

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. М.В. ЛОМОНОСОВА  
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ

В.Ю. Королев

# ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УЧЕБНИК



• ПРОСПЕКТ •

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. М.В. ЛОМОНОСОВА  
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ

---

В.Ю. Королев

# ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УЧЕБНИК

---

*Допущено Министерством образования и науки Российской Федерации  
в качестве учебника для студентов высших учебных заведений,  
обучающихся по экономическим и инженерным специальностям*

---



ПРОСПЕКТ

Москва  
2006

УДК 519.2(075.8)  
ББК 22.17я73  
К68

ЗБС  
К-682

**Королев В. Ю.**

**К68** Теория вероятностей и математическая статистика : учеб. – М. : ТК Велби, Изд-во Проспект, 2006. – 160 с.

ISBN 5-482-00274-8

Учебник содержит изложение основных разделов теории вероятностей и математической статистики. Основное внимание уделено особенностям практического применения описываемых моделей и методов теории вероятностей и математической статистики. Учебник ориентирован на формирование у читателя понимания возможностей теории вероятностей и математической статистики и ошибок, происходящих при неправильном применении вероятностных моделей и статистических процедур. Наряду с разделами, традиционно включаемыми в стандартные курсы теории вероятностей и математической статистики, содержатся некоторые специальные разделы, в частности, связанные со свойствами равномерного распределения и его ролью. Для понимания основных частей курса достаточно хорошего знания математики в рамках школьной программы.

Для студентов, обучающихся по экономическим, инженерным, психологическим и другим специальностям, использующим методы теории вероятностей и процедуры статистического анализа в качестве инструмента для решения конкретных практических задач.

Научная библиотека МГУ



**62005083**

УДК 519.2(075.8)  
ББК 22.17я73

Учебное издание

**Королев Виктор Юрьевич**

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Учебник

Подписано в печать 01.09.05. Формат 60×90<sup>1</sup>/<sub>16</sub>. Печать офсетная.  
Бумага газетная. Печ. л. 10,0. Тираж 3000 экз. Заказ № 0512560.

ООО «ТК Велби»

107120, г. Москва, Хлебников пер., д. 7, стр. 2.

Отпечатано в полном соответствии  
с качеством предоставленных диапозитивов  
в ОАО «Ярославский полиграфкомбинат»  
150049, г. Ярославль, ул. Свободы, 97

ISBN 5-482-00274-8



9 785482 002742



© В. Ю. Королев, 2006

© ООО «Издательство Проспект», 2006

НАУЧНАЯ БИБЛИОТЕКА МГУ

**28**

# Содержание

Предисловие	6
1. Основные понятия прикладной теории вероятностей	9
1.1. Вероятность как мера случайности	9
1.1.1. Понятие о предмете теории вероятностей. Стохастические ситуации	9
1.1.2. Классические вероятностные модели	12
1.1.3. Дискретные вероятностные модели	13
1.1.4. Геометрические вероятностные модели	17
1.1.5. Структура вероятностных моделей стохастических ситуаций	19
1.2. Независимость	23
1.2.1. Независимые события	23
1.2.2. Действия над событиями	24
1.2.3. Условные вероятности	27
1.2.4. Формула полной вероятности	28
1.2.5. Формула Байеса	30
1.3. Случайные величины	31
1.3.1. Определение случайной величины	31
1.3.2. Дискретные и непрерывные случайные величины. Функции распределения, Плотности вероятностей	32
1.4. Числовые характеристики случайных величин	38
1.4.1. Характеристики центра случайных величин	38
1.4.2. Характеристики разброса случайных величин	46
1.4.3. Неравенства для характеристик центра и разброса случайных величин	48

1.4.4.	Аналитические средние . . . . .	50
1.4.5.	Числовые характеристики формы распределения случайных величин . . . . .	52
1.5.	Независимость случайных величин . . . . .	54
1.5.1.	Независимые случайные величины . . . . .	54
1.5.2.	Ковариация. Коэффициент корреляции . . . . .	55
1.6.	Испытания Бернулли . . . . .	58
1.6.1.	Биномиальное распределение . . . . .	61
1.6.2.	Геометрическое распределение . . . . .	63
1.7.	Предельные теоремы, связанные с испытаниями Бернулли . . . . .	64
1.7.1.	Теорема Пуассона. Распределение Пуассона . . . . .	64
1.7.2.	Теорема Муавра-Лапласа. Нормальное распределение . . . . .	71
1.7.3.	Теорема Реньи. Показательное распределение . . . . .	76
1.8.	Закон больших чисел . . . . .	78
1.9.	Центральная предельная теорема . . . . .	79.
1.10.	Равномерное распределение. Закон Бенфорда . . . . .	88
1.11.	Многомерные случайные величины (случайные векторы) . . . . .	94
1.12.	Формулы свертки . . . . .	103
1.13.	Условные распределения . . . . .	108
1.13.1.	Дискретные условные распределения . . . . .	108
1.13.2.	Дискретные условные математические ожидания . . . . .	109
1.13.3.	Условные плотности . . . . .	111
1.13.4.	Условные математические ожидания непрерывных случайных величин . . . . .	114
<b>2.</b>	<b>Основные понятия прикладной статистики . . . . .</b>	<b>117</b>
2.1.	Задачи математической и прикладной статистики . . . . .	117
2.2.	Выборочные характеристики . . . . .	119
2.2.1.	Выборка. Вариационный ряд. Порядковые статистики . . . . .	119
2.2.2.	Статистические выводы о параметрах положения, разброса и формы распределения . . . . .	121
2.2.3.	Непараметрическое оценивание распределения генеральной совокупности . . . . .	130
2.2.4.	Репрезентативность выборки . . . . .	134
2.3.	Статистический анализ нормальных выборок . . . . .	140

---

2.3.1.	Распределения вероятностей, связанные с нормальным законом. Распределения хи-квадрат, Стьюдента, Фишера–Снедекора . . . . .	140
2.3.2.	Статистические выводы о параметрах нормального распределения . . . . .	142
2.4.	Методы построения оценок неизвестных параметров . . . . .	148
2.4.1.	Метод наибольшего правдоподобия . . . . .	148
2.4.2.	Метод моментов . . . . .	151
2.5.	Проверка согласия экспериментальных данных с теоретической моделью распределения генеральной совокупности . . . . .	153
2.5.1.	Критерий согласия хи-квадрат . . . . .	153
2.5.2.	Критерий согласия Колмогорова . . . . .	156
2.6.	Исследование стохастических зависимостей . . . . .	158
2.6.1.	Выборочный коэффициент корреляции . . . . .	158
2.6.2.	Линейная регрессия. Метод наименьших квадратов . . . . .	159
	<b>Рекомендуемая литература . . . . .</b>	<b>160</b>

# Предисловие

Теория вероятностей есть врожденная категорическая функция, мысленно превосходящая сменные явления природы и многообразно согласующаяся с функциями души и тела. Роль вероятностей (т. е. условных и безусловных достоверностей) по вопросам жизни и многосторонне посредническая, между субъективной в регулировании течений блага, строящая систему посредствующих, ограждающих и искупающих запасов, залогов, божков для умного выпуска явлений против многообразных (типических) "огневых" народных бедствий, и устанавливающая исчислением, измерением, формулами и словом или иными знаками и графиками критерии средства и соотношение или связь (интеграцию, интерполяцию) между составными частями, секциями и самостоятельными органами живого Всего (Целого) и их функциями.

П. А. Некрасов. *Теория вероятностей*. Санкт-Петербург, 1912.

Теория вероятностей – математическая наука, позволяющая по вероятностям одних случайных событий находить вероятности других случайных событий, связанных каким-либо образом с первыми.

*Математическая энциклопедия*. "Наука", Москва, 1977.

Идея данной книги возникла в результате частого общения автора с инженерами, экономистами, медиками, юристами, физиками и специалистами в других областях, которые используют аппарат теории вероятностей и математической статистики для решения своих специфических задач.

Можно критиковать определение, приведенное в первом эпиграфе, или подшучивать над ним, но оно по сути правильно отражает прикладную направленность теории вероятностей. Среди математиков за теорией вероятностей прочно закрепился статус прикладной науки. Благодаря этому, к сожалению, несмотря на большое число выдающихся математических результатов, полученных специалистами в области теории вероятностей и совершенных ими действительно фундаментальных математических от-

крытий, ни один из них не удостоен (и, по-видимому, не будет удостоен) премии Филдса, которая вручается за выдающийся вклад в фундаментальную математику. Тем не менее, несмотря на этот статус прикладной математической дисциплины, между “академической” теорией вероятностей и практическим применением ее результатов и методов возникла огромная пропасть.

В классической теории вероятностей и математической статистике много очень математически красивых и мощных результатов. Однако каждый такой результат начинается со слов “предположим, что ...” после которых идет перечень условий, которые должны выполняться, чтобы этот результат был справедлив. Очень часто чрезвычайно сложно или просто невозможно проверить, выполнены ли эти условия в конкретной реальной ситуации. Поэтому правильное применение результатов теории вероятностей и математической статистики на практике сопряжено с глубоким осмыслением предпосылок и реальных возможностей применения тех или иных методов.

К сожалению, этот важнейший этап очень часто игнорируется. Вместо такого анализа, во многих случаях исследователи-нематематики просто принимают на веру многие рекомендации математиков, находясь в плену стереотипов о всемогуществе “точнейшей из наук”.

Этим обстоятельством часто вольно или невольно пользуются некоторые авторы статей, в которых приводятся результаты статистической обработки тех или иных данных. Почти все читатели наверняка неоднократно встречали в различных публикациях фразы типа “вычисления на компьютере показывают, что ...”, “применяя методы математической статистики, мы убеждаемся, что ...”, “согласно закону больших чисел, ...” или “основываясь на центральной предельной теореме теории вероятностей, мы заключаем, что ...”, где вместо многоточий приводятся удобные авторам выводы. Такой прием рассчитан на читателя, привыкшего слепо доверять печатному слову и/или слепо верящего в непогрешимость компьютера, результатов теории вероятностей или методов статистического анализа. К большому сожалению, в действительности компьютер является лишь орудием в руках его пользователя, равно как и подготовка данных, выбор статистического метода для их анализа и интерпретация результатов применения этого метода всегда были, есть и будут прерогативой исследователя.



В данной книге сделана попытка, не углубляясь в дебри математических выкладок, но сохраняя логическую строгость, ввести читателей в курс основных понятий теории вероятностей. При этом автор старался сделать особый акцент на особенностях применения результатов и методов теории вероятностей и математической статистики к анализу случайных явлений, наблюдаемых на практике, и по возможности разрушить некоторые вредные стереотипы, связанные с практическим применением результатов и методов этих наук.

Книга в первую очередь рассчитана на тех читателей, которые не получили высшего математического образования, но которым приходится сталкиваться с необходимостью применения методов теории вероятностей и математической статистики для решения их практических проблем, скажем, в области экономики, биологии, психологии, физики, юриспруденции.

Данная книга, естественно, не дает готовых рецептов на все случаи жизни. Она предназначена для заинтересованных читателей, которым она может помочь самим разобраться в том, что теория вероятностей может и чего не может в их конкретных задачах, связанных с теми или иными конкретными ситуациями. В данной книге разобрано много примеров. В отличие от многих учебников по теории вероятностей и математической статистике, автор старался по возможности максимально приблизить приводимые в данном курсе примеры к практике. При этом многие примеры взяты из реальных прикладных задач.

Для понимания материала книги достаточно хорошего знания математики в пределах школьной программы и желания разобраться в предмете, которому данная книга посвящена. Необходимые понятия из области высшей математики вводятся по ходу изложения на том уровне строгости, который не должен испугать читателя, не знакомого с дифференциальным и интегральным исчислением. Для лучшего понимания некоторых разделов книги желательны графики или рисунки. Однако в книге их нет. Это сделано умышленно. Изложение построено так, что в случае необходимости читателю не составит большого труда самостоятельно выполнить соответствующие упражнения и нарисовать нужные графики или чертежи.

Для дальнейшего ознакомления с предметом заинтересованному читателю рекомендуются книги, включенные в список литературы.

# Основные понятия прикладной теории вероятностей

## 1.1. Вероятность как мера случайности

### 1.1.1. Понятие о предмете теории вероятностей.

#### Стохастические ситуации

Окружающая нас действительность постоянно порождает неопределенные ситуации, исходы которых невозможно заранее предсказать с исчерпывающей точностью. Иногда это связано просто с недостатком информации. В таких случаях получение дополнительной информации может существенно уменьшить неопределенность и даже совсем ее устранить. Однако иногда неопределенность принципиально нельзя устранить совсем, например, в лотереях или биржевых играх. Но даже в тех ситуациях, в которых неопределенность принципиально не устранима полностью, ее часто можно существенно уменьшить за счет лучшего понимания, уточнения самих механизмов проявления неопределенности. В частности, для этих целей можно использовать математические методы.

Математика предоставляет средства описания окружающей действительности, которые являются универсальными в том смысле, что они с одинаковым успехом могут быть использованы в самых разных областях — от физики, техники, биологии и медицины до страхования, финансов и юриспруденции. К разделам математики, изучающим механизмы проявления принципиально неустранимой неопределенности, можно отнести и теорию вероятностей.

Теория вероятностей изучает свойства математических моделей случайных явлений или процессов. Под *случайностью* мы будем понимать принципиально неустранимую неопределенность. С помощью понятий и утверждений теории вероятностей можно описать сами механизмы проявления неопределенности, выявить закономерности в проявлениях случайности.

Любая математическая теория устроена следующим образом. Фундаментом каждой такой теории является набор аксиом, то есть не противоречащих друг другу утверждений или принципов, заведомо считающихся верными и принимаемых без доказательств. Из этих аксиом с помощью логических переходов конструируются понятия и утверждения соответствующей теории. Разные наборы аксиом ведут к разным математическим теориям, которые могут описывать одни и те же процессы и явления. При этом практическая полезность или эффективность той или иной математической теории определяется удобством ее применения и ее адекватностью, то есть степенью согласованности получаемых с ее помощью выводов со свойствами описываемой ею реальности, наблюдаемыми на практике.

Имеется довольно много математических теорий, описывающих свойства математических моделей случайных явлений или процессов. В каждой из этих теорий так или иначе присутствует понятие вероятности как числового выражения меры возможности осуществления того или иного события, связанного с неопределенной ситуацией. Другими словами, имеется несколько теорий вероятностей. В нашем курсе мы будем иметь дело с теорией вероятностей, основанной на системе аксиом, которая была предложена в 20 – 30-х годах XX столетия великим русским математиком Андреем Николаевичем Колмогоровым<sup>1</sup>. Как правило, именно эта теория и называется собственно теорией вероятностей. За

---

<sup>1</sup>[1] А. Н. Колмогоров. Общая теория меры и исчисление вероятностей. *Труды Коммунистической академии. Раздел математики*. 1929, т. 1, с. 8-21; также см. А. Н. Колмогоров. *Теория вероятностей и математическая статистика*. “Наука”, Москва, 1986, с. 48-58. [2] A. Kolmogoroff. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin, Springer, 1933; также см. А. Н. Колмогоров. *Основные понятия теории вероятностей*. ОНТИ, Москва–Ленинград, 1936; 2-е издание: “Наука”, Москва, 1974; 3-е издание: “Фазис”, Москва, 1998.

другими теориями вероятностей закреплены особые названия, например, теория субъективных вероятностей, интервальная теория вероятностей и т. п.

Этот курс предназначен для нематематиков, и потому мы, естественно, не будем скрупулезно описывать сами колмогоровские аксиомы. Мы лишь опишем те свойства, которые должны быть присущи реальной неопределенной ситуации, чтобы ее можно было успешно математически описать на языке теории вероятностей. Другими словами, мы выделим те неопределенные ситуации, описание которых с помощью теории вероятностей ведет к адекватным выводам.

Итак, назовем *стохастической* такую ситуацию, которая характеризуется следующими свойствами или условиями:

- **непредсказуемость:** исход ситуации невозможно заранее предсказать с абсолютной точностью;
- **воспроизводимость:** имеется по крайней мере теоретическая возможность воспроизвести рассматриваемую ситуацию как угодно много раз в остающихся неизменными условиях;
- **устойчивость частот:** каким бы ни было интересующее нас событие, связанное с рассматриваемой ситуацией, при многократном воспроизведении этой ситуации частота события (то есть отношение количества случаев, в которых наблюдалось рассматриваемое событие, к общему числу воспроизведений ситуации) колеблется возле некоторого числа, приближаясь к нему все ближе и ближе по мере увеличения числа воспроизведений ситуации.

Поясним сказанное. Свойство непредсказуемости довольно очевидно. Если исход ситуации прогнозируем однозначно, то вообще нет никакой необходимости в привлечении аппарата теории вероятностей.

Свойство воспроизводимости ситуации является ключевым для того, чтобы быть уверенным в успехе применения аппарата теории вероятностей к ее описанию. Именно это свойство имеют в виду, когда говорят, что теория вероятностей и математическая статистика направлены на изучение *массовых* явлений. В связи с условием воспроизводимости следует весьма осторожно относиться к попыткам применения теории вероятностей к анализу *уникальных* явлений или систем. Например, известны многочисленные попытки дать количественный ответ на вопрос о том,

какова вероятность существования во Вселенной других планет, населенных разумными существами. Однако пока нет достаточных оснований считать, что наличие других планет и, тем более, существование на них разумной жизни является массовым явлением. Поэтому существующие прогнозы весьма разноречивы и потому неадекватны.

Наконец, свойство устойчивости частот позволяет связать математическое определение вероятности события с интуитивным представлением о ней как о понимаемом в определенном смысле пределе частоты осуществления события при неограниченном воспроизведении соответствующей ситуации.

Перейдем к рассмотрению примеров стохастических ситуаций и описывающих их вероятностных моделей.

### 1.1.2. Классические вероятностные модели

Начнем со следующей классической задачи.

**Задача 1.2.1.** Предположим, что в некотором сосуде (урне) имеется  $N$  шаров, различающихся лишь цветом: из этих шаров  $M$  — красные ( $M \leq N$ ) и  $N - M$  — белые. Из сосуда случайно извлекаются  $n$  шаров так, что каждый вынутый шар не возвращается обратно. Пусть  $m$  — некоторое целое число,  $0 \leq m \leq n$ . Какова вероятность того, что среди  $n$  вынутых шаров окажется ровно  $m$  красных?

Для решения этой задачи нам понадобятся некоторые понятия из комбинаторики. Рассмотрим совокупность, содержащую  $k$  элементов. Сколько существует способов упорядочить эти  $k$  элементов, то есть перенумеровать их? Первый номер может быть присвоен любому из  $k$  элементов. После того, как первый номер присвоен, второй номер может быть присвоен любому из оставшихся  $k - 1$  элементов. После того, как присвоены первый и второй номера, третий номер может быть присвоен любому из оставшихся  $k - 2$  элементов и так далее. Таким образом, общее число упорядоченных перестановок  $k$  элементов, составляющих рассматриваемую совокупность, равно

$$k \cdot (k - 1) \cdot (k - 2) \cdot \dots \cdot 2 \cdot 1 \equiv k!$$

(читается “ $k$  факториал”). Только что рассмотренный вопрос является

частным случаем другой чуть более общей задачи. Из совокупности, содержащей  $k$  элементов, по одному извлекаются без возвращения  $r$  элементов, образующие упорядоченную выборку объема  $r$ . Сколько существует возможных различных упорядоченных выборок объема  $r$  из совокупности, состоящей из  $k$  элементов? Рассуждая как и выше, мы замечаем, что первым может быть извлечен любой из  $k$  элементов. Вторым может быть извлечен любой из оставшихся  $k - 1$  элементов. Третьим — любой из оставшихся  $k - 2$  элементов и так далее до выбора  $r$ -го элемента. Таким образом, мы приходим к заключению, что число упорядоченных выборок объема  $r$  из совокупности, состоящей из  $k$  элементов, равно

$$k \cdot (k - 1) \cdot (k - 2) \cdot \dots \cdot (k - r + 1) = \frac{k!}{(k - r)!} \equiv (k)_r.$$

Очевидно, что  $(k)_k = k!$  (по определению мы считаем, что  $0! = 1$ ). Число  $(k)_r$  называется *числом перестановок* (или *числом размещений*) из  $k$  элементов по  $r$  элементов.

В только что рассмотренной задаче две выборки одинакового объема  $r$ , содержащие одни и те же элементы, но отличающиеся их порядком, считались разными. Рассмотрим еще одну комбинаторную задачу. Теперь, пренебрегая порядком элементов в выборке, мы будем любые две выборки, содержащие одни и те же элементы, но отличающиеся, быть может, их порядком, считать одинаковыми. Такие выборки мы будем называть *подмножествами* совокупности, состоящей из  $k$  элементов. Рассмотрим вопрос о том, сколько различных подмножеств по  $r$  элементов имеет совокупность, содержащая  $k \geq r$  элементов. Используя приведенные выше рассуждения, мы заметим, что, так как элементы подмножества, содержащего  $r$  элементов, можно упорядочить  $r!$  способами, то число упорядоченных выборок объема  $r$  из совокупности, содержащей  $k$  элементов, в  $r!$  раз больше, чем число ее подмножеств объема  $r$ , которое мы будем обозначать  $C_k^r$ . Другими словами,  $(k)_r = r! \cdot C_k^r$ , откуда

$$C_k^r = \frac{(k)_r}{r!} = \frac{k!}{r!(k - r)!}.$$

Число  $C_k^r$  называется *числом сочетаний* из  $k$  элементов по  $r$  элементов. Легко видеть, что  $C_k^r = C_k^{k-r}$ .

Теперь у нас есть все необходимое для решения Задачи 1.2.1.

