяют системе дифференциальных уравнений, которые могут быть выведены с помощью рассуждений предыдущего параграфа, с тем только изменением, что (2.2) заменяется на

$$P_n(t+h) = P_n(t)(1-\lambda_n h) + P_{n-1}(t)\lambda_{n-1}h + o(h). \qquad (3.1)$$

Таким образом, мы получаем *основную систему дифференциальных уравнений*:

$$\begin{aligned} P_n'(t) &= -\lambda_n P_n(t) + \lambda_{n-1}P_{n-1}(t) \qquad (n \geqslant 1), \\ P_0'(t) &= -\lambda_0 P_0(t). \end{aligned} \qquad (3.2)$$

Мы можем вычислить $P_0(t)$ и затем последовательно все $P_n(t)$. Если состояние системы представляет собой число изменений за время $(0, t)$, то начальным состоянием является E_0, так что $P_0(0) = 1$ и, следовательно, $P_0(t) = e^{-\lambda_0 t}$. Однако не обязательно, чтобы система исходила из состояния E_0 (см. пример 3, б). Если в момент 0 система находится в состоянии E_i, то

$$P_i(0) = 1, \quad P_n(0) = 0 \text{ для } n \neq i. \qquad (3.3)$$

Эти *начальные условия* единственным образом определяют решения $\{P_n(t)\}$ уравнений (3.2). (В частности, $P_0(t) = P_1(t) = \ldots$ $\ldots = P_{i-1}(t) = 0$.) Точные формулы для $P_n(t)$ были получены независимо многими авторами, но они для нас неинтересны. Легко проверить, что при любых λ_n решения $\{P_n(t)\}$ обладают всеми требуемыми свойствами, за исключением того, что при некоторых условиях может оказаться $\sum P_n(t) < 1$. Это явление будет изучено в § 4.

Примеры. а) *Радиоактивные превращения*. Радиоактивный атом может в результате испускания частиц или γ-лучей превратиться в атом другого вида. Каждый вид атома представляет собой возможное состояние системы, и если процесс продолжается и дальше, то мы получаем последовательность переходов $E_0 \to E_1 \to E_2 \to \ldots \to E_m$.

В соответствии с принятыми физическими теориями вероятность перехода $E_n \to E_{n+1}$ остается неизменной, как бы долго атом ни находился в состоянии E_n; эта гипотеза выражается нашим исходным предположением. Следовательно, этот процесс описывается дифференциальными уравнениями (3.2) (факт, хорошо известный физикам). Если E_n — конечное состояние, из которого невозможны дальнейшие переходы, то $\lambda_m = 0$, и система (3.2) обрывается при $n = m$. (При $n > m$ автоматически получается $P_n(t) = 0$).

б) *Процесс Юла.* Рассмотрим совокупность элементов, которые могут (путем деления или другим способом) порождать новые элементы, но не могут исчезать. Предположим, что за короткий промежуток времени h каждый элемент с вероятностью $\lambda h + o(h)$ производит новый элемент; постоянная λ определяет скорость разрастания нашей совокупности. Если не имеется никакого взаимодействия между элементами и в момент t объем совокупности равен n, то вероятность увеличения объема за время $(t, t+h)$ равна $n\lambda h + o(h)$. Вероятность $P_n(t)$ того, что объем совокупности равен ровно n, удовлетворяет, следовательно, уравнению (3.2) с $\lambda_n = n\lambda$, т. е.

$$P'_n(t) = -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t) \qquad (n \geqslant 1). \tag{3.4}$$

Если i — объем совокупности в момент $t = 0$, то применимы начальные условия (3.3). Легко найти решение, которое равно

$$P_n(t) = \binom{n-1}{n-i} e^{-i\lambda t}(1 - e^{-\lambda t})^{n-i} \tag{3.5}$$

при $n \geqslant i$, в то время как, конечно, $P_n(t) = 0$ при $n < i$.

Это распределение является частным случаем отрицательного биномиального распределения: используя определение (8.1) гл. VI, мы можем записать формулу (3.5) в виде $P_n(t) = f(n-i;\ i,\ e^{-\lambda t})$. Это означает (см. пример 3, в гл. IX), что объем совокупности в момент t есть сумма i независимых случайных величин, каждая из которых имеет распределение, получающееся из (3.5) заменой i на 1. Эти i величин представляют потомство каждой из i начальных частиц нашей совокупности.

Этот тип процесса был впервые изучен Юлом [1]) в связи с математической теорией эволюции. Популяция состоит из видов в пределах рода, и появление нового элемента обязано мутациям. Пред-

---

[1]) Yule G. U., A mathematical theory of evolution, based on the Conclusions of Dr. J. C. Willis, F. R. S., *Philosophical Transactions of the Royal Society, London*, серия B, **213** (1924), 21—87. Юл не вводил дифференциальных уравнений (3.4), а получал $P_n(t)$ предельным переходом, аналогичным тому, который был использован в § 5 гл. VI для процесса Пуассона.

Значительно более общая и более гибкая модель такого рода была построена и применена к анализу эпидемий и роста популяций в скромной, но в высшей степени интересной работе подполковника A.G. M'Kendrick, Applications of mathematics to medical problems, *Proceedings Edinburgh Mathematical Society*, **44** (1925), 1—34. К большому сожалению, эта замечательная работа осталась практически незамеченной. В частности, она была неизвестна автору настоящей книги, когда он посвятил различным моделям роста популяций статью Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung, *Acta Biotheoretica*, **5** (1939), 11—40.

положение о том, что каждый вид имеет одну и ту же вероятность произвести новый вид, не учитывает различия в объемах видов. Так как мы также пренебрегаем возможностью гибели данного вида, то следует ожидать, что формула (3.5) дает довольно грубое приближение к действительности.

Фарри[1]) использовал эту модель для описания процессов, связанных с космическими лучами, но приближение здесь также довольно грубое. Дифференциальные уравнения (3.4) применимы, строго говоря, только к совокупности частиц, которые могут делиться, образуя точные копии самих себя, в предположении, конечно, что не имеется взаимодействия между частицами.

§ 4 *). Расходящийся процесс размножения

Решение $\{P_n(t)\}$ бесконечной системы дифференциальных уравнений (3.2), соответствующее начальным условиям (3.3), может быть получено последовательно, начиная с $P_i(t) = e^{-\lambda_i t}$. Распределение $\{P_n(t)\}$, следовательно, определяется однозначно. Из обычных формул для решения линейных уравнений следует, что $P_n(t) \geqslant 0$. Остается открытым лишь вопрос, является ли $\{P_n(t)\}$ настоящим распределением вероятностей, т. е. выполняется ли условие

$$\sum P_n(t) = 1 \tag{4.1}$$

при всех t. Мы увидим, что это верно не всегда: если коэффициенты λ_n возрастают достаточно быстро, то может случиться, что

$$\sum P_n(t) < 1. \tag{4.2}$$

На первый взгляд эта возможность кажется удивительной; однако она имеет простое объяснение. Левую часть (4.2) можно интерпретировать как вероятность того, что в течение времени t произойдет только *конечное число* изменений. Согласно этому разность между правой и левой частями неравенства (4.2) соответствует возможности бесконечного числа скачков или своего рода взрыва. Для лучшего понимания этого явления сравним нашу вероятностную модель с тем, что получается при обычном детерминированном подходе.

Величина λ_n в (3.2) может быть названа средней скоростью роста в момент, когда объем совокупности равен n. Например, в частном случае (3.4) $\lambda_n = n\lambda$, так что средняя скорость роста пропорциональна объему имеющейся в данный момент совокупности. Если рост не

---

[1]) Furry, On fluctuation phenomena in the passage of high-energy electrons through lead, *Physical Reviews*, 52 (1937), 569.

*) См. примечание на стр. 388.

подвержен случайным колебаниям и скорость возрастания пропорциональна объему имеющейся в данный момент совокупности, то закон изменения объема совокупности $x(t)$ описывается дифференциальным уравнением

$$\frac{dx(t)}{dt} = \lambda x(t). \tag{4.3}$$

Отсюда следует, что в момент t объем совокупности равен

$$x(t) = ie^{\lambda t}, \tag{4.4}$$

где $i = x(0)$ — объем исходной совокупности. Связь между (3.4) и (4.3) не только формальна; легко видеть, что (4.4) дает на самом деле математическое ожидание распределения (3.5), так что (4.3) описывает средний объем совокупности, в то время как (3.4) отражает влияние случайных колебаний.

Рассмотрим теперь детерминированный процесс роста, в котором скорость роста увеличивается быстрее, чем объем совокупности. Если скорость роста пропорциональна $x^2(t)$, то мы получим дифференциальное уравнение

$$\frac{dx(t)}{dt} = \lambda x^2(t), \tag{4.5}$$

решение которого имеет вид

$$x(t) = \frac{i}{1 - \lambda it}. \tag{4.6}$$

Заметим, что если $t \to 1/\lambda i$, то $x(t)$ стремится к бесконечности. Другими словами, из предположения, что скорость роста возрастает пропорционально квадрату объема совокупности, вытекает, что совокупность за конечный промежуток времени достигнет бесконечного объема. Точный ответ на вопрос о том, при каких условиях происходит такой бесконечный рост, дает следующая теорема.

Теорема. *Для того чтобы условие* (4.1) *выполнялось при всех* t, *необходимо и достаточно, чтобы ряд*

$$\sum 1/\lambda_n \tag{4.7}$$

расходился.

Доказательство. Полагая

$$S_k(t) = P_0(t) + \ldots + P_k(t), \tag{4.8}$$

получим из (3.2)

$$S'_k(t) = -\lambda_k P_k(t), \tag{4.9}$$

и, следовательно, при $k \geqslant i$

$$1 - S_k(t) = \lambda_k \int_0^t P_k(\tau)\, d\tau. \tag{4.10}$$

Так как все члены (4.8) неотрицательны, последовательность $S_k(t)$ (при фиксированном t) может только возрастать при возрастании k, и, следовательно, правая часть (4.10) монотонно убывает. Обозначим ее предел через $\mu(t)$. Тогда при $k \geqslant i$

$$\lambda_k \int_0^t P_k(\tau)\,d\tau \geqslant \mu(t), \tag{4.11}$$

и, следовательно,

$$\int_0^t S_n(\tau)\,d\tau \geqslant \mu(t)\left(\frac{1}{\lambda_i}+\frac{1}{\lambda_{i+1}}+\ldots+\frac{1}{\lambda_n}\right). \tag{4.12}$$

Но из (4.10) мы имеем $S_n(t) \leqslant 1$, так что левая часть (4.12) не превосходит t. Если ряд (4.7) расходится, то второй множитель в правой части (4.12) стремится к бесконечности, и неравенство может быть выполнено лишь тогда, когда $\mu(t)=0$ при всех t. В этом случае правая часть равенства (4.10) стремится при $k \to \infty$ к 0 и, следовательно, $S_n(t) \to 1$, так что (4.1) выполняется.

Обратно[1]), интегрируя (4.8) и используя (4.10), нетрудно установить, что левая часть в (4.12) меньше, чем $\lambda_0^{-1}+\lambda_1^{-1}+\ldots+\lambda_n^{-1}$. Если ряд (4.7) сходится, то последнее выражение ограничено, и поэтому невозможно, чтобы $S_n(t) \to 1$ при всех t.

§ 5. Процесс размножения и гибели

Процесс чистого размножения, описанный в § 3, дает удовлетворительное описание радиоактивных превращений, но, очевидно, не может быть верной моделью изменений объема совокупности, члены которой могут погибать (или исчезать любым другим способом). Это наталкивает на мысль обобщить нашу модель, допустив переходы из E_n не только в ближайшее сверху состояние E_{n+1}, но и в ближайшее снизу состояние E_{n-1}. (Еще более общий процесс будет определен в § 9.) В соответствии с этим введем следующие

Постулаты. *Система изменяется, лишь переходя из некоторого состояния в ближайшие соседние (из E_n в E_{n+1} или E_{n-1}, если $n \geqslant 1$, а из E_0 только в E_1). Если в некоторый момент система находится в состоянии E_n, то вероятность того, что за время $(t, t+h)$ осуществится переход $E_n \to E_{n+1}$, равна*

[1]) Было обнаружено, что три последующие строки в первом выпуске первого издания оказались неверными. Вместо доказательства необходимости были частично повторены предшествующие рассуждения. Ошибка была исправлена через несколько месяцев. (Обсуждение результатов настоящего параграфа продолжается в § 10.)

$\lambda_n h + o(h)$, *а вероятность перехода* $E_n \to E_{n-1}$ (*при* $n \geqslant 1$) *равна* $\mu_n h + o(h)$. *Вероятность того, что за время* $(t, t+h)$ *осуществится более чем одно изменение, имеет порядок малости, более высокий, чем* h.

Легко приспособить метод § 2 к выводу дифференциальных уравнений для вероятностей $P_n(t)$ того, что в момент t система находится в состоянии E_n. Для вычисления $P_n(t+h)$ заметим, что система может находиться в момент $t+h$ в состоянии E_n тогда и только тогда, когда выполнено одно из следующих условий: 1) в момент t система находится в E_n и за время $(t, t+h)$ не происходит никаких изменений; 2) в момент t система находится в E_{n-1} и затем переходит в E_n; 3) в момент t система находится в E_{n+1} и затем переходит в E_n; 4) за время $(t, t+h)$ осуществляются два или более переходов. По предположению, вероятность последней ситуации стремится к нулю быстрее, чем h. Первые три возможности взаимно исключают друг друга, так что их вероятности складываются. Поэтому получаем

$$P_n(t+h) = P_n(t)\{1 - \lambda_n h - \mu_n h\} + \\ + \lambda_{n-1} h P_{n-1}(t) + \mu_{n+1} h P_{n+1}(t) + o(h). \quad (5.1)$$

Перенеся направо слагаемое $P_n(t)$ и разделив на h, получим в правой части отношение $P_n(t+h) - P_n(t)$ к h. Положив $h \to 0$, получим

$$P'_n(t) = -(\lambda_n + \mu_n) P_n(t) + \lambda_{n-1} P_{n-1}(t) + \mu_{n+1} P_{n+1}(t). \quad (5.2)$$

Эти уравнения выполняются при $n \geqslant 1$. При $n = 0$ аналогичным образом выводится уравнение

$$P'_0(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t). \quad (5.3)$$

Если в момент 0 система находится в состоянии E_i, то должны выполняться начальные условия:

$$P_i(0) = 1, \ P_n(0) = 0 \text{ при } n \neq i. \quad (5.4)$$

Таким образом, процесс гибели и размножения зависит от бесконечной системы дифференциальных уравнений (5.2) — (5.3) вместе с начальными условиями (5.4). Вопрос о существовании и единственности решения этой системы отнюдь не тривиален. В случае процесса чистого размножения система дифференциальных уравнений (3.2) была также бесконечной; однако она имела вид рекуррентных соотношений; $P_0(t)$ определялось первым уравнением и $P_n(t)$ можно было вычислить через $P_{n-1}(t)$. Новая система (5.2) имеет иную структуру, и все функции $P_n(t)$ должны находиться одновременно. Здесь (как и в других подобных случаях в этой главе) мы будем формулировать свойства решений без доказательства[1]).

Для произвольных коэффициентов $\lambda_n \geqslant 0$ *и* $\mu_n \geqslant 0$ *существует положительное решение системы* (5.2) — (5.4), *такое, что* $\sum P_n(t) \leqslant 1$. *Если коэффициенты ограничены* (*или возрастают достаточно медленно*), *это решение единственно и удовлетворяет условию регулярности* $\sum P_n(t) = 1$. Однако коэффициенты можно выбрать таким образом, чтобы $\sum P_n(t) < 1$ и чтобы существовало бесконечное число решений. В последнем случае мы сталкиваемся с явлениями, аналогичными тем, которые изучались в предыдущей главе для случая процессов чистого размножения. Эта ситуация представляет значительный теоретический интерес[2]), но читатель может без опасения считать, что во всех практически интересных случаях условия единственности выполняются и тогда автоматически $\sum P_n(t) = 1$ (см. § 10).

Если $\lambda_0 = 0$, то переход $E_0 \to E_1$ невозможен. В терминологии цепей Маркова E_0 — *поглощающее состояние*, уйти из которого невозможно; если система находится в E_0, она и останется там навсегда. Из (5.3) следует, что в этом случае $P_0'(t) \geqslant 0$, так что $P_0(t)$ монотонно возрастает. Предел $P_0(\infty)$ есть вероятность того, что в конце концов поглощение произойдет.

Вообще можно показать, что *пределы*

$$\lim_{t \to \infty} P_n(t) = p_n \tag{5.5}$$

---

[1]) Простое доказательство теоремы существования и критерия единственности, применимое к самым общим уравнениям настоящей главы (но требующее, впрочем, использования преобразования Лапласа), содержится в § 4 работы Феллера. On boundary conditions for the Kolmogorov differential equations, *Annals of Mathematics*, **65** (1957), 527—570. Первое доказательство теоремы существования было приведено в статье того же автора The integrodifferential equations of completely discontinuous Markov processes, *Transactions American Mathematical Society*, **48** (1940), 488 — 515. К сожалению, в последней работе рассматривается общий случай несчетных пространств элементарных событий и коэффициентов, зависящих от времени, но, по-видимому, специализация на случай обыкновенных дифференциальных уравнений с постоянными коэффициентами, которые изучаются в этой главе, приводит к простому доказательству теоремы существования.

[2]) Решения уравнений, связанных с процессами размножения и гибели, для которых $\sum P_n(t) < 1$, недавно привлекли большое внимание. См. W. Ledermann, G. E. Reuter, Spectral theory for the differential equations of simple birth and death processes, *Philosophical Transactions Royal Society*, London, Ser. A, **246** (1954), 321—369; S. Karlin and J. McGregor, Representation of a class of stochastic processes, *Proceedings National Academy Sciences*, USA (6) 4 (1955), 387 — 391; последующие работы тех же авторов, а также статью Купмана, см. *Transactions American Mathematical Society*.

существуют и не зависят от начальных условий (5.4); они удовлетворяют системе линейных уравнений, получаемых из (5.2) — (5.3), если положить в них $P_n'(t)=0$.