По смыслу задачи ясно, что нам безразличен порядок шаров в выборке. Поэтому общее число разных способов, которыми мы можем извлечь  $n$  шаров из совокупности, содержащей  $N$  шаров, то есть общее число возможных исходов ситуации, равно  $C_N^n$ . При этом все эти исходы равновероятны. Найдем число исходов, в результате каждого из которых наступит интересующее нас событие. Выбрать  $m$  шаров из имеющихся  $M$  красных можно  $C_M^m$  способами. При этом остальные  $n - m$  шаров в выборке должны быть белыми. Выбрать  $n - m$  шаров из имеющихся  $N - M$  белых можно  $C_{N-M}^{n-m}$  способами. Ясно, что каждому набору из  $m$  красных шаров в выборке может соответствовать любой из  $C_{N-M}^{n-m}$  наборов белых шаров. Таким образом, общее число “благоприятных” исходов равно  $C_M^m \cdot C_{N-M}^{n-m}$ . Поскольку все возможные исходы равновероятны, в качестве искомой вероятности разумно взять отношение числа благоприятных исходов к общему числу исходов, а именно

$$p_m = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}. \quad (1.2.1)$$

Казалось бы, данная задача в том виде, как она сформулирована, имеет чисто познавательный интерес. Однако, как только мы зададим конкретные значения для параметров  $N$ ,  $M$ ,  $n$  и  $m$ , то сразу можем заметить, что эта задача очень важна для практики. Пусть, например,  $N = 49$ ,  $M = 6$ ,  $n = 6$ . Тогда при  $m = 4$ ,  $m = 5$  и  $m = 6$  с помощью формулы (1.2.1) мы получаем правило вычисления вероятностей угадать соответственно 4, 5 и 6 номеров в лотерее типа Спортлото “6 из 49”.

Набор чисел  $p_m$ , определяемых по формуле (1.2.1) для  $m = 0, 1, \dots, n$ , называется *гипергеометрическим распределением* вероятностей.

Задача 1.2.1 является примером использования так называемой *классической вероятностной модели*, которая характеризуется тем, что рассматриваемая стохастическая ситуация имеет конечное число равновероятных исходов. В такой модели под вероятностью события понимается отношение количества исходов, каждый из которых влечет осуществление этого события, к общему числу возможных исходов.

### 1.1.3. Дискретные вероятностные модели

Классические вероятностные модели приводят к самому простому и хорошо согласующемуся с интуитивными представлениями определению вероятности. Однако далеко не каждая стохастическая ситуация имеет *равновероятные* исходы, не говоря уже о том, что исходов может быть бесконечно много. Поэтому нужны более общие и более гибкие подходы к определению вероятности. Для иллюстрации рассмотрим еще одну классическую задачу, известную как *задача о справедливом дележе ставки*. Решение именно этой задачи было найдено в ходе знаменитой переписки великих ученых XVII столетия Блеза Паскаля (1623–1662) и Пьера Ферма (1601–1665), в которой, как считает большинство историков науки, и были заложены основы теории вероятностей.

**ЗАДАЧА 1.3.1.** Два игрока (назовем их, скажем,  $A$  и  $B$ ), поставив одинаковые суммы денег, играют в игру, состоящую из отдельных последовательных партий. В каждой партии возможны лишь два исхода: выигрыш игрока  $A$  или выигрыш игрока  $B$ . Вся игра считается выигранной тем из игроков, кто первым выиграет 6 партий (не обязательно подряд). Однако при счете 5:3 в пользу игрока  $A$  игра прервана и не может быть продолжена. В какой пропорции следует разделить ставку между  $A$  и  $B$ ?

Ясно, что речь идет о наиболее справедливом способе дележа, ведь всего способов довольно много. Например, поровну — 1:1. Однако такой способ не устраивает первого игрока, который к моменту прекращения игры уже выиграл больше партий, чем игрок  $B$ . Можно было бы разделить ставку в отношении 1:0, отдав все игроку  $A$ , который выиграл больше партий. Но такой способ, очевидно, не устраивает игрока  $B$ , ибо такой дележ предусмотрен лишь для случая, когда весь матч закончен, в то время как матч не закончен, и игрок  $B$  все-таки выиграл 3 партии. Кажется, что дележ ставки в отношении 5:3 будет справедливым, ведь такой дележ пропорционален количеству выигранных партий. Однако и он не является идеальным, поскольку на самом деле ничем не лучше дележа в отношении 3:1, обратном пропорциональном числу партий, которые осталось выиграть игрокам до окончательной победы.

Паскаль и Ферма пришли к обоснованному выводу, что делить ставку следует пропорционально шансам игроков на окончательную победу, то



есть пропорционально вероятностям того, что матч закончится победой соответствующего игрока, если бы он был продолжен. Найдем эти вероятности. Сначала опишем все возможные исходы данной стохастической ситуации. Победу игрока  $A$  в отдельной партии условно обозначим  $a$ , победу игрока  $B$  в отдельной партии обозначим  $b$ . Если первую после возобновления игры партию выигрывает игрок  $A$ , уже выигравший до этого пять партий, то весь матч (играющийся до шестой победы одного из игроков) заканчивается победой игрока  $A$ , выигравшего нужные шесть партий. Если эту партию выигрывает игрок  $B$ , то матч продолжается. Если следующую партию выигрывает игрок  $A$ , то весь матч заканчивается победой игрока  $A$ . Если же следующую партию выигрывает  $B$ , то в матче должна быть сыграна еще одна, решающая, партия. Если эту партию выигрывает  $A$ , то он побеждает в матче. Но если же эту партию выигрывает  $B$ , то весь матч заканчивается победой игрока  $B$ . Таким образом, возможные пути развития событий в матче, то есть возможные исходы рассматриваемой стохастической ситуации имеют вид  $a$ ,  $ba$ ,  $bba$  и  $bbb$ . Особо отметим, что здесь мы используем термин “исход” не для описания *результата* партии или матча, но для описания того *процесса*, который привел к тому или иному результату матча. Будем считать, что игроки одинаково сильны в этой игре и, отложив математическую конкретизацию сказанного до следующих разделов, предположим, что результат каждой последующей партии не зависит от того, как закончились предыдущие. Тогда, очевидно, вероятность исхода  $a$  равна  $\frac{1}{2}$ , вероятность исхода  $ba$  равна  $\frac{1}{4}$ , а вероятности каждого из исходов  $bba$  и  $bbb$  равны  $\frac{1}{8}$ . Событие “игра выиграна игроком  $A$ ” состоит из исходов  $a$ ,  $ba$  и  $bba$ . В качестве вероятности этого события следует взять сумму вероятностей элементарных исходов, в результате каждого из которых наступает рассматриваемое событие. Таким образом, вероятность события “игра выиграна игроком  $A$ ” равна  $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$ . В то же время, вероятность события “игра выиграна игроком  $B$ ” равна  $\frac{1}{8}$ . Следовательно, оптимальная пропорция, в которой следует поделить ставку, как это ни странно, составляет 7:1. (Выбранные для рассмотренного примера числа соответствуют памятного любителям шахмат матчу между Карповым и Каспаровым, скандально прерванному в 1985 г. Этот матч игрался по регламенту Фишера: до шести побед одного из соперников и без учета ничьих.)

Рассмотренная задача приводит нас к следующей конструкции *дискретной вероятностной модели*. Пусть  $\Omega$  — совокупность всех исходов рассматриваемой стохастической ситуации, элементарных в том смысле, что каждый из них нельзя разбить на более мелкие исходы. Под *событием* мы будем понимать любое подмножество  $A$  элементов совокупности  $\Omega$ . При этом вероятность события  $A$  определим как суммарную вероятность элементарных исходов, его составляющих. Более формально, пусть  $\Omega = \{\omega_1, \dots, \omega_n\}$  и вероятность исхода  $\omega_i$  равна  $p_i, i = 1, \dots, n$ . Пусть  $A$  — некоторое событие. Тогда вероятность события  $A$ , обозначаемая  $P(A)$ , определяется как

$$P(A) = \sum p_i, \quad (1.3.1)$$

где суммирование идет по всем номерам  $i$  элементарных исходов, составляющих событие  $A$ . При этом должно выполняться естественное условие нормировки

$$\sum_{i=1}^n p_i = 1.$$

Заметим, что в этой модели допускается возможность бесконечного числа элементарных исходов стохастической ситуации ( $n = \infty$ ), а вероятности  $p_i$  совсем не обязаны быть равны между собой.

Классическая вероятностная модель, рассмотренная в п. 1.1.2, является частным случаем дискретной модели, в котором  $p_i = \frac{1}{n}, i = 1, \dots, n$ .

#### 1.1.4. Геометрические вероятностные модели

Обе вероятностные модели, рассмотренные в предыдущих подразделах, характеризуются тем, что *все* элементарные исходы соответствующих стохастических ситуаций можно занумеровать, поставив каждому исходу в соответствие натуральное число так, что разным исходам будут соответствовать разные числа. Именно это свойство позволяет определить вероятность события по формуле (1.3.1). Однако зачастую исходы стохастической ситуации не могут быть перенумерованы разумным образом. В качестве примера рассмотрим следующую задачу.

**ЗАДАЧА 1.4.1.** Два человека, назовем их, скажем,  $X$  и  $Y$ , договорились встретиться в условленном месте между  $12^{00}$  и  $13^{00}$ . Согласно обоюдной

договоренности, пришедший первым ждет второго лишь в течение 20 минут (например, чтобы не привлекать излишнее внимание окружающих). Считая, что каждый из  $X$  и  $Y$  может прийти на место встречи в любой момент указанного интервала времени с одинаковой возможностью, найти вероятность того, что встреча состоится.

Чтобы решить эту задачу, сначала построим модель множества возможных элементарных исходов данной стохастической ситуации. Совокупность возможных моментов появления  $X$  на месте встречи — это отрезок длиной 1 час. Аналогично, другой точно такой же отрезок является совокупностью моментов появления  $Y$  на месте встречи. Отложим эти отрезки на соответствующих перпендикулярных координатных осях, резервируя горизонтальную ось за временем прихода  $X$ , а вертикальную — за временем прихода  $Y$ . Пусть  $x$  — момент появления  $X$ ,  $y$  — момент появления  $Y$ . Тогда каждый исход стохастической ситуации характеризуется парой чисел  $(x, y)$ . Если мы отождествим каждую такую пару с точкой, имеющей координаты  $(x, y)$  на плоскости с указанными координатными осями, то станет ясно, что совокупность исходов данной стохастической ситуации можно представить в виде квадрата с вершинами  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 0)$  и  $(1, 1)$ . Теперь найдем совокупность тех исходов, в результате каждого из которых наступает встреча  $X$  и  $Y$ . Если  $x$  отстоит от  $y$  более чем на  $\frac{1}{3}$  (20 минут составляют  $\frac{1}{3}$  часа), то встреча  $X$  с  $Y$  не может состояться, и наоборот. Таким образом, совокупность исходов, “благоприятствующих” встрече  $X$  с  $Y$ , содержит те и только те точки с координатами  $(x, y)$ , которые, во-первых, попадают внутрь указанного квадрата и, во-вторых, удовлетворяют условию  $|x - y| \leq \frac{1}{3}$ . С учетом того, что по условию все исходы равновозможны, искомая вероятность будет равна отношению площади шестиугольника, задаваемого условиями  $0 \leq x \leq 1$ ,  $0 \leq y \leq 1$ ,  $|x - y| \leq \frac{1}{3}$ , и равной, как несложно подсчитать,  $\frac{5}{9}$ , к площади всего квадрата, равной, очевидно, единице. Таким образом, искомая вероятность равна  $\frac{5}{9}$ .

Подобные рассуждения вполне применимы и в других задачах, где оказывается возможным отождествить совокупность всех элементарных исходов стохастической ситуации с каким-либо одно-, дву- или трехмерным множеством и все исходы считаются *равновозможными*. Тогда вероятность события, отождествляемого с каким-либо подмножеством сово-

купности всех элементарных исходов, будет определяться как отношение длин в первом случае, площадей во втором и объемов в третьем случае. Такие вероятностные модели принято называть *геометрическими*. Эти модели являются частными случаями так называемых непрерывных вероятностных моделей, связанных с описанием поведения непрерывных случайных величин, см. раздел 3.

### 1.1.5. Структура вероятностных моделей стохастических ситуаций

В каждом из примеров, рассмотренных в этом и предыдущих разделах, каждая *вероятностная модель* той или иной стохастической ситуации имеет вид тройки объектов  $(\Omega, \mathcal{A}, P)$ , где  $\Omega$  — совокупность всех элементарных исходов стохастической ситуации, каждое возможное событие является некоторым множеством элементарных исходов, то есть подмножеством множества  $\Omega$ , так что  $\mathcal{A}$  — совокупность всех возможных событий (которая в первых двух примерах совпадала с совокупностью всех подмножеств совокупности  $\Omega$ ),  $P$  — неотрицательная функция, областью определения которой является множество  $\mathcal{A}$ , а областью значений — отрезок  $[0, 1]$ . Каждому событию  $A \in \mathcal{A}$  функция  $P$  ставит в соответствие число  $P(A)$ , понимаемое как вероятность события  $A$ . При этом  $P(\Omega) = 1$ .

Каждая такая тройка является математическим описанием (математической моделью) соответствующей стохастической ситуации. Условимся в дальнейшем отождествлять термин *вероятностная модель* с конкретной тройкой  $(\Omega, \mathcal{A}, P)$ .

Выбор правильной, адекватной вероятностной модели в конкретном случае очень важен и иногда очень не прост. Дело в том, что иногда стохастическая ситуация, которая, казалось бы, вполне конкретно, однозначно описывается словами, допускает несколько разных вероятностных моделей, как показывает следующий пример.

**ПРИМЕР 1.5.1** (парадокс Бертрана<sup>2</sup>). Рассмотрим следующую задачу. Пусть имеется окружность, в которую вписан равносторонний треугольник с вершинами, скажем,  $A$ ,  $B$  и  $C$ . Спрашивается: какова вероятность

<sup>2</sup>J. Bertrand. *Calcul des Probabilités*. Paris, 1899.

того, что наугад (случайно) выбранная хорда данной окружности окажется длиннее стороны вписанного равностороннего треугольника?

*Первое решение.* Так как хорда выбирается случайно, то каждый из ее концов может попасть в любую точку окружности. Предположим, что один из концов хорды попал в точку  $A$ , являющуюся одной из вершин треугольника. Это предположение никак не ограничивает общность наших рассуждений, поскольку в противном случае мы всегда можем так повернуть треугольник, чтобы его вершина  $A$  совместилась с одним из концов хорды (вопрос, сформулированный в задаче, никак не ограничивает взаимное расположение хорды и треугольника). Теперь ясно, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы другой ее конец (напомним, что один из ее концов совмещен с точкой  $A$ ) попал на дугу  $BC$ , соединяющую две другие вершины треугольника. Вспоминая геометрическое определение вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то каждый ее конец может с равными возможностями попасть в любую точку окружности. Следовательно, искомая вероятность будет равна отношению длины дуги  $BC$ , каждая точка которой “благоприятствует” наступлению интересующего нас события, к длине всей окружности. А так как треугольник  $ABC$  — равносторонний, то это отношение равно  $\frac{1}{3}$ .

*Второе решение.* Так как хорда выбирается случайно, то она может быть сориентирована произвольным образом по отношению к стороне  $AC$  треугольника. Предположим, что хорда оказалась параллельна этой стороне. Это предположение никак не ограничивает общность наших рассуждений, поскольку в противном случае мы всегда можем так повернуть треугольник, чтобы его сторона  $AC$  оказалась параллельна хорде (вопрос, сформулированный в задаче, никак не ограничивает взаимное расположение хорды и треугольника). Заметим, что середина всех хорд, параллельных стороне  $AC$ , обязательно попадает на диаметр окружности, перпендикулярный стороне  $AC$ . Более того, так как треугольник  $ABC$  — правильный, то точка пересечения этого диаметра со стороной  $AC$ , скажем, точка  $D$ , делит радиус пополам. Пусть  $D'$  — точка, симметричная точке  $D$  относительно центра окружности. Теперь ясно, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы ее середина попала внутрь отрезка  $DD'$ . Вспоминая геометрическое опре-

деление вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то ее середина может с равными возможностями попасть в любую точку диаметра, перпендикулярного стороне  $AC$ . Следовательно, искомая вероятность будет равна отношению длины отрезка  $DD'$ , каждая точка которого “благоприятствует” наступлению интересующего нас события, к длине всего диаметра. Но из сказанного выше вытекает, что длина отрезка  $DD'$  составляет ровно половину длины диаметра окружности. Таким образом, искомая вероятность равна  $\frac{1}{2}$ .

*Третье решение.* Так как хорда выбирается случайно, то ее середина может попасть в любую точку круга, ограниченного рассматриваемой окружностью. Используя построения, аналогичные проведенным в ходе второго решения, легко убедиться, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы ее середина попала внутрь окружности, вписанной в треугольник  $ABC$ , радиус которой, как несложно видеть, равен половине радиуса окружности, описанной около этого же треугольника. Вспоминая геометрическое определение вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то ее середина может с равными возможностями попасть в любую точку круга, ограничиваемого описанной окружностью. Следовательно, искомая вероятность будет равна отношению площади круга, ограниченного вписанной окружностью, каждой точке которого “благоприятствует” наступлению интересующего нас события, к площади круга, ограничиваемого описанной окружностью. А так как это отношение равно отношению квадратов радиусов окружностей, то искомая вероятность равна  $\frac{1}{4}$ .

Налицо явный парадокс: одна задача имеет сразу три решения, и каждое из этих решений, будучи полученным логически непротиворечивым путем, является верным! Однако с математической точки зрения здесь нет парадокса. Дело в том, что исходная задача была сформулирована довольно нестрого, неоднозначно, так сказать, некорректно. Это проявилось в том, что понятие “случайного выбора хорды” допускает несколько математических конкретизаций.

Данная стохастическая ситуация может быть описана несколькими разными вероятностными моделями. При первом решении была использована модель  $(\Omega, \mathcal{A}, P)$ , в которой множество элементарных исходов  $\Omega$  отождествлено с окружностью, совокупность событий  $\mathcal{A}$  включает в се-

бя все дуги, а функция  $P$ , определенная на  $\mathcal{A}$ , каждой дуге приписывает число, равное отношению длины этой дуги к длине всей окружности.

При втором решении была использована модель  $(\Omega, \mathcal{A}, P)$ , в которой множество элементарных исходов  $\Omega$  отождествлено с диаметром окружности, совокупность событий  $\mathcal{A}$  включает в себя все отрезки, лежащие на этом диаметре, а функция  $P$ , определенная на  $\mathcal{A}$ , каждому отрезку приписывает число, равное отношению длины этого отрезка к диаметру.

При третьем же решении была использована модель  $(\Omega, \mathcal{A}, P)$ , в которой множество элементарных исходов  $\Omega$  отождествлено с кругом, ограниченным описанной окружностью, совокупность событий  $\mathcal{A}$  включает в себя все подмножества этого круга, площадь которых можно измерить, а функция  $P$ , определенная на  $\mathcal{A}$ , каждому подмножеству указанного круга ставит в соответствие число, равное отношению площади этого подмножества к площади круга.

Кстати, автор регулярно разбирает эту задачу на вводной лекции в курсе “вероятностные модели”, читаемом им студентам третьего курса факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова. Обсуждение задачи завершается “социологическим опросом” аудитории, в ходе которого из года в год с большим отрывом “наиболее правильным” из приведенных выше трех правильных решений признается первое решение. По-видимому, соответствующая вероятностная модель (согласно которой случайный выбор хорды отождествляется со случайным, равновероятным размещением двух точек на окружности) наиболее адекватно соответствует психологическому восприятию понятия случайного выбора. Кстати, справедливости ради можно отметить, что приведенные выше три решения отнюдь не исчерпывают все возможные подходы к задаче о случайном выборе хорды. На самом деле, каким бы ни было наперед заданное число  $p$  из интервала  $(0, 1)$ , можно указать такую (разумную!) вероятностную модель, решая вышеуказанную задачу, в соответствии с которой в качестве ответа можно получить число  $p$ .

В данной задаче самым сложным является адекватный выбор конкретной вероятностной модели. Далее уже, как говорят, дело техники. Такая ситуация типична для практики. Большинство задач, выдвигаемых жизнью, являются некорректными, допускают неоднозначную математическую интерпретацию. Ниже мы еще столкнемся с примерами подобных задач.

Можно сказать, что настоящий парадокс заключается в том, что выбор адекватной *математической*, вероятностной модели часто связан с *нематематическими* обстоятельствами. Еще в XIX веке великий русский математик П. Л. Чебышев говорил, что правильно поставить математическую задачу — значит наполовину ее решить.

## 1.2. Независимость

### 1.2.1. Независимые события

Рассмотрим какие-либо два события  $A$  и  $B$ , связанные с некоторой вероятностной моделью  $(\Omega, \mathcal{A}, P)$ .

**ОПРЕДЕЛЕНИЕ 2.1.1.** События  $A$  и  $B$  называются *независимыми*, если вероятность их одновременного осуществления равна произведению вероятностей каждого из них.

Событие, заключающееся в одновременном осуществлении событий  $A$  и  $B$ , принято называть *пересечением* событий  $A$  и  $B$  и обозначать  $A \cap B$ , или просто  $AB$ . Тогда Определение 2.1.1 можно формально записать так: события  $A$  и  $B$  называются независимыми, если

$$P(A \cap B) = P(A) \cdot P(B).$$

Пусть теперь  $A_1, A_2, \dots, A_n$  — некоторые события в вероятностной модели  $(\Omega, \mathcal{A}, P)$  ( $n \geq 2$ ).