Обычно считают, что соотношение (5.5) выражает «тенденцию к устойчивому состоянию», и это соблазнительное наименование не раз приводило к недоразумениям. Следует понимать, что, за исключением того случая, когда E_0 — поглощающее состояние, случайные флуктуации вовсе не ослабляются и формула (5.5) показывает только то, что через долгое время влияние начальных условий исчезает. Замечание, сделанное в § 6 гл. XV относительно понятия статистического равновесия, применимо без изменений и в этом случае.

Справедливость соотношения (5.5) может быть доказана, исходя или из явных формул для $P_n(t)$, или общей эргодической теории. Интуитивно теорема становится почти очевидной после сравнения нашего процесса с простой цепью Маркова с вероятностями перехода

$$p_{n,\,n+1}=\frac{\lambda_n}{\lambda_n+\mu_n}; \quad p_{n,\,n-1}=\frac{\mu_n}{\lambda_n+\mu_n}. \tag{5.6}$$

Для этой цепи единственно возможными являются переходы $E_n\to E_{n+1}$ и $E_n\to E_{n-1}$, и они имеют те же условные вероятности, что и в нашем процессе. Разница между цепью и нашим процессом состоит только в том, что в нашем случае изменения могут осуществляться в любые моменты, так что число переходов за время t является случайной величиной. Однако при больших t это число будет, несомненно, большим, и, следовательно, весьма правдоподобно, что при $t\to\infty$ вероятности $P_n(t)$ ведут себя так же, как соответствующие вероятности для простой цепи.

Основной областью приложений процессов размножения и гибели являются задачи о времени ожидания, телефонные задачи и т. д. Такие приложения будут рассмотрены в § 6 и 7.

Пример. а) *Линейный рост.* Предположим, что наша совокупность состоит из элементов, которые могут делиться и могут погибать. За короткий интервал времени h для любого (не погибшего) элемента вероятность разделиться на два равна $\lambda h+o(h)$, а соответствующая вероятность погибнуть равна $\mu h+o(h)$. Здесь λ и μ — постоянные, характеризующие элементы. Если между элементами не имеется никакого взаимодействия, то мы приходим к процессу размножения и гибели с $\lambda_n=n\lambda$, $\mu_n=n\mu$. Основные дифференциальные уравнения принимают вид

$$P_0'(t)=\mu P_1(t), \tag{5.7}$$

$$P_n'(t)=-(\lambda+\mu)\,nP_n(t)+\lambda(n-1)\,P_{n-1}(t)+\mu(n+1)\,P_{n+1}(t)$$
$$(n=1,\ 2,\ \ldots).$$

Могут быть найдены явные решения[1]) этих уравнений (см. задачи 9—11), но мы этим заниматься не будем. Можно показать что пределы (5.5) существуют. Они, очевидно, удовлетворяют уравнениям (5.7) при $P'_n(t) = 0$. Из первого уравнения получаем $p_1 = 0$, и по индукции из второго уравнения следует, что $p_n = 0$ при всех $n \geqslant 1$. Если $p_0 = 1$, то можно сказать, что вероятность вымирания равна единице. Если $p_0 < 1$, то из соотношения $p_1 = p_2 = \ldots = 0$ вытекает, что с вероятностью $1 - p_0$ объем совокупности безгранично возрастает; в конце концов или должно произойти вымирание, или же совокупность должна бесконечно возрастать. Чтобы найти вероятность p_0 вымирания, мы должны сравнить наш процесс с соответствующей цепью Маркова. В нашем случае вероятности перехода (5.7) не зависят от n, и мы имеем поэтому обыкновенное случайное блуждание, в котором шаг направо и шаг налево имеют соответственно вероятности $p = \lambda/(\lambda + \mu)$ и $q = \mu/(\lambda + \mu)$. Состояние E_0 (или $x = 0$) — поглощающий экран. Мы знаем из классической задачи о разорении (§ 2 гл. XIV), что вероятность вымирания равна единице, если $p \leqslant q$, и $(q/p)^r$, если $q < p$, а r — исходное состояние. Мы заключаем, что *в нашем процессе вероятность* $p_0 = \lim P_0(t)$ *вымирания равна единице, если* $\lambda \leqslant \mu$, *и равна* $(\lambda/\mu)^r$, *если* $\lambda > \mu$. (Это нетрудно проверить с помощью явного решения; см. задачу 10.)

Во многих аналогичных случаях явные решения (5.8) довольно сложны, и желательно вычислить лишь математическое ожидание и дисперсию распределения $\{P_n(t)\}$. Запишем математическое ожидание

$$M(t) = \sum_{n=1}^{\infty} nP_n(t). \tag{5.8}$$

Мы опустим доказательство того, что $M(t)$ конечно и что следующие формальные операции законны (это легко проверить, исходя из явного вида $P_n(t)$, см. задачу 10). Умножив второе уравнение (5.8) на n и суммируя по $n = 1, 2, \ldots$, убеждаемся, что члены, содержащие n^2, сокращаются, и получаем

$$M'(t) = \lambda \sum (n-1) P_{n-1}(t) - \mu \sum (n+1) P_{n+1}(t) = (\lambda - \mu) M(t). \tag{5.9}$$

[1]) Основной метод состоит в выводе дифференциальных уравнений в частных производных для производящей функции $\sum P_n(t) s^n$. Более общий процесс [в котором коэффициенты λ и μ в (5.8) могут зависеть от времени] рассмотрен детально в статье David G. Kendall, The Generalized «Birth and Death» Process, *Annals of Mathematical Statistics*, 19 (1948), 1—15. См. также новую статью того же автора Stochastic Processes and Population Growth (появившуюся в 1950 г. в *Journal of the Royal statistical Society*), где теория обобщается так, чтобы она учитывала распределение возраста элементов в биологических популяциях.

Это — дифференциальное уравнение для $M(t)$. В момент $t=0$ объем совокупности равен i, и, следовательно, $M(0)=i$. Поэтому

$$M(t)=ie^{(\lambda-\mu)t}. \tag{5.10}$$

Мы видим, что математическое ожидание стремится к 0 или к бесконечности в зависимости от того, будет ли $\lambda<\mu$ или $\lambda>\mu$. Дисперсия $\{P_n(t)\}$ может быть вычислена аналогичным образом (см. задачу 12).

б) *Время ожидания в случае одной телефонной линии.* В простейшем частном случае постоянных коэффициентов $\lambda_n=\lambda$ и $\mu_n=\mu$ задача изучения процесса гибели и размножения сводится к специальному случаю (при $a=1$) задачи об очередях примера 7, б.

§ 6. Показательное время обслуживания

Основная область приложения процессов размножения и гибели связана с расчетами числа занятых телефонных линий и различных типов очередей для телефонов, машин и т. п. Этот тип задач может изучаться с различными степенями математической идеализации. Схема процессов размножения и гибели служит первым приближением. Эта модель основана на математическом упрощении, известном как *предположение о показательном времени обслуживания.* Мы начнем с обсуждения этого основного предположения.

Для конкретности будем рассматривать телефонные разговоры. Предположим, что их длина измеряется обязательно целым числом секунд. Будем рассматривать длительность разговора как случайную величину X и будем считать известным ее распределение вероятностей $p_n=P\{X=n\}$. Телефонная линия представляет собой тогда физическую систему с двумя возможными состояниями: «занято» (E_0) и «свободно» (E_1). Если в некоторый момент линия занята, то вероятность изменения состояния в течение следующей секунды зависит от того, как долго уже идет разговор. Другими словами, прошлое влияет на будущее, и наш процесс не является марковским (см. § 10 гл. XV). Это обстоятельство является источником большинства трудностей в более сложных задачах. Однако имеется один простой исключительный случай.

Представим, что вопрос о том, продолжать ли дальше разговор или нет, решается каждую секунду случайно при помощи бросания несимметричной монеты. Другими словами, производится последовательность испытаний Бернулли по одному испытанию в секунду до тех пор, пока не произойдет первый успех. Разговор оканчивается, когда осуществится первый успех. В этом случае общая длина разговора или, как мы будем говорить, «время обслуживания» имеет геометрическое распределение $p_n=q^{n-1}p$. Если в некоторый момент t линия занята, то вероятность того, что она останется еще занятой более чем одну секунду, равна q, а вероятность перехода $E_0 \to E_1$ за следующий шаг равна p; эти вероятности теперь не зависят от того, как долго была занята линия.

Если не считать временной параметр дискретным, то нам придется иметь дело с непрерывными случайными величинами. Роль геометрического распределения для времени ожидания играет в этом случае *показательное распределение*. Это единственное непрерывное распределение, обладающее марковским свойством, т. е. полным отсутствием последействия. Другими словами, вероятность того, что разговор, который происходит в момент t, продлится до момента $x+h$, не зависит от продолжительности предыдущего разговора тогда и только тогда, когда вероятность того, что продолжительность разговора больше, чем t, равна $e^{-\lambda t}$. «Показательное время обслуживания» уже появлялось у нас, как нулевой член распределения Пуассона, т. е. как время ожидания первого изменения.

Схема процесса размножения и гибели применима, лишь если все вероятности перехода не зависят от прошлого. Для задач о числе занятых линий и об очередях это означает, что время обслуживания должно быть показательным. С практической точки зрения это предположение может на первый взгляд показаться довольно искусственным, но опыт показывает, что оно удовлетворительно описывает действительные явления. В частности, многочисленные измерения показали, что телефонные разговоры внутри одного города[1]) удовлетворяют показательному закону с удивительной степенью точности.

Предыдущие замечания относятся к временам обслуживания (например, к длительности телефонного разговора, к продолжительности ремонта машины и т. п.). Мы должны теперь охарактеризовать так называемую приходящую нагрузку (поступающие вызовы, поломки машин и т. д.). Предположим, что вероятность поступления вызова за любой интервал времени длины h равна λh плюс слагаемое, которым можно пренебречь, причем вероятностью того, что произойдет два или больше вызовов, также можно в пределе пренебречь. Согласно результатам § 2, это означает, что число поступивших вызовов имеет распределение Пуассона с математическим ожиданием λt. Мы можем описать это положение, говоря, что *поступающая нагрузка имеет пуассоновский тип с интенсивностью* λ.

Нетрудно проверить описанное свойство показательного времени обслуживания. Обозначим через $u(t)$ вероятность того, что продолжительность разговора составляет по крайней мере t единиц времени. Вероятность

[1]) Для междугородных переговоров обычно устанавливается продолжительность разговора 3 минуты, и поэтому правдоподобно, что время обслуживания будет кратно 3 минутам. Это представляет собой систематическое отклонение от показательного закона, и наша теория здесь неприменима.

$u(t+s)$ того, что разговор, начавшийся в момент 0, окончится после момента $t+s$, равна вероятности того, что он окончится после момента t, умноженной на условную вероятность того, что он продлится еще s единиц времени, при условии, что его длина превосходит t. Если продолжительность предшествующего разговора не оказывает влияния, то последняя условная вероятность должна равняться $u(s)$, т. е. мы приходим к соотношению

$$u(t+s)=u(t)\,u(s). \tag{6.1}$$

Остается доказать следующую теорему.

Теорема. *Пусть функция $u(t)$ определена при $t>0$ и ограничена в каждом конечном интервале. Если $u(t)$ удовлетворяет уравнению* (6.1), *то или $u(t)=0$ при всех t, или найдется такая постоянная λ, что $u=e^{-\lambda t}$.*

Доказательство. Если функция $u(t)$ не равна нулю тождественно, то найдется такая точка x, что $u(x)>0$. Пусть $\lambda=-\log u(x)$ и $v(t)=e^{\lambda t}u(xt)$. Тогда

$$v(t+s)=v(t)\,v(s),\quad v(1)=1, \tag{6.2}$$

и мы покажем, что $v(t)=1$ при всех $t>0$. Очевидно, что $v^2\left(\frac{1}{2}\right)=v(1)=1$ и вообще $v^n\left(\frac{1}{n}\right)=v(1)=1$ для любого целого $n>0$. Поэтому $v\left(\frac{1}{n}\right)=1$ и, значит, $v\left(\frac{m}{n}\right)=v^m\left(\frac{1}{n}\right)=1$ для любой пары целых чисел $m>0$, $n>0$. Итак, $v(r)=1$ при любом рациональном r. Предположим, что существует точка τ, для которой $v(\tau)=c\neq 1$. Тогда $v(\tau^{-1})=c^{-1}$, и поэтому без ограничения общности можно считать, что $c>1$. В этом случае $v(N\tau)=v^N(\tau)=c^N$ может при соответствующем выборе N быть сделано сколь угодно большим. Выберем теперь рациональное r, лежащее в интервале

$$N\tau-1<r<N\tau.$$

Тогда

$$v(N\tau-r)=v(N\tau-r)\,v(r)=v(N\tau)=c^N.$$

Отсюда следует, что в интервале $0<a<1$ существуют точки $a=N\tau-r$, такие, что $v(a)>c^N$, а это противоречит предположению о том, что функция $u(t)$ [а значит, и $v(t)$] ограничена в каждом конечном интервале.

§ 7. Очереди и задачи обслуживания

а) Простейшая задача о телефонных линиях[1] [2]. Предположим, что имеется бесконечное число телефонных линий и что вероятность окончания разговора в течение времени $(t,\ t+h)$ равна μh плюс

[1] C. Palm, Intensitätsschwankungen im Fernsprechverkehr, *Ericsson Thechnics* (Stockholm), 44 (1943), 1—189, в частности, стр. 57. Очереди и задачи обслуживания для телефонных линий изучались задолго до того, как была создана общая теория случайных процессов, и это изучение оказало стимулирующее влияние на развитие теории. В частности, замечательные работы Пальма, спустя много лет после их опубликования, оказались полезными нескольким авторам. Самые первые работы в этой области принадлежат Эрлангу. См. книгу E. Brockmeyer, H. L. Halström, Arne Jensen, The Life and Works of A. K. Erlang, *Transactions of the Danish*

слагаемые, которыми при $h \to 0$ можно пренебречь (показательное время обслуживания). Поступающие вызовы образуют нагрузку пуассоновского типа с параметром λ. Система находится в состоянии E_n, если занято n линий.

Конечно, предполагается, что продолжительности разговоров взаимно независимы. Если занято n линий, то вероятность того, что одна из них освободится за промежуток времени h, равна $n\mu h + o(h)$. Вероятность того, что за это время окончатся два или больше разговоров, имеет, очевидно, порядок малости h^2, и ею можно пренебречь. Вероятность поступления нового вызова равна $\lambda h + o(h)$. Вероятность поступления за это время нескольких вызовов или поступления вызова и окончания разговора имеет порядок малости $o(h)$. Таким образом, в обозначениях § 5

$$\lambda_n = \lambda, \quad \mu_n = n\mu. \tag{7.1}$$

Основные дифференциальные уравнения (5.2), (5.3) имеют теперь вид

$$\begin{aligned} P_0'(t) &= -\lambda P_0(t) + \mu P_1(t), \\ P_n'(t) &= -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t) \quad (n \geqslant 1). \end{aligned} \tag{7.2}$$

Явное решение может быть получено выводом уравнения в частных производных для производящей функции (см. задачу 11). Мы определим сейчас только пределы (5.6). Они удовлетворяют уравнениям

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ (\lambda + n\mu) p_n &= \lambda p_{n-1} + (n+1)\mu p_{n+1}. \end{aligned} \tag{7.3}$$

Последовательно находим из уравнений, что $p_n = p_0 (\lambda/\mu)^n / n!$, и, следовательно,

$$p_n = e^{-\lambda/\mu} \frac{(\lambda/\mu)^n}{n!}. \tag{7.4}$$

Таким образом, *предельное распределение есть распределение Пуассона с параметром* λ/μ. *Оно не зависит от начального состояния.*

Легко найти математическое ожидание $M(t) = \sum n P_n(t)$. Умножив n-е уравнение (7.2) на n и суммируя по n, получим [учитывая, что сумма $P_n(t)$ равна единице]

$$M'(t) = \lambda - \mu M(t). \tag{7.5}$$

Academy Technical Sciences, № 2, Copenhagen, 1948. Ценная основополагающая работа была проделана независимо Фраем. Его книга, цитированная в примечании на стр. 155, много сделала для развития технических приложений теории вероятностей.

[2]) Для случая одной линии и произвольного закона распределения для времени обслуживания задача решена А. Я. Хинчиным в работе «Математическая теория стационарной очереди», *Математический сборник*, **39**:**4** (1932), 73—84. — *Прим. ред.*

Если начальным состоянием является E_i, то $M(0)=i$ и

$$M(t)=(\lambda/\mu)(1-e^{-\mu t})+ie^{-\mu t}. \tag{7.6}$$

Мы видим, что $M(t)$ при $t\to\infty$ приближается к математическому ожиданию распределения Пуассона, найденному выше. Читатель может проверить, что в частном случае $i=0$ вероятности $P_n(t)$ задают точно распределение Пуассона с математическим ожиданием $M(t)$.

б) Очереди в случае конечного числа телефонных линий[1]. Рассмотрим теперь последний пример в более реальной обстановке. Предположения будут теми же самыми, за исключением того, что *число a телефонных линий конечно. Если же все линии заняты, то каждый новый вызов становится в очередь и ожидает, пока освободится какая-нибудь линия.* Это значит, что все линии имеют *общую* очередь.

Слово «вызов» может быть заменено на «окошко» в почтовом отделении, а «разговор» — на «*обслуживание*». Мы, по существу, рассматриваем общую задачу об очереди в предположении, что ждать приходится только тогда, когда все a обслуживающих аппаратов заняты.

Мы говорим, что *система находится в состоянии E_n, если n — общее число лиц, которые обслуживаются или ожидают обслуживания.* Очередь имеется только тогда, когда система находится в состоянии E_n при $n>a$, и тогда в очереди стоит $n-a$ человек.

До тех пор пока хотя бы один аппарат свободен, мы имеем точно такое же положение, как и в предшествующем примере. Однако если система находится в состоянии E_n при $n>a$, то ведутся только a разговоров, и мы имеем, следовательно, $\mu_n=a\mu$ при $n\geqslant a$. Основная система дифференциальных уравнений будет иметь вид (7.2) лишь при $n\leqslant a$, но при $n\geqslant a$ уравнения имеют вид

$$P_n'(t)=-(\lambda+a\mu)P_n(t)+\lambda P_{n-1}(t)+a\mu P_{n+1}(t). \tag{7.7}$$

Мы снова исследуем пределы (5.6) (которые, как можно показать, существуют). Они должны удовлетворять уравнениям (7.3), если $n\leqslant a$, и уравнениям

$$(\lambda+a\mu)p_n=\lambda p_{n-1}+a\mu p_{n+1}, \tag{7.8}$$

если $n\geqslant a$. Снова последовательно находим, что при $n\leqslant a$

$$p_n=p_0\frac{(\lambda/\mu)^n}{n!}, \tag{7.9}$$

[1] А. Н. Колмогоров, О проблеме ожидания, *Математический сборник*, **38**: 1—2 (1931), 101—106.

в то время как при $n \geqslant a$

$$p_n = \frac{(\lambda/\mu)^n}{a!\,a^{n-a}}\,p_0. \tag{7.10}$$

Ряд $\sum (p_n/p_0)$ сходится, лишь если

$$\lambda/\mu < a. \tag{7.11}$$

Следовательно, если (7.11) не выполняется, то предельное распределение $\{p_k\}$ не существует. *В этом случае $p_n = 0$ при всех n; это означает, что очередь будет безгранично возрастать.* С другой стороны, если (7.11) выполняется, то можно определить p_0 так, чтобы сумма выражений (7.9) и (7.10) равнялась единице. Из явных выражений для $P_n(t)$ (которых мы, однако, выводить не будем) можно показать, что числа p_n, получаемые таким образом, действительно представляют собой *предельное распределение* для $P_n(t)$. Табл. 1 дает численную иллюстрацию для случая $a = 3$, $\lambda/\mu = 2$.

Таблица 1

Предельные вероятности в случае $a = 3$ линий и $\lambda/\mu = 2$

n	0	1	2	3	4	5	6	7
Число занятых линий	0	1	2	3	3	3	3	3
Число ожидающих линий	0	0	0	0	1	2	3	4
p_n	0,1111	0,2222	0,2222	0,1481	0,0988	0,0658	0,0439	0,0293

в) Обслуживание машин[1][2]. Результаты, которые мы выведем в этом и следующем параграфах, успешно применяются в шведской промышленности. Для ориентировки мы начнем с простейшего случая и затем обобщим его в следующем примере. Задача состоит в следующем.

Рассмотрим автомат, который при нормальной работе не требует вмешательства человека. Однако в любой момент автомат может разладиться и потребовать обслуживания. Время, необходимое для обслуживания машины, рассматривается снова как случайная величина

---

[1]) Примеры в) и г), включая численные иллюстрации, взяты из статьи C. Palm, Распределение числа рабочих, необходимых для обслуживания автоматов (по-шведски), *Industritidningen Norden*, **75** (1947), 75—80, 90—94, 119—123. Пальм приводит таблицы и графики для определения наиболее выгодного числа обслуживающих рабочих.

[2]) См. также А. Я. Хинчин, О среднем времени простоя станков, *Математический сборник*, **42** (1933), 119. — *Прим. ред.*

с показательным распределением. Другими словами, машина характеризуется двумя постоянными λ и μ, определяемыми следующим образом. Если в момент t автомат работает, то вероятность того, что он потребует обслуживания раньше момента $t+h$, равна λh плюс слагаемые, которыми можно пренебречь в пределе при $h \to 0$. Обратно, если в момент t автомат обслуживается, то вероятность того, что обслуживание закончится раньше, чем в момент $t+h$, и автомат начнет работать, равна $\mu h + o(h)$. Для надежного автомата λ будет относительно мало, а μ относительно велико. Отношение λ/μ называется *коэффициентом обслуживания*.

Предположим, что m автоматов с одинаковыми параметрами λ и μ обслуживаются одним рабочим. Если автомат испортился, он обслуживается немедленно, если только рабочий не ремонтирует в это время другой автомат; в этом последнем случае образуется очередь. Мы будем говорить, что система находится *в состоянии E_n, если не работают n автоматов*. При $1 \leqslant n \leqslant m$ это означает, что один автомат обслуживается, а $n-1$ стоят в очереди; в состоянии E_0 все автоматы работают, и обслуживающий рабочий отдыхает. Все m автоматов предполагаются работающими независимо друг от друга.

Переход $E_n \to E_{n+1}$ возникает при поломке одного из $m-n$ работающих автоматов, в то время как переход $E_n \to E_{n-1}$ осуществляется, если один из разладившихся автоматов вернулся в рабочее состояние. Следовательно, мы имеем процесс размножения и гибели с коэффициентами

$$\begin{aligned} \lambda_0 &= m\lambda, & \mu_0 &= 0, \\ \lambda_n &= (m-n)\lambda, & \mu_n &= \mu \end{aligned} \quad (0 < n \leqslant m) \tag{7.12}$$

и основные дифференциальные уравнения (5.2) и (5.3) принимают вид $(1 \leqslant n \leqslant m-1)$

$$\begin{aligned} P_0'(t) &= -m\lambda P_0(t) + \mu P_1(t), \\ P_n'(t) &= -\{(m-n)\lambda + \mu\} P_n(t) + \\ &\qquad + (m-n+1)\lambda P_{n-1}(t) + \mu P_{n+1}(t), \\ P_m'(t) &= -\mu P_m(t) + \lambda P_{m-1}(t). \end{aligned} \tag{7.13}$$

Это конечная система дифференциальных уравнений, которая может быть решена обычными методами. Пределы (5.6) существуют и удовлетворяют уравнениям

$$\begin{aligned} m\lambda p_0 &= \mu p_1, \\ \{(m-n)\lambda + \mu\} p_n &= (m-n+1)\lambda p_{n-1} + \mu p_{n+1}, \\ \mu p_m &= \lambda p_{m-1}. \end{aligned} \tag{7.14}$$

Отсюда легко вытекает рекуррентная формула

$$(m-n)\lambda p_n = \mu p_{n+1}. \tag{7.15}$$

Полагая последовательно $n=m-1,\ m-2,\ \ldots,\ 1,\ 0$, получаем

$$p_{m-k} = \frac{1}{k!}\left(\frac{\mu}{\lambda}\right)^k p_m. \tag{7.16}$$

Неизвестную постоянную p_m можно определить из условия, что сумма p_j равна единице. Легко видеть, что

$$p_m = \left\{1 + \frac{1}{1!}\left(\frac{\mu}{\lambda}\right)^1 + \ldots + \frac{1}{m!}\left(\frac{\mu}{\lambda}\right)^m\right\}^{-1}. \tag{7.16a}$$

Формула (7.16) хорошо известна инженерам-связистам как *формула Эрланга*. Типичные числовые значения приведены в табл 2.

Таблица 2

Вероятности p_n
для случая $\lambda/\mu = 0{,}1$, $m=6$

n	Число машин, ожидающих обслуживания	p_n/p_0	p_n
0	0	1	0,4845
1	0	0,6	0,2907
2	1	0,3	0,1454
3	2	0,12	0,0582
4	3	0,036	0,0175
5	4	0,0072	0,0035
6	5	0,00072	0,0003

$$1/p_0 = \sum p_n/p_0 = 2{,}06392$$

Вероятность p_0 может быть интерпретирована как вероятность того, что обслуживающий рабочий не занят (в примере табл. 2 он будет, вероятно, не занят почти половину всего времени). *Математическое ожидание числа машин, стоящих в очереди, равно*

$$w = \sum_{k=1}^{m}(k-1)\ p_k = \sum_{k=1}^{m} kp_k - (1-p_0). \tag{7.17}$$

Эта величина может быть найдена суммированием уравнений (7.15) по $n=0,\ 1,\ \ldots,\ m$. Используя тот факт, что p_n дают в сумме единицу, получаем

$$m\lambda - \lambda w - \lambda(1-p_0) = \mu(1-p_0)$$

или

$$w = m - \frac{\lambda+\mu}{\lambda}(1-p_0). \tag{7.18}$$

В примере табл. 2 имеем $w = 6 \cdot 0{,}0549$. Таким образом, отношение числа автоматов, ожидающих обслуживания, к общему числу автоматов имеет среднее значение 0,0549.

г) Продолжение. *Несколько обслуживающих рабочих.* Мы не изменим основных предположений предыдущей задачи, за исключением того, что *m автоматов обслуживаются теперь r рабочими* ($r < m$). Таким образом, при $n \leqslant r$ состояние E_n означает, что $r - n$ рабочих свободны, n автоматов обслуживаются, и ни один автомат не стоит в очереди. При $n > r$ состояние E_n означает, что r автоматов обслуживаются и $n - r$ автоматов ожидают обслуживания. Мы можем использовать результаты предыдущего примера, за исключением того, что (7.12) заменяется на

$$\begin{aligned} &\lambda_0 = m\lambda, && \mu_0 = 0; && \\ &\lambda_n = (m-n)\lambda, && \mu_n = n\mu && (1 \leqslant n \leqslant r); \\ &\lambda_n = (m-n)\lambda, && \mu_n = r\mu && (r \leqslant n \leqslant m). \end{aligned} \tag{7.19}$$

Не будем выписывать основной системы дифференциальных уравнений; напишем только уравнения для предельных вероятностей

$$\begin{aligned} &m\lambda p_0 = \mu p_1, \\ &\{(m-n)\lambda + n\mu\}\, p_n = (m-n+1)\lambda p_{n-1} + (n+1)\mu p_{n+1} \\ &\qquad\qquad (1 \leqslant n < r), \\ &\{(m-n)\lambda + r\mu\}\, p_n = (m-n+1)\lambda p_{n-1} + r\mu p_{n+1} \\ &\qquad\qquad (r \leqslant n \leqslant m). \end{aligned} \tag{7.20}$$

Из первого уравнения находим отношение p_1/p_0. Из второго получаем по индукции, что при $n < r$

$$(n+1)\mu p_{n+1} = (m-n)\lambda p_n; \tag{7.21}$$

наконец, при $n \geqslant r$ получаем из последнего уравнения (7.20)

$$r\mu p_{n+1} = (m-n)\lambda p_n. \tag{7.22}$$

Эти уравнения позволяют вычислить последовательно отношения p_n/p_0. Наконец, p_0 находится из условия $\sum p_k = 1$. Значения, указанные в табл. 3, получены именно таким способом.

Сравнение табл. 2 и 3 дает любопытные результаты. Отметим, что обе таблицы относятся к одним и тем же автоматам ($\lambda/\mu = 0{,}1$), но во втором случае мы имеем $m = 20$ автоматов и $r = 3$ рабочих. Число автоматов, приходящееся на одного рабочего, возросло с 6 до $6\frac{2}{3}$, но в то же время автоматы обслуживаются более эффективно. Определим *коэффициент простоя автоматов* формулой

$$\frac{w}{m} = \frac{\text{Среднее число автоматов в очереди}}{\text{Общее число автоматов}} \tag{7.23}$$

Таблица 3

Вероятности p_n
для случая $\lambda/\mu = 0,1$, $m = 20$, $r = 3$

n	Число обслуживаемых автоматов	Число ожидающих автоматов	Число незанятых рабочих	p_n
0	0	0	3	0,13625
1	1	0	2	0,27250
2	2	0	1	0,25888
3	3	0	0	0,15533
4	3	1	0	0,08802
5	3	2	0	0,04694
6	3	3	0	0,02347
7	3	4	0	0,01095
8	3	5	0	0,00475
9	3	6	0	0,00190
10	3	7	0	0,00070
11	3	8	0	0,00023
12	3	9	0	0,00007

и *коэффициент простоя обслуживающих рабочих* формулой

$$\frac{\rho}{r} = \frac{\text{Среднее число незанятых рабочих}}{\text{Общее число рабочих}}. \tag{7.24}$$

Для практических целей отождествим вероятности $P_n(t)$ с их пределами p_n. Из табл. 3 получаем $w = p_4 + 2p_5 + 3p_6 + \ldots + 17p_{20}$ и $\rho = 3p_0 + 2p_1 + p_2$. Табл. 4 показывает, что для автоматов рассматриваемого типа (с характеристикой $\lambda/\mu = 0,1$) *иметь трех*

Таблица 4

Сравнение эффективности двух систем, рассмотренных в примерах в) и г)

	I	II
Число автоматов	6	20
Число рабочих	1	3
Число автоматов на одного рабочего	6	$6^2/_3$
Коэффициент простоя рабочих	0,4845	0,4042
Коэффициент простоя автоматов	0,0549	0,01694

рабочих на двадцать автоматов много выгоднее, чем одного рабочего на шесть автоматов. Таблицы Пальма, на которые мы указывали в примечании на стр. 447, позволяют найти наиболее выгодное отношение числа рабочих к числу машин.

д) Задача о снабжении энергией[1]. Электрическая линия обслуживает a сварщиков, которые используют электроэнергию только время от времени. Если в момент t сварщик использует энергию, то вероятность того, что он прекратит пользование ею к моменту $t+h$, равна $\mu h+o(h)$; если в момент t сварщик не пользуется электроэнергией, то вероятность того, что энергия потребуется ему раньше момента $t+h$, равна $\lambda h+o(h)$. Сварщики работают независимо друг от друга.

Будем говорить, что система находится в состоянии E_n, если электроэнергию используют n сварщиков. Таким образом, мы имеем только конечное число состояний $E_0, \ldots, E_a$.

Если система находится в состоянии E_n, то $a-n$ сварщиков не используют электроэнергию, и вероятность, что возникнет новая потребность в энергии, равна $(a-n)\lambda h+o(h)$. С другой стороны, вероятность того, что один из n сварщиков прекратит использование энергии, равна $n\mu h+o(h)$. Следовательно, мы имеем процесс размножения и гибели с параметрами

$$\lambda_n=(a-n)\lambda, \quad \mu_n=n\mu \quad 0\leqslant n\leqslant a. \tag{7.25}$$

Основные дифференциальные уравнения принимают вид

$$\begin{aligned} P_0'(t)&=-a\lambda P_0(t)+\mu P_1(t),\\ P_n'(t)&=-\{n\mu+(a-n)\lambda\}P_n(t)+(n+1)\mu P_{n+1}(t)+\\ &\qquad+(a-n+1)\lambda P_{n-1}(t),\\ P_a'(t)&=-a\mu P_a(t)+\lambda P_{a-1}(t) \end{aligned} \tag{7.26}$$

(при $1\leqslant n\leqslant a-1$). Легко проверить, что *предельные вероятности задаются биномиальным распределением*

$$p_n=\binom{a}{n}\left(\frac{\lambda}{\lambda+\mu}\right)^n\left(\frac{\mu}{\lambda+\mu}\right)^{a-n}, \tag{7.27}$$

результат, которого можно было ожидать из элементарных соображений.

---

[1] Этот пример подсказан задачей, изученной в статье H. A. Adler and K. W. Miller, A New Approach to Probability Problems in Electrical Engineering, *Transactions of the American Institute of Electrical Engineers*, 65 (1946), 630—632.

§ 8. Обратные уравнения (уравнения, «обращенные в прошлое»)

В предыдущих параграфах мы изучали вероятности $P_n(t)$ пребывания системы в момент t в состоянии E_n. Такое обозначение хотя и удобно, но может привести к недоразумениям, так как здесь опущено указание на исходное состояние E_i системы в момент нуль. Для теоретических целей более естественно поэтому ввести обозначение $P_{in}(t)$. *Это — вероятность того, что система находится в момент t в состоянии E_n, при условии, что в момент* 0 *она была в состоянии E_i.* Величины $P_{in}(t)$ называются *вероятностями перехода.*

Надо подчеркнуть, что мы все время изучали именно эти переходные вероятности и что ничего, кроме обозначений, не изменилось. Если известно, что исходным состоянием было E_i, то $\{P_{in}(t)\}$ будет абсолютным распределением вероятностей в момент времени t. Если в момент 0 мы имели для исходного состояния распределение вероятностей $\{q_i\}$, то вероятность того, что в момент t система будет находиться в состоянии E_n, равна

$$Q_n(t) = \sum_i q_i P_{in}(t). \tag{8.1}$$

В случае процесса чистого размножения и процесса размножения и гибели мы убедились, что при произвольном фиксированном i вероятности перехода $P_{in}(t)$ удовлетворяют основным дифференциальным уравнениям (3.2) и (5.2). От индекса i зависят только начальные условия, которые могут быть записаны в виде

$$P_{in}(0) = \begin{cases} 1, & \text{если } n = i; \\ 0 & \text{в других случаях.} \end{cases} \tag{8.2}$$

Эти основные дифференциальные уравнения были выведены с помощью увеличения интервала времени $(0, t)$ до $(0, t+h)$ и рассмотрения изменений, которые могут произойти за малый промежуток времени $(t, t+h)$. Мы можем продолжить интервал $(0, t)$ в другом направлении и рассмотреть изменения за время $(-h, 0)$. В этом случае мы получим новую систему дифференциальных уравнений, в которой (вместо i) остается фиксированным n.

Рассмотрим сначала случай процесса чистого размножения и пренебрежем событиями, вероятности которых стремятся к нулю быстрее, чем h. Если система переходит из состояния $E_i (i > 0)$ в момент $-h$ в состояние E_n в момент t, то в момент 0 она должна быть или в E_i, или в E_{i+1}. Методами, изложенными в § 2 и 3, получаем следующее соотношение:

$$P_{in}(t+h) = P_{in}(t)(1 - \lambda_i h) + P_{i+1, n}(t)\lambda_i h + o(h). \tag{8.3}$$

Следовательно, при $i > 0$ новая основная система имеет вид

$$P'_{in}(t) = -\lambda_i P_{in}(t) + \lambda_i P_{i+1,\,n}(t) \tag{8.4}$$

и

$$P'_{0n}(t) = -\lambda_0 P_{0n}(t). \tag{8.5}$$

Эти уравнения называются *обратными* уравнениями, а уравнения (3.2) называются для различения *прямыми уравнениями*[1]). Начальными условиями являются соотношения (8.2).

В случае процесса размножения и гибели, если система находилась в момент $-h$ в состоянии E_i, в момент 0 она может быть в состояниях E_{i+1}, E_i или E_{i-1}; те же самые соображения приводят нас *к обратным уравнениям*

$$P'_{in}(t) = -(\lambda_i + \mu_i) P_{i,\,n}(t) + \lambda_i P_{i+1,\,n}(t) + \mu_i P_{i-1,\,n}(t). \tag{8.6}$$

Эти уравнения соответствуют (5.2).