**ОПРЕДЕЛЕНИЕ 2.1.2.** События  $A_1, A_2, \dots, A_n$  называются *независимыми в совокупности*, если для любого  $k \leq n$  и любых индексов  $i_1, \dots, i_k$  ( $i_p \neq i_q$  при  $p \neq q$  и  $1 \leq i_p \leq n$ ,  $p = 1, \dots, k$ )

$$P\left(\bigcap_{p=1}^k A_{i_p}\right) = \prod_{p=1}^k P(A_{i_p}).$$

Определение 2.1.1 при этом можно трактовать как определение *попарной* независимости событий. Очевидно, что из независимости событий в совокупности вытекает попарная независимость. Обратное утверждение,



вообще говоря, неверно. Чтобы в этом убедиться, рассмотрим следующий пример.

**ПРИМЕР 2.1.1 (тетраэдр Бернштейна).** На горизонтальную плоскость случайно бросается тетраэдр (правильная треугольная пирамида), грани которого раскрашены следующим образом. Одна грань — красная, вторая — синяя, третья — белая, а четвертая выкрашена в красно-сине-белую полосу. Пусть  $R$  — событие, заключающееся в том, что на нижней грани тетраэдра есть красный цвет,  $B$  — событие, заключающееся в том, что на нижней грани тетраэдра есть синий цвет,  $W$  — событие, заключающееся в том, что на нижней грани тетраэдра есть белый цвет. Убедимся, что события  $R$ ,  $B$ ,  $W$  попарно независимы, но не являются независимыми в совокупности. Действительно, события  $R$  и  $B$  могут одновременно произойти лишь в том случае, когда тетраэдр упал полосатой гранью вниз. Вероятность этого, очевидно, равна  $\frac{1}{4}$ . В то же время, событие  $R$  может осуществиться в результате двух исходов, когда тетраэдр падает либо красной, либо полосатой гранью вниз. Поэтому  $P(R) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ . Аналогично, событие  $B$  может осуществиться в результате двух исходов, когда тетраэдр падает либо синей, либо полосатой гранью вниз. Поэтому также  $P(B) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ . Таким образом,

$$P(R \cap B) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(R) \cdot P(B),$$

то есть события  $R$  и  $B$  независимы. Аналогично проверяется попарная независимость остальных событий. В то же время, события  $R$ ,  $B$  и  $W$  могут одновременно произойти лишь тогда, когда тетраэдр упал полосатой гранью вниз. Поэтому

$$P(R \cap B \cap W) = \frac{1}{4} \neq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(R) \cdot P(B) \cdot P(W),$$

то есть события  $R$ ,  $B$  и  $W$  не являются независимыми в совокупности.

### 1.2.2. Действия над событиями

Сначала введем некоторые обозначения, которые мы будем использовать в дальнейшем. Тот факт, что элементарный исход  $\omega$  влечет наступление

события  $A$ , мы будем записывать в виде  $\omega \in A$ . Это по сути означает, что исход  $\omega$  является элементом совокупности (множества)  $A$ , или, как еще говорят, исход  $\omega$  принадлежит к событию  $A$ . Если каждый элементарный исход события  $A$  влечет также наступление события  $B$ , то мы пишем  $A \subseteq B$ . Это фактически означает, что событие  $A$  является частью события  $B$ . Два события  $A$  и  $B$  эквивалентны (мы будем обозначать этот факт  $A = B$ ), если одновременно  $A \subseteq B$  и  $B \subseteq A$ . (В дальнейшем символы  $\in$  и  $\subseteq$  мы будем использовать не только для элементарных исходов и/или событий. Например, мы будем писать  $x \in [a, b]$ , подразумевая при этом, что число  $x$  принадлежит к интервалу  $[a, b]$ .)

События  $A$  и  $B$ , которые не имеют общих элементарных исходов, называются *несовместными*. Другими словами, несовместные события не могут произойти одновременно. Если символом  $\emptyset$  обозначить так называемое *пустое множество*, не содержащее никаких элементов, то можно формально записать, что события  $A$  и  $B$  несовместны, если  $A \cap B = \emptyset$ .

Событие, заключающееся в том, что произойдет хотя бы одно из событий  $A$  или  $B$ , называется *объединением* событий  $A$  и  $B$  и обозначается  $A \cup B$ .

Во всех вероятностных моделях, рассмотренных в предыдущих разделах, вероятность обладала следующим свойством. Если  $A$  и  $B$  — два несовместных события в вероятностной модели  $(\Omega, \mathcal{A}, P)$ , то

$$P(A \cup B) = P(A) + P(B). \quad (2.2.1)$$

Свойство (2.2.1), называемое *аддитивностью* вероятности, хорошо согласуется с интуитивными представлениями и является характеристическим.

Если события  $A$  и  $B$  могут иметь общие элементарные исходы, то

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

Пусть  $A \subseteq B$ . Символом  $B \setminus A$  обозначают совокупность всех элементарных исходов, которые составляют событие  $B$ , но не влекут событие  $A$ . Так как при этом  $B = A \cup (B \setminus A)$ , причем  $A \cap (B \setminus A) = \emptyset$ , то по свойству аддитивности вероятности  $P(B) = P(A) + P(B \setminus A)$ . Отсюда вытекает, что если  $A \subseteq B$ , то  $P(A) \leq P(B)$ .

Событие, заключающееся в том, что событие  $A$  не произойдет, называется *событием, противоположным  $A$* , или *дополнением к  $A$*  и обозначается  $\bar{A}$ . Так как  $\Omega = A \cup \bar{A}$  и  $A \cap \bar{A} = \emptyset$ , каково бы ни было событие

$A$ , то по свойству (2.2.1) аддитивности вероятности мы имеем

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

Очевидно, что  $\Omega = \Omega \cup \emptyset$ , причем  $\Omega \cap \emptyset = \emptyset$ . Поэтому по свойству аддитивности вероятности мы имеем  $P(\emptyset) = 1 - P(\Omega) = 1 - 1 = 0$ .

Чтобы не путать понятия независимых и непересекающихся событий, рассмотрим следующую задачу.

**ЗАДАЧА 2.2.1.** В каком случае несовместные события  $A$  и  $B$  независимы?

С одной стороны, события  $A$  и  $B$  независимы, то есть  $P(A \cap B) = P(A) \cdot P(B)$ . С другой стороны, поскольку события  $A$  и  $B$  несовместны,  $P(A \cap B) = P(\emptyset) = 0$ . Таким образом,  $P(A) \cdot P(B) = 0$ . Другими словами, несовместные события независимы тогда и только тогда, когда хотя бы одно из них имеет нулевую вероятность.

В связи с понятием независимости рассмотрим еще одну полезную задачу.

**ЗАДАЧА 2.2.2.** Пусть  $A_1, A_2, \dots, A_n$  — независимые события в вероятностной модели  $(\Omega, \mathcal{A}, P)$  ( $n \geq 2$ ). Какова вероятность того, что произойдет хотя бы одно из этих событий?

Эту задачу удобно решать в терминах противоположных событий. Обозначим  $A = A_1 \cup A_2 \cup \dots \cup A_n$ . Событие  $A$  как раз и заключается в том, что произойдет хотя бы одно из событий  $A_1, A_2, \dots, A_n$ . Событие, противоположное  $A$ , заключается в том, что не произойдет ни одно из событий  $A_1, A_2, \dots, A_n$ , то есть

$$\bar{A} = \overline{A_1 \cup A_2 \cup \dots \cup A_n} = \bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n. \quad (2.2.2)$$

Формулу (2.2.2) называют *законом двойственности*. Так как события  $A_1, A_2, \dots, A_n$  независимы, то независимы и события  $\bar{A}_1, \bar{A}_2, \dots, \bar{A}_n$ . Поэтому мы окончательно получаем

$$\begin{aligned} P(A) &= 1 - P(\bar{A}) = 1 - P(\bar{A}_1 \cap \bar{A}_2 \cap \dots \cap \bar{A}_n) = \\ &= 1 - \prod_{i=1}^n P(\bar{A}_i) = 1 - \prod_{i=1}^n [1 - P(A_i)]. \end{aligned}$$

Весьма часто в нематематической (художественной или популярной) литературе встречаются фразы вроде “по закону больших чисел это должно было обязательно произойти”. Под этим обычно подразумевается, что, как бы мала ни была вероятность некоторого события, скажем,  $A$ , связанного с некоторой стохастической ситуацией, при многократном (неограниченном) воспроизведении этой стохастической ситуации событие  $A$  обязательно произойдет хотя бы раз. На самом деле это утверждение не имеет никакого отношения к закону больших чисел. Оно является следствием Задачи 2.2.2. Действительно, пусть  $n \geq 1$  — некоторое число. Тогда вероятность того, что событие  $A$  произойдет хотя бы один раз при  $n$  (независимых) воспроизведениях стохастической ситуации, равна  $1 - (1 - P(A))^n$ . Эта вероятность не стремится к единице при  $n \rightarrow \infty$  только тогда, когда  $P(A) = 0$ . Легко видеть, что во всех остальных случаях, какой бы маленькой ни была *положительная* вероятность события  $A$ , мы имеем

$$\lim_{n \rightarrow \infty} [1 - (1 - P(A))^n] = 1.$$

### 1.2.3. Условные вероятности

Иногда бывает полезно вычислить вероятность события с учетом информации о том, что осуществилось какое-либо другое событие.

**ОПРЕДЕЛЕНИЕ 2.3.1.** Пусть  $A$  и  $B$  — события, причем  $P(B) \neq 0$ . *Условной вероятностью  $A$  при условии  $B$*  называется величина

$$P_B(A) = \frac{P(A \cap B)}{P(B)}.$$

С помощью Определения 2.3.1 мы можем перейти от вероятностной модели  $(\Omega, \mathcal{A}, P)$  к вероятностной модели  $(B, \mathcal{A}_B, P_B)$ . Здесь символом  $\mathcal{A}_B$  обозначена совокупность всех событий вида  $A \cap B$ , где  $A \in \mathcal{A}$ . Такой переход означает, что мы как бы считаем, что событие  $B$  уже произошло, и учитываем это обстоятельство при вычислении вероятностей, относящихся к рассматриваемой ситуации.

Если события  $A$  и  $B$  независимы, то из Определения 2.3.1 следует, что  $P_B(A) = P(A)$ .

Из Определения 2.3.1 сразу вытекает формула

$$P(A \cap B) = P(B) \cdot P_B(A).$$

Эту формулу иногда называют *законом умножения вероятностей*.

### 1.2.4. Формула полной вероятности

В некоторой вероятностной модели  $(\Omega, \mathcal{A}, P)$  рассмотрим события  $A_1, A_2, \dots, A_n$  ( $n \geq 2$ ), которые обладают следующими свойствами:

- a) события  $A_1, A_2, \dots, A_n$  несовместны, то есть никакие два из них не могут произойти одновременно;
- b) одно из событий  $A_1, A_2, \dots, A_n$  обязательно произойдет, то есть  $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$ , причем  $P(A_i) > 0$ ,  $i = 1, \dots, n$ .

Если события  $A_1, A_2, \dots, A_n$  обладают свойствами a) и b), то говорят, что они образуют *полную группу*.

Пусть  $B$  — некоторое событие, а события  $A_1, A_2, \dots, A_n$  образуют полную группу. В таком случае справедлива формула

$$P(B) = \sum_{k=1}^n P_{A_k}(B)P(A_k). \quad (2.4.1)$$

Соотношение (2.4.1) называется *формулой полной вероятности*.

В качестве примера применения формулы полной вероятности рассмотрим следующую задачу.

**ЗАДАЧА 2.4.1.** Допустим, что вероятность попадания в цель при одном выстреле равна  $p$  ( $0 < p < 1$ ). Предположим, что “живучесть” цели описывается числом  $q$  ( $0 < q < 1$ ): при попадании в нее она с вероятностью  $q$  сохраняет свои свойства, то есть не поражается. Какова вероятность того, что цель поражена, если в нее произведено  $n$  выстрелов?

Пусть  $A_k$  — событие, заключающееся в том, что ровно  $k$  выстрелов попали в цель,  $k = 0, 1, \dots, n$ . Несложно убедиться, что события  $A_0, A_1, \dots, A_n$  образуют полную группу. Будем считать, что выстрелы производятся независимо друг от друга. Можно подсчитать, что

$$P(A_0) = (1 - p)^n, \quad P(A_k) = C_n^k p^k (1 - p)^{n-k}, \quad k = 1, \dots, n$$

(о том, как вычислить эти вероятности, мы подробно поговорим в разделе 1.6). Пусть  $B$  — событие, заключающееся в том, что цель поражена. Вероятность поражения цели при  $k$  попаданиях в нее равна  $1 - q^k$  (см. Задачу 2.2.2), то есть

$$P_{A_k}(B) = 1 - q^k, \quad k = 1, \dots, n,$$

и очевидно, что  $P_{A_0}(B) = 0$ . При этом по формуле (2.4.1) мы получаем

$$\begin{aligned} P(B) &= \sum_{k=0}^n P_{A_k}(B)P(A_k) = \sum_{k=1}^n P_{A_k}(B)P(A_k) \\ &= \sum_{k=1}^n (1 - q^k)C_n^k p^k (1 - p)^{n-k} \\ &= \sum_{k=1}^n C_n^k p^k (1 - p)^{n-k} - \sum_{k=1}^n C_n^k (pq)^k (1 - p)^{n-k}. \end{aligned}$$

Первая сумма в правой части здесь равна

$$1 - P(A_0) = 1 - (1 - p)^n.$$

Используя формулу бинома Ньютона, для второй суммы мы получаем

$$\begin{aligned} \sum_{k=1}^n C_n^k (pq)^k (1 - p)^{n-k} &= (1 - p)^n \left[ \left( 1 + \frac{pq}{1 - p} \right)^n - 1 \right] \\ &= [1 - p(1 - q)]^n - (1 - p)^n. \end{aligned}$$

Подставив найденные выражения в (2.4.1), мы окончательно получаем

$$P(B) = 1 - [1 - p(1 - q)]^n.$$

**ЗАДАЧА 2.4.2.** Известно, что 5% всех мужчин и 0,25% всех женщин страдают дальтонизмом. Какова вероятность того, что выбранное наугад лицо является дальтоником? Другими словами, какова доля лиц, страдающих дальтонизмом, во всем населении?

В качестве полной группы событий возьмем события  $A_1 =$  “выбран мужчина” и  $A_2 =$  “выбрана женщина”. Будем считать, что мужчины составляют 48% населения, а женщины, соответственно, — 52%, то есть  $P(A_1) = 0,48$ ,  $P(A_2) = 0,52$ . Пусть  $D$  — событие, заключающееся в том, что выбранное наугад лицо является дальтоником. По условию  $P_{A_1}(D) = 0,05$  и  $P_{A_2}(D) = 0,0025$ . Тогда по формуле полной вероятности

$$\begin{aligned} P(D) &= P_{A_1}(D)P(A_1) + P_{A_2}(D)P(A_2) \\ &= 0,05 \cdot 0,48 + 0,0025 \cdot 0,52 = 0,0253 = 2,53\%. \end{aligned}$$

### 1.2.5. Формула Байеса

Пусть  $B$  — некоторое событие, имеющее положительную вероятность ( $P(B) > 0$ ), а события  $A_1, A_2, \dots, A_n$  образуют полную группу. В таком случае справедлива формула

$$P_B(A_k) = \frac{P_{A_k}(B)P(A_k)}{\sum_{k=1}^n P_{A_k}(B)P(A_k)}, \quad k = 1, \dots, n. \quad (2.5.1)$$

Соотношение (2.5.1) называется *формулой Байеса*. Формула Байеса позволяет уточнить представление о вероятности любого из событий, составляющих полную группу, с учетом информации об осуществлении некоторого события.

В качестве примера применения формулы Байеса рассмотрим следующую задачу, являющуюся продолжением Задачи 2.4.2.

**ЗАДАЧА 2.5.1.** Известно, что 5% всех мужчин и 0,25% всех женщин страдают дальтонизмом. Выбранное наугад лицо страдает дальтонизмом. Какова вероятность того, что это мужчина?

В обозначениях, введенных при решении Задачи 2.4.2, искомая вероятность — это  $P_D(A_1)$ . Используя данные, полученные при решении Задачи 2.4.2, по формуле Байеса мы получим

$$P_D(A_1) = \frac{P_{A_1}(D)P(A_1)}{P_{A_1}(D)P(A_1) + P_{A_2}(D)P(A_2)} = \frac{0,05 \cdot 0,48}{0,0253} \approx 0,9486.$$

Следующая задача весьма примечательна.

**ЗАДАЧА 2.5.2.** При контроле правдивости показаний подозреваемого на “детекторе лжи” вероятность признать ложью ответ, не соответствующий действительности, равна 0,99, вероятность ошибочно признать ложью правдивый ответ равна 0,01. Известно, что ответы, не соответствующие действительности, составляют 1% всех ответов подозреваемого. Какова вероятность того, что ответ, признанный ложью, и в самом деле не соответствует действительности?

Пусть  $\alpha$  — вероятность признать ложью ответ, не соответствующий действительности,  $\beta$  — вероятность ошибочно признать ложью правдивый ответ,  $\gamma$  — доля ответов, не соответствующих действительности, среди всех ответов подозреваемого. Тогда по формуле Байеса искомая условная вероятность  $p$  равна

$$p = \frac{\alpha\gamma}{\alpha\gamma + \beta(1 - \gamma)} = \frac{0,99 \cdot 0,01}{0,99 \cdot 0,01 + 0,01 \cdot 0,99} = \frac{1}{2}.$$

Другими словами, на самом деле лишь половина признанных ложью ответов не соответствуют действительности, хотя, казалось бы, выбранные значения параметров контроля  $\alpha$  и  $\beta$  являются довольно жесткими. Более того, легко подсчитать, что если не соответствующим действительности является в среднем лишь один ответ подозреваемого из двухсот, то на самом деле не соответствующими действительности будет менее трети ответов, признанных ложью.

## 1.3. Случайные величины

### 1.3.1. Определение случайной величины

Большинство стохастических ситуаций характеризуются тем обстоятельством, что каждому их элементарному исходу соответствует некоторое число. Например, число угаданных номеров в лотерее “Спортлото”, количество пострадавших при очередном пожаре и т. п. Даже тогда, когда фактически числа не участвуют в определении исходов (например, диагноза, который врач поставит своему очередному пациенту) (в таком случае говорят о *номинальной шкале* наблюдаемых характеристик), исходы можно искусственно оцифровать (например, в предыдущем примере о врачах — занумеровав все известные болезни). Если каждому элементу



некоторой совокупности ставится в однозначное соответствие число. Говорят, что на данной совокупности определена *функция*. При этом, вообще говоря, соответствие не обязано быть взаимно однозначным, т.е. разным элементам может соответствовать одно и то же число, но одному элементу может соответствовать лишь одно число.

**ОПРЕДЕЛЕНИЕ 3.1.1.** *Случайной величиной* называется функция, заданная на множестве элементарных исходов в вероятностной модели, описывающей стохастическую ситуацию.

Из этого довольно формального определения вытекает, что конкретное значение случайной величины определяется тем, какой исход стохастической ситуации будет реализован.

### 1.3.2. Дискретные и непрерывные случайные величины. Функции распределения. Плотности вероятностей

Рассмотрим вероятностную модель  $(\Omega, \mathcal{A}, P)$ , описывающую некоторую стохастическую ситуацию, с которой связана некоторая случайная величина. Случайную величину как функцию элементарного исхода  $\omega$  мы будем обозначать  $X(\omega)$ , иногда для краткости опуская аргумент у функции  $X$ . Сначала предположим, что все возможные значения случайной величины  $X$  можно перенумеровать. В такой ситуации случайная величина  $X$  называется *дискретной*. Возможные значения случайной величины  $X$  мы обозначим  $x_1, x_2, \dots$ .

Заметим, что, вообще говоря, случайная величина может принимать одно и то же значение для разных элементарных исходов. Например, студенту предстоит сдавать четыре конкурсных экзамена, каждый из которых оценивается по пятибалльной системе, то до начала экзаменов возможную сумму набранных баллов разумно считать случайной величиной, реализующейся в результате стохастической ситуации, элементарным исходом которой является цепочка  $(\omega_1, \dots, \omega_4)$ , где  $\omega_j$  — результат  $j$ -го экзамена. При этом такая случайная величина — сумма баллов — может принять, скажем, значение 19 как на цепочке  $(\omega_1 = 4, \omega_2 = 5, \omega_3 = 5, \omega_4 = 5)$ , так и на цепочке  $(\omega_1 = 5, \omega_2 = 5, \omega_3 = 5, \omega_4 = 4)$ .

Под символом  $P(X = x_k)$  мы будем понимать вероятность события  $\{\omega: X(\omega) = x_k\}$ , состоящего из всех элементарных исходов, в результате каждого из которых случайная величина  $X$  принимает значение  $x_k$ .

Для краткости обозначим  $p_k = P(X = x_k)$ .

**ОПРЕДЕЛЕНИЕ 3.2.1.** Набор пар  $\{(x_k, p_k), k = 1, 2, \dots\}$  называется *распределением* (или *распределением вероятностей*) дискретной случайной величины  $X$ .

Однако далеко не всегда множество всех возможных значений случайной величины можно перенумеровать разумным образом. Во многих ситуациях эти значения заполняют *непрерывное* или, как еще говорят, *континуальное* множество, например, отрезок (конечный или бесконечный) на прямой или часть плоскости или пространства. В таких случаях говорят о *непрерывных* случайных величинах. Для таких случайных величин, вообще говоря, нельзя ввести понятие распределения вероятностей по схеме, использовавшейся нами выше для дискретных случайных величин, поскольку, как мы увидим впоследствии, как это ни парадоксально, почти каждое свое конкретное значение любая непрерывная случайная величина принимает с вероятностью, равной нулю.