Ясно, что прямые и обратные уравнения не независимы друг от друга: решения обратных уравнений с начальными условиями (8.2) автоматически удовлетворяют прямым уравнениям. Эти связи указаны здесь только в качестве подготовки к общей теории следующего параграфа.

Пример. *Процесс Пуассона.* В § 2 мы интерпретировали выражение Пуассона (1.1) как вероятность того, что за интервал времени длины t поступит ровно n вызовов. Теперь будем измерять время, начиная с произвольного момента, и будем говорить, что система находится в состоянии E_n, если ровно n вызовов произошло до момента t. Тогда переход из E_i (в момент t_1) в E_n (в момент t_2) означает, что за время (t_1, t_2) произошло $n - i$ вызовов. Это возможно, лишь если $n \geqslant i$, и, следовательно, вероятности перехода в процессе Пуассона определяются формулами

$$\begin{aligned} P_{in}(t) &= e^{-\lambda t}\frac{(\lambda t)^{n-i}}{(n-i)!}, \quad \text{если } n \geqslant i; \\ P_{in}(0) &= 0, \qquad \text{если } n < i. \end{aligned} \tag{8.7}$$

Прямые и обратные уравнения имеют соответственно вид

$$P'_{in}(t) = -\lambda P_{in}(t) + \lambda P_{i,\,n-1}(t), \tag{8.8}$$

$$P'_{in}(t) = -\lambda P_{in}(t) + \lambda P_{i+1,\,n}(t), \tag{8.9}$$

и легко проверить, что (8.7) есть решение обеих систем, удовлетворяющее начальным условиям (8.2).

---

[1]) Английские термины — «backward equations» и «forward equations», т. е., буквально, «уравнения, направленные назад», и «уравнения, направленные вперед». Мы предпочли менее точный, но более короткий перевод: «обратные» и «прямые» уравнения. — *Прим. ред.*

§ 9. Обобщение; уравнения Колмогорова

До сих пор мы рассматривали только такие процессы, при которых непосредственные переходы из состояния E_n возможны только в соседние состояния E_{n+1} и E_{n-1}. Кроме того, процессы были однородными по времени, т. е. вероятности перехода $P_{in}(t)$ были одними и теми же для всех интервалов времени длины t. Рассмотрим теперь более общий процесс, в котором оба эти предположения отброшены.

Как и в теории обычных цепей Маркова (гл. XV), разрешим непосредственные переходы из любого состояния E_i в любое состояние E_n. Вероятности перехода могут изменяться с течением времени. Это делает необходимым указание обеих конечных точек интервалов времени, а не только длины этих интервалов. В соответствии с этим будем обозначать *через $P_{in}(\tau, t)$ условную вероятность пребывания системы в момент t в состоянии E_n, при условии, что в предшествующий момент τ она находилась в состоянии E_i.* Символ $P_{in}(\tau, t)$ бессмыслен, если $\tau < t$. Если процесс однороден по времени, то $P_{in}(\tau, t)$ зависит только от разности $t-\tau$, и можно написать $P_{in}(t)$ вместо вероятности $P_{in}(\tau, \tau+t)$ (которая не зависит тогда от τ).

Основным свойством нашего процесса является свойство Маркова, обсуждавшееся в § 10 гл. XV. Оно состоит в том, что если задано состояние системы в некоторый момент, то будущие изменения не зависят от прошедшего. Более точно: рассмотрим три момента времени $\tau < s < t$ и предположим, что в момент τ система находится в состоянии E_i, а в момент s — в состоянии E_ν. Для общего случайного процесса (условная) вероятность пребывания системы в момент t в состоянии E_n зависит и от i, и от ν. Другими словами, не только «настоящее состояние» E_ν, но также и прошедшее состояние E_i влияет на «будущее» состояние в момент t. Однако для марковского процесса это уже не так. Для него рассматриваемая вероятность равна вероятности $P_{\nu n}(s, t)$ перехода из E_ν в момент s в E_n в момент t. Знание того, что в момент $\tau < s$ система была в состоянии E_i, не дает возможности каких-либо новых выводов о будущем. Это предположение приводит немедленно к важным заключениям. Переход из E_j в момент τ в E_n в момент t происходит через некоторое состояние E_ν в момент s, и для марковского процесса вероятность того, что этот переход произойдет через данное состояние E_ν, равна $P_{i\nu}(\tau, s)P_{\nu n}(s, t)$. Отсюда следует, что

$$P_{in}(\tau, t) = \sum_\nu P_{i\nu}(\tau, s)P_{\nu n}(s, t) \tag{9.1}$$

тождественно при всех $\tau < s < t$. Это — уравнение Колмогорова — Чэпмена. Оно является аналогом уравнения (10.3) *гл. XV, которое выполняется, когда временной параметр принимает только целые значения.*

В § 10 гл. XV было показано, что уравнение Колмогорова — Чэпмена выполняется не для всех стохастических процессов. Примем уравнение (9.1) *за определение того класса процессов, которыми мы будем заниматься*[1]). Действительно, прибавив только некоторые ограничения регулярности, мы выведем из (9.1) наши основные дифференциальные уравнения. Уравнение Колмогорова — Чэпмена имеет вероятностный смысл; однако мы не должны все время к этому возвращаться: если (9.1) уже дано, мы можем легко вывести дифференциальные уравнения, позволяющие определить вероятности $P_{in}(t)$, действуя чисто аналитическим способом.

В случае однородного по времени процесса (9.1) принимает простой вид

$$P_{in}(t+s)=\sum_{\nu} P_{i\nu}(t)\,P_{\nu n}(s). \tag{9.2}$$

Для процесса Пуассона это уравнение сводится к теореме о композиции распределений Пуассона, установленной в гл. XI [пример 2.6.]

Введем теперь основные предположения регулярности, которые очевидным образом обобщают исходные предположения о процессе размножения и гибели.

Предположение 1. *Каждому состоянию E_n соответствует такая непрерывная функция $c_n(t)\geqslant 0$, что при $h\to 0$*

$$\frac{1-P_{nn}(t,\,t+h)}{h}\to c_n(t) \tag{9.3}$$

равномерно по t.

Вероятностная интерпретация (9.3) очевидна: если в момент t система находится в состоянии E_n, то вероятность того, что за время $(t,\,t+h)$ произойдет изменение, равна $c_n(t)\,h+o(h)$. Аналитически (9.3) означает, что $P_{nn}(t,\,s)\to 1$ при $s\to t$ и что $P_{nn}(t,\,x)$ имеет при $x=t$ производную по переменной x. Функция $c_n(t)$ играет роль $\lambda_n+\mu_n$ в процессе размножения и гибели. В случае однородного по времени процесса c_n постоянно.

Предположение 2. *Каждой паре состояний E_j и E_k, $j\neq k$, соответствуют такие вероятности перехода $p_{jk}(t)$ (зависящие от времени), что при $h\to 0$*

$$\frac{P_{jk}(t,\,t+h)}{h}\to c_j(t)\,p_{jk}(t) \qquad (j\neq k) \tag{9.4}$$

[1]) Вопрос о том, насколько уравнение Колмогорова характеризует марковский процесс, приводит к трудным проблемам, связанным с изучением случайных функций $X(t)$. Напомним, что мы используем кратчайший путь для получения дифференциальных уравнений относительно некоторых вероятностей и не анализируем процесс во всех его аспектах.

равномерно по t. Величины $p_{jk}(t)$ непрерывны по t, и при любых фиксированных t, j

$$\sum_k p_{jk}(t)=1, \quad p_{jj}(t)=0. \tag{9.5}$$

Здесь $p_{jk}(t)$ могут быть интерпретированы как условные вероятности того, что если переход из E_j осуществится за время $(t, t+h)$, то при этом система перейдет из E_j в E_k. В процессе размножения и гибели

$$p_{j,\,j+1}(t)=\frac{\lambda_j}{\lambda_j+\mu_j}, \qquad p_{j,\,j-1}(t)=\frac{\mu_j}{\lambda_j+\mu_j} \tag{9.6}$$

и $p_{jk}(t)=0$ для всех других комбинаций j и k. При каждом фиксированном t величины $p_{jk}(t)$ могут быть интерпретированы, как вероятности перехода цепи Маркова. Этих двух условий достаточно для вывода системы обратных уравнений для $P_{jk}(\tau, t)$, но для вывода прямых уравнений потребуется дополнительное

Предположение 3. *При фиксированном k переход к пределу в* (9.4) *равномерен по j.*

Необходимость этих предположений представляет значительный интерес для теории бесконечных систем дифференциальных уравнений; мы займемся этим в следующем параграфе.

Выведем теперь систему дифференциальных уравнений для $P_{ik}(\tau, t)$ как функций от t и n (прямые уравнения). Из (9.1) имеем

$$P_{ik}(\tau, t+h)=\sum_j P_{ij}(\tau, t)\,P_{jk}(t, t+h). \tag{9.7}$$

Если мы выразим слагаемое $P_{kk}(t, t+h)$ в правой части, воспользовавшись (9.3), то получим

$$\frac{P_{ik}(\tau, t+h)-P_{ik}(\tau, t)}{h}=$$

$$=-c_k(t)\,P_{ik}(t, \tau)+\frac{1}{h}\sum_{j\neq k} P_{ij}(\tau, t)\,P_{jk}(t, t+h)+\ \ldots, \tag{9.8}$$

где опущенные члены стремятся к нулю вместе с h и суммирование производится по всем j, за исключением $j=k$. Мы можем теперь применить (9.4) к слагаемым этой суммы. Так как (по предположению 3) переход к пределу равномерен по j, правая часть имеет предел. Следовательно, имеет предел и левая часть; это означает, что $P_{ik}(\tau, t)$ имеет частную производную по t, и

$$\frac{\partial P_{ik}(\tau, t)}{\partial t}=-c_k(t)\,P_{ik}(\tau, t)+\sum_i P_{ij}(\tau, t)\,c_j(t)\,p_{jk}(t). \tag{9.9}$$

Это — основная система прямых дифференциальных уравнений. Заметим, что в ней i и τ фиксированы, так что мы имеем (несмотря на формально появившуюся частную производную) систему обыкновенных дифференциальных уравнений для бесконечного числа функций $P_{ik}(\tau, t)$, $k=0, 1, 2, \ldots$. Параметры i и τ входят только в начальные условия

$$P_{ik}(\tau, \tau)=\begin{cases} 1, \text{ если } k=i; \\ 0 \text{ в остальных случаях.}\end{cases} \tag{9.10}$$

Система обратных уравнений может быть получена аналогичным образом. Вывод этих уравнений даже проще, так как теперь мы можем полностью освободиться от предположения 3. Что касается уравнений (9.3) и (9.4), то их более естественно использовать в форме

$$\frac{1-P_{nn}(t-h, t)}{h} \to c_n(t), \tag{9.3a}$$

$$\frac{P_{jk}(t-h, t)}{h} \to c_j(t)\, p_{jk}(t) \qquad (j \neq k). \tag{9.4a}$$

Можно показать, что эти уравнения эквивалентны (9.3) и (9.4), но мы будем просто исходить из соотношений (9.3a), (9.4a) и (9.5) как наших основных предположений. Переписывая уравнение Колмогорова — Чэпмена (9.1) в виде

$$P_{ik}(\tau-h, t)=\sum_{\nu} P_{i\nu}(\tau-h, \tau)\, P_{\nu k}(\tau, t) \tag{9.11}$$

и используя (9.3a) при $n=i$, получаем

$$\frac{P_{ik}(\tau-h, t)-P_{ik}(\tau, t)}{h}=-c_i(\tau)\, P_{ik}(\tau, t)+$$
$$+\frac{1}{h}\sum_{\nu\neq i} P_{i\nu}(\tau-h, \tau)\, P_{\nu k}(\tau, t)+\frac{o(h)}{h}. \tag{9.12}$$

Но $h^{-1}P_{i\nu}(\tau-h, \tau) \to c_i(\tau)\, p_{i\nu}(\tau)$ и переход к пределу всегда равномерен. Действительно, при $N>i$

$$0 \leqslant h^{-1}\sum_{\nu=N+1}^{\infty} P_{i\nu}(\tau-h, \tau)\, P_{\nu k}(\tau, t) \leqslant h^{-1}\sum_{\nu=N+1}^{\infty} P_{i\nu}(\tau-h, \tau) \leqslant$$
$$\leqslant h^{-1}\left\{1-\sum_{\nu=0}^{N} P_{i\nu}(\tau-h, \tau)\right\} \to c_i(\tau)\left\{1-\sum_{\nu=0}^{N} p_{i\nu}(\tau)\right\}. \tag{9.13}$$

В силу соотношения (9.5) правая часть может быть сделана сколь угодно малой при достаточно большом N. Отсюда вытекает, что в правой части уравнения (9.12) возможен почленный предельный переход. Итак, имеем

$$\frac{\partial P_{ik}(\tau, t)}{\partial \tau} = c_i(\tau) P_{ik}(\tau, t) - c_i(\tau) \sum_{\nu} p_{i\nu}(\tau) P_{\nu k}(\tau, t). \qquad (9.14)$$

Это соотношение вместе с начальными условиями (9.10) и есть *основная система обратных дифференциальных уравнений.*

Эти две системы были впервые выведены А. Колмогоровым в его знаменитой работе[1]), в которой он заложил основания теории марковских процессов (много более общего типа, чем рассмотренные здесь). Можно показать[2]), что всегда существует общее решение $\{P_{ik}(\tau, t)\}$ обеих систем, которое удовлетворяет уравнению Колмогорова — Чэпмена (9.1) и неравенствам

$$P_{ik}(\tau, t) \geqslant 0, \qquad \sum_{k} P_{ik}(\tau, t) \leqslant 1. \qquad (9.15)$$

Мы знаем, что для случая процесса чистого размножения (§ 4) вероятности $P_{ik}(\tau, t)$ не всегда дают в сумме единицу. Разность $1 - \sum P_{ik}(\tau, t)$ учитывает возможность бесконечного числа переходов за конечный интервал времени (τ, t). Если $\sum P_{ik}(\tau, t) = 1$, то решение единственно, но, вообще говоря, различные процессы могут удовлетворять одним и тем же прямым и обратным уравнениям (см. § 10). С точки зрения приложений возможностью $\sum P_{ik}(\tau, t) < 1$ можно без всякой опасности пренебречь.

Пример. *Сложный процесс Пуассона.* Рассмотрим случай, когда все $c_i(t)$ равны одной и той же постоянной

$$c_i(t) = \lambda \qquad (9.16)$$

и когда p_{jk} не зависят от t. В этом случае вероятности p_{jk} определяют обычную цепь Маркова, и мы обозначим (так же, как и в гл. XV) соответствующие вероятности перехода за n шагов через $p_{jk}^{(n)}$.

Из (9.16) следует, что вероятность изменения состояния системы за время $(t, t+h)$ не зависит от состояния системы в момент t и равна $\lambda h + o(h)$. В соответствии с этим, число переходов за интервал времени (τ, t) имеет распределение Пуассона с параметром $\lambda(t-\tau)$.

---

[1]) Об аналитических методах в теории вероятностей, *Mathematische Annalen*, **104** (1931), 415—458, *Успехи математических наук*, вып. 5 (38) (1938), 5—41.

[2]) См. примечание 1 на стр. 439.

Условные вероятности переходов при условии, что сделано n переходов, равны $p_{jk}^{(n)}$. Поэтому имеем

$$P_{jk}(\tau, t) = e^{-\lambda(t-\tau)} \sum_{n=0}^{\infty} \frac{\{\lambda(t-\tau)\}^n}{n!} p_{ik}^{(n)} \tag{9.17}$$

(где $p_{jk}^{(0)}$ равно 0 или 1 в зависимости от того, будет ли $j \neq k$ или $j = k$). Легко проверить, что (9.18) действительно есть решение наших двух систем дифференциальных уравнений.

Если, в частности,

$$p_{jk} = 0 \quad \text{при} \quad k < j, \quad p_{jk} = f_{k-j} \quad \text{при} \quad k \geqslant j, \tag{9.18}$$

то формула (9.17) дает *сложное распределение Пуассона* (см. § 1 гл. XII).

§ 10. Процессы, уходящие в бесконечность

Пример процесса чистого размножения (§ 3 и 4) показывает, что вероятности перехода $P_{ik}(t)$, определяемые дифференциальными уравнениями Колмогорова, не всегда дают в сумме единицу. Может случиться, что

$$\sum_k P_{ik}(t) < 1. \tag{10.1}$$

Когда это явление было впервые обнаружено в 1940 г., то оно воспринималось как нарушающий гармонию сюрприз. Исследованию этого явления и связанного с ним факта, что система уравнений Колмогорова не всегда однозначно определяет набор вероятностей перехода $P_{ik}(t)$, была посвящена огромная литература. Процессы с такими свойствами обычно называются нерегулярными[1]). Постепенно пришло понимание того, что в действительности мы имеем дело с простой и естественной аналогией хорошо известной ситуации в теории диффузии. Появление соотношения (10.1) больше не кажется разрушительным, а приводит к доставляющему удовлетворение открытию, что теория дифференциальных уравнений Колмогорова разделяет основные черты теории диффузии. Несмотря на совершенно различные методы вывода основных уравнений и используемый аналитический аппарат, как в той, так и в другой теории мы сталкиваемся с общими типами граничных условий и другими аналогиями. Любая из этих теорий позволяет в новом свете взглянуть на другую, и никакой резкой грани между ними в действительности провести нельзя. На этом пути теория марковских про-

[1]) Аналог этого параграфа в первом издании назывался «Вырожденные процессы».

цессов достигает неожиданного и весьма приятного внутреннего единства.

Еще раз рассмотрим простой процесс чистого размножения из § 3. Система проводит некоторое время в начальном состоянии E_0, затем переходит в состояние E_1, оттуда в E_2 и т. д. Вероятность того, что время пребывания в E_0 превосходит t, получается по формуле (3.2), именно $P_0(t)=e^{-\lambda_0 t}$. Время пребывания T_0 является случайной величиной, но так как она может принять любое неотрицательное значение t, то, формально говоря, его рассмотрение выходит за рамки этой книги. Однако переход от геометрического распределения к показательному так тривиален, что мы можем безнаказанно немного нарушить границы. Приближение T_0 дискретной случайной величиной с геометрическим распределением показывает, что среднее значение времени пребывания в E_0 естественно определить формулой

$$E(T_0)=\int_0^\infty te^{-\lambda_0 t}\,dt=\lambda_0^{-1}. \tag{10.2}$$

В момент, когда система впервые достигнет E_j, это состояние начинает играть роль начального состояния, и те же заключения применимы к времени пребывания T_j в E_j. *Среднее время пребывания в состоянии* E_j равно $E(T_j)=\lambda_j^{-1}$. Отсюда вытекает, что $\lambda_0^{-1}+\lambda_1^{-1}+\ldots+\lambda_n^{-1}$ есть математическое ожидание времени, которое система проводит в множестве состояний $E_0, E_1, \ldots, E_n$. Учитывая это, критерий § 4 можно сформулировать так:

Для того чтобы $\sum P_n(t)=1$ *при всех* t, *необходимо и достаточно, чтобы*

$$\sum E(T_j)=\sum \lambda_j^{-1}=\infty, \tag{10.3}$$

т. е. среднее время, проведенное в $E_0, E_1, E_2, \ldots$, *должно быть бесконечным.* Очевидно, $L_0(t)=1-\sum P_n(t)$ означает вероятность того, что система побывает во *всех* состояниях до момента t.

В такой форме теорема становится очень наглядной. Если среднее время пребывания в E_j равно 2^{-j}, то вероятность того, что система побывает во всех состояниях до момента $1+2^{-1}+2^{-2}+\ldots=2$, должна быть положительной. Аналогично этому частица, движущаяся с экспоненциально возрастающей скоростью вдоль оси x, пройдет всю ось за конечное время.

Если процесс чистого размножения служит моделью роста некоторой совокупности, то состояние E_n соответствует объему, равному n, и достижение бесконечности за конечное время означает своего рода взрыв. В этом случае (10.1) представляет действительно некоторую аномалию, но в других приложениях это неравенство появляется вполне регулярным образом. Говоря геометрически, нет никакой нужды размещать состояния E_0, E_1, ... в точках 0, 1, ... оси x. Вместо этого можно представлять, что состояния расположены в точках $0 = x_0 < x_1 < x_2 < \ldots$, таких, что $x_n \to 1$. Процесс размножения можно тогда описать как движение «частицы», которая отправляется из точки $x_0 = 0$, прыгает в x_1, спустя некоторое время прыгает в x_2 и т. д. При такой интерпретации частица, побывавшая во всех состояниях, достигает предельной точки 1, и естественно, что это может произойти за конечное время. Рассмотрим для сравнения детерминированное движение частицы, выходящей из начала и имеющей скорость $f(x)$ в точке x. Пусть $x(t)$ — положение частицы в момент t. Тогда очевидно выполнение дифференциального уравнения $x'(t) = f(x(t))$, и время τ достижения точки $x = 1$ определяется формулой

$$\tau = \int_0^1 \frac{dx}{f(x)} \leqslant \infty. \tag{10.4}$$

Достигнет ли частица точки 1 за конечное время или будет лишь асимптотически к ней приближаться, зависит от сходимости интеграла (10.4). В вероятностной модели движение происходит скачками, но λ_n^{-1} есть *среднее* время, затраченное на переход из x_n в x_{n+1}. С этой точки зрения критерии (10.3) и (10.4) являются близнецами.

Продолжим рассмотрение процессов чистого размножения и покажем, как критерий (10.3) связан с вопросом о *единственности* решений дифференциальных уравнений Колмогорова.

Вероятности $P_n(t)$, введенные в § 3, мы будем теперь рассматривать как вероятности перехода и обозначать их через $P_{in}(t)$. Основное дифференциальное уравнение (3.2) применимо в равной мере и к $P_{ik}(t)$ при произвольном (но фиксированном) i. Система *прямых уравнений* принимает вид

$$P'_{i0}(t) = -\lambda_0 P_{i0}(t), \quad P'_{ik}(t) = -\lambda_k P_{ik}(t) + \lambda_{k-1} P_{i,k-1}(t). \tag{10.5}$$

(Индекс $i \geqslant 0$ фиксирован, а $k = 1, 2, \ldots$.) Из (8.4) и (8.5) находим соответствующие *обратные уравнения*

$$P'_{ik}(t) = -\lambda_i P_{ik}(t) + \lambda_i P_{i+1,k}(t). \tag{10.6}$$

Эти уравнения справедливы при фиксированном $k \geqslant 0$ и $i = 0, 1, \ldots$. Начальные условия, очевидно, имеют один и тот же вид

$$P_{ii}(0) = 1, \quad P_{ik}(0) = 0, \quad k \neq i \tag{10.7}$$

Взглянув на (10.5), нетрудно убедиться в том, что первое дифференциальное уравнение вместе с начальными условиями (10.7) однозначно определяют $P_{i0}(t)$. Мы можем тогда последовательно вычислить все вероятности $P_{ik}(t)$. Как решать линейные уравнения первого порядка хорошо известно, и мы получаем легко проверяемую формулу для единственного решения системы прямых уравнений (10.5) с начальными условиями (10.7):

$$\begin{gathered} P_{ik}(t)=0 \quad \text{при} \quad k<i, \quad P_{ii}(t)=e^{-\lambda_i t}, \\ P_{ik}(t)=\lambda_{k-1}\int_0^t e^{-\lambda_k s}P_{i,\,k-1}(t-s)\,ds \quad (k>i). \end{gathered} \tag{10.8}$$

В случае обратных уравнений ситуация совершенно отлична от описанной. Мы покажем, что если $\sum \lambda_n^{-1}<\infty$, то решение не единственно.

Лемма. *Единственное решение системы прямых уравнений* (10.5), *определяемое формулой* (10.8), *автоматически является решением системы обратных уравнений* (10.6). *Если* $\bar{P}_{ik}(t)$ — *любое неотрицательное решение системы* (10.6) — (10.7), *то*

$$\bar{P}_{ik}(t)\geqslant P_{ik}(t). \tag{10.9}$$

Доказательство. Рассмотрим уравнения (10.6), ставя черту над всеми $P_{ik}(t)$. i-е уравнение можно разрешить относительно $\bar{P}_{ik}(t)$. Получаем

$$\begin{gathered} \bar{P}_{ik}(t)=\lambda_i\int_0^t e^{-\lambda_i s}\bar{P}_{i+1,\,k}(t-s)\,ds, \qquad k\neq i, \\ \bar{P}_{kk}(t)=e^{-\lambda_k t}+\lambda_k\int_0^t e^{-\lambda_k s}\bar{P}_{k+1,\,k}(t-s)\,ds. \end{gathered} \tag{10.10}$$

[Заметим, что эта система соотношений не является рекуррентной и не может быть использована для решения системы уравнений (10.6).]

Пусть $P_{ik}(t)$ — решение прямых уравнений, определяемое формулами (10.8). При любом k и $i>k$ положим $\bar{P}_{ik}(t)=P_{ik}(t)=0$. Эти функции удовлетворяют соотношениям (10.10). Далее (10.10) определяет $\bar{P}_{kk}(t)=e^{-\lambda_k t}=P_{kk}(t)$. Полагая в (10.10) последовательно $i=k-1,\ k-2,\ \ldots$, мы получаем $\bar{P}_{ik}(t)$, определенные при всех i и являющиеся решением обратных уравнений (10.6) с начальными условиями (10.7). Очевидно, $\bar{P}_{k-1,\,k}(t)=P_{k-1,\,k}(t)$. Проверим по индукции, что $\bar{P}_{ik}(t)=P_{ik}(t)$. Предположим, что это тождество выполнено для всех комбинаций индексов i, k, таких, что $k-i\leqslant r$, где $r\geqslant 1$ — целое число (мы знаем, что оно справедливо при $r=1$). Пусть $k-i=r+1$. Интеграл в (10.10) выражает $\bar{P}_{ik}$ через $\bar{P}_{i+1,\,k}=P_{i+1,\,k}$, а в (10.8) в свою очередь $P_{i+1,\,k}$.

представляется в виде интеграла, содержащего $P_{i+1,\,k-1}=\overline{P}_{i+1,\,k-1}$. Выполнив подстановку, мы получим $\overline{P}_{ik}$ в виде двойного интеграла, зависящего от $\overline{P}_{i+1,\,k-1}$. Изменяя порядок интегрирования, приходим к формуле

$$\overline{P}_{ik}(t)=\lambda_i\lambda_{k-1}\int\limits_0^t e^{-\lambda_k x}\,dx\int\limits_0^{t-x} e^{-\lambda_i s}\overline{P}_{i+1,\,k-1}(t-x-s)\,ds=$$

$$=\lambda_{k-1}\int\limits_0^t e^{-\lambda_k x}\overline{P}_{i,\,k-1}(t-x)\,dx. \qquad (10.11)$$

Но, по предположению индукции, $\overline{P}_{i,\,k-1}(t)=P_{i,\,k-1}(t)$ и сравнение соотношений (10.8) и (10.11) показывает, что $\overline{P}_{ik}(t)=P_{ik}(t)$. Первая часть леммы доказана.

Остается проверить, что для произвольного неотрицательного решения системы обратных уравнений (10.6) выполняется неравенство (10.9). Как $\overline{P}_{ik}(t)$, так и $P_{ik}(t)$ удовлетворяют соотношениям (10.10). При $i>k$ имеем $\overline{P}_{ik}(t)\geqslant P_{ik}(t)=0$. Полагая последовательно $i=k,\ k-1,\ k-2,\ \ldots$, нетрудно убедиться в том, что неравенство (10.9) имеет место для любой пары индексов $i,\ k$, что доказывает лемму окончательно.

Дальнейшие рассуждения зависят от того, какая из двух следующих возможностей имеет место.

а) *Случай* $\sum\lambda_n^{-1}=\infty$. Мы знаем из § 4, что в этом случае $\sum\limits_k P_{ik}(t)=1$. Отсюда следует, что для любого другого решения обратных уравнений, сумма $\sum\limits_k P'_{ik}(t)$ должна превосходить единицу, что невозможно. Итак, в случае (а) *системы прямых и обратных уравнений имеют общее и единственное решение. Это решение представляет вероятности перехода процесса чистого размножения, причем* $\sum P_{ik}(t)=1$. [Прямым дифференцированием нетрудно доказать, что уравнение Колмогорова — Чэпмена (9.1) выполняется.]

б) *Случай* $\sum\lambda_n^{-1}<\infty$. Мы знаем, что в этом случае $\sum P_{ik}(t)<1$. Тогда

$$L_i(t)=1-\sum_{k=0}^{\infty}P_{ik}(t) \qquad (10.12)$$

— *вероятность того, что, отправляясь из состояния* E_i, *система достигнет «бесконечности» до момента* t. Мы знаем, что соотношения (10.10) выполняются при $\bar{P}_{ik}(t) = P_{ik}(t)$. Суммируя эти соотношения, убеждаемся, что

$$L_i(t) = \lambda_i \int_0^t e^{-\lambda_i s} L_{i+1}(t-s)\,ds \tag{10.13}$$

или

$$L'_i(t) = -\lambda_i L_i(t) + \lambda_i L_{i+1}(t), \quad L_i(0) = 0. \tag{10.14}$$

Итак, *в случае* (б) *бесконечная система дифференциальных уравнений* (10.14) *имеет ненулевое решение* $\{L_i(t)\}$, *для которого* $L_i(0) = 0$. *При произвольном* $A_k(t)$ *матрица*

$$\bar{P}_{ik}(t) = P_{ik}(t) + L_i(t)\,A_k(t) \tag{10.15}$$

представляет решение обратных уравнений (10.6), *удовлетворяющее начальным условиям* (10.7).

Возникает вопрос о том, можно ли $A_k(t)$ определить таким образом, чтобы $\bar{P}_{ik}(t)$ стали вероятностями перехода, удовлетворяющими уравнению Колмогорова—Чэпмена (9.1). Ответ утвердительный. Мы воздержимся от доказательства этого утверждения, однако дадим вероятностную интерпретацию.

Вероятности $P_{ik}(t)$ определяют так называемый *процесс с поглощающим экраном: когда система достигает бесконечности, процесс останавливается.* Дуб[1]) был первым, кто рассмотрел процессы с возвращением, когда система, достигнув бесконечности, мгновенно возвращается в E_0 (или некоторое другое фиксированное состояние и процесс начинается заново). В этом случае система может перейти из состояния E_0 в состояние E_5 или за пять шагов, или за бесконечное число шагов, уходя один или несколько раз в бесконечность и возвращаясь каждый раз снова в E_0. Переходные вероятности такого процесса имеют вид (10.15). Они *удовлетворяют системе обратных уравнений* (10.6), *но не системе* (10.5) *прямых уравнений.*

Этим объясняется, почему при выводе прямых уравнений потребовалось странное на первый взгляд предположение 3, не являющееся необходимым при выводе обратных уравнений. Простые и наглядные с вероятностной точки зрения предположения 1—2 относятся к процессам с возвращением, для которых прямые уравнения (10.5) не выполняются. Другими словами, если мы исходим лишь из предположений 1—2, то нетрудно вывести обратные уравнения Колмогорова,

---

[1]) J. L. Doob, Markoff chains — denumerable case, *Trans. Amer. Math. Soc.*, 58 (1945), 455—473.

но к прямым уравнениям должен быть прибавлен дополнительный член [1]).

Процессы чистого размножения, конечно, слишком тривиальны, чтобы представлять действительный интерес, но ситуация, описанная здесь, является типичной для случая общих дифференциальных уравнений Колмогорова. Однако при этом возникают два существенно новых явления. Во-первых, в случае процессов размножения единственная особенность поведения системы состоит в уходе на бесконечность, или, говоря абстрактно, существует только одна *граничная* точка. В противоположность этому, общие процессы характеризуются тем, что соответствующая граница имеет сложную топологическую структуру. Во-вторых, в случае процессов чистого размножения движение частицы происходит по направлению к границе, так как возможны только переходы вида $E_n \to E_{n+1}$. Можно построить процессы другого типа; например, изменив направление движения, мы получим процесс, при котором возможны только переходы $E_{n+1} \to E_n$. Такие процессы могут *выходить* из граничных точек, вместо того, чтобы их достигать. Для процессов гибели и размножения переходы возможны в обоих направлениях, как и при одномерной диффузии. Оказывается, что в этом случае существуют процессы, аналогичные процессам с поглощающими и отражающими экранами теории диффузии, но их описание вывело бы нас за рамки настоящей книги.

§ 11. Задачи

1. В процессе чистого размножения, определяемом уравнением (3.2), пусть $\lambda_n > 0$ при всех n. Доказать, что при каждом фиксированном $n \geqslant 1$ функция $P_n(t)$ сначала возрастает, а потом убывает до 0. Если t_n — точка максимума $P_n(t)$, то $t_1 < t_2 < t_3 < \ldots$.

Указание. Использовать метод индукции. Продифференцировать (3.2).

2. Продолжение. Если $\sum \lambda_n^{-1} = \infty$, показать, что $t_n \to \infty$.

Указание. Если $t_n \to \tau$, то при фиксированном $t > \tau$ последовательность $\lambda_n P_n(t)$ возрастает. Использовать (4.10).

3. Процесс Юла. Вывести математическое ожидание и дисперсию распределения, определяемого формулой (3.4). [Использовать только дифференциальные уравнения, а не явные формулы (3.5).]

4. Процесс чистой гибели. Найти дифференциальные уравнения процесса типа Юла с переходами только из E_n в E_{n-1}. Найти распределение $P_n(t)$, его математическое ожидание и дисперсию, предполагая, что исходным состоянием является E_i.

5. Автопарк. В автопарк, рассчитанный на N мест, прибывает пуассоновский поток машин с интенсивностью λ до тех пор, пока имеются свободные места. Найти дифференциальные уравнения для вероятностей $P_n(t)$ того, что ровно n мест заняты.

---

[1]) Его вид приводится в сравнительно недавней работе, цитированной в примечании 1 на стр. 439, где изучаются различные типы процессов и соответствующие граничные условия.

6. Очередь обладает тем свойством, что *клиент, пришедший последним, обслуживается первым*[1]). Найти дифференциальные уравнения для вероятностей $P_n(t)$ того, что ровно n вновь прибывших клиентов будет обслужено в течение времени ожидания клиента, выбранного случайно.

7. Процесс Пойа[2]). Это неоднородный процесс чистого размножения с λ_n, зависящими от времени:

$$\lambda_n(t) = \frac{1+an}{1+at}. \tag{11.1}$$

Показать, что решение с начальным условием $P_0(0) = 1$ имеет вид

$$\begin{gathered} P_0(t) = (1+at)^{-1/a}, \\ P_n(t) = t^n (1+at)^{-n-1/a} \frac{(1+a)(1+2a)\ldots\{1+(n-1)a\}}{n!}. \end{gathered} \tag{11.2}$$

Показать (с помощью дифференциального уравнения), что математическое ожидание и дисперсия равны соответственно t и $t(1+at)$.

8. Продолжение. Процесс Пойа может быть получен предельным переходом из урновой схемы Пойа примера 2, в гл. V. Если состояние системы определяется как число красных шаров, то вероятность перехода $E_k \to E_{k+1}$ при $(n+1)$-м извлечении равна

$$p_{k,n} = \frac{r+kc}{r+b+nc} = \frac{p+k\gamma}{1+n\gamma}, \tag{11.3}$$

где $p = r/(r+b)$, $\gamma = c/(r+b)$. Так же, как и при переходе от испытаний Бернулли к распределению Пуассона, испытания производятся со скоростью по одному испытанию за время h, и пусть $h \to 0$, $n \to \infty$ так, что $np \to t$, $n\gamma \to at$. Показать, что в пределе (11.3) приводит к (11.1). Показать также, что распределение Пойа (2.3) гл. V переходит в (11.2).

9. Линейный рост. Если для процесса, определяемого уравнениями (5.7), $\lambda = \mu$ и $P_1(0) = 1$, то

$$P_0(t) = \frac{\lambda t}{1+\lambda t}, \qquad P_n(t) = \frac{(\lambda t)^{n-1}}{(1+\lambda t)^{n+1}}. \tag{11.4}$$

Вероятность вымирания равна 1.