Однако для описания вероятностно-статистических свойств как дискретных, так и непрерывных случайных величин можно использовать следующий подход, основанный на рассмотрении не вероятностей того, что случайная величина примет значение, *равное* некоторому числу, а вероятностей того, что случайная величина примет значение, *меньшее* (или *большее*) некоторого числа.

Пусть  $f(x)$  — непрерывная неотрицательная функция вещественного аргумента  $x$ . Пусть  $a$  и  $b$  — два числа,  $-\infty < a \leq b < \infty$ . Символом

$$\int_a^b f(x) dx$$

мы будем обозначать площадь фигуры на плоскости, ограниченной снизу горизонтальной координатной осью (осью иксов), сверху — графиком функции  $f(x)$ , слева — перпендикуляром к оси иксов, опущенным из точки с координатами  $(a, f(a))$ , и справа — перпендикуляром к оси иксов, опущенным из точки с координатами  $(b, f(b))$ . Это обозначение для указанной площади называется *интегралом функции  $f(x)$  по  $dx$  от  $a$  до  $b$* .

Для бесконечных значений  $a$  и (или)  $b$  понятие интеграла можно ввести следующим образом. Пусть, к примеру,  $a = -\infty$ , но  $|b| < \infty$ . Тогда для каждого натурального  $n = 1, 2, \dots$  положим

$$A_n = \int_{b-n}^b f(x) dx$$

и интеграл

$$\int_{-\infty}^b f(x) dx$$

определим как предел числовой последовательности  $\{A_n\}_{n=1}^{\infty}$ :

$$\int_{-\infty}^b f(x) dx = \lim_{n \rightarrow \infty} A_n.$$

Последовательность  $A_n$ , как легко видеть, монотонно не убывает, так как при каждом  $n \geq 1$  фигура, ограниченная сверху графиком неотрицательной функции  $f(x)$ , справа — прямой  $x = b$ , снизу — осью иксов, а слева — прямой  $x = b - n$  целиком вложена в фигуру, ограниченную сверху графиком той же функции  $f(x)$ , справа — прямой  $x = b$ , снизу — осью иксов, а слева — прямой  $x = b - (n + 1)$ . Поэтому указанный предел существует, но может быть бесконечным. Аналогично, если  $|a| < \infty$ , но  $b = \infty$ , то для каждого натурального  $n = 1, 2, \dots$  положим

$$B_n = \int_a^{a+n} f(x) dx$$

и интеграл

$$\int_a^{\infty} f(x) dx$$

определим как предел числовой последовательности  $\{B_n\}_{n=1}^{\infty}$ :

$$\int_a^{\infty} f(x) dx = \lim_{n \rightarrow \infty} B_n.$$

В силу приведенных выше соображений этот предел также всегда существует, но может оказаться бесконечным. Если же оба предела интегрирования  $a$  и  $b$  бесконечны, то для каждого натурального  $n = 1, 2, \dots$

положим

$$D_n = \int_{-n}^n f(x) dx$$

и интеграл

$$\int_{-\infty}^{\infty} f(x) dx$$

определим как предел числовой последовательности  $\{D_n\}_{n=1}^{\infty}$ :

$$\int_{-\infty}^{\infty} f(x) dx = \lim_{n \rightarrow \infty} D_n.$$

Этот предел опять-таки существует, но может быть бесконечным.

По аналогии со связанными с дискретными случайными величинами вероятностями  $P(X = x)$  событий  $\{\omega: X(\omega) = x\}$ , включающих все элементарные исходы  $\omega$ , для которых  $X(\omega) = x$ , где  $x$  — произвольное число, рассматривая непрерывные случайные величины, для произвольных чисел  $a$  и  $b$  таких, что  $a \leq b$ , символом  $P(a \leq X \leq b)$  мы будем обозначать вероятность события  $\{\omega: a \leq X(\omega) \leq b\}$ , включающего все элементарные исходы  $\omega$ , для которых  $a \leq X(\omega) \leq b$ . Другими словами, символ  $P(a \leq X \leq b)$  — это краткое обозначение вероятности  $P(\{\omega: a \leq X(\omega) \leq b\})$ . Во всех указанных выражениях знаки неравенств могут быть заменены на строгие неравенства. При этом  $P(X < x) = P(-\infty \leq X < x)$ .

В подавляющем числе встречающихся на практике стохастических ситуаций для каждой непрерывной случайной величины  $X$  существует такая неотрицательная функция  $f(x)$ , что, какими бы ни были числа  $a$  и  $b$  ( $a \leq b$ ), вероятность  $P(a \leq X \leq b)$  того, что случайная величина примет значение из интервала  $[a, b]$ , может быть записана в виде

$$P(a \leq X \leq b) = \int_a^b f(x) dx. \quad (3.2.1)$$

**ОПРЕДЕЛЕНИЕ 3.2.2.** Если соотношение (3.2.1) выполняется для любых  $a$  и  $b$  ( $a \leq b$ ), то функция  $f(x)$ , фигурирующая в этом соотношении, называется *плотностью вероятностей* или *плотностью распределения вероятностей* или *плотностью распределения* или просто *плотностью* случайной величины  $X$ .

Из соотношения (3.2.1) вытекает, что поскольку  $P(\Omega) = 1$ , мы обязательно имеем

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Условимся в дальнейшем называть непрерывными только те случайные величины, которые удовлетворяют соотношению (3.2.1) с некоторой функцией  $f(x)$ .

Вероятностно-статистические свойства любой дискретной или непрерывной случайной величины полностью определяются ее функцией распределения  $F(x)$ , которая для каждого значения ее аргумента  $x$  определяется как

$$F(x) = P(X < x) = \begin{cases} \sum_{k: x_k < x} P(X = x_k), & \text{если случайная величина } X \\ & \text{дискретна,} \\ \int_{-\infty}^x f(x) dx, & \text{если случайная величина } X \\ & \text{непрерывна.} \end{cases}$$

Любая функция распределения  $F(x)$  обладает следующими свойствами:

- $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$ ;
- функция  $F(x)$  монотонно не убывает: если  $x_1 < x_2$ , то  $F(x_1) \leq F(x_2)$ ;
- функция  $F(x)$  непрерывна слева, то есть если  $x$  — произвольное число и при  $n \rightarrow \infty$  числовая последовательность  $\{x_n\}_{n \geq 1}$  неограниченно приближается к  $x$ , не превосходя  $x$ , то последовательность соответствующих значений  $F(x_n)$  неограниченно приближается к  $F(x)$ .

Разным случайным величинам могут соответствовать одинаковые функции распределения.

Мы можем теперь переформулировать определение непрерывной случайной величины в терминах функции распределения. А именно, случайная величина с функцией распределения  $F(x)$  непрерывна, если для любого  $x$  справедливо равенство

$$F(x) = \int_{-\infty}^x f(x) dx.$$

При этом из геометрической интерпретации интеграла ясно, что плотность  $f(x)$  характеризует скорость возрастания  $F(x)$  при увеличении  $x$ :

Если  $x$  — произвольное число и при  $n \rightarrow \infty$  числовая последовательность  $\{x_n\}_{n \geq 1}$  неограниченно приближается к  $x$ , оставаясь больше или равной  $x$ , то предельное значение последовательности соответствующих значений  $F(x_n)$  мы будем обозначать символом  $F(x+0)$ . Из определения функции распределения вытекает, что  $F(x+0) = P(X \leq x)$ .

Для любой случайной величины справедливо соотношение

$$P(X = x) = F(x+0) - F(x). \quad (3.2.2)$$

Чтобы убедиться в его справедливости, заметим, что

$$\{\omega: X(\omega) \leq x\} = \{\omega: X(\omega) < x\} \cup \{\omega: X(\omega) = x\},$$

причем события в правой части не имеют общих элементарных исходов. Поэтому

$$F(x+0) = P(X \leq x) = P(X < x) + P(X = x) = F(x) + P(X = x),$$

откуда вытекает требуемое.

При этом, если случайная величина  $X$  дискретна и  $x$  — одно из ее возможных значений, то есть  $P(X = x) > 0$ , то  $F(x+0) - F(x) > 0$ . Если же случайная величина  $X$  непрерывна и  $y$  — произвольное число, то

$$F(y+0) - F(y) = \int_y^y f(x) dx = 0,$$

а значит,

$$P(X = y) = 0.$$

Другими словами, как мы уже отмечали, непрерывная случайная величина принимает каждое конкретное свое значение с вероятностью, равной нулю. Поэтому для таких случайных величин имеет смысл говорить о вероятностях их попадания в интервалы.

Вероятности попадания любой случайной величины в любые интервалы могут быть записаны в терминах ее функции распределения: если  $a$  и  $b$  — произвольные числа,  $a \leq b$ , то

$$P(a \leq X \leq b) = F(b+0) - F(a), \quad P(a < X \leq b) = F(b+0) - F(a+0), \\ P(a \leq X < b) = F(b) - F(a), \quad P(a < X < b) = F(b) - F(a+0).$$

Справедливость этих соотношений устанавливается точно так же, как мы доказали равенство (3.2.2).

## 1.4. Числовые характеристики случайных величин

### 1.4.1. Характеристики центра случайных величин

Весьма часто при анализе той или иной стохастической ситуации основной интерес представляет даже не распределение вероятностей наблюдаемой случайной величины, а лишь такое число, вокруг которого концентрируются возможные значения этой случайной величины при многократном воспроизведении стохастической ситуации. Примером может служить непосредственное измерение некоторого параметра, когда вследствие случайных погрешностей каждое новое измерение отличается от предыдущего, но вместе с тем ясно, что все результаты измерения группируются вокруг некоторого “центрального” значения. Возможными синонимами понятия “центрального” значения случайной величины являются “среднее”, “ожидаемое”, “наиболее вероятное” значение. Мы увидим, что, несмотря на интуитивно кажущуюся простоту, понятие “центрального” значения случайной величины является довольно тонким, и каждому из словесных описаний, приведенных выше, можно придать свое математически строгое определение. Эти определения, вообще говоря, приводят к *разным* понятиям и, соответственно, к *разным* числам.

**ОПРЕДЕЛЕНИЕ 4.1.1.** Пусть  $X$  – дискретная случайная величина, принимающая значения  $x_1, x_2, \dots$  с вероятностями соответственно  $p_1, p_2, \dots$ . *Математическим ожиданием дискретной случайной величины* называется число

$$EX = x_1 p_1 + x_2 p_2 + \dots \quad (4.1.1)$$

Смысл математического ожидания становится ясным, если мы предположим, что случайная величина может с равными вероятностями принять одно из конечного числа значений  $x_1, \dots, x_n$ . Тогда  $p_k = \frac{1}{n}$ ,  $k = 1, \dots, n$ , и математическое ожидание такой случайной величины совпадает со сред-

ним арифметическим чисел  $x_1, \dots, x_n$ :

$$EX = \frac{1}{n} \sum_{k=1}^n x_k.$$

Таким образом, мы можем заключить, что в общем случае математическое ожидание дискретной случайной величины, определенное формулой (4.1.1), является обобщением понятия среднего арифметического на случай (возможно, бесконечного числа) неравновероятных возможных значений.

Чтобы определить математическое ожидание непрерывной случайной величины, нам понадобится обобщение понятия интеграла, определенное выше. Напомним, что выше мы определили интеграл от неотрицательной функции. Сейчас, используя это уже известное нам определение, мы определим интеграл от произвольной функции  $g(x)$ , принимающей как положительные, так и отрицательные значения.

С функцией  $g(x)$  мы свяжем два множества

$$G_+ = \{x: g(x) \geq 0\} \quad \text{и} \quad G_- = \{x: g(x) < 0\}.$$

Множество  $G_+$  содержит те и только те значения аргумента  $x$ , для которых функция  $g(x)$  неотрицательна, а множество  $G_-$  содержит те и только те значения аргумента  $x$ , для которых функция  $g(x)$  отрицательна. Если  $A$  — некоторое множество, то символом  $1_A(x)$  мы будем обозначать индикаторную функцию множества  $A$ :

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

Тогда функции  $g_+(x) = g(x)1_{G_+}(x)$  и  $g_-(x) = |g(x)|1_{G_-}(x)$  неотрицательны и, стало быть, определены интегралы  $\int_{-\infty}^{\infty} g_+(x) dx$  и  $\int_{-\infty}^{\infty} g_-(x) dx$ . Более того, справедлива следующая формула для функции  $g(x)$

$$g(x) \equiv g_+(x) - g_-(x).$$

Поэтому мы естественно приходим к определению интеграла от произвольной функции  $g(x)$ , обозначаемого  $\int_{-\infty}^{\infty} g(x) dx$ , как разности интегралов от функций  $g_+(x)$  и  $g_-(x)$ :

$$\int_{-\infty}^{\infty} g(x) dx = \int_{-\infty}^{\infty} g_+(x) dx - \int_{-\infty}^{\infty} g_-(x) dx.$$



Теперь мы можем определить математическое ожидание непрерывной случайной величины.

**ОПРЕДЕЛЕНИЕ 4.1.2.** Пусть  $X$  – непрерывная случайная величина, имеющая плотность  $f(x)$ . Математическим ожиданием непрерывной случайной величины называется число

$$EX = \int_{-\infty}^{\infty} x f(x) dx. \quad (4.1.2)$$

Свойства математических ожиданий дискретных и непрерывных случайных величин совпадают. Основным из них можно считать его линейность: если  $a$  и  $b$  – два числа, а  $X$  и  $Y$  – две случайные величины, то

$$E(aX + bY) = aEX + bEY.$$

Отсюда, в частности, вытекает, что:

- математическое ожидание суммы случайных величин равно сумме математических ожиданий этих случайных величин;
- неслучайный коэффициент при случайной величине можно выносить за знак математического ожидания.

К числу свойств математического ожидания можно отнести и следующий факт. Пусть  $X$  – случайная величина,  $A$  – некоторое подмножество ее значений. Пусть  $1_A(x)$  – индикаторная функция множества  $A$ . Тогда  $1_A(X)$  – случайная величина, принимающая значение 1, если  $X \in A$ , и значение 0, если  $X \notin A$ , причем

$$P(1_A(X) = 1) = P(X \in A) = 1 - P(1_A(X) = 0).$$

Тогда в соответствии с определением математического ожидания дискретной случайной величины мы имеем

$$E1_A(X) = 1 \cdot P(X \in A) + 0 \cdot P(X \notin A) = P(X \in A).$$

Смысл математического ожидания поясняет следующая задача.

**ЗАДАЧА 4.1.1.** Предположим, что некий человек на перемещение из точки  $a$  в точку  $b$  затрачивает усилия, равные  $(b - a)^2$ . Это означает,

что он легко перемещается на малые расстояния, но большие перемещения даются ему с большим трудом. Скажем, мы рассматриваем пожилого немощного человека. Пусть  $X$  – случайная величина, характеризующая координату точки, в которую необходимо переместиться этому пожилому человеку (скажем, он может пойти или поехать в поликлинику, в собес, в магазин, к внукам и т. п.). Спрашивается: где должен жить этот человек, чтобы минимизировать свои ожидаемые усилия, связанные с перемещениями?

Решение этой задачи таково. Интуитивно ясно, что если этот человек будет жить в “средней”, “центральной” точке между возможными конечными пунктами своих перемещений, то его усилия в среднем будут невелики. Подкрепим эту догадку строгими рассуждениями. Средние усилия, затрачиваемые этим человеком на перемещение из неслучайной точки  $x$  в случайную точку  $X$ , имеют вид

$$E(X - x)^2 = E(X^2 - 2xX + x^2) = x^2 - 2xE(X) + E(X^2).$$

Правая часть этого выражения как функция от  $x$  представляет собой хорошо знакомую квадратичную параболу. Ее минимум находится в точке  $x_0 = E(X)$ . Таким образом, математическое ожидание характеризует центр распределения случайной величины  $X$  (или, в терминах данной задачи, определяет “среднюю” или “центральную” точку населенного пункта) в смысле квадратичной функции потерь.

Математическое ожидание не всегда определено, но если оно определено, то всегда единственно. Поясним сказанное на примере случайной величины  $X$ , имеющей так называемое *распределение Коши*, задаваемое плотностью

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty.$$

Тогда в соответствии с определением интеграла, приведенным выше, мы имеем

$$I_- = \int_{-\infty}^0 \frac{|x| dx}{\pi(1+x^2)} = \infty, \quad I_+ = \int_0^{\infty} \frac{x dx}{\pi(1+x^2)} = \infty$$

и следовательно, в соответствии с определением математического ожидания непрерывной случайной величины мы получаем неопределенность

$$\int_{-\infty}^{\infty} \frac{x dx}{\pi(1+x^2)} = I_+ - I_- = \infty - \infty.$$

Таким образом, математическое ожидание случайной величины, имеющей распределение Коши, не определено.

**ОПРЕДЕЛЕНИЕ 4.1.3.** Пусть  $q$  – некоторое число из интервала  $(0, 1)$ . *Квантилью порядка  $q$*  (или  *$q$ -квантилью*) случайной величины  $X$  называется число  $x_q$  такое, что одновременно

$$P(X \leq x_q) \geq q \quad \text{и} \quad P(X \geq x_q) \geq 1 - q.$$

**ОПРЕДЕЛЕНИЕ 4.1.4.** *Медианой* случайной величины  $X$  называется ее квантиль порядка  $\frac{1}{2}$ .

Медиана случайной величины  $X$  обозначается  $\text{med } X$ . По определению медианы, одновременно выполняются неравенства

$$P(X \leq \text{med } X) \geq \frac{1}{2} \quad \text{и} \quad P(X \geq \text{med } X) \geq \frac{1}{2}.$$

Другими словами, медиана случайной величины  $X$  характеризует “центр” ее распределения в том смысле, что случайная величина  $X$  может с равными возможностями быть как меньше медианы, так и больше нее.

Медиану *непрерывной* случайной величины можно вычислить как корень уравнения

$$F(x) = \frac{1}{2},$$

где  $F(x)$  – функция распределения случайной величины  $X$ .

Смысл понятия медианы поясняет следующая задача, являющаяся в некотором смысле дополнением к Задаче 4.1.1.

**ЗАДАЧА 4.1.2.** Предположим, что некий человек на перемещение из точки  $a$  в точку  $b$  затрачивает усилия, равные  $|b - a|$ . Это означает, что его затраты на перемещение пропорциональны длине пути. Скажем, мы рассматриваем делового человека, для которого фактор времени имеет решающее значение и который может перемещаться в любую точку с постоянной скоростью. Пусть  $X$  – случайная величина, характеризующая

координату точки, куда необходимо переместиться этому человеку. Спрашивается: где должен жить этот человек, чтобы минимизировать свои ожидаемые усилия, связанные с перемещениями?

Сведем решение этой задачи к доказательству неравенства

$$E|X - \text{med } X| \leq E|X - d|, \quad (4.1.3)$$

справедливого для любого числа  $d$ . Это будет означать, что решением задачи является медиана случайной величины  $X$ . Сначала рассмотрим случай  $d > \text{med } X$ . Тогда

$$|X - d| - |X - \text{med } X| = \begin{cases} \text{med } X - d & \text{при } X \in [d, \infty), \\ d + \text{med } X - 2X & \text{при } X \in (\text{med } X, d), \\ d - \text{med } X & \text{при } X \in (-\infty, \text{med } X]. \end{cases}$$

Так как  $d + \text{med } X - 2X > \text{med } X - d$  при  $\text{med } X < X < d$ , то

$$\begin{aligned} E|X - d| - E|X - \text{med } X| &= E[|X - d| - |X - \text{med } X|] \\ &\geq (\text{med } X - d)E1_{[d, \infty)}(X) + (\text{med } X - d)E1_{(\text{med } X, d)}(X) \\ &\quad + (d - \text{med } X)E1_{(-\infty, \text{med } X]}(X) \\ &= (\text{med } X - d)P(X \geq d) + (\text{med } X - d)P(\text{med } X < X < d) \\ &\quad + (d - \text{med } X)P(X \leq \text{med } X) \\ &= (d - \text{med } X)[P(X \leq \text{med } X) - P(X > \text{med } X)] \geq 0. \end{aligned}$$

Последнее неравенство вытекает из определения медианы. Но это и означает справедливость (4.1.3) для случая  $d > \text{med } X$ . Случай  $d < \text{med } X$  рассматривается аналогично.

Таким образом, медиана характеризует центр распределения случайной величины  $X$  (определяет центр населенного пункта в терминах Задачи 4.1.2) в смысле функции потерь, имеющей вид абсолютной величины.

В отличие от математического ожидания, медиана всегда определена, но не всегда единственна. Поясним сказанное на примере случайной величины  $X$ , принимающей значения 0 и 1 с равными вероятностями. Тогда, как легко видеть, определению медианы удовлетворяет любая точка отрезка  $[0, 1]$  (включая концы). В то же время, очевидно,  $EX = \frac{1}{2}$ .

ОПРЕДЕЛЕНИЕ 4.1.3. Модой непрерывной случайной величины  $X$ , имеющей плотность распределения  $f(x)$ , называется число  $\text{mod } X$ , удовлетворяющее условию

$$f(x) \leq f(\text{mod } X)$$

для любого  $x$ .

Другими словами, мода непрерывной случайной величины  $X$  — это такая точка, в которой плотность распределения случайной величины  $X$  принимает наибольшее значение.

ОПРЕДЕЛЕНИЕ 4.1.4. Модой дискретной случайной величины  $X$ , принимающей значения  $x_1, x_2, \dots$ , называется такое ее значение  $x_{k_0}$ , которое удовлетворяет условию

$$P(X = x_k) \leq P(X = x_{k_0})$$

для любого  $k = 1, 2, \dots$ .