10. Продолжение. Подставляя в уравнения (5.8) $P_n(t) = A(t)B^n(t)$, доказать, что решение с $P_1(0) = 1$ имеет вид

$$P_0(t) = \mu B(t), \quad P_n(t) = \{1-\lambda B(t)\}\{1-\mu B(t)\}\{\lambda B(t)\}^{n-1}, \tag{11.5}$$

$$B(t) = \frac{1-e^{(\lambda-\mu)t}}{\mu-\lambda e^{(\lambda-\mu)t}}. \tag{11.6}$$

---

[1]) E. Vaulot, Delais d'attente des appels téléphoniques dans l'ordre inverse de leur arrivée, *Comptes Rendues Académie des Sciences*, Paris, 238 (1954), 1188—1189.

[2]) O. Lundberg, On random processes and their applications to sickness and accident statistics, Uppsala, 1940. В этой книге многие свойства процесса Пойа рассматриваются главным образом в связи со сложным процессом Пуассона.

11. Продолжение. Производящая функция $P(s, t) = \sum P_n(t) s^n$ удовлетворяет уравнению в частных производных

$$\frac{\partial P}{\partial t} = \{\mu - (\lambda + \mu) s + \lambda s^2\} \frac{\partial P}{\partial s}. \tag{11.7}$$

12. Продолжение. Пусть $M_2(t) = \sum n^2 P_n(t)$ и $M(t) = \sum n P_n(t)$ (как и в § 5). Показать, что

$$M_2'(t) = 2(\lambda - \mu) M_2(t) + (\lambda + \mu) M(t). \tag{11.8}$$

Вывести отсюда, что при $\lambda > \mu$ *дисперсия* распределения $\{P_n\}$ определяется формулой

$$e^{2(\lambda-\mu)t}\{1 - e^{(\mu-\lambda)t}\}(\lambda + \mu)/(\lambda - \mu). \tag{11.9}$$

13. Для процесса, определяемого формулой (7.2), производящая функция $P(s, t) = \sum P_n(t) s^n$ удовлетворяет уравнению в частных производных

$$\frac{\partial P}{\partial t} = (1 - s)\left\{-\lambda P + \mu \frac{\partial P}{\partial s}\right\}. \tag{11.10}$$

Решение этого уравнения имеет вид

$$P = e^{-\lambda(1-s)(1-e^{-\mu t})/\mu}\{1 - (1 - s) e^{-\mu t}\}^i.$$

При $i = 0$ это просто распределение Пуассона с параметром $\lambda(1 - e^{-\mu t})/\mu$. При $t \to \infty$ *распределение* $\{P_n(t)\}$ *стремится к распределению Пуассона с параметром* λ/μ.

14. Для процесса, определяемого формулой (7.26), производящая функция $P(s, t) = \sum P_n(t) s^n$ удовлетворяет уравнению в частных производных

$$(\mu + \lambda s) \frac{\partial P}{\partial s} = a\lambda P, \tag{11.11}$$

с решением $P = \{(\mu + \lambda s)/(\lambda + \mu)\}^a$.

15. В «простейшей задаче о телефонных линиях», пример 7, а, пусть $Q_n(t)$ — вероятность того, что исходя из E_n система достигнет E_0 до момента t. Доказать, что выполняются следующие дифференциальные уравнения:

$$\begin{aligned} Q_n'(t) &= -(\lambda + n\mu) Q_n(t) + \lambda Q_{n+1}(t) + n\mu Q_{n-1}(t) \quad (n \geqslant 2) \\ Q_1'(t) &= -(\lambda + \mu) Q_1(t) + \lambda Q_2(t) + \mu \end{aligned} \tag{11.12}$$

с начальными условиями $Q_n(0) = 0$.

16. Продолжение. Рассмотрим аналогичную задачу для процесса, определяемого произвольной системой прямых уравнений. Показать, что $Q_n(t)$ удовлетворяют соответствующим *обратным уравнениям* (при фиксированном k) с заменой $P_{0k}(t)$ на 1.

17. Показать, что вероятности перехода в процессе чистого размножения и в процессе размножения и гибели удовлетворяют уравнениям Колмогорова — Чэпмена (9.1).

18. Пусть $P_{ik}(t)$ удовлетворяют уравнению Колмогорова — Чэпмена (9.1). Допустим, что $P_{ik}(t) > 0$ и что $S_i(t) = \sum_k P_{ik}(t) \leqslant 1$. Доказать, что или $S_i(t) = 1$ при всех t, или $S_i(t) < 1$ при всех t.

19. Эргодическое свойство. Рассмотрим однородный процесс с конечным числом состояний, т. е. предположим, что система дифференциальных уравнений (9.9) конечна и коэффициенты c_j и p_{jk} постоянны. Доказать, что решения — линейные комбинации показательных членов $e^{\lambda(t-\tau)}$, где действительная часть λ отрицательна, если только $\lambda = 0$. Вывести отсюда, что асимптотическое поведение переходных вероятностей такое же, как и в случае конечных цепей Маркова, с той лишь разницей, что периодический случай невозможен.

ОТВЕТЫ К ЗАДАЧАМ

Глава I

1. (а) $\frac{3}{5}$; (б) $\frac{3}{5}$; (в) $\frac{3}{10}$.

2. События S_1, S_2, $S_1 \cup S_2$ и S_1S_2 содержат соответственно 12, 12, 18 и 6 точек.

4. Пространство элементарных событий содержит две точки *ГГ* и *РР* с вероятностями $\frac{1}{4}$, две точки *ГРР* и *РГГ* с вероятностями $\frac{1}{8}$ и вообще две точки с вероятностями 2^{-n} при $n \geqslant 2$. Эти вероятности дают в сумме 1, так что нет необходимости рассматривать возможность бесконечной серии бросаний. Искомые вероятности равны $\frac{15}{16}$ и $\frac{2}{3}$ соответственно.

9. $P\{AB\} = \frac{1}{6}$, $P(A \cup B) = \frac{23}{36}$, $P\{A\bar{B}\} = \frac{1}{3}$.

12. $x = 0$ в событиях (а), (б) и (ж).
$x = 1$ в событиях (д) и (е).
$x = 2$ в событии (г).
$x = 4$ в событии (в).

15. (а) A; (б) AB; (в) $B \cup (AC)$.

16. Верны соотношения (в), (г), (д), (е), (з), (л), (м). Утверждение (а) бессмысленно, исключая тот случай, когда $C \subset B$. Вообще говоря, оно не верно и в этом случае, но имеет место при дополнительном ограничении $AC = 0$. Утверждение (б) верно, если $C \supset AB$. Утверждение (ж) следует читать $(A \cup B) - A = \bar{A}B$. Наконец, (л) есть правильный вариант (к).

17. (а) $A\bar{B}\bar{C}$; (б) $AB\bar{C}$; (в) ABC; (г) $A \cup B \cup C$.
(д) $AB \cup AC \cup BC$; (е) $A\bar{B}\bar{C} \cup \bar{A}B\bar{C} \cup \bar{A}\bar{B}C$;
ж) $\bar{A}BC \cup A\bar{B}C \cup AB\bar{C} = (AB \cup AC \cup BC) - ABC$;
(з) $\overline{\bar{A}\bar{B}C}$; (и) $\overline{ABC}$.

18. $A \cup B \cup C = A \cup (B - AB) \cup \{C - C(A \cup B)\} = A \cup B\bar{A} \cup C\bar{A}\bar{B}$.

Глава II

1. (а) 26^3; (б) $26^2 + 26^3 = 18\,252$; (в) $26^2 + 26^3 + 26^4$. В городе с 20 000 семей или некоторые люди имеют общие инициалы или по крайней мере 1748 людей имеют инициалы более чем из трех букв.

2. $64 \cdot 14 = 896$. Для доски с n^2 полями формула имеет вид $n^2(2n - 2)$.

3. $2(2^{10} - 1) = 2046$.

4. $\binom{n}{2} + n = \frac{n(n+1)}{2}$. 5. (а) $\frac{1}{n}$; (б) $\frac{1}{n(n-1)}$.

6. (а) $p_1 = 0{,}01$, $p_2 = 0{,}27$, $p_3 = 0{,}72$; (б) $p_1 = 0{,}001$, $p_2 = 0{,}063$, $p_3 = 0{,}432$, $p_4 = 0{,}504$.

7. $p_r = (10)_r 10^{-r}$. Например, $p_3 = 0{,}72$, $p_{10} = 0{,}00036288$. Формула Стирлинга дает $p_{10} = 0{,}0003598\ldots$.

8. (а) $(9/10)^k$; (б) $(9/10)^k$; (в) $(8/10)^k$; (г) $2(9/10)^k - (8/10)^k$; (д) AB и $A \cup B$.

9. $\binom{n}{2} n!\, n^{-n}$. 10. $9 : \binom{12}{8} = \frac{1}{55}$.

11. Вероятность того, что потребуется ровно r испытаний, равна $(n-1)_{r-1} : (n)_r = n^{-1}$.

12. (а) $1/1 \cdot 3 \cdot 5 \ldots (2n-1) = 2^n n!/(2n)!$; (б) $(n!)/1 \cdot 3 \ldots (2n-1) = 2^n / \binom{2n}{n}$.

13. В предположении случайности вероятность того, что все двенадцать штрафов придутся или на вторники или на четверги равна $\left(\frac{2}{7}\right)^{12} = 0{,}0000003\ldots$. Имеется лишь $\binom{7}{2} = 21$ пара дней, так что вероятность остается неожиданно малой даже для любой пары дней. Значит, разумно предположить, что полиция действует по системе.

14. В предположении случайности вероятность события равна $\left(\frac{6}{7}\right)^{12}$, или приблизительно $\frac{1}{6}$. Никакого уверенного заключения сделать нельзя.

15. $(90)_{10} : (100)_{10} = 0{,}330476\ldots$.

16. $25!\,(5!)^{-5}\, 5^{-25} = 0{,}00209\ldots$.

17. $\frac{2(n-2)_r (n-r-1)!}{n!} = \frac{2(n-r-1)}{n(n-1)}$.

18. (а) $\frac{1}{216}$; (б) $\frac{83}{3888}$.

19. Вероятности равны $1 - \left(\frac{5}{6}\right)^4 = 0{,}517747\ldots$ и $1 - \left(\frac{35}{36}\right)^{24} = 0{,}491404\ldots$.

20. (а) $(n-N)_r : (n)_r$; (б) $(1 - N/n)^r$. При $r = N = 3$ вероятности равны (а) $0{,}911812\ldots$; (б) $0{,}912673\ldots$. При $r = N = 10$ они равны (а) $0{,}330476$; (б) $0{,}348678\ldots$.

21. (а) $(1 - N/n)^{r-1}$; (б) $(n)_{Nr} : ((n)_N)^r$.

22. $(1 - 2/n)^{2^r - 2}$; для медианы распределения имеет место приближенное равенство $2^{r+1} = 0{,}7n$.

23. В предположении случайности вероятности того, что три из четырех тарелок разбиты (а) одной девочкой, (б) самой младшей из девочек равны соответственно $\frac{13}{64} \approx 0{,}2$ и $\frac{13}{256} \approx 0{,}05$.

24. (а) $12!/12^{12} = 0{,}000054$; (б) $\binom{12}{2}(2^6 - 2)\, 12^{-6} = 0{,}00137\ldots$.

25. $\frac{30!}{2^6 \cdot 6^6} \binom{12}{6} 12^{-30} \approx 0{,}00035\ldots$.

26. (а) $\binom{n}{2r} 2^{2r} : \binom{2n}{2r}$; (б) $n \binom{n-1}{2r-2} 2^{2r-2} : \binom{2n}{2r}$; (в) $\binom{n}{2}\binom{n-2}{2r-4} 2^{2r-4} : \binom{2n}{2r}$.

27. $\binom{N-3}{r-1} : \binom{N-1}{r-1}$.

28. $p=\binom{2N}{N}^2:\binom{4N}{2N}\approx\{2/(N\pi)\}^{\frac{1}{2}}$.

29. $p=\dfrac{\binom{4}{k}\binom{48}{13-k}\binom{39}{13}\binom{26}{13}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}}=\dfrac{\binom{4}{k}\binom{48}{13-k}}{\binom{52}{13}}$.

30. См. задачу 29. Вероятность равна

$$\binom{13}{m}\binom{39}{13-m}\binom{13-m}{n}\binom{26+m}{13-n}:\binom{52}{13}\binom{39}{13}.$$

31. $\binom{4}{k}\binom{48}{26-k}:\binom{52}{26}$.

32. $\dfrac{\binom{13}{a}\binom{39}{13-a}\binom{13-a}{b}\binom{26+a}{13-b}\binom{13-a-b}{c}\binom{13+a+b}{13-c}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}}$.

33. (а) $24p(5, 4, 3, 1)$; (б) $4p(4, 4, 4, 1)$; (в) $12p(4, 4, 3, 2)$.

34. $\dfrac{\binom{13}{a}\binom{13}{b}\binom{13}{c}\binom{13}{d}}{\binom{52}{13}}$. (См. задачу 33, где найдена вероятность того, что набор первого игрока содержит a карт одной масти, b — другой и т. д.

35. $p_0(r)=(52-r)_4:(52)_4$; $p_1(r)=4r(52-r)_3:(52)_4$;
$p_2(r)=6r(r-1)(52-r)_2:(52)_4$;
$p_3(r)=4r(r-1)(r-2)(52-r):(52)_r$; $p_4(r)=(r)_4:(52)_4$.

36. Вероятности того, что времена ожиданий первого, ..., четвертого тузов превосходят r, равны $w_1(r)=p_0(r)$; $w_2(r)=p_0(r)+p_1(r)$; $w_3(r)=p_0(r)+p_1(r)+p_2(r)$; $w_4(r)=1-p_4(r)$. Далее $f_i(r)=w_i(r)-w_i(r+1)$. Медианы равны 8, 20, 32, 44.

37. (а) $\binom{4}{k}\binom{4-k}{k}\binom{48}{r-k}\binom{48-r+k}{r-k}:\binom{52}{r}\binom{52-r}{r}$ при $k\leqslant 2$;
(б) $\left\{\binom{4}{k}\binom{48}{r-k}:\binom{52}{r}\right\}^2$ при $k\leqslant 4$.

39. $\binom{r_1+n-1}{r_1}\binom{r_2+n-1}{r_2}$. 40. $\binom{r_1+5}{5}(r_2+1)$.

41. $\dfrac{(r_1+r_2+r_3)!}{r_1!\,r_2!\,r_3!}$. 42. $(49)_4:(52)_4$.

43.
$$P\{(7)\}=10\cdot 10^{-7}=0{,}000001.$$
$$P\{(6,1)\}=\frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{1!\,6!}\cdot 10^{-7}=0{,}000063.$$
$$P\{(5,2)\}=\frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{2!\,5!}\cdot 10^{-7}=0{,}000189.$$
$$P\{(5,1,1)\}=\frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{1!\,1!\,5!}\cdot 10^{-7}=0{,}001512.$$
$$P\{(4,3)\}=\frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{3!\,4!}\cdot 10^{-7}=0{,}000315.$$

$$P\{4, 2, 1)\} = \frac{10!}{7!\,1!\,1!\,1!}\cdot\frac{7!}{1!\,2!\,4!}\cdot 10^{-7} = 0{,}007560.$$

$$P\{(4, 1, 1, 1)\} = \frac{10!}{6!\,3!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,4!}\cdot 10^{-7} = 0{,}017640.$$

$$P\{(3, 3, 1)\} = \frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{1!\,3!\,3!}\cdot 10^{-7} = 0{,}005040.$$

$$P\{(3, 2, 2)\} = \frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{2!\,2!\,3!}\cdot 10^{-7} = 0{,}007560.$$

$$P\{(3, 2, 1, 1)\} = \frac{10!}{6!\,2!\,1!\,1!}\cdot\frac{7!}{1!\,1!\,2!\,3!}\cdot 10^{-7} = 0{,}105840.$$

$$P\{(3, 1, 1, 1, 1)\} = \frac{10!}{5!\,4!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,1!\,3!}\cdot 10^{-7} = 0{,}105840.$$

$$P\{(2, 2, 2, 1)\} = \frac{10!}{6!\,3!\,1!}\cdot\frac{7!}{1!\,2!\,2!\,2!}\cdot 10^{-7} = 0{,}052920.$$

$$P\{(2, 2, 1, 1, 1)\} = \frac{10!}{5!\,3!\,2!}\cdot\frac{7!}{1!\,1!\,1!\,2!\,2!}\cdot 10^{-7} = 0{,}317520.$$

$$P\{(2, 1, 1, 1, 1, 1)\} = \frac{10!}{4!\,5!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,1!\,1!\,2!}\cdot 10^{-7} = 0{,}317520.$$

$$P\{(1, 1, 1, 1, 1, 1, 1)\} = \frac{10!}{3!\,7}\cdot 7!\cdot 10^{-7} = 0{,}060480.$$

44. Пусть S, D, T, Q означает, что в какой-то день родился один, два, три и четыре человека соответственно. Тогда

$$P\{22S\} = \frac{365!}{22!\,343!}\cdot 365^{-22} = 0{,}52430.$$

$$P\{20S+1D\} = \frac{365!}{20!\,1!\,344!}\cdot\frac{22!}{20!\,2!}\cdot 365^{-22} = 0{,}35208.$$

$$P\{18S+2D\} = \frac{365!}{18!\,2!\,345!}\cdot\frac{22!}{18!\,2!\,2!}\cdot 365^{-22} = 0{,}09695.$$

$$P\{16S+3D\} = \frac{365!}{16!\,3!\,346!}\cdot\frac{22!}{16!\,2!\,2!\,2!}\cdot 365^{-22} = 0{,}01429.$$

$$P\{19S+1T\} = \frac{365!}{19!\,1!\,345!}\cdot\frac{22!}{19!\,3!}\cdot 365^{-22} = 0{,}00680.$$

$$P\{17S+1D+1T\} = \frac{365!}{17!\,1!\,1!\,346!}\cdot\frac{22!}{17!\,2!\,3!}\cdot 365^{-22} = 0{,}00336.$$

$$P\{14S+4D\} = \frac{365!}{14!\,4!\,347!}\cdot\frac{22!}{14!\,2!\,2!\,2!\,2!}\cdot 365^{-22} = 0{,}00124.$$

$$P\{15S+2D+1T\} = \frac{365!}{15!\,2!\,1!\,347!}\cdot\frac{22!}{15!\,2!\,2!\,3!}\cdot 365^{-22} = 0{,}00066.$$

$$P\{18S+1Q\} = \frac{365!}{18!\,1!\,346!}\cdot\frac{22!}{18!\,4!}\cdot 365^{-22} = 0{,}00009.$$

45. Пусть $q = \binom{52}{5} = 2598960$. Вероятности равны:

(а) $4/q$; (б) $13\cdot 12\cdot 4\cdot q^{-1} = \frac{1}{4165}$;

(в) $13\cdot 12\cdot 4\cdot 6\cdot q^{-1} = \frac{6}{4165}$;

(г) $9 \cdot 4^5 q^{-1} = \frac{768}{216580}$;

(д) $13 \cdot \binom{12}{2} \cdot 4 \cdot 4^2 q^{-1} = \frac{88}{4165}$;

(е) $\binom{13}{2} \cdot 11 \cdot 6 \cdot 6 \cdot 4 \cdot q^{-1} = \frac{198}{4165}$;

(ж) $13 \cdot \binom{12}{3} \cdot 6 \cdot 4^3 \cdot q^{-1} = \frac{1760}{4165}$.

Глава IV

1. 99/323. 2. 0,21 ... . 3. 1/4. 4. $7/2^6$. 5. 1/81 и $31/6^6$.

6. Если A_k есть событие, что (k, k) не появится, то из (1.5) следует, что

$$1 - p_r = 6\left(\frac{35}{36}\right)^r - \binom{6}{2}\left(\frac{34}{36}\right)^r + \binom{6}{3}\left(\frac{33}{36}\right)^r - \binom{6}{4}\left(\frac{32}{36}\right)^r + 6\left(\frac{31}{36}\right)^r - \left(\frac{30}{36}\right)^r.$$

7. Положим $p^{-1} = \binom{52}{13}$. Тогда $S_1 = 13\binom{48}{9} p$; $S_2 = \binom{13}{2}\binom{44}{5} p$; $S_3 = 40\binom{13}{3} p$. Приближенные числовые значения: $P_{[0]} = 0{,}9658$; $P_{[1]} = 0{,}0341$; $P_{[2]} = 0{,}0001$.

8. $$u_r = \sum_{k=0}^{N} (-1)^k \binom{N}{k} \left(1 - \frac{k}{n}\right)^r.$$

9. $$p_r = \sum_{k=0}^{N} (-1)^k \binom{N}{k} \frac{(n-k)_r}{(n)_r}.$$ См. (12.18) гл. II, где доказывается, что эти две формулы совпадают.

10. Общий член имеет вид $a_{1k_1} a_{2k_2} \ldots a_{Nk_N}$, где $(k_1, k_2, \ldots, k_N)$ есть перестановка $(1, 2, \ldots, N)$. Для диагональных элементов $k_\nu = \nu$.

12. $$u_r = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{(ns-ks)_r}{(ns)_r}.$$

14. Заметить, что, по определению, $u_r = 0$ при $r < n$ и $u_n = n!\, s^n/(ns)_n$.

15. $$u_r - u_{r-1} = \sum_{k=1}^{n} (-1)^{k-1} \binom{n-1}{k-1} \frac{(ns-ks)_{r-1}}{(ns-1)_{r-1}} \to$$
$$\to \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} \left(1 - \frac{k+1}{n}\right)^{r-1}.$$

16. $$\binom{N}{2}^{-r} \binom{N}{m} \sum_{k=2}^{m} (-1)^{m-k} \binom{m}{k} \binom{k}{2}^r.$$

17. Воспользоваться тем, что $\binom{52}{5} S_k = \binom{4}{k}\binom{52-13k}{5}$. $P_{[0]} = 0{,}264$, $P_{[1]} = 0{,}588$, $P_{[2]} = 0{,}146$, $P_{[3]} = 0{,}002$ приблизительно.

18. Воспользоваться тем, что $\binom{52}{13} S_k = \binom{4}{k}\binom{52-2k}{13-2k}$. $P_{[0]} = 0{,}780217$, $P_{[1]} = 0{,}204606$, $P_{[2]} = 0{,}014845$, $P_{[3]} = 0{,}000330$, $P_{[4]} = 0{,}000002$ приблизительно.

19. $m!\,N!\,u_m = \sum\limits_{k=0}^{N-m} (-1)^k (N-m-k)!/k!$.

20. См. следующую формулу при $r=2$.

21. $(rN)!\ x = \binom{N}{2} r^2 (rN-2)! - \binom{N}{3} r^3 (rN-3)! + - \ldots +$
$+ (-1)^N r^N (rN-N)!$.

24. $$P_{[m]} = \frac{\binom{n}{m}}{\binom{n+r-1}{r}} \sum_{k=0}^{n-m} (-1)^k \binom{n-m}{k}\binom{n-m+r-1-k}{r}.$$

25. Воспользоваться формулами (12.16) и (12.4) гл. II.

26. Положить $U_N = A_1 \cup \ldots \cup A_N$ и заметить затем, что
$$U_{N+1} = U_N \cup A_{N+1} \quad \text{и} \quad U_N A_{N+1} = (A_1 A_{N+1}) \cup \ldots \cup (A_N A_{N+1}).$$

Глава V

1. $1 - \frac{(5)_3}{(6)_3} = \frac{1}{2}$. 2. $p = 1 - \frac{10 \cdot 5^9}{6^{10} - 5^{10}} = 0{,}61\ \ldots$.

3. (а) $\binom{35}{13} : \binom{39}{13} = 0{,}182\ \ldots$. Вероятность появления в точности одного туза равна $4 \cdot \binom{35}{12} : \binom{39}{13} = 0{,}411\ \ldots$. (б) $1 - 0{,}182 - 0{,}411 = 0{,}407$ приблизительно.

4. (а) $2 \cdot \frac{\binom{23}{10}}{\binom{26}{13}} = \frac{11}{50}$; (б) $2 \cdot \frac{\binom{23}{12}}{\binom{26}{13}} = \frac{13}{50}$.

6. $\frac{125}{345}$; $\frac{140}{345}$; $\frac{80}{345}\ \ldots$. 7. $\frac{20}{21}$. 9. $\left(\frac{5}{6}\right)^2$.

10. $1 - \left(\frac{5}{6}\right)^2$. 12. $\frac{p}{2-p}$. 13. (б) $\frac{3}{5}$; (в) $2^n \cdot (1+2^n)^{-1}$.

14. (г) Положим $a_n = x_n - \frac{4}{7}$; $b_n = y_n - \frac{1}{7}$; $c_n = z_n - \frac{2}{7}$. Тогда $|a_n| + |b_n| + |c_n| = \frac{1}{2}\{|a_{n+1}| + |b_{n+1}| + |c_{n+1}|\}$. Значит, $|a_n| + |b_n| + |c_n|$ возрастает по геометрической прогрессии.

15. $p = (1-p_1)(1-p_2)\ldots(1-p_n)$.

16. Воспользоваться тем, что $1 - x < e^{-x}$ при $0 < x < 1$, или рядом Тейлора для $\log(1-x)$; см. (8.12) гл. II.

18. $\frac{b+c}{b+c+r}$.

19. Если утверждение справедливо для n-го извлечения, каковы бы ни были b, r и c, то вероятность появления черного шара при $(n+1)$-м испытании равна
$$\frac{b}{b+r} \cdot \frac{b+c}{b+r+c} + \frac{r}{b+r} \cdot \frac{b}{b+r+c} = \frac{b}{b+r}.$$

20. Предыдущая задача устанавливает, что утверждение справедливо при $m=1$ и любом n. Для индукции рассмотреть две возможности при первом испытании.

23. Воспользоваться формулой (12.9) гл. II.

26. Из (5.2) следует $2v=2p(1-p)\leqslant\frac{1}{2}$.

28. (а) u^2; (б) $u^2+uv+v^2/4$; (в) $u^2+(25uv+9v^2+vw+2uw)/16$.

33. $p_{11}=p_{32}=2p_{21}=p$, $p_{12}=p_{33}=2p_{23}=q$, $p_{13}=p_{31}=0$, $p_{22}=\frac{1}{2}$.

Глава VI

1. $\frac{5}{16}$. 2. Вероятность равна 0,02804 .... 3. $(0,9)^x\leqslant 0,1$, $x\geqslant 22$.

4. $q^x\leqslant\frac{1}{2}$ и $(1-4q)^x\leqslant\frac{1}{2}$, где $p=\binom{48}{9}:\binom{52}{13}$. Следовательно, $x\geqslant 263$ и $x\geqslant 66$ соответственно.

5. $1-(0,8)^{10}-2(0,8)^9=0,6242\ldots$.

6. $\{1-(0,8)^{10}-2(0,8)^9\}/\{1-(0,8)^{10}\}=0,6993\ldots$.

7. $\binom{26}{2}\binom{26}{11}:\binom{52}{13}=0,003954\ldots$, и $\binom{13}{2}\frac{1}{2^{13}}=0,00952\ldots$.

8. $\binom{12}{2}\{6^{-6}-2\cdot 12^{-6}\}$.

9. Истинные значения: 0,6651 ..., 0,40187 ..., и 0,2009 .... Приближенная формула Пуассона дает $1-e^{-1}=0,6321\ldots$, 0,3679 ... и 0,1839 ....

10. $e^{-2}\sum_{4}^{\infty}2^k/k!=0,143\ldots$. 11. $e^{-1}\sum_{3}^{\infty}1/k!=0,080\ldots$.

12. $e^{-x/100}\leqslant 0,05$ или $x\geqslant 300$.

13. $e^{-1}=0,3679\ldots$, $1-2e^{-1}=0,264\ldots$.

14. $e^{-x}\leqslant 0,01$, $x\geqslant 5$. 15. $1/p=649\,740$.

16. $1-p^n$, где $p=p(0;\lambda)+\ldots+p(k;\lambda)$.

18. q^3 при $k=0$; pq^3 при $k=1, 2, 3$ и pq^3-pq^6 при $k=4$.

19. $\sum_{k=0}^{n}\binom{n}{k}^2 2^{-2n}=\binom{2n}{n}2^{-2n}\approx\left(\frac{1}{\pi n}\right)^{1/2}$ при больших n.

20. $\sum_{k=a}^{a+b-1}\binom{a+b-1}{k}p^k q^{a+b-1-k}$. Эту формулу можно переписать в другом виде: $p^a\sum_{k=0}^{b-1}\binom{a+k-1}{k}q^k$, где k-й член равен вероятности того, что a-му успеху предшествует $k\leqslant b-1$ неудач.

21. $x_r=\binom{2N-1-r}{r-1}\cdot 2^{-2N+r+1}$.

22. (а) $x=\sum_{r=1}^{N}x_r 2^{-r-1}=2^{-2N}\sum_{r=1}^{N}\binom{2N-1-r}{N-1}$. (б) Воспользоваться формулой (12.6) гл. II.

23. $k_i\approx np_i$, $k_{12}\approx np_{12}$, где $n\approx k_1k_2/k_{12}$.

24. $\binom{n}{n_1}\cdot\binom{n-s_1}{n_2}\cdots\binom{n-s_{r-1}}{n_r}\cdot q^{s_r}p^{(rn-s_1-\ldots-s_r)}$,

где
$$s_i = n_1 + \ldots + n_i.$$

25. $p = p_1q_2(p_1q_2 + p_2q_1)^{-1}$.

31. Из разложения Тейлора для логарифма следует
$$b(0;\ n,\ p) = q^n = \left(1 - \frac{\lambda}{n}\right)^n < e^{-\lambda} = p(0;\ \lambda).$$

Члены каждого распределения дают в сумме единицу, и поэтому невозможно, чтобы *все* члены одного распределения были больше соответствующих членов другого.

32. Имеется только конечное число членов распределения Пуассона, которые больше, чем ε, и оставшиеся члены превосходят соответствующие члены биномиального распределения.

Глава VII

1. Продолжить рассуждения § 1.
2. Воспользоваться формулой (1.7).
3. $\Phi\left(-\frac{32}{30}\right) = 0{,}143\ldots$. 4. 0,99. 5. 500. 6. 66 400.

7. Это можно утверждать с полной уверенностью. Из неравенств гл. VI следует, что вероятность превзойти восьмикратное стандартное уклонение исчезающе мала.

8. $(2\pi n)^{-1}\{p_1p_2(1-p_1-p_2)\}^{-1/2}$.

Глава VIII

1. $\beta = 21$.
2. $x = pu + qv + rw$, где u, v, w — решения системы
$$u = p^{\alpha-1} + (qv + rw)\frac{1-p^{\alpha-1}}{1-p}, \quad v = (pu + rw)\frac{1-q^{\beta-1}}{1-q},$$
$$w = pu + qv + rw = x.$$

3. $u = p^{\alpha-1} + (qv + rw)\dfrac{1-p^{\alpha-1}}{1-p}$,
$$v = (pu + rw)\frac{1-q^{\beta-1}}{1-q}, \quad w = (pu + qv)\frac{1-r^{\gamma-1}}{1-r}.$$

4. Заметить, что $P\{A_n\} < (2p)^n$, но
$$P\{A_n\} > 1 - (1-p^n)^{2^n/2n} > 1 - e^{-(2p)^n/2n}.$$

Если $p = \frac{1}{2}$, то последняя величина эквивалентна $\frac{1}{2}n$; если $p > 1$, то $P\{A_n\}$ даже не стремится к нулю.

Глава IX

1. Возможны комбинации (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (2, 0), (2, 1) (3, 0). Их вероятности равны 0,047539, 0,108883, 0,017850, 0,156364, 0,214197, 0,321295, 0,026775, 0,107098.

2. (а) Совместное распределение задается матрицей вида 6×6. Главная диагональ состоит из элементов q, $2q$, ..., $6q$, где $q=\frac{1}{36}$. По одну сторону от главной диагонали все элементы равны 0, по другую q. (б) $E(X)=\frac{7}{2}$, $D(X)=\frac{35}{12}$, $E(Y)=\frac{161}{36}$, $D(Y)=\frac{2555}{1296}$, $\operatorname{Cov}(X, Y)=\frac{105}{72}$.

3. Перед всеми строками в дальнейшем подразумевается множитель 32^{-1}. Строки совместного распределения X, Y равны (1, 0, 0, 0, 0, 0), (0, 5, 4, 3, 2, 1), (0, 0, 6, 6, 3, 0), (0, 0, 0, 1, 0, 0); для X, Z — (1, 0, 0, 0, 0, 0), (0, 5, 6, 1, 0, 0), (0, 0, 4, 6, 1, 0), (0, 0, 0, 3, 2, 0), (0, 0, 0, 0, 2, 0), (0, 0, 0, 0, 0, 1); для Y, Z — (1, 0, 0, 0), (0, 5, 6, 1), (0, 4, 7, 0), (0, 3, 2, 0), (0, 2, 0, 0), (0, 1, 0, 0). Распределение $X+Y$ — (1, 0, 5, 4, 9, 8, 5), значения $X+Y$ меняются от 0 до 6. Распределение $X\cdot Y$ — (1, 5, 4, 3, 8, 1, 6, 0, 3, 1), значения меняются от 0 до 9.

$$E(X)=\frac{5}{2},\quad E(Y)=\frac{3}{2},\quad E(Z)=\frac{31}{16},\quad D(X)=\frac{5}{4},$$
$$D(Y)=\frac{3}{8},\quad D(Z)=\frac{303}{256}.$$

4. $P\{Z=i,\ X=j\}=q^{i+j}p^2$, если $i>j$ и $=(1-q^{i+1})\,q^ip$, если $i=j$; другие значения невозможны. $P\{Z=i\}=2q^ip-q^{2i}p-q^{2i+1}p$.

8. Распределение V_n задается формулой (3.5), для U_n оно получается по симметрии.

9. $P\{X\leqslant r,\ Y\geqslant s\}=\left(\frac{r-s+1}{N}\right)^n$ при $r\geqslant s$;

$$P\{X=r,\ Y=s\}=\begin{cases} N^{-n}\left\{(r-s+1)^{-n}-2(r-s)^n+(r-s-1)^n\right\}, & \text{если } r>s;\\ N^{-n}, & \text{если } r=s.\end{cases}$$

10. $x=\frac{r^{n-2}-(r-1)^{n-2}}{r^n-(r-1)^n}$, если $j<r$ и $k<r$; $x=\frac{r^{n-2}}{r^n-(r-1)^n}$, если $j\leqslant r$ и $k=r$, или $j=r$ и $k\leqslant r$; $x=0$, если $j>r$ или $k>r$.

11. $\sigma^2\approx\frac{nN^2}{(n+1)^2(n+2)}$.