Из этих определений вытекает, что моду случайной величины  $X$  можно интерпретировать как наиболее вероятное ее значение.

Каждая из введенных выше числовых характеристик — математическое ожидание, медиана и мода — описывает “центральную” тенденцию в поведении случайной величины. Однако эти характеристики совсем не обязаны совпадать. Чтобы в этом убедиться, рассмотрим следующую весьма поучительную задачу.

ЗАДАЧА 4.1.3. Предположим, что вероятность того, что среднегодовой доход наугад выбранного жителя некоторой страны превосходит уровень  $x$ , равна  $\alpha^x$ , где  $\alpha$  — некоторое число,  $0 < \alpha < 1$ . Другими словами, доля индивидуумов, чей среднегодовой доход превосходит уровень  $x$ , во всем населении страны, составляет  $\alpha^x \cdot 100\%$ . Каков средний доход жителей этой страны?

Эта кажущаяся очень простой задача весьма знакома. Однако на самом деле эта задача не является тривиальной, но все затруднения не являются математическими. Главная особенность связана с интерпретацией понятия среднего дохода.

Пусть  $X$  — случайная величина, равная среднегодовому доходу наугад выбранного жителя. Тогда  $F(x) = P(X < x) = 1 - \alpha^x$ . Такая функция

распределения соответствует непрерывной случайной величине с плотностью вероятностей  $f(x) = C \cdot \alpha^x$ , где число  $C > 0$  определяется из условия

$$\int_0^{\infty} f(x) dx = C \int_0^{\infty} \alpha^x dx = 1$$

(см. соотношение 3.2.1), то есть

$$C = \frac{1}{\int_0^{\infty} \alpha^x dx}.$$

Можно показать, что  $C = \ln \frac{1}{\alpha}$ , где  $\ln$  – это так называемый натуральный логарифм (логарифм по основанию  $e$ , см. ниже). Тогда

$$EX = C \int_0^{\infty} x \alpha^x dx = \frac{1}{C}.$$

Медиана случайной величины  $X$  находится как решение уравнения  $1 - \alpha^x = \frac{1}{2}$  и оказывается равной

$$\text{med } X = \log_{\alpha} \frac{1}{2}.$$

В то же время, мода случайной величины  $X$ , то есть та точка, в которой плотность  $f(x) = C \cdot \alpha^x$  максимальна, равна нулю, то есть

$$\text{mod } X = 0.$$

Пусть, к примеру,  $\alpha = \frac{1}{2}$ . Тогда

$$EX \approx 1,4427, \quad \text{med } X = 1, \quad \text{mod } X = 0.$$

Таким образом, если мы по какой-либо причине желаем приукрасить ситуацию (скажем, мы готовим отчет правительства перед парламентом), но *не хотим подтасовывать данные*, то, конечно же, мы будем характеризовать средний доход как математическое ожидание, которое в данном случае почти в полтора раза (а это весьма существенно!) превосходит медиану. Если же мы хотим дать более пессимистический ответ (скажем, готовя аналитическую записку для умеренной оппозиции в парламенте), то мы будем характеризовать средний доход как медиану. Ну а если мы,

к примеру, представляем ультрарадикальную оппозицию, то на митингах мы будем, опять-таки, *не подтасовывая данные*, говорить о среднем доходе, имея в виду моду случайной величины  $X$ .

Анализируя решение приведенной выше задачи, следует иметь в виду следующее обстоятельство. В определении математического ожидания в равной степени учитываются как вероятности значений случайной величины, так и *сами значения*. Поэтому наличие одного-двух мультимиллиардеров довольно заметно повышает средний доход граждан страны, характеризуемый математическим ожиданием. В определении же медианы, грубо говоря, величина больших значений рассматриваемой характеристики не играет большой роли. Важно лишь их количество. Поэтому наличие нескольких мультимиллиардеров практически не влияет на средний доход граждан страны, характеризуемый медианой.

### 1.4.2. Характеристики разброса случайных величин

Наряду с характеристикой “центра” распределения вероятностей случайной величины, определенную информацию об этом распределении несет параметр, характеризующий степень разброса возможных значений случайной величины вокруг “центрального” значения. Как и “центр”, разброс могут описывать различные характеристики.

**ОПРЕДЕЛЕНИЕ 4.2.1.** *Дисперсией* случайной величины  $X$  называется число  $DX$ , определяемое как

$$DX = E(X - EX)^2.$$

Словами дисперсию можно определить как среднее значение квадрата отклонения случайной величины от своего математического ожидания. Дисперсия характеризует “центр” распределения квадратов отклонения случайной величины  $X$  от ее “центра” при многократном воспроизведении стохастической ситуации, с которой связана случайная величина  $X$ .

Дисперсию иногда бывает удобно вычислять по формуле

$$DX = EX^2 - (EX)^2.$$

Перечислим основные свойства дисперсии.

- Неслучайный коэффициент можно выносить из-под знака дисперсии, возведя его в квадрат:

$$D(cX) = c^2 \cdot DX.$$

- Дисперсия любой случайной величины неотрицательна.
- Дисперсия случайной величины равна нулю тогда и только тогда, когда эта случайная величина всегда (с вероятностью единица) принимает одно и то же значение (естественно, равное ее математическому ожиданию).

Для унификации единиц измерения степени разброса с единицами измерения самой случайной величины используют понятие *среднеквадратичного отклонения*, определяемого как квадратный корень из дисперсии:  $\sigma = \sqrt{DX}$ .

Несложно видеть, что

$$DX = E(X - EX)^2 = \min_a E(X - a)^2$$

(см. Задачу 4.1.1). Поэтому, вспоминая Задачу 4.1.2, мы заключаем, что наряду с дисперсией степень разброса значений случайной величины может характеризовать также такой параметр, как

$$E|X - \text{med } X| = \min_a E|X - a|.$$

Однако такая характеристика используется редко из-за отсутствия у нее удобных аналитических свойств.

Однако в описательной статистике довольно часто используется еще одна характеристика степени разбросанности возможных значений случайной величины.

**ОПРЕДЕЛЕНИЕ 4.2.2.** Пусть  $x_{1/4}$  и  $x_{3/4}$  – квантили случайной величины  $X$  порядков  $\frac{1}{4}$  и  $\frac{3}{4}$  соответственно. Разность  $x_{3/4} - x_{1/4}$  называется *интерквартильным размахом* случайной величины  $X$ .

Можно сказать, что интерквартильный размах в той же мере характеризует степень разброса возможных значений случайной величины, в какой медиана характеризует ее центральное значение: внутрь интервала



$[x_{1/4}, x_{3/4}]$  случайная величина  $X$  попадает с вероятностью  $\frac{1}{2}$ . При этом среди всех возможных интервалов с таким свойством (легко видеть, что все такие интервалы имеют своими концами квантили случайной величины порядков  $q - \frac{1}{4}$  и  $q + \frac{1}{4}$  при  $\frac{1}{4} \leq q \leq \frac{3}{4}$ ) интервал  $[x_{1/4}, x_{3/4}]$  занимает “срединное” положение, поскольку вероятность того, что случайная величина  $X$  попадет левее этого интервала равна вероятности того, что случайная величина  $X$  попадет правее этого интервала (при этом обе эти вероятности равны  $\frac{1}{4}$ ). Поэтому длина этого интервала вполне может быть принята в качестве меры разброса случайной величины.

### 1.4.3. Неравенства для характеристик центра и разброса случайных величин

**Неравенство Маркова.** Пусть  $X$  – неотрицательная случайная величина. Тогда, используя свойства математического ожидания, для произвольного положительного числа  $x$  мы можем записать:

$$\begin{aligned} E \min\{x, X\} &= E [X1_{[0,x)}(X) + x1_{[x,\infty)}(X)] \\ &= E [X1_{[0,x)}(X)] + E [x1_{[x,\infty)}(X)] \\ &= xE1_{[x,\infty)}(X) + E [X1_{[0,x)}(X)] \\ &\geq xE1_{[x,\infty)}(X) = xP(X \geq x), \end{aligned}$$

откуда, сравнивая левую и правую части этого соотношения, мы получаем неравенство

$$P(X \geq x) \leq \frac{E \min\{x, X\}}{x}, \quad (4.3.1)$$

справедливое при любом  $x > 0$ .

Вытекающее из (4.3.1) неравенство

$$P(X \geq x) \leq \frac{EX}{x}$$

называется *неравенством Маркова*.

**Неравенство Чебышева.** Пусть  $X$  – произвольная случайная величина. Применим неравенство Маркова к случайной величине  $Z = (X - EX)^2$  и

получим

$$P(|X - EX| \geq x) = P\left((X - EX)^2 \geq x^2\right) \leq \frac{E(X - EX)^2}{x^2} = \frac{DX}{x^2}.$$

Соотношение между левой и правой частями этого неравенства называют *неравенством Чебышева*:

$$P(|X - EX| \geq x) \leq \frac{DX}{x^2}. \quad (4.3.2)$$

С помощью неравенств Маркова и Чебышева можно оценивать вероятности больших значений неотрицательных случайных величин или вероятности больших отклонений случайных величин от своего среднего значения.

### Неравенство Иенсена и его частные случаи.

**ОПРЕДЕЛЕНИЕ 4.3.1.** Функция  $g(x)$  называется *выпуклой*, если ее график лежит *выше* касательной к нему, проведенной в любой его точке. Функция  $h(x)$  называется *вогнутой*, если ее график лежит *ниже* касательной к нему, проведенной в любой его точке.

Примерами выпуклых функций служат функции  $g(x) = x^2$  ( $-\infty < x < \infty$ ),  $g(x) = 1/x$  ( $x > 0$ ). Примерами вогнутых функций служат функции  $h(x) = \sqrt{x}$  ( $x > 0$ ),  $h(x) = \log_a x$  ( $a > 1, x > 0$ ).

Пусть  $X$  – случайная величина,  $g(x)$  – произвольная выпуклая функция, а  $h(x)$  – произвольная вогнутая функция. *Неравенство Иенсена* утверждает, что

$$Eg(X) \geq g(EX), \quad Eh(X) \leq h(EX).$$

В частности, применяя неравенство Иенсена для выпуклых функций к произвольной случайной величине  $Y$  и функции  $g(x) = |x|$ , мы получаем

$$|EY| \leq E|Y|. \quad (4.3.3)$$

С помощью неравенства Иенсена мы можем получить следующие полезные соотношения между характеристиками “центра” случайной величины и характеристиками ее разброса. Пусть  $X$  – произвольная случайная величина. Применим неравенство (4.3.3) к случайной величине

$Y = X - \text{med } X$ . Вследствие линейности математического ожидания мы при этом получим

$$|EY| = |E(X - \text{med } X)| = E|X - \text{med } X|. \quad (4.3.4)$$

Пусть  $X$  — произвольная случайная величина. Применим теперь неравенство Йенсена для выпуклых функций к случайной величине  $Y = |X - EX|$  и функции  $g(x) = x^2$ . Мы получим

$$(E|X - EX|)^2 \leq E(X - EX)^2$$

или, что то же самое,

$$E|X - EX| \leq \sqrt{DX}. \quad (4.3.5)$$

Соотношение (4.3.5) принято называть *неравенством Ляпунова*.

Вспомним Задачу 4.1.2, согласно которой  $E|X - \text{med } X| \leq E|X - a|$  для любого  $a$ , в частности, при  $a = EX$  мы получаем  $E|X - \text{med } X| \leq E|X - EX|$ . Используя последнее неравенство как связующее звено между (4.3.4) и (4.3.5), мы получаем неравенство

$$|EX - \text{med } X| \leq \sqrt{DX}.$$

#### 1.4.4. Аналитические средние

Помимо введенных выше характеристик “центра” распределения случайной величины (математическое ожидание, медиана, мода), которые мы будем условно называть *структурными* средними, в эконометрических задачах традиционно используют так называемые *аналитические* средние значения, примерами которых являются среднее арифметическое, среднее гармоническое и среднее геометрическое. Все аналитические средние определяются единообразно. А именно, пусть  $X$  — положительная случайная величина, а  $Q(x)$  — некоторая функция, имеющая обратную функцию  $Q^{-1}(x)$  (если  $y = Q(x)$ , то  $x = Q^{-1}(y)$ ). С помощью функции  $Q(x)$  и операции вычисления математического ожидания  $Q$ -среднее значение случайной величины  $X$ , обозначаемое  $M_Q X$ , определяется как

$$M_Q X = Q^{-1}(EQ(X)).$$

В частности, если  $Q(x) \equiv x$ , то  $Q^{-1}(x) \equiv x$  и мы получаем “обобщенное среднее арифметическое”

$$M_Q X \equiv m_A(X) = EX.$$

Если  $Q(x) \equiv 1/x$ , то  $Q^{-1}(x) \equiv 1/x$  и мы получаем “обобщенное среднее гармоническое”

$$M_Q X \equiv m_H(X) = \frac{1}{E[1/X]}.$$

Если  $Q(x) \equiv \log_a x$ , то  $Q^{-1}(x) \equiv a^x$  и мы получаем “обобщенное среднее геометрическое”

$$M_Q X \equiv m_G(X) = a^{E \log_a X}.$$

Несложно убедиться, что, если случайная величина  $X$  принимает положительные значения  $x_1, \dots, x_n$  с равными вероятностями  $P(X = x_k) = \frac{1}{n}$ ,  $k = 1, \dots, n$ , то мы соответственно получаем традиционные средние арифметическое, гармоническое и геометрическое:

$$m_A(X) = \frac{1}{n} \sum_{k=1}^n x_k, \quad m_H(X) = \frac{1}{\frac{1}{n} \sum_{k=1}^n \frac{1}{x_k}}, \quad m_G(X) = \sqrt[n]{x_1 \cdot \dots \cdot x_n}.$$

С помощью неравенства Иенсена мы получаем следующее соотношение между средним арифметическим, средним гармоническим и средним геометрическим:

$$m_H(X) \leq m_G(X) \leq m_A(X). \quad (4.3.6)$$

Действительно, если основание логарифма в определении среднего геометрического больше единицы (то есть  $a > 1$ ), то функция  $\log_a x$  вогнута и по неравенству Иенсена для вогнутых функций мы имеем

$$E \log_a X \leq \log_a EX.$$

Поэтому

$$m_G(X) = a^{E \log_a X} \leq a^{\log_a EX} = EX = m_A(X), \quad (4.3.7)$$

то есть имеет место правое неравенство в (4.3.6). Обозначим  $Y = 1/X$ . Тогда по неравенству Иенсена для вогнутых функций

$$m_H(X) = \frac{1}{EY} = a^{-E \log_a Y} \leq a^{-E \log_a Y} = a^{E \log_a X} = m_G(X), \quad (4.3.8)$$

то есть верно левое неравенство в (4.3.6).

Если же основание логарифма в определении  $m_G(X)$  меньше единицы, то функция  $\log_a x$  выпукла и по неравенству Йенсена

$$E \log_a X \geq \log_a EX. \quad (4.3.9)$$

Следовательно,

$$a^{E \log_a X} \leq a^{\log_a EX},$$

то есть имеет место (4.3.7). Из (4.3.9) вытекает, что

$$-E \log_a X \leq -\log_a EX,$$

откуда мы заключаем, что и соотношение (4.3.8) сохраняет свою силу при  $a < 1$ .

### 1.4.5. Числовые характеристики формы распределения случайных величин

Пусть  $X$  – случайная величина и  $EX$  и  $DX$  – соответственно ее математическое ожидание и дисперсия. Рассмотрим случайную величину

$$\tilde{X} = \frac{X - EX}{\sqrt{DX}}.$$

Несложно убедиться, что  $E\tilde{X} = 0$ ,  $D\tilde{X} = E\tilde{X}^2 = 1$ . Эти соотношения выполняются для любой случайной величины  $X$ , у которой существует (конечна) дисперсия. Однако математические ожидания более высоких степеней случайной величины  $\tilde{X}$  уже будут разными для разных случайных величин  $X$  и несут в себе некоторую информацию о форме распределения случайной величины  $X$ .

**ОПРЕДЕЛЕНИЕ 4.5.1.** Предположим, что существует  $E|X|^3$ . Число

$$\alpha_3 = E\tilde{X}^3 = E\left(\frac{X - EX}{\sqrt{DX}}\right)^3$$

называется *коэффициентом асимметрии* случайной величины  $X$ .

Если  $\kappa_3 < 0$ , то левый “хвост” распределения случайной величины  $X$  тяжелее правого. Если при этом случайная величина  $X$  непрерывна, то в этом случае левый “склон” ее плотности более пологий, нежели правый. Если  $\kappa_3 > 0$ , то левый “хвост” распределения случайной величины  $X$  легче правого. Если при этом случайная величина  $X$  непрерывна, то в этом случае левый “склон” ее плотности более крут, нежели правый. Если случайная величина  $X$  симметрична, то есть распределения величин  $X$  и  $-X$  совпадают, то  $\kappa_3 = 0$ .

ОПРЕДЕЛЕНИЕ 4.5.2. Предположим, что существует  $EX^4$ . Число

$$\kappa_4 = E\tilde{X}^4 = E\left(\frac{X - EX}{\sqrt{DX}}\right)^4 = \frac{E(X - EX)^4}{(DX)^2}$$

называется *коэффициентом эксцесса* или *коэффициентом островершинности* случайной величины  $X$ .

Чем больше коэффициент островершинности, тем более остра вершина плотности соответствующей случайной величины. Наоборот, если коэффициент островершинности мал, то график плотности соответствующей случайной величины имеет тупую вершину.

В математике известны многие замечательные числа. Например, всем хорошо знакомо число  $\pi = 3,141592654\dots$ , равное отношению длины окружности к диаметру этой окружности. Не менее замечательную роль играет также число  $e = 2,718281829\dots$ , равное так называемому второму замечательному пределу: можно показать, что  $(1 + \frac{1}{n})^n \rightarrow e$  при  $n \rightarrow \infty$ . Это число будет играть очень важную роль во многих приводимых ниже формулах. При этом мы часто будем использовать обозначение  $\exp\{x\}$  для числа  $e^x$ .

В качестве эталона “островершинности” традиционно принимают так называемое нормальное распределение, задаваемое плотностью

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

Коэффициент эксцесса случайной величины, имеющей “эталонную” стандартную нормальную плотность  $\varphi(x)$ , равен 3.

## 1.5. Независимость случайных величин

### 1.5.1. Независимые случайные величины

Напомним, что в разделе 1.2.1 мы ввели понятие независимых событий. Мы назвали два события  $A$  и  $B$ , связанные с некоторой стохастической ситуацией, независимыми, если  $P(A \cap B) = P(A) \cdot P(B)$ . Это понятие, равно как и понятие независимости в совокупности можно распространить и на случайные величины.

В некоторых стохастических ситуациях с каждым элементарным исходом связано значение не одной, а нескольких случайных величин. К примеру, при проведении социологического опроса из интересующей категории населения случайно выбирается некоторое заранее заданное число, скажем,  $n$  респондентов. Для простоты предположим, что каждому из них задается один и тот же единственный вопрос, например, предлагается по десятибалльной шкале оценить эффективность работы какого-либо государственного органа, скажем, органов охраны общественного порядка. Тогда формальный результат опроса имеет вид  $X_1, X_2, \dots, X_n$ , где  $X_i$  — ответ  $i$ -го респондента на заданный вопрос. Здесь элементарный исход — это выбор  $n$  конкретных респондентов. Каждому такому элементарному исходу соответствует не одно, а  $n$  чисел, то есть с рассматриваемой стохастической ситуацией связаны  $n$  случайных величин, значения которых зависят от выбора конкретных  $n$  респондентов.

**ОПРЕДЕЛЕНИЕ 5.1.1.** Случайные величины  $X$  и  $Y$ , связанные с некоторой стохастической ситуацией, называются *независимыми*, если, какими бы ни были числа  $x$  и  $y$ , независимы события  $\{\omega: X(\omega) < x\}$  и  $\{\omega: Y(\omega) < y\}$ .

Для дискретных случайных величин критерий независимости принимает более простой вид. А именно, пусть  $X$  и  $Y$  — дискретные случайные величины, принимающие соответственно значения  $x_1, x_2, \dots$  и  $y_1, y_2, \dots$ . Тогда случайные величины  $X$  и  $Y$  *независимы* тогда и только тогда, когда для всех  $i$  и  $j$  независимы события  $\{\omega: X(\omega) = x_i\}$  и  $\{\omega: Y(\omega) = y_j\}$ .

Можно показать, что если случайные величины независимы, то

$$EXY = EX \cdot EY.$$

**ОПРЕДЕЛЕНИЕ 5.1.2.** Случайные величины  $X_1, X_2, \dots, X_n$  называются *независимыми в совокупности*, если, какими бы ни были числа  $x_1, x_2, \dots, x_n$ , события  $\{\omega: X_i(\omega) < x_i\}, i = 1, \dots, n$ , независимы в совокупности.

**ОПРЕДЕЛЕНИЕ 5.1.3.** Пусть  $X_1, X_2, \dots$  – бесконечная последовательность случайных величин. Мы будем говорить о ней как о *последовательности независимых случайных величин*, если при каждом  $n \geq 2$  составляющие ее случайные величины  $X_1, \dots, X_n$  независимы в совокупности.

Вообще говоря, понятие независимости применительно к разным объектам можно понимать и трактовать по-разному. Например, независимость может быть политической, экономической и т.п. Более того, это понятие интуитивно связано с понятием зависимости. В приведенных выше определениях мы говорим о *стохастической независимости*. При этом понятие *зависимости* мы не конкретизируем и не будем называть *зависимыми* те случайные величины, которые не являются стохастически независимыми. Дело в том, что, как мы увидим, в теории вероятностей известны случайные величины (см. раздел 2.3.2), которые стохастически независимы, но при этом связаны функциональной зависимостью.