12. $P\{N=n,\ K=k\}=\binom{n}{k}p^{n-k}(qq')^k qp'$,

$$P\{N=n\}=(1-qp')^n qp',$$
$$P\{K=k\}=(qq')^k qp'\sum\binom{-k-1}{\nu}(-p)^\nu=p'q'^k.$$

13. $E\left(\frac{K}{N+1}\right)=\sum kp_{k,\,n}/(n+1)=$

$$=q^2p'q'\sum_{n=1}^{\infty}\left(1-\frac{1}{n+1}\right)(p+qq')^{n-1}=\frac{qq'}{1-qp'}-\frac{p^2p'q'}{(1-qp')^2}\log\frac{1}{qp'}.$$

$$E(K)=\frac{q'}{p'};\quad E(N)=\frac{1-qp'}{qp'};\quad \operatorname{Cov}(K, N)=\frac{q}{qp'^2}.$$
$$\rho(K, N)=\left\{\frac{q'}{(1-qp')}\right\}^2.$$

14. $p_k = p^k q + q^k p$; $E(X) = pq^{-1} + qp^{-1}$; $D(X) = pq^{-2} + qp^{-2} - 2$.
15. $q_k = p^2 q^{k-1} + q^2 p^{k-1}$; $P\{X = m,\ Y = n\} = $
$= p^{m+1}q^n + q^{m+1}p^n$, где m и $n \geqslant 1$;
$E(Y) = 2$; $\sigma^2 = 2(pq^{-1} + qp^{-1} - 1)$.
17. $\binom{n}{k} 364^{n-k}\, 365^{1-n}$.
18. (а) $365\{1 - 364^n \cdot 365^{-n} - n\, 364^{n-1} \cdot 365^{-n}\}$; (б) $n \geqslant 28$.
19. (а) $\mu = n$, $\sigma^2 = (n-1)n$; (б) $\mu = (n+1)/2$, $\sigma^2 = (n^2-1)/12$.
20. $E(X) = np_1$; $D(X) = np_1(1-p_1)$; $\mathrm{Cov}(X, Y) = -np_1p_2$.
21. $-n/36$. Это есть частный случай задачи 20.
25. $E(Y_r) = \sum_{k=1}^{r} \frac{N}{r-k+1}$; $D(Y_r) = \sum_{k=1}^{r} \frac{N(N-r+k-1)}{(r-k+1)^2}$.
26. (а) $1 - q^k$; (б) $E(X) = N\left\{1 - q^k + \frac{1}{k}\right\}$; (в) $\frac{dE(X)}{dk} = 0$.
27. $\sum (1-p_j)^n$. Пусть $X_j = 0$ или 1, в соответствии с тем представлен или не представлен j-й класс.
28. $E(X) = \frac{r_1(r_2+1)}{r_1+r_2}$; $D(X) = \frac{r_1r_2(r_1-1)(r_2+1)}{(r_1+r_2-1)(r_1+r_2)^2}$.
30. $E(S_n) = \frac{nb}{b+r}$; $D(S_n) = \frac{nbr\{b+r+nc\}}{(b+r)^2(b+r+c)}$.
33.
$$E\left(\frac{r}{X}\right) = r\sum_{k=r}^{\infty} k^{-1}\binom{k-1}{r-1} p^r q^{k-r} =$$
$$= \sum_{k=1}^{r-1} (-1)^{k-1}\frac{r}{r-k}\left(\frac{p}{q}\right)^k + \left(\frac{-p}{q}\right)^r r\log p.$$

Чтобы вывести последнюю формулу из первой, положим $f(q) = r\sum k^{-1}\binom{k-1}{r-1} q^k$. Используя (12.4) гл. II, найдем $f'(q) = rq^{r-1}(1-q)^{-r}$. Утверждение теперь доказывается повторным интегрированием по частям.

Глава XI

1. $sP(s)$ и $P(s^2)$.
2. (а) $(1-s)^{-1}P(s)$; (б) $(1-s)^{-1}sP(s)$; (в) $\{1 - sP(s)\}/(1-s)$; (г) $p_0 s^{-1} + \{1 - s^{-1}P(s)\}/(1-s)$; (д) $\frac{1}{2}\{P(s^{1/2}) + P(-s^{1/2})\}$.
3. $U(s) = pqs^2/(1-ps)(1-qs)$. Математическое ожидание равно $1/pq$; $D = (1-3pq)/p^2q^2$.
6. Нуль может быть первым, вторым, ..., и поэтому $U(s) = \sum F^k(s)$.
7. Производящая функция равна $\{1 - F(s)\}(1-s)^{-1} = (1+s)U(s)$.
8. Производящая функция равна $\sum\left\{\frac{1}{2}F(s)\right\}^k = 2F(s)s^{-2} - 1$.
9. Та же самая производящая функция.
10. k-й нуль должен появиться при испытании с номером $2r \leqslant n$, и последующие $n - 2r$ испытаний не должны иметь своим результатом нуль.
11. Воспользоваться очевидной аналогией с (1.6) для случая, когда $P(1) < 1$.

12. Используя производящую функцию для геометрического распределения X_ν, мы получаем без всяких вычислений

$$P_r(s) = s^r\left(\frac{N-1}{N-s}\right)\left(\frac{N-2}{N-2s}\right)\cdots\left(\frac{N-r+1}{N-(r-1)s}\right).$$

13. $P_r(s)\{N-(r-1)s\} = P_{r-1}(s)(N-r-1)s.$

14. $P_r(s) = \dfrac{s}{N-(N-1)s}\cdot\dfrac{2s}{N-(N-2)s}\cdots\dfrac{rs}{N-(N-r)s}.$

15. S_r есть сумма r независимых случайных величин с общим геометрическим распределением. Поэтому

$$P_r(s) = \left(\frac{q}{1-ps}\right)^r, \quad p_{r,k} = q^r p^k\binom{r+k-1}{k}.$$

16. $P(R=r) = \sum_{k=0}^{\nu-1} P\{S_{r-1}=k\}\, P\{X_r \geqslant \nu-k\} =$

$$= \sum_{k=0}^{\nu-1} q^{r-1}p^k\binom{r+k-2}{k}p^{\nu-k} = p^\nu q^{r-1}\binom{r+\nu-2}{\nu-1}.$$

$$E(R) = 1+\frac{q\nu}{p}, \quad D(R) = \frac{\nu q}{p^2}.$$

21. $u_n = q^n + \sum_{k=3}^{n}\binom{k-1}{2}p^3q^{k-3}u_{n-k}$, где $u_0=1$, $u_1=q$, $u_2=q^2$, $u_3=p^3+q^3$. Используя тот факт, что это рекуррентное соотношение типа свертки, получим,

$$U(s) = \frac{1}{1-qs} + \frac{(ps)^3}{(1-qs)^3}U(s).$$

22. $u_n = pw_{n-1}+qu_{n-1}$, $v_n = pu_{n-1}+qv_{n-1}$, $w_n = pv_{n-1}+qw_{n-1}$. Значит, $U(s)-1 = psW(s)+qsU(s)$; $V(s) = psU(s)+qs\cdot V(s)$; $W(s) = psV(s)+qsW(s)$.

Глава XIII

1. Достаточно показать, что для всех корней $s \neq 1$ уравнения $F(s)=1$ выполнено неравенство $|s| \geqslant 1$ и что $|s|=1$ возможно только в периодическом случае.

2. $u_{2n} = \left\{\binom{2n}{n}2^{-2n}\right\}^r \sim (\pi n)^{-\frac{1}{2}r}$. Поэтому $\mathfrak{E}$ невозвратно только при $r=2$. При $r=3$ метод касательных для численного интегрирования дает

$$\sum_{n=1}^{\infty} u_{2n} \sim \pi^{-\frac{3}{2}}\int_{\frac{1}{2}}^{\infty} x^{-\frac{3}{2}}\,dx = \left(\frac{2}{\pi}\right)^{\frac{3}{2}} \approx \frac{1}{2}.$$

Поэтому из (3.5) следует, что вероятность того, что $\mathfrak{E}$ когда-нибудь произойдет, равна приблизительно $\frac{1}{3}$. Более точное значение суммы 0,47, что дает $x=0{,}32$.

3. $S_n=0$ только в том случае, когда $n=k(a+b)$ и биномиальное распределение показывает, что для таких n мы имеем $P(S_n=0)\sim$ $\sim(a+b)^{\frac{1}{2}}(2\pi abk)^{-\frac{1}{2}}$. Ряд расходится.

4. Из того, что $\sum f_k+P\{X_1>0\}\leqslant 1$, заключить, что $f<1$, кроме того случая, когда $P\{X_1>0\}=0$. В этом случае все $X_k\leqslant 0$ и $\mathfrak{E}$ произойдет при первом испытании или никогда.

5. Пусть $u_1,\ldots,u_N$ заданы и $u_n=p_1u_{n-1}+p_2u_{n-2}+\ldots+p_Nu_{n-N}$ при $n\geqslant N$.
Тогда

$$\lim u_n=\frac{u_1p_N+u_2(p_N+p_{N-1})+\ldots+u_N(p_1+p_2+\ldots+p_N)}{p_1+2p_2+3p_3+\ldots}.$$

Если $u_k=N^{-1}$, то $\lim u_n=N^{-1}(u_1+2u_2+\ldots+Nu_N)$.

6. $U_1(s)=\frac{1}{2}+\frac{1}{2}U(s),\ F_1=\sum_{k=1}^{\infty}\left(\frac{1}{2}F\right)^k=s^{-2}F^2(s)$.

7. $F_2(s)=\frac{1}{2}s\{1+F_1(s)\}=s^{-1}F(s)$.

8. $U_2(s)=\{1-F_2(s)\}^{-1}=\frac{1}{2}+\frac{1}{2}(1+s)(1-s^2)^{-\frac{1}{2}}$. Это показывает, что вероятность в момент $2n$ впервые достигнуть *положительной* точки равна половине вероятности того, что $S_{2n}=0$.

9. (а) $F(s)=qs(1-qs^r)^{-1},\ \mu=1+rpq^{-1},\ \sigma^2=r^2pq^{-2}$; (б) $Z_n=n-N_n$, $E(Z_n)\sim npq(q+rp)^{-1},\ D(Z_n)=nr^2pq(q+pr)^{-2}$.

10. $U(s)=1+qs+\ldots+q^{r-1}s^{r-1}+q^rs^r(1-s)^{-1},\ \mu^{-1}=q^r$.

11. $N_n^*\approx(N_n-714{,}3)/22{,}75$; $\Phi\left(\frac{2}{3}\right)-\Phi\left(-\frac{2}{3}\right)\approx\frac{1}{2}$.

12. $r_n=r_{n-1}-\frac{1}{4}r_{n-2}+\frac{1}{8}r_{n-3}$, причем $r_0=r_1=r_2=1$; $R(s)=$ $=(8+2s^2)(8-8s+2s^2-s^3)^{-1}$; $r_n\sim 1{,}444248\,(1{,}139680)^{-n-1}$.

14. Если a_n есть вероятность того, что A-серия длины r появится при n-м испытании, то $A(s)$ задается формулой (7.5) с заменой p на α и q на $1-\alpha$. Пусть $B(s)$ и $C(s)$ — соответствующие функции для B и C-серий. Искомые производящие функции равны $F(s)=1-U^{-1}(s)$, где в случае (а) $U(s)=A(s)$; в случае (б) $U(s)=A(s)+B(s)-1$ и в случае (в) $U(s)=A(s)+B(s)+C(s)-2$.

15. Скомбинировать методы примера 8, б и задачи 14.

17. $u_n=Np,\ v_k(\infty)=Npq^k$.

19. Заметим, что $1-F(s)=(1-s)Q(s)$ и $\mu-Q(s)=(1-s)R(s)$, поэтому $Q(1)=\mu,\quad 2R(1)=\sigma^2-\mu+\mu^2$. Степенной ряд $Q^{-1}(s)=$ $=\sum(u_n-u_{n-1})s^n$ при $s=1$ сходится.

Глава XIV

1. Вероятность разорения задается формулой (2.4), где $p=\alpha(1-\gamma)^{-1}$, $q=\beta(1-\gamma)^{-1}$. Средняя продолжительность игры равна $D_z(1-\gamma)^{-1}$, где D_z определяется из (3.4) или (3.5).

2. Граничные условия (2.2) заменятся на $q_0 - \delta q_1 = 1 - \delta$, $q_a = 0$. Новое решение имеет вид

$$q_z = \{(q/p)^a - (q/p)^z\}(1-\delta) : \{(q/p)^a(1-\delta) + \delta q/p - 1\}.$$

Граничные условия для (3.2) принимают вид $D_0 = \delta D_1$, $D_a = 0$.

3. Формуле (2.1) соответствует $q_2 = pq_{z+2} + qq_{z-1}$ и $q_z = \lambda^z$ есть частное решение, если $\lambda = p\lambda^3 + q$, т. е. если $\lambda = 1$ или $\lambda^2 + \lambda = qp^{-1}$. Вероятность разорения равна

$$q_z = \begin{cases} 1, & \text{если } q \geqslant 2p, \\ \left\{\left(\dfrac{1}{4} + \dfrac{q}{p}\right)^{\frac{1}{2}} - \dfrac{1}{2}\right\}^z, & \text{если } q \leqslant 2p. \end{cases}$$

5. $w_{z,n+1}(x) = pw_{z+1,n}(x) + qw_{z-1,n}(x)$ с граничными условиями $w_{0,n}(x) = w_{a,n}(x) = 0$ при $n \geqslant 1$; (2) $w_{z,0}(x) = 0$ при $z \neq x$ и $w_{x,0}(x) = 1$.

6. Заменить (1) на $w_{0,n}(x) = w_{1,n}(x)$ и $w_{a,n}(x) = w_{a-1,n}(x)$.

10. $P\{M_n < z\} = \sum\limits_{x=1}^{\infty} (v_{x-z,n} - v_{x+z,n})$,

$P\{M_n = z\} = P\{M_n < z+1\} - P\{M_n < z\}$.

11. Первое достижение точки x происходит в момент $k \leqslant n$, и частица возвращается из x в x в последующие $n-k$ шагов.

Глава XV

1. P имеет строки $(p, q, 0, 0)$, $(0, 0, p, q)$, $(p, q, 0, 0)$ и $(0, 0, p, q)$. При $n > 1$ строки равны (p^2, pq, pq, q^2).

2. (а) Цепь неприводима и эргодична; $p_{jk}^{(n)} \to \frac{1}{3}$ для любых j, k. (Заметить, что матрица P — двойная стохастическая.) (б) Цепь имеет период 3, множество G_1 состоит из E_1 и E_2, состояния E_4 образует G_2 и E_3 образует G_3. Имеем $u_1 = u_2 = \frac{1}{2}$, $u_3 = u_4 = 1$. (в) Состояния E_1 и E_3 образуют замкнутое множество S_1, а E_4, E_5 — другое замкнутое множество S_2, в то время как E_2 невозвратно. Матрицы, отвечающие замкнутым множествам, имеют вид (2×2)-матриц с элементами $\frac{1}{2}$. Поэтому $p_{jk}^{(n)} \to \frac{1}{2}$, если E_j и E_k принадлежат одному S_r; $p_{j2}^{(n)} \to 0$; наконец $p_{2k}^{(n)} \to \frac{1}{2}$, если $k = 1, 3$ и $p_{2k}^{(n)} \to 0$, если $k = 2, 4, 5$. (г) Цепь имеет период 3. Положив $a = \left(0, 0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$, $b = (1, 0, 0, 0, 0, 0)$, $c = \left(0, \frac{1}{2}, \frac{1}{2}, 0, 0, 0\right)$, мы найдем, что строки $P^2 = P^5 = \ldots$ равны a, b, b, c, c, c; строки $P^3 = P^6 = \ldots$ равны b, c, c, a, a, a, и, наконец, строки $P = P^4 = \ldots$ равны c, a, a, b, b, b.

3. $p_{jj}^{(n)} = (j/6)^n$, $p_{jk}^{(n)} = (k/6)^n - ((k-1)/6)^n$, если $k > j$, и $p_{jk}^{(n)} = 0$, если $k < j$.

4. $x_k = \left(\frac{3}{4}, \frac{1}{2}, \frac{1}{4}, \frac{1}{2}\right)$, $y_k = \left(\frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{1}{2}\right)$.

6. Положим $\mu = \sum np_n$. Состояние является нулевым, если $\mu = \infty$. Стационарное распределение имеет вид $u_k = \mu^{-1}(p_k + p_{k+1} + \ldots)$.

7. Состояния эргодичны, если $\sum (1-q_0)(1-q_1)\ldots(1-q_{n-1}) < \infty$. Стационарное распределение пропорционально членам ряда.

8. $u_r = (p/q)^r (q-p)/p$.

9. $u_r = \{1-p/q\}(p/q)^{r-1} : \{1-(p/q)^a\}$.

10. $p_{jj} = 2j(N-j)/N^2$, $p_{j,\,j+1} = (N-j)^2/N^2$,

$$p_{j,\,j-1} = j^2/N^2, \quad u_k = \binom{N}{k}^2 : \binom{2N}{N}.$$

13.
$$P = \begin{Vmatrix} q & p & & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 \ldots & 0 & 1 \\ q & p & 0 & 0 \ldots & 0 & 0 \end{Vmatrix}.$$

14. Заметить, что матрица двойная стохастическая; использовать пример 6, б.

15. Положим $p_{k,\,k+1} = 1$ при $k = 1, \ldots, N-1$ и $p_{Nk} = p_k$.

16. $\sum u_j p_{jk} = u_k$, тогда $U(s) = u_0(1-s)\{P(s)-s\}^{-1}$. Для эргодичности необходимо и достаточно, чтобы $P'(1) < 1$.

23. Пусть M — максимум x_j. Рассмотреть состояние E_r, для которого $x_r = M$.

26. Если $N \geqslant m-2$, то случайные величины $X^{(m)}$ и $X^{(n)}$ независимы и, следовательно, три строки матрицы $p_{jk}^{(m,\,n)}$ совпадают с распределением $X^{(n)}$, а именно $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$. При $n = m+1$ три строки равны соответственно $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$, $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$, $\left(0, \frac{1}{2}, \frac{1}{2}\right)$.

Глава XVII

3. $E(X) = ie^{\lambda t}$; $D(X) = ie^{\lambda t}(e^{\lambda t}-1)$.

4. $P_n' = -\lambda n P_n + \lambda(n+1)P_{n+1}$.

$$P_n = \binom{i}{n} e^{-i\lambda t}(e^{\lambda}-1)^{i-nt} \qquad (n \leqslant i).$$

$$E(X) = \quad {}^{-\lambda t}; \; D(X) = ie^{-\lambda t}(1-e^{-\lambda t}).$$

5. $P_n'(t) = -(\lambda + n\mu)P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t)$ при $n \leqslant N-1$ и $P_N'(t) = -N\mu P_N(t) + \lambda P_{N-1}(t)$.

6. Процесс гибели и размножения $\lambda_n = \lambda$ и $\mu_n = \mu$.

19. Стандартный метод решения линейных дифференциальных уравнений приводит к системе линейных алгебраических уравнений. См. указание содержащееся в примечании 3 к гл. XVI.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

В. ФЕЛЛЕР

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИЛОЖЕНИЯ

Редактор *А. А. Бряндинская*
Художник *Н. А. Зарин*
Художественный редактор *В. И. Шаповалов*
Технический редактор *Т. Л. Сухорукова*
Корректор *Т. А. Палладина*

Подписано к печати 28/IV 1967 г.
Бумага тип. № 3. $60\times90/_{16}$=15,6 бум. л.
Усл. печ. л. 31,2. Уч.-изд. л. 31,5.
Изд. № 1/4386. Цена 2 р. 37 к. Зак. 805.

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Ленинградская типография № 2
имени Евгении Соколовой Главполиграфпрома
Комитета по печати при Совете Министров СССР.
Измайловский проспект, 29.

3 р. 23 к.