## 1.5.2. Ковариация. Коэффициент корреляции

Наряду с числовыми характеристиками центра и разброса можно ввести числовые характеристики взаимозависимости случайных величин.

**ОПРЕДЕЛЕНИЕ 5.2.1.** Пусть  $X$  и  $Y$  – случайные величины. Число

$$\text{cov}(X, Y) = E[(X - EX)(Y - EY)]$$

называется *ковариацией* случайных величин  $X$  и  $Y$ .

Ковариацию удобно вычислять по формуле

$$\text{cov}(X, Y) = EXY - EX \cdot EY.$$

Если ковариация случайных величин  $X$  и  $Y$  положительна, то в среднем, при многократном воспроизведении стохастической ситуации, отклонение случайной величины  $X$  от своего среднего значения в большую



сторону влечет одновременное отклонение случайной величины  $Y$  от своего среднего значения также в большую сторону и отклонение случайной величины  $Y$  от своего среднего значения в большую сторону влечет одновременное отклонение случайной величины  $X$  от своего среднего значения также в большую сторону. В то же время отклонение случайной величины  $X$  от своего среднего значения в меньшую сторону влечет одновременное отклонение случайной величины  $Y$  от своего среднего значения также в меньшую сторону и отклонение случайной величины  $Y$  от своего среднего значения в меньшую сторону влечет одновременное отклонение случайной величины  $X$  от своего среднего значения также в меньшую сторону. В таком случае говорят, что случайные величины  $X$  и  $Y$  *положительно коррелированы*.

Если же ковариация случайных величин  $X$  и  $Y$  отрицательна, то эти случайные величины в среднем, при многократном воспроизведении стохастической ситуации, отклоняются от своих средних значений в разные стороны. В таком случае говорят, что случайные величины  $X$  и  $Y$  *отрицательно коррелированы*.

Если  $\text{cov}(X, Y) = 0$ , то говорят, что случайные величины  $X$  и  $Y$  *некоррелированы*. Некоррелированность является более слабым свойством случайных величин, чем их независимость. Действительно, легко видеть, что, если случайные величины  $X$  и  $Y$  независимы, то  $\text{cov}(X, Y) = 0$ . Однако обратное утверждение неверно: из равенства нулю ковариации случайных величин не следует их независимость. В этом нас убеждает следующий пример.

**ПРИМЕР 5.2.1.** Рассмотрим случайные величины  $X$  и  $Y$ , совместное распределение вероятностей которых задается следующим соотношением

$$\begin{aligned}
 P(X = 0; Y = 1) &= P(X = 0; Y = -1) \\
 &= P(X = -1; Y = 0) = P(X = 1; Y = 0) \\
 &= P(X = \sqrt{2}/2; Y = \sqrt{2}/2) \\
 &= P(X = -\sqrt{2}/2; Y = \sqrt{2}/2) \\
 &= P(X = \sqrt{2}/2; Y = -\sqrt{2}/2) \\
 &= P(X = -\sqrt{2}/2; Y = -\sqrt{2}/2) = \frac{1}{8}.
 \end{aligned}$$

Несложно видеть, что при этом

$$EXY = 0 \cdot \frac{1}{8} + 0 \cdot \frac{1}{8} + 0 \cdot \frac{1}{8} + 0 \cdot \frac{1}{8} + \frac{1}{2} \cdot \frac{1}{8} - \frac{1}{2} \cdot \frac{1}{8} - \frac{1}{2} \cdot \frac{1}{8} + \frac{1}{2} \cdot \frac{1}{8} = 0.$$

В то же время

$$P(X = -1) = P(X = 1) = P(Y = -1) = P(Y = 1) = \frac{1}{8},$$

$$P(X = 0) = P(Y = 0) = \frac{1}{4},$$

$$P(X = \sqrt{2}/2) = P(X = -\sqrt{2}/2) = P(Y = \sqrt{2}/2) = P(Y = -\sqrt{2}/2) = \frac{1}{4},$$

откуда

$$EX = 0, \quad EY = 0.$$

Таким образом,

$$\text{cov}(X, Y) = EXY - EX \cdot EY = 0,$$

то есть случайные величины  $X$  и  $Y$  некоррелированы. Однако мы замечаем, что, например,

$$0 = P(X = 0; Y = 0) \neq P(X = 0) \cdot P(Y = 0) = \frac{1}{8},$$

то есть случайные величины  $X$  и  $Y$  не являются независимыми. Более того, случайные величины  $X$  и  $Y$  связаны жесткой функциональной зависимостью

$$X^2 + Y^2 = 1.$$

Для дисперсии суммы двух случайных величин справедлива формула

$$D(X + Y) = DX + DY + 2 \text{cov}(X, Y).$$

Отсюда вытекает, что дисперсия суммы независимых случайных величин равна сумме дисперсий слагаемых.

Ковариация не является безразмерной и ее величина зависит от того, какими единицами измеряются рассматриваемые случайные величины. Другими словами, по величине ковариации нельзя судить о том, насколько тесна взаимосвязь (взаимозависимость) случайных величин. Чтобы обойти это препятствие, рассмотрим еще одну характеристику.

## ОПРЕДЕЛЕНИЕ 5.2.2. Число

$$\rho = \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DX \cdot DY}}$$

называется коэффициентом корреляции случайных величин  $X$  и  $Y$ .

Можно показать, что всегда  $|\rho(X, Y)| \leq 1$ . Более того, если  $|\rho(X, Y)| = 1$ , то случайные величины  $X$  и  $Y$  связаны линейной зависимостью: существуют числа  $a$  и  $b$  такие, что  $P(Y = aX + b) = 1$ . При этом, если  $\rho = 1$ , то  $a > 0$ , а если  $\rho = -1$ , то  $a < 0$ . Если случайные величины  $X$  и  $Y$  независимы, то, очевидно,  $\rho = 0$ . Таким образом, можно сказать, что величина коэффициента корреляции характеризует степень линейной зависимости.

Близость коэффициента корреляции к единице означает лишь наличие статистической взаимосвязи, но никак не может быть использована для выявления направления причинно-следственной взаимосвязи. Например, коэффициент корреляции между числом пожаров в наугад выбранном районе большого города за год и числом пожарных частей, расположенном в этом районе, как правило, довольно велик (например, для Нью-Йорка он приблизительно равен 0,8). Но это совсем не означает, что чем больше пожарных частей, тем больше пожаров. Другой пример. В Голландии значение коэффициента корреляции между числом новорожденных за год и числом гнезд аистов в соответствующем районе довольно близко к единице ( $\approx 0,8$ ). Однако мы не можем использовать это обстоятельство для обоснования теории о том, что детей приносят аисты. Объяснение здесь, увы, другое. Дело в том, что, в Голландии вообще рождаемость низка и, как правило, дети рождаются в молодых семьях, а молодые семьи поселяются в отдельных новых домах, на крышах которых и вьют свои гнезда аисты.

## 1.6. Испытания Бернулли

Рассмотрим последовательность однотипных случайных экспериментов (испытаний), в каждом из которых возможны лишь два взаимоисключающих исхода. Условимся называть эти исходы “успех” и “неудача”. Пусть вероятность успеха одинакова для каждого испытания и равна

$p$ ,  $0 \leq p \leq 1$ . С этими испытаниями свяжем случайные величины  $X_1, \dots, X_n$  по следующему правилу. Если в  $k$ -м испытании зафиксирован успех, то  $X_k = 1$ , а если в  $k$ -м испытании зафиксирована неудача, то  $X_k = 0$ ,  $k = 1, \dots, n$ . При этом, очевидно, все случайные величины  $X_1, \dots, X_n$  имеют одинаковое распределение:

$$P(X_k = 1) = p = 1 - P(X_k = 0), \quad k = 1, \dots, n. \quad (6.1.1)$$

Предположим, что рассматриваемые дихотомические испытания независимы в том смысле, что случайные величины  $X_1, \dots, X_n$  независимы в совокупности. Испытания, удовлетворяющие указанным условиям, называются *испытаниями Бернулли*.

**ЗАДАЧА 6.1.1.** Набранную книгу независимо друг от друга вычитывали два корректора. Первый корректор обнаружил  $m_1$  опечаток, второй заметил  $m_2$  опечаток. При этом  $m$  опечаток оказались обнаруженными и первым, и вторым корректорами. Сколько опечаток остались незамеченными?

Во-первых, казалось бы, имеющихся данных недостаточно, чтобы ответить на поставленный вопрос. Действительно, как же можно подсчитать число объектов (в данном случае — оставшихся опечаток), о которых достоверно не известно, существуют они или нет!

Во-вторых, ясно, что формулировка задачи не имеет строгого математического вида, заранее предполагающего выбор конкретной математической модели, описывающей процесс отыскания опечаток. Такая ситуация, в отличие от подавляющего большинства задач, включаемых в задачки по теории вероятностей, типична для практики, когда перед тем, как решить задачу, исследователь должен ее аккуратно сформулировать на математическом языке, выбрав конкретную математическую модель.

Сформулированная задача интересна тем, что показывает, как за счет выбора довольно разумной и естественной математической модели можно компенсировать недостаток информации и найти приемлемое решение задачи, то есть сделать на первый взгляд неразрешимую задачу вполне решаемой.

Сначала заметим, что общее число замеченных опечаток равно  $m_1 + m_2 - m$  (так как опечатки, замеченные обоими корректорами одновременно, входят как в число опечаток, замеченных первым корректором,

так и число опечаток, замеченных вторым корректором, то, чтобы не считать каждую такую опечатку дважды, из суммы  $m_1 + m_2$  нужно вычесть  $m$ , сравните это выражение с формулой для объединения множеств в разделе 1.2.2). Чтобы решить задачу, мы должны сделать дополнительные предположения и попытаться построить математическую модель (которая не должна входить в противоречие с представлением о процессе работы корректоров, основанным на здравом смысле). Итак, предположим, что первый корректор находит очередную опечатку вне зависимости, нашел он предыдущую опечатку или пропустил ее. При этом будем считать, что для первого корректора возможности обнаружения каждой опечатки одинаковы и зависят от его квалификации. Таким образом, процесс вычитки книги первым корректором может быть описан последовательностью испытаний Бернулли с вероятностью успеха в отдельном испытании (обнаружения очередной опечатки) равной, скажем,  $p_1$ . Аналогично, процесс вычитки книги вторым корректором может быть описан последовательностью испытаний Бернулли с вероятностью успеха в отдельном испытании равной  $p_2$ . С этими двумя последовательностями испытаний Бернулли свяжем третью, описывающую процесс вычитки книги обоими корректорами. Для каждой опечатки возможны следующие варианты:

- ее не заметил ни первый, ни второй корректор;
- ее заметил первый корректор, но не заметил второй;
- ее заметил второй корректор, но не заметил первый;
- ее заметили оба корректора.

Так как по условию задачи корректоры работают независимо друг от друга, то перечисленные выше варианты имеют вероятности  $(1 - p_1)(1 - p_2)$ ,  $p_1(1 - p_2)$ ,  $(1 - p_1)p_2$ ,  $p_1p_2$  соответственно. Так как в условии задачи фигурирует число опечаток, замеченных обоими корректорами одновременно, наряду с описанными выше двумя последовательностями испытаний Бернулли рассмотрим третью, в которой вероятность успеха в отдельном испытании (обнаружение опечатки обоими корректорами) равна  $p_1p_2$ .

Общее количество опечаток в книге обозначим  $n$ . Понятно, что это число равно количеству испытаний Бернулли в каждой из трех указанных выше серий.

Интуитивно ясно, что теоретическая вероятность успеха в каждом из  $n$  испытаний Бернулли примерно равна отношению числа зарегистрированных успехов к общему числу испытаний. Более строго в этом можно убедиться на основании закона больших чисел (см. раздел 1.8); также см. задачу о рейтинге в разделе 2.2.4. Более того, можно заметить, что указанное отношение является наиболее правдоподобной оценкой теоретической вероятности успеха (см. раздел 2.4.1).

Таким образом, используя условия задачи и для удобства записи заменяя приближенные равенства строгими, мы приходим к системе следующих уравнений:

$$\begin{cases} p_1 = \frac{m_1}{n} \\ p_2 = \frac{m_2}{n} \\ p_1 p_2 = \frac{m}{n} \end{cases}$$

В этой системе три уравнения и три неизвестных:  $n$ ,  $p_1$  и  $p_2$ . Решая эту систему, мы получаем

$$\begin{cases} n = \frac{m_1 m_2}{m} \\ p_1 = \frac{m}{m_2} \\ p_2 = \frac{m}{m_1} \end{cases}$$

Таким образом, в книге осталось

$$\frac{m_1 m_2}{n} - m_1 - m_2 + m$$

опечаток.

### 1.6.1. Биномиальное распределение

Рассмотрим стохастическую ситуацию, заключающуюся в осуществлении  $n$  испытаний Бернулли ( $n \geq 1$ ). Построим соответствующую вероятностную модель. С этой целью заметим, что каждый элементарный исход

последовательности  $n$  испытаний Бернулли представляет собой цепочку вида  $\omega = (\omega_1, \dots, \omega_n)$ , где  $\omega_k$  равно либо 1 либо 0 в зависимости от того, какое значение случайной величины  $X_k$  наблюдается в  $k$ -м испытании, то есть в зависимости от того, закончилось ли  $k$ -е испытание успехом или неудачей. Несложно убедиться, что всего имеется  $N = 2^n$  разных цепочек вида  $\omega = (\omega_1, \dots, \omega_n)$ , в которых  $\omega_k$  равно либо 1 либо 0. Отсюда мы соответственно заключаем, что совокупность всех элементарных исходов рассматриваемой стохастической ситуации состоит из  $N = 2^n$  элементов указанного выше вида.

Так как испытания независимы, то вероятность каждой элементарной цепочки  $\omega = (\omega_1, \dots, \omega_n)$  оказывается равной произведению вероятностей соответствующих исходов дихотомических испытаний, составляющих эту цепочку:

$$P(\{\omega\}) = \prod_{k=1}^n p^{\omega_k} (1-p)^{1-\omega_k} = p^{\sum_{k=1}^n \omega_k} (1-p)^{n-\sum_{k=1}^n \omega_k} \quad (6.1.2)$$

(число единиц в цепочке  $\omega = (\omega_1, \dots, \omega_n)$  равно  $\sum_{k=1}^n \omega_k$  и, соответственно, число нулей в такой цепочке равно  $n - \sum_{k=1}^n \omega_k$ ).

Пусть  $X$  — случайная величина, равная количеству успехов в  $n$  испытаниях Бернулли. Найдем ее распределение. Очевидно, что возможными значениями случайной величины  $X$  являются числа  $0, 1, \dots, n$ . Несложно видеть, что для каждого элементарного исхода  $\omega = (\omega_1, \dots, \omega_n)$  случайная величина  $X$  равна

$$X(\omega) = \sum_{k=1}^n \omega_k. \quad (6.1.3)$$

Пусть  $m$  — целое число,  $0 \leq m \leq n$ . В разделе 1.1.2 мы установили, что существует  $C_n^m$  разных цепочек  $\omega = (\omega_1, \dots, \omega_n)$ , удовлетворяющих условию

$$\sum_{k=1}^n \omega_k = m. \quad (6.1.4)$$

С учетом соотношения (6.1.2) мы замечаем, что каждая из цепочек  $\omega = (\omega_1, \dots, \omega_n)$ , удовлетворяющих условию (6.1.4), имеет вероятность

$P(\{\omega\}) = p^m(1-p)^{n-m}$ . Событие  $\{\omega: X(\omega) = m\}$  состоит из всех цепочек, удовлетворяющих условию (6.1.4). Поэтому

$$p_m = P(X = m) = \sum_{\substack{\omega=(\omega_1, \dots, \omega_n): \\ \omega_1 + \dots + \omega_n = m}} P(\{\omega\}) = C_n^m p^m (1-p)^{n-m}, \quad m = 0, \dots, n.$$

Используя формулу бинома Ньютона, несложно убедиться, что  $p_0 + p_1 + \dots + p_m = 1$ .

Набор пар  $\{(m, p_m), m = 0, \dots, n\}$  называется *биномиальным распределением вероятностей*.

Мы можем заметить, что случайную величину  $X$ , равную числу успехов в  $n$  испытаниях Бернулли, можно представить в виде

$$X = X_1 + \dots + X_n, \quad (6.1.5)$$

где  $X_1, \dots, X_n$  – введенные в самом начале раздела 1.6 независимые случайные величины. На основании (6.1.1) мы заключаем, что

$$EX_k = 1 \cdot p + 0 \cdot (1-p) = p,$$

$$DX_k = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p(1-p), \quad k = 1, \dots, n.$$

Поэтому в силу справедливости представления (6.1.5) мы заключаем, что если случайная величина  $X$  имеет биномиальное распределение, то

$$EX = EX_1 + \dots + EX_n = n \cdot EX_1 = np,$$

$$DX = DX_1 + \dots + DX_n = n \cdot DX_1 = np(1-p).$$

## 1.6.2. Геометрическое распределение

Рассмотрим стохастическую ситуацию, состоящую в проведении испытаний Бернулли до первой неудачи. В такой ситуации совокупность элементарных исходов состоит из цепочек  $\omega = (\omega_1, \dots, \omega_n)$ , в которых  $\omega_1 = \dots = \omega_{n-1} = 1$ , а  $\omega_n = 0$ ,  $n = 1, 2, \dots$ . Таких цепочек бесконечно много. Так как испытания независимы, то цепочка  $\omega = (\omega_1, \dots, \omega_n)$  указанного вида имеет вероятность

$$p_n = P(\{\omega\}) = p^{n-1}(1-p). \quad (6.1.6)$$



По формуле суммы членов бесконечно убывающей геометрической прогрессии можно убедиться, что

$$\sum_{n=1}^{\infty} p^n = (1-p) \sum_{n=1}^{\infty} p^{n-1} = 1.$$

Набор пар  $\{(n, p_n), n = 1, 2, \dots\}$ , где  $p_n$  имеет вид (6.1.6), называется *геометрическим* распределением вероятностей.

С рассматриваемой стохастической ситуацией свяжем случайную величину  $X$ , определив ее для каждого элементарного исхода  $\omega = (\omega_1, \dots, \omega_n)$  указанного выше вида как  $X(\omega) = n$ . Случайная величина  $X$  принимает значения  $1, 2, \dots$  и имеет смысл количества испытаний Бернулли, осуществленных до первой неудачи. В таком случае говорят, что случайная величина  $X$  имеет геометрическое распределение.

Можно показать, что если случайная величина  $X$  имеет геометрическое распределение, то

$$EX = \frac{1}{1-p}, \quad DX = \frac{1}{(1-p)^2}.$$

## 1.7. Предельные теоремы, связанные с испытаниями Бернулли

Вычисление биномиальных вероятностей  $C_n^k p^k (1-p)^{n-k}$  при больших  $n$  представляет собой очень сложную задачу. Однако оказывается возможным использовать приближенные формулы, которые являются следствиями предельных теорем, о которых пойдет речь ниже.

### 1.7.1. Теорема Пуассона. Распределение Пуассона

Напомним, что в разделе 1.4 мы упомянули число  $e = 2,718281829 \dots$ , равное так называемому второму замечательному пределу. Мы договорились использовать обозначение  $\exp\{x\}$  для числа  $e^x$ .

Пусть  $X$  – случайная величина, имеющая биномиальное распределение с параметрами  $n$  и  $p$ , то есть

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Можно показать, что если  $k - 1 \leq n/4$  и  $p \leq 1/4$ , то

$$P(X = k) = \frac{(np)^k}{k!} \exp\{-np + r_n(k)\}, \quad (7.1.1)$$

где

$$\frac{k(1-k) + 6knp - 4(np)^2}{6n} \leq r_n(k) \leq \frac{k(1-k) + 2knp}{2n}.$$

Из соотношения (7.1.1) вытекает возможность использования для указанных выше значений параметров приближенной формулы

$$P(X = k) \approx e^{-np} \frac{(np)^k}{k!}, \quad (7.1.2)$$

которая имеет хорошую точность даже при умеренно больших  $n$ . Из этого же соотношения вытекает, что бóльшую точность дает приближенная формула

$$P(X = k) \approx \frac{(np)^k}{k!} \exp\left\{-np + \frac{k(1-k)}{2n} + kp\right\}. \quad (7.1.3)$$

Однако соотношение (7.1.1) показывает, что если  $n \rightarrow \infty$ , но при этом значение  $p$  остается постоянным, то точность приведенных выше аппроксимаций становится неудовлетворительной. Для того, чтобы при  $n \rightarrow \infty$  формулы (7.1.2) и (7.1.3) давали приемлемую точность, нужно, чтобы одновременно с неограниченным увеличением  $n$  параметр  $p$  стремился к нулю. При этом из соотношения (7.1.1) вытекает следующее утверждение.

Пусть  $n \rightarrow \infty$  и одновременно  $p \rightarrow 0$  так, что  $np \rightarrow \lambda$  для некоторого  $\lambda > 0$ . Тогда для каждого  $k = 0, 1, 2, \dots$

$$P(X = k) = C_n^k p^k (1-p)^{n-k} \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}. \quad (7.1.4)$$

Это утверждение называется *теоремой Пуассона, теоремой о редких событиях*, или *законом малых чисел*.

Условия теоремы Пуассона можно интерпретировать следующим образом. Если число испытаний Бернулли велико, а вероятность успеха в одном отдельном испытании мала и при этом ожидаемое число успехов

умеренно (напомним, что  $np$  – это математическое ожидание числа успехов в  $n$  испытаниях Бернулли с вероятностью успеха в одном отдельном испытании, равной  $p$ ), то биномиальные вероятности можно вычислять по приближенной формуле (7.1.2):

$$C_n^k p^k (1-p)^{n-k} \approx e^{-np} \frac{(np)^k}{k!}.$$

Если мы просуммируем числа в правой части соотношения (7.1.4) по всем  $k$ , то получим

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1,$$

то есть набор пар  $\{(k, e^{-\lambda} \lambda^k / k!), k = 0, 1, \dots\}$  представляет собой распределение вероятностей. Это распределение называется *распределением Пуассона*. Если случайная величина  $X$  имеет распределение Пуассона, то

$$EX = DX = \lambda.$$

Распределение Пуассона часто используется, скажем, в страховании для описания распределения числа страховых случаев в течение некоторого промежутка времени.

Пример применения теоремы Пуассона содержится в следующей задаче.

**Задача 7.1.1.** В двух городах, скажем,  $A$  и  $B$ , примерно одинаковое и довольно большое число жителей и, соответственно, примерно одинаковый уровень системы здравоохранения. В течение года в городе  $A$  зарегистрировано  $n_A$  заболеваний лейкемией, а в городе  $B$  зарегистрировано  $n_B$  заболеваний, причем  $n_A > n_B$ . Свидетельствует ли это обстоятельство о наличии в городе  $A$  некоторого систематического фактора, увеличивающего риск заболевания по сравнению с городом  $B$ ?

Ясно, что конкретный ответ на поставленный вопрос можно дать, только имея конкретные числовые значения  $n_A$  и  $n_B$ . Однако с помощью теоремы Пуассона можно построить довольно осмысленный *метод* получения подобных выводов. Интуитивно ясно, что если  $n_A$  намного больше, чем  $n_B$ , то дополнительный фактор риска в городе  $A$  присутствует. Однако, в то же время,  $n_A$  может оказаться бóльшим  $n_B$  чисто случайно.

Поэтому задача сводится к оценке того, случайно ли наблюдаемое отклонение  $n_A$  от  $n_B$  или оно имеет систематическую причину. Однако и в такой постановке, кажется, не хватает информации для осмысленного решения. Действительно, в условиях ничего не говорится о том, каковы распределения случайных величин, наблюдаемыми значениями которых являются  $n_A$  и  $n_B$ . К тому же в процедуре сравнения случайных величин довольно много сложностей. Намного проще сравнивать конкретные неслучайные числа.

Приведенная выше постановка задачи типична для практики. Она существенно отличается от типичных формулировок проблем в многочисленных учебниках и задачниках. Дело в том, что приведенная выше задача не сформулирована математически, то есть, строго говоря, решить ее нельзя. Однако жизнь устроена так, что, как правило, приходится решать не те задачи, которые можно решить, а те, которые нужно решить. Руководствуясь уже цитированным выше высказыванием П. Л. Чебышева о том, что правильно поставить задачу математически — это значит наполовину ее решить, мы сейчас попробуем использовать имеющийся у нас багаж сведений и попытаемся найти подходящую математическую формулировку приведенной выше задачи.

Пусть  $N$  — количество жителей в городе  $A$ . Пусть  $X_k$  — случайная величина, равная 1, если у  $k$ -го жителя города  $A$  обнаружена лейкемия, и 0, если у этого жителя нет данного заболевания. Поскольку лейкемия не является инфекционным заболеванием, мы можем считать, что случайные величины  $X_1, \dots, X_N$  независимы. Пусть  $N_A$  — случайная величина, равная числу жителей города  $A$ , больных лейкемией. Тогда

$$N_A = X_1 + \dots + X_N.$$

Допустим, что риск заболевания лейкемией одинаков для всех жителей города. Тогда мы приходим к заключению о том, что случайная величина  $N_A$  имеет биномиальное распределение. К счастью, лейкемия не является очень распространенным заболеванием, так что вероятность того, что случайно выбранный житель города  $A$  или города  $B$  страдает этой болезнью, мала. Далее, как сказано выше, число  $N$  жителей города  $A$  довольно велико. Все это означает, что мы находимся в условиях теоремы Пуассона, которая позволяет нам прийти к выводу о том, что вполне

разумно предположить, что случайная величина  $N_A$  имеет распределение Пуассона с некоторым параметром  $\lambda_A$ .

Точно такими же рассуждениями мы приходим к выводу о том, что и случайная величина  $N_B$  имеет распределение Пуассона с некоторым параметром  $\lambda_B$ . Более того, если города  $A$  и  $B$  находятся на существенном расстоянии друг от друга, то мы можем предположить, что пуассоновские случайные величины  $N_A$  и  $N_B$  независимы.

Как мы убедились выше, для пуассоновских случайных величин параметр является их математическим ожиданием, то есть “средним” значением:

$$EN_A = \lambda_A, \quad EN_B = \lambda_B.$$

Поэтому, если мы, зная, что в эксперименте случайные величины  $N_A$  и  $N_B$  соответственно приняли значения  $n_A$  и  $n_B$ , причем  $n_A > n_B$ , сумеем сделать вывод о том, что  $\lambda_A > \lambda_B$ , то это и будет означать, что мы установили наличие дополнительного систематического фактора риска в городе  $A$ . Покажем, как это можно сделать.

Хотя мы уже многое знаем о распределениях случайных величин  $N_A$  и  $N_B$ , мы пока, к сожалению, не можем выполнить указанное сравнение, так как сами параметры  $\lambda_A$  и  $\lambda_B$  нам неизвестны. Однако это обстоятельство не является непреодолимым препятствием.

Пусть  $n$  и  $k$  – целые числа,  $0 \leq k \leq n$ ,  $n \geq 1$ . Найдем условную вероятность того, что  $N_A = k$  при условии  $N_A + N_B = n$ . По определению условной вероятности мы имеем

$$P_{\{N_A+N_B=n\}}(N_A = k) = \frac{P(N_A = k, N_A + N_B = n)}{P(N_A + N_B = n)}. \quad (7.1.5)$$

Найдем по отдельности вероятности, стоящие в числителе и знаменателе дроби в правой части этой формулы. Так как случайные величины  $N_A$  и  $N_B$  независимы, то

$$\begin{aligned} P(N_A = k, N_A + N_B = n) &= P(N_A = k, N_B = n - k) \\ &= P(N_A = k)P(N_B = n - k) \\ &= \frac{e^{-(\lambda_A+\lambda_B)} \lambda_A^k \lambda_B^{n-k}}{k! \cdot (n - k)!}. \end{aligned}$$

Рассмотрим вероятность, стоящую в знаменателе дроби в правой части формулы (7.1.5). Событие  $\{N_A + N_B = n\}$  может быть представлено в виде

$$\{N_A + N_B = n\} = \bigcup_{j=0}^n \{N_A = j, N_B = n - j\}.$$

При этом, очевидно, что, если  $i \neq j$ , то

$$\{N_A = j, N_B = n - j\} \cap \{N_A = i, N_B = n - i\} = \emptyset.$$

Поэтому вследствие аддитивности вероятности и независимости случайных величин  $N_A$  и  $N_B$  мы имеем

$$\begin{aligned} P(N_A + N_B = n) &= \sum_{j=0}^n P(N_A = j, N_B = n - j) \\ &= \sum_{j=0}^n P(N_A = j)P(N_B = n - j) \\ &= \sum_{j=0}^n \frac{e^{-\lambda_A} \lambda_A^j}{j!} \cdot \frac{e^{-\lambda_B} \lambda_B^{n-j}}{(n-j)!} \\ &= \frac{e^{-(\lambda_A + \lambda_B)}}{n!} \sum_{j=0}^n \frac{n!}{j!(n-j)!} \lambda_A^j \lambda_B^{n-j} \\ &= \frac{e^{-(\lambda_A + \lambda_B)} (\lambda_A + \lambda_B)^n}{n!}. \end{aligned}$$

Обратим внимание, что в правой части последнего соотношения стоит пуассоновская вероятность. Таким образом, мы попутно установили, что распределение суммы независимых пуассоновских случайных величин снова является пуассоновским с параметром, равным сумме параметров слагаемых.

Подставляя найденные выражения в формулу (7.1.5), мы получаем

$$\begin{aligned}
 P_{\{N_A+N_B=n\}}(N_A = k) &= \frac{e^{-(\lambda_A+\lambda_B)} \lambda_A^k \lambda_B^{n-k}}{k! \cdot (n-k)!} \cdot \frac{n!}{e^{-(\lambda_A+\lambda_B)} (\lambda_A + \lambda_B)^n} \\
 &= \frac{n!}{k!(n-k)!} \left( \frac{\lambda_A}{\lambda_A + \lambda_B} \right)^k \left( \frac{\lambda_B}{\lambda_A + \lambda_B} \right)^{n-k} \\
 &= C_n^k \left( \frac{\lambda_A}{\lambda_A + \lambda_B} \right)^k \left( 1 - \frac{\lambda_A}{\lambda_A + \lambda_B} \right)^{n-k}. \quad (7.1.6)
 \end{aligned}$$

Но в правой части последнего соотношения стоит биномиальная вероятность. Другими словами, условное распределение случайной величины  $N_A$  при фиксированном значении суммы  $N_A + N_B = n$  является биномиальным с параметрами  $n$  и  $\lambda_A/(\lambda_A + \lambda_B)$ .

Следующим важным логическим шагом на пути решения исходной задачи является замечание о том, что в терминах параметров  $\lambda_A$  и  $\lambda_B$  наши возможные выводы имеют вид

- а)  $\lambda_A > \lambda_B$ , что означает наличие дополнительного фактора риска, и
- б)  $\lambda_A = \lambda_B$ , что означает отсутствие дополнительного фактора риска

(зная, что  $n_A > n_B$ , неразумно рассматривать возможность  $\lambda_A < \lambda_B$ ). Более того, эти два возможных вывода являются взаимоисключающими, то есть, не согласившись с одним из утверждений а) или б), мы автоматически соглашаемся с другим. Таким образом, грубо говоря, с точки зрения логики нам все равно, справедливость какого из этих двух утверждений нам надо установить. Однако с математической точки зрения нам намного удобнее проверять утверждение б).

Дело в том, что, вероятность, фигурирующая в формуле (7.1.6), вычисленная в предположении о том, что верно утверждение б), имеет вид

$$P_{\{N_A+N_B=n\}}(N_A = k) = \frac{C_n^k}{2^n},$$

то есть не зависит от  $\lambda_A$  и  $\lambda_B$ . Теперь в предположении о том, что верно утверждение б), мы можем вычислить вероятность того, что случайная

величина  $N_A$  примет значение, большее или равное  $n_A$ :

$$P_{\{N_A+N_B=n_A+n_B\}}(N_A \geq n_A) = \frac{(n_A + n_B)!}{2^{n_A+n_B}} \sum_{k=n_A}^{n_A+n_B} \frac{1}{k!(n_A + n_B - k)!}. \quad (7.1.7)$$

Другими словами, если бы никакого дополнительного фактора риска в городе  $A$  не было, то есть если бы условия жизни в обоих городах были одинаковы, то вероятность случайно наблюдать такое же или еще большее число больных лейкемией в городе  $A$  была бы равна правой части формулы (7.1.7). При этом, если вероятность (7.1.7), вычисленная при конкретных  $n_A$  и  $n_B$ , оказывается очень маленькой, то это означает, что мы наблюдаем то событие, которое в предположении о справедливости предположения  $b$ ) (то есть о том, что условия в обоих городах одинаковы) практически невозможно. Но это, в свою очередь, означает, что предположение  $b$ ) не соответствует действительности, поскольку мы наблюдаем именно значение  $n_A$ . Если вероятность (7.1.7), вычисленная при конкретных  $n_A$  и  $n_B$ , не оказывается пренебрежимо малой, то мы не можем отвергнуть предположение о том, что условия в городах  $A$  и  $B$  одинаковы, так как наблюдаемое превышение  $n_A$  над  $n_B$  может быть обусловлено чистой случайностью.

К сожалению, вычисление правой части (7.1.7) представляет собой отдельную довольно непростую задачу. О том, как упростить ее решение мы расскажем в следующем разделе.

## 1.7.2. Теорема Муавра–Лапласа.

### Нормальное распределение

Если нет оснований считать вероятность успеха в отдельном испытании Бернулли малой, то можно воспользоваться следующим утверждением.

Пусть  $X$  — число успехов в  $n$  испытаниях Бернулли. Если число испытаний неограниченно увеличивается, а вероятность успеха в одном испытании остается той же самой, то, какими бы ни были числа  $a$  и  $b$ ,  $a \leq b$ ,

$$P\left(a \leq \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Это утверждение называется *теоремой Муавра–Лапласа*.



Можно убедиться, что

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1,$$

то есть функция  $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  представляет собой плотность распределения. Распределение вероятностей, соответствующее плотности  $\varphi(x)$ , называется *стандартным нормальным*. Соответствующую функцию распределения обозначают  $\Phi(x)$  и называют *стандартной нормальной*. Значения функции  $\Phi(x)$  можно найти в специальных таблицах. Если случайная величина  $X$  имеет стандартное нормальное распределение, то  $EX = 0$ ,  $DX = 1$ . Можно показать, что если случайная величина  $X$  имеет стандартное нормальное распределение, то случайная величина  $Y = \sigma X + a$  имеет плотность

$$\varphi(x; a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

и функцию распределения  $\Phi\left(\frac{x-a}{\sigma}\right)$ . В таком случае  $EY = a$ ,  $DY = \sigma^2$  и говорят, что случайная величина  $Y$  имеет *нормальное распределение с параметрами  $a$  и  $\sigma^2$* .

Приведем примеры применения теоремы Муавра–Лапласа. Сначала вернемся к задаче 7.1.1.

**Задача 7.2.1 (ПРОДОЛЖЕНИЕ ЗАДАЧИ 7.1.1).** Напомним, что при решении задачи о том, подтверждают ли данные  $n_A$  и  $n_B$  о числе больных лейкемией в примерно однотипных городах  $A$  и  $B$  ( $n_A > n_B$ ) наличие в городе  $A$  дополнительного фактора риска, мы столкнулись с необходимостью вычислить вероятность  $P_{(N_A+N_B=n_A+n_B)}(N_A \geq n_A)$ . В силу того, что, как мы установили по ходу решения Задачи 7.1.1, условное распределение случайной величины  $N_A$  при фиксированном значении  $n = N_A + N_B$  является биномиальным с параметрами  $n$  и  $\lambda_A/(\lambda_A + \lambda_B)$ , теорема Муавра–Лапласа

позволяет нам записать

$$\begin{aligned} & P_{\{N_A+N_B=n_A+n_B\}}(N_A \geq n_A) \\ &= P_{\{N_A+N_B=n_A+n_B\}} \left( \frac{N_A - \frac{1}{2}(n_A + n_B)}{\sqrt{\frac{1}{4}(n_A + n_B)}} \geq \frac{n_A - \frac{1}{2}(n_A + n_B)}{\sqrt{\frac{1}{4}(n_A + n_B)}} \right) \\ &\approx 1 - \Phi \left( \frac{n_A - \frac{1}{2}(n_A + n_B)}{\sqrt{\frac{1}{4}(n_A + n_B)}} \right) = 1 - \Phi \left( \frac{n_A - n_B}{\sqrt{n_A + n_B}} \right). \end{aligned}$$

Например, если  $n_A = 125$ ;  $n_B = 100$ , то последняя вероятность равна  $1 - \Phi\left(\frac{25}{15}\right) < 1 - \Phi(1,67) = 0,0475$ .

Можно поставить обратную задачу: сколь большим должно быть превышение  $n_A$  над  $n_B$ , чтобы с уверенностью можно было говорить о наличии в городе  $A$  дополнительного фактора риска. Ясно, что ответ зависит от общего числа больных лейкемией. Упростим задачу и будем считать, что в городе  $B$  зарегистрировано 100 случаев лейкемии, то есть  $n_B = 100$ . Предположим, что мы хотим быть на 99% уверены в правильности вывода. Тогда критический порог  $n_A$  для количества случаев лейкемии в городе  $A$  мы найдем из условия

$$P_{\{N_A+N_B=n_A+100\}}(N_A \geq n_A) \approx 1 - \Phi \left( \frac{n_A - 100}{\sqrt{n_A + 100}} \right) \leq 0,01,$$

что эквивалентно условию

$$\Phi \left( \frac{n_A - 100}{\sqrt{n_A + 100}} \right) \geq 0,99. \quad (7.2.1)$$

Пусть  $\alpha$  — произвольное число из интервала  $[0, 1]$ . Символом  $u_\alpha$  мы обозначим  $\alpha$ -квантиль стандартного нормального распределения, то есть число, удовлетворяющее соотношению  $\Phi(u_\alpha) = \alpha$ . Так как любая функция распределения не убывает, то тогда неравенство (7.2.1) оказывается эквивалентным неравенству

$$\frac{n_A - 100}{\sqrt{n_A + 100}} \geq u_{0,99}. \quad (7.2.2)$$

По таблицам нормального распределения мы находим  $u_{0,99} = 2,33$ . Следовательно, неравенство (7.2.2) эквивалентно неравенству

$$n_A - 100 \geq 2,33\sqrt{n_A + 100}.$$

Наименьшим натуральным числом, удовлетворяющим последнему неравенству, является  $n_A = 136$ . Таким образом, если в городе А зарегистрировано более 135 случаев лейкемии, то с 99-процентной уверенностью можно говорить о том, что в городе А имеется дополнительный фактор риска.

В общем случае для определения порога  $n_A$ , превышение которого числом больных лейкемией в городе А позволяет нам с надежностью  $\gamma \cdot 100\%$  сделать вывод о наличии в городе А дополнительного фактора риска ( $0 < \gamma < 1$ ), вместо (7.2.1) надо воспользоваться соотношением

$$\frac{n_A - n_B}{\sqrt{n_A + n_B}} \geq u_\gamma,$$

откуда несложно получить неравенство

$$n_A - n_B \geq \frac{u_\gamma}{2} \left( u_\gamma + \sqrt{u_\gamma^2 + 8n_B} \right). \quad (7.2.3)$$

Вычисления по последней формуле приводят нас к выводу о том, что при фиксированном значении  $\gamma$  относительное критическое превышение  $n_A$  над  $n_B$  заметно убывает с ростом  $n_B$ . Это иллюстрирует следующая таблица, в которой символом  $n_A^*$  обозначено минимальное значение  $n_A$ , удовлетворяющее неравенству (7.2.3) при  $\gamma = 0.99$ .

$n_B$	$n_A^*$	$\frac{n_A^* - n_B}{n_B} \cdot 100\%$
100	135	35%
200	250	25%
300	360	20%
400	469	17.25%
500	577	15.4%
1000	1108	10.8%
2000	2151	7.55%
5000	5236	4.72%
10000	10333	3.33%

При осмыслении этой таблицы необходимо учитывать, что она имеет смысл лишь тогда, когда пуассоновская аппроксимация для распределения количества больных лейкемией имеет приемлемую точность, то есть когда число жителей в каждом городе не менее чем в 100 раз превышает число случаев лейкемии.

**ЗАДАЧА 7.2.2.** Общеизвестны книги американского врача Бенджамена Спока о воспитании детей. Доктор Спок известен не только своими книгами, но и своими антивоенными убеждениями. Он неоднократно участвовал в антивоенных акциях, в том числе во время войны, которую США вели во Вьетнаме, и был судим по обвинению в антиконституционной деятельности. Однажды он был арестован в Бостоне за участие в очередной антивоенной демонстрации. Его дело должен был рассматривать суд присяжных. Присяжные отбираются из дееспособного населения Бостона с помощью многоступенчатой процедуры, на очередном этапе которой было отобрано 300 человек (из числа которых потом и должны были быть выбраны 12 присяжных). Однако среди этих 300 человек оказалось лишь 90 женщин. Это обстоятельство послужило причиной того, что адвокаты доктора Спока подали протест (воспитанием детей в США, как и в России, в основном занимаются женщины и поэтому ясно, что в рассматриваемой ситуации при прочих равных женщины с большим сочувствием отнесутся к подсудимому). Насколько обоснован протест?

Доля женщин в дееспособном населении Бостона составляет примерно 50%. Поэтому если вероятность случайного выбора женщины в число 300 кандидатов обозначить  $p$ , то при абсолютно беспристрастной процедуре выбора должно быть  $p = \frac{1}{2}$ . Проверим, так ли это на самом деле. Для этого, предполагая, что  $p = \frac{1}{2}$ , найдем вероятность того, что будет отобрано 90 или еще меньше женщин. Другими словами, найдем вероятность того, что наблюдаемое отклонение числа отобранных женщин от ожидаемого их числа 150 обусловлено чистой случайностью. Если эта вероятность окажется исчезающе малой, то мы придем к выводу о том, что осуществилось событие, которое в сделанном предположении произойти практически не может. Это будет означать, что сделанное предположение не верно. (Такая логика рассуждений характерна для проверки *статистических гипотез*, о которой фактически речь шла в Задаче 7.1.1 и о которой мы будем говорить в разделах 2.3.2, 2.5.1 и 2.5.2.)

Итак, пусть  $X$  – количество отобранных женщин. Мы можем считать  $X$  числом успехов в 300 испытаниях Бернулли с вероятностью успеха в одном испытании, равной  $p$ . В предположении полной случайности выбора ( $p = \frac{1}{2}$ ) искомая вероятность имеет вид

$$P(X \leq 90) = \sum_{k=0}^{90} C_{300}^k p^k (1-p)^{n-k} = \frac{1}{2^{300}} \sum_{k=0}^{90} C_{300}^k.$$

Вычисление этой вероятности затруднено. Чтобы приблизительно оценить эту вероятность, применим теорему Муавра–Лапласа и получим

$$\begin{aligned} P(X \leq 90) &= P\left(\frac{X - 150}{\sqrt{75}} \leq \frac{90 - 150}{\sqrt{75}}\right) \approx \Phi\left(\frac{90 - 150}{\sqrt{75}}\right) \\ &= \Phi(-6,9282) < 0,06e^{-24} < 2,3 \cdot 10^{-12}. \end{aligned} \quad (7.2.4)$$

Эта вероятность может быть расценена как ничтожно малая. Так что, скорее всего, исходное предположение о том, что отбор присяжных беспристрастен, то есть что  $p = \frac{1}{2}$  не является справедливым. Более того, мы можем заключить, что и вероятность ошибочно отвергнуть предположение о беспристрастности отбора по имеющимся данным не превосходит  $2,3 \cdot 10^{-12}$ . К сожалению, книга Г. Кимбла “Как правильно пользоваться статистикой”, из которой взят только что рассмотренный пример, умалчивает о том, чем закончилось рассмотрение протеста.

К сожалению, ответ на важнейший в рассматриваемом случае вопрос о том, можно ли считать эту вероятность пренебрежимо малой или нет, не является прерогативой теории вероятностей.

Следует иметь в виду, что пользоваться теоремой Муавра–Лапласа для вычисления малых вероятностей, подобных той, которая фигурирует в цепочке (7.2.4), следует весьма осторожно, так как погрешность соответствующего приближенного равенства может во много раз превышать оцениваемую вероятность. К вопросу о точности нормальной аппроксимации мы вернемся в разделах 1.9 и 2.2.4.

### 1.7.3. Теорема Реньи. Показательное распределение

Пусть  $X$  – случайная величина с геометрическим распределением

$$P(X = n) = p^{n-1}(1-p), \quad k = 1, 2, \dots,$$

$0 < p < 1$ . В этом случае, как мы уже видели в Разделе 1.6.2,  $EX = 1/(1 - p)$ . Поэтому, если  $p \rightarrow 1$ , то  $EX \rightarrow \infty$ . Однако если одновременно со стремлением  $p$  к нулю производить нормировку рассматриваемой случайной величины, деля ее на ее математическое ожидание, то можно убедиться, что, каково бы ни было  $x > 0$ ,

$$P\left(\frac{X}{EX} < x\right) = P((1 - p)X < x) \rightarrow 1 - e^{-x}.$$

Это утверждение является частным случаем так называемой *теоремы Реньи*. Легко видеть, что предельная функция является функцией распределения. Такое распределение называется *стандартным показательным*. В общем случае, вводя параметр масштаба  $\lambda > 0$ , можно определить *показательное распределение с параметром  $\lambda$*  как задаваемое соответствующей функцией распределения  $F(x)$ , равной 0 при  $x < 0$  и равной  $1 - e^{-\lambda x}$  при  $x \geq 0$ .

Если  $Y$  – случайная величина с показательным распределением с параметром  $\lambda$ , то

$$EY = \frac{1}{\lambda}, \quad DY = \frac{1}{\lambda^2}.$$

Показательное распределение часто используется в качестве модели распределения продолжительности жизни или длительности безотказной работы. Однако такие модели весьма грубы и, вообще говоря, являются идеальными, поскольку не учитывают эффект старения. Действительно, в предположении о том, что время  $X$  безотказной работы некоторого устройства имеет показательное распределение с некоторым параметром  $\lambda$ , рассмотрим условную вероятность того, что после того как устройство проработало в течение времени  $t$ , устройство проработает еще как минимум время  $\tau$ . Эта вероятность имеет вид

$$\begin{aligned} P(X > t + \tau | X > t) &= \frac{P(X > t + \tau; X > t)}{P(X > t)} = \frac{P(X > t + \tau)}{P(X > t)} \\ &= \frac{e^{-(t+\tau)}}{e^{-t}} = e^{-\tau} = P(X > \tau), \end{aligned}$$

то есть эта условная вероятность не зависит от  $t$  и совпадает с безусловной вероятностью того, что устройство проработает как минимум время  $\tau$ .

## 1.8. Закон больших чисел

Пусть  $X$  — число успехов в  $n$  испытаниях Бернулли с вероятностью успеха в отдельном испытании, равной  $p$ . Основываясь на результатах п. 1.6.1, мы можем сказать, что случайная величина  $X$  имеет биномиальное распределение с параметрами  $n$  и  $p$ . Тогда под *частотой успеха* следует понимать дробь  $X/n$ . Еще очень давно было замечено, что при увеличении  $n$  частота успеха сближается с вероятностью  $p$ . Формально это сближение можно установить, например, с помощью неравенства Чебышева.

Зафиксируем какое-либо произвольно малое положительное число  $\varepsilon$  и рассмотрим вероятность  $P(|X/n - p| > \varepsilon)$  того, что отклонение частоты успеха от вероятности успеха превзойдет  $\varepsilon$  по абсолютной величине. Так как  $EX = np$  и  $DX = np(1-p)$ , то  $E(X/n) = p$  и  $D(X/n) = n^{-2}DX = p(1-p)/n$ . Тогда по неравенству Чебышева

$$P\left(\left|\frac{X}{n} - p\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^2} D\left(\frac{X}{n}\right) = \frac{p(1-p)}{\varepsilon^2 n} \rightarrow 0$$

при  $n \rightarrow \infty$ . Это соотношение является частным (и, по-видимому, самым простым) случаем *закона больших чисел*, который устанавливает сближение среднего арифметического случайных величин со средним арифметическим их математических ожиданий.

Известны многие варианты закона больших чисел. Мы приведем наиболее наглядную формулировку.

Пусть  $X_1, X_2, \dots$  — независимые одинаково распределенные случайные величины и  $EX_1 = a$ . Тогда с вероятностью единица выполняется соотношение

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = a.$$

Закон больших чисел является основанием для использования среднего арифметического наблюдаемых значений одинаково распределенных случайных величин в качестве приближения для их (общего) математического ожидания. Более подробно об этом и, в частности, о точности подобной аппроксимации мы поговорим во второй главе.

## 1.9. Центральная предельная теорема

Теорема Муавра–Лапласа является частным случаем так называемой *центральной предельной теоремы*, устанавливающей асимптотическую нормальность распределения суммы независимых случайных величин. Известны многие варианты центральной предельной теоремы. Мы приведем наиболее наглядную формулировку.

Напомним, что в разделе 1.6.2 была введена стандартная нормальная функция распределения

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

Пусть  $X_1, X_2, \dots$  — независимые одинаково распределенные случайные величины,  $EX_1 = a$  и  $DX_1 = \sigma^2$ , причем  $0 < \sigma^2 < \infty$ . Тогда при  $n \rightarrow \infty$  наибольшее по  $x$  значение величины

$$\left| P\left(\frac{\sum_{k=1}^n X_k - na}{\sigma\sqrt{n}} < x\right) - \Phi(x) \right|$$

стремится к нулю.

Из центральной предельной теоремы вытекает, что при больших  $n$  и любых  $x$  можно пользоваться приближением

$$P\left(\sum_{k=1}^n X_k < x\right) \approx \Phi\left(\frac{x - na}{\sigma\sqrt{n}}\right). \quad (9.1.1)$$

Другими словами, при больших  $n$  распределение суммы  $n$  независимых случайных величин можно аппроксимировать нормальным распределением с параметрами  $na$  и  $\sigma\sqrt{n}$ .

Центральную предельную теорему часто используют для обоснования того, что распределение погрешностей измерений предполагается нормальным. Действительно, погрешность измерения является результатом суммарного воздействия большого числа случайных факторов. Поэтому вследствие центральной предельной теоремы погрешность можно считать случайной величиной, распределенной по нормальному закону.

С помощью центральной предельной теоремы можно получить представление о том, какова скорость сходимости в законе больших чисел, то



есть о том, как быстро средние арифметические независимых одинаково распределенных случайных величин сближаются с математическим ожиданием. Действительно, пусть, как и ранее,  $X_1, X_2, \dots$  — независимые одинаково распределенные случайные величины с  $EX_1 = a$  и  $DX_1 = \sigma^2$ , причем  $0 < \sigma^2 < \infty$ . В силу закона больших чисел при этом последовательность неотрицательных случайных величин

$$Z_n = \left| \frac{1}{n} \sum_{j=1}^n X_j - a \right| \quad (9.1.2)$$

с вероятностью единица стремится к нулю при  $n \rightarrow \infty$ . Под *скоростью* сходимости последовательности случайных величин  $Z_n$  при  $n \rightarrow \infty$  условимся понимать такую неслучайную функцию  $r(n)$ , которая ведет себя при  $n \rightarrow \infty$  в определенном смысле так же, как *случайная* последовательность  $\{Z_n\}$ . Построим функцию  $r(n)$  для последовательности  $\{Z_n\}$ , определяемой соотношением (9.1.2). Пусть  $r(n)$  — *некоторая* последовательность положительных чисел. Рассмотрим поведение случайных величин

$$Y_n = \frac{Z_n}{r(n)}$$

при  $n \rightarrow \infty$ . Интуитивно ясно, что функция  $r(n)$  характеризует скорость сходимости последовательности  $\{Z_n\}$  к нулю, если последовательность  $\{Y_n\}$  имеет отличный от нуля и от бесконечности предел. Но мы рассматриваем последовательность *случайных* величин, а случайные величины являются не числами, но *функциями* элементарных исходов, и потому понятие предела такой последовательности отличается от понятия предела числовой последовательности. Предел последовательности случайных величин можно трактовать по-разному. Не ставя цели перечислить все подходы, мы будем следовать такому, при котором рассматриваются не сами случайные величины, а их распределения. Соотношение (9.1.2) означает, что

$$\lim_{n \rightarrow \infty} P(Z_n < x) = F(x) = \begin{cases} 0, & x \leq 0, \\ 1, & x > 0. \end{cases} \quad (9.1.3)$$

Предельная функция распределения  $F(x)$  в (9.1.3) имеет единственный единичный скачок в нуле. Функции распределения, имеющие единствен-

ный единичный скачок, принято называть *вырожденными*. Если вырожденная функция распределения имеет скачок в некоторой точке  $a$ , то говорят, что соответствующее распределение вероятностей вырождено в точке  $a$ . Несложно видеть, что распределение, вырожденное в точке  $a$ , соответствует “неслучайной” случайной величине  $Y_a$  такой, что  $P(Y_a = a) = 1$ . В частности, функция распределения, стоящая в правой части (9.1.3), вырождена в нуле и соответствует случайной величине  $Y_0$  такой, что  $P(Y_0 = 0) = 1$ .

Под упомянутым выше желательным отличием случайной величины, предельной для последовательности  $\{Y_n\}$ , от нуля или бесконечности мы будем понимать то, что ее функция распределения отлична от вырожденной в нуле или бесконечности.

Положим

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

Пусть  $x > 0$  — произвольное число. Тогда согласно центральной предельной теореме

$$\begin{aligned} P(Y_n < x) &= P(\sqrt{n}Z_n < x) = P\left(\sqrt{n}\left|\frac{1}{n}\sum_{j=1}^{n^*} X_j - a\right| < x\right) \\ &= P\left(\left|\frac{\sum_{j=1}^n X_j - na}{\sqrt{n}}\right| < x\right) = P\left(-x < \frac{\sum_{j=1}^n X_j - na}{\sigma\sqrt{n}} < \frac{x}{\sigma}\right) \\ &= P\left(\frac{\sum_{j=1}^n X_j - na}{\sigma\sqrt{n}} < \frac{x}{\sigma}\right) - P\left(\frac{\sum_{j=1}^n X_j - na}{\sigma\sqrt{n}} \leq -\frac{x}{\sigma}\right) \\ &\rightarrow \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(-\frac{x}{\sigma}\right). \end{aligned} \tag{9.1.4}$$

Так как функция  $e^{-x^2/2}$ , стоящая под интегралом в определении функции распределения  $\Phi(x)$ , четна (ее график симметричен относительно вертикальной оси координат), то  $\Phi(-x) = 1 - \Phi(x)$  для любого  $x > 0$ . Поэтому выражение в правой части (9.1.4) оказывается равным

$$G_\sigma(x) = \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(-\frac{x}{\sigma}\right) = \Phi\left(\frac{x}{\sigma}\right) - \left[1 - \Phi\left(\frac{x}{\sigma}\right)\right] = 2\Phi\left(\frac{x}{\sigma}\right) - 1.$$

Эта функция распределения, очевидно, не является вырожденной (она соответствует случайной величине  $Z = |Y|$ , где  $Y$  — случайная величина, имеющая нормальное распределение с нулевым ожиданием и дисперсией  $\sigma^2$ ).

Таким образом, мы можем заключить, что разность средних арифметических независимых одинаково распределенных случайных величин и математического ожидания стремится к нулю примерно как  $1/\sqrt{n}$ . Другими словами, скорость сходимости в законе больших чисел равна  $1/\sqrt{n}$ .

Обратим внимание, что такое заключение мы можем сделать только если в дополнение к условию справедливости закона больших чисел (существованию математического ожидания) выполнено условие справедливости центральной предельной теоремы (существование дисперсии случайной величины  $X_1$ ). Можно показать, что из существования дисперсии вытекает существование математического ожидания, но не наоборот. Ответ на вопрос о скорости сходимости в законе больших чисел, когда математическое ожидание случайной величины  $X_1$  существует, а дисперсия — нет, требует привлечения более тонкой техники.

Точность приближенного равенства (9.1.1) можно оценить. А именно, если известны значения величин  $a = EX_1$ ,  $\sigma = \sqrt{DX_1}$  и  $\mu = E|X_1 - a|^3$ , то наибольшее по  $x$  значение величины

$$\left| P\left(\sum_{k=1}^n X_k < x\right) - \Phi\left(\frac{x - na}{\sigma\sqrt{n}}\right) \right|$$

не превосходит так называемой *дроби Ляпунова*

$$L_3 = \frac{C_0\mu}{\sigma^3\sqrt{n}},$$

где  $C_0$  — абсолютная положительная постоянная,  $0,4 < C_0 \leq 0,7655$ . Другими словами, точность приближения (9.1.1) обратно пропорциональна квадратному корню из числа слагаемых в сумме. Кстати, такая закономерность типична для большинства статистических задач. К обсуждению точности нормальной аппроксимации мы вернемся в разделе 2.2.4.

Нормальное (гауссово) распределение вероятностей хорошо изучено. Термин “нормальное распределение вероятностей” впервые возник в конце XIX в. Его предложил Чарлз Пирсон (1857–1936) для описания распределения вероятностей, впервые встретившегося в работах Абрахама де Муавра (1667–1754). Это распределение связывают также с именами

Пьера Симона Лапласа (1749–1827), Карла Фридриха Гаусса (1777–1855) и Роберта Эдрейна (1755–1843), чьи исследования дали первые обоснования важности применения этого распределения в прикладных исследованиях.

В некоторых нематематических областях знания, например, в физике, экономике, биологии или социологии, нормальное распределение часто называют *нормальным законом*. С таким термином, к сожалению, связано одно довольно серьезное заблуждение. К сожалению, нормальному закону часто ошибочно придают некий универсальный характер, что проявляется в том, что если результаты наблюдений (статистические данные) согласуются с нормальным распределением, то считается, что все хорошо, все правильно, а если же распределение данных не соответствует нормальному, то делается вывод о том, что эксперимент поставлен неудачно, данные имеют неправильный характер. Возможно, это заблуждение имеет “лингвистические” причины, которые можно описать следующим образом. Нормальный закон — это закон природы, вроде закона всемирного тяготения. А закон природы *должен* выполняться. К тому же это не просто закон, а *нормальный* закон. Более того, плохое хорошим словом не назовут, а “нормальное” — конечно же, “хорошее” слово, ибо “не нормальное” — значит “ненормальное”, а со словом “ненормальное” у большинства нормальных людей связаны нехорошие ассоциации. Ясно, что такие “лингвистические” рассуждения нельзя считать серьезными. В математике известны многие понятия, связанные с термином “нормальное”. Например, нормальным называется направление, перпендикулярное (ортогональное) к поверхности в данной точке. При этом ясно, что категории “хорошее” и/или “плохое” являются относительными и, вообще говоря, без дополнительных пояснений к направлению, равно как и к распределению вероятностей, вообще говоря, применены быть не могут. Точно так же можно заключить, что нормальный закон является правильным в не большей степени, нежели другие корректно доказанные математические утверждения. Точно так же, другие, не нормальные, распределения вероятностей, возникающие в соответствующих математических моделях стохастических ситуаций, являются не менее правильными, чем нормальное.

Доверие большинства серьезных исследователей к нормальному закону имеет, естественно, не лингвистические, а математические корни,

базирующиеся на некоторых известных свойствах нормального распределения и соответствующих математических утверждениях, главным из которых, конечно же, является центральная предельная теорема теории вероятностей (опять-таки обратим внимание на лингвистическую особенность: именно *центральная*, то есть *основная* предельная теорема). Это утверждение описывает асимптотическую нормальность распределений сумм случайных величин при неограниченно увеличивающемся числе слагаемых. Часто для обоснования использования нормального распределения на центральную предельную теорему ссылаются чисто механически, не задумываясь о возможных обстоятельствах, препятствующих адекватности нормального распределения в тех или иных конкретных условиях. Здесь необходимо быть чрезвычайно осторожным. Дело в том, что иногда (например, когда используется теорема Муавра–Лапласа, являющаяся частным случаем центральной предельной теоремы) при использовании нормального распределения для аппроксимации дискретного распределения происходит подмена распределения *дискретной* характеристики (например, биномиально распределенной случайной величины) *непрерывным* распределением, то есть истинное распределение подменяется таким, каким оно *не может быть никогда по самой своей сути*, то есть заведомо неадекватной моделью. При этом одновременно возникает “проблема хвостов”. Эта проблема наглядно проявляется, например, в теореме Муавра–Лапласа: вероятность того, что аппроксимируемая *ограниченная* биномиальная случайная величина примет отрицательное значение, равно как и значение, превосходящее число испытаний Бернулли, с которыми связана эта биномиальная случайная величина, очевидно, *равна нулю*. В то же время, согласно нормальному распределению, вероятности любых как угодно больших отклонений аппроксимирующей *неограниченной* нормальной случайной величины от среднего значения *положительны*. А это обстоятельство очень серьезно, например, в юриспруденции, где действует правило, гласящее, что если вероятность события строго положительна, то это событие *возможно*, как бы мала ни была его вероятность. Поэтому иногда возникают курьезные ситуации, “казусы”, подобные следующему.

С помощью “высоконаучных” рассуждений, основанных на центральной предельной теореме, можно сделать вывод о том, что продолжи-

тельность беременности у человека имеет нормальное распределение с математическим ожиданием 268 дней (действительно, на продолжительность беременности влияют многие факторы, и отклонение действительной продолжительности беременности от ожидаемого срока в 268 дней обусловлено суммарным их воздействием). Основываясь на подобных рассуждениях адвокатов и медицинском заключении, один из судов Нью-Йорка в целях установления законности рождения ребенка у некой истицы признал, что беременность у нее длилась 365 дней. И действительно, ведь при нормальном распределении длительности беременности вероятность такого отклонения случайной величины от ожидаемого значения *положительна* (см. (Кимбл, 1982)). При этом, естественно, адвокаты истицы скорее всего умолчали о том, что в таком случае положительными следует считать и вероятности *любых* сроков беременности, в том числе и отрицательных!

(Справедливости ради необходимо отметить, что, как бы абсурдно иногда ни проявлялось упомянутое правило, совсем отказаться от него также нельзя. Например, недавно один из английских судов приговорил к тюремному заключению по обвинению в убийстве троих своих детей некую женщину. Прямых улик по этому делу не было. Более того, дети погибали в разное время. Единственным основанием для обвинительного решения суда стало заключение некоего “специалиста” по статистике, который, “используя методы теории вероятностей и статистический анализ данных” о детской смертности подсчитал, что вероятность случайных смертей троих детей в одной семье чрезвычайно мала, и, стало быть, их смерть не случайна, так как никаких официально зарегистрированных жалоб на их здоровье не было. Естественно, что принятое решение по меньшей мере небесспорно, так как вполне возможно, что у всех смертей есть одна глубинная причина. Эта причина вполне может иметь генетический характер, и тогда, наоборот, упомянутые трагические события в одной семье вполне вероятны.)

Подобный прием, основанный на апелляциях к результатам теории вероятностей и/или математической статистики, далеко не нов. К сожалению, очень часто за упомянутые выше “высокonaучные фразы” авторы некоторых выводов или статей прячут либо недобросовестно подтасованные выводы, либо свою некомпетентность.

Однако имеются ситуации, когда о недобросовестности или некомпетентности говорить не приходится. В финансовой математике нормальное распределение является официально признанным фетишем. “Проблема хвостов” проявляется в финансовой математике как проблема распределений приращений биржевых цен. Ниже мы обсудим эту проблему, весьма сходную с проблемой приращений процессов, наблюдаемых в турбулентной плазме. Здесь же мы лишь упомянем, что более ста лет назад, основываясь на рассуждениях, связанных с центральной предельной теоремой, Л. Башелье построил основы теории вычисления так называемой рациональной или справедливой цены некоторых ценных бумаг, связанных с акциями. Ключевым принципом в этой теории является нормальность распределения приращений биржевых цен. Более того, за большой вклад в развитие теории Башелье в 1997 г. Нобелевская премия в области экономики была присуждена М. Шоулзу и Р. Мертону, чьи работы были основаны опять-таки на уже упоминавшемся принципе нормальности. Эта теория широко признана и активно используется, хотя уже с 1915 г. многие исследователи обратили внимание на то, что реальное распределение приращений биржевых цен довольно далеко от нормального, имея более острую вершину и более тяжелые хвосты. Ясно, что, зная хвосты этих распределений, можно точно оценить вероятности больших изменений биржевых цен (которые на самом деле больше тех, которые соответствуют нормальному закону), а значит, более аккуратно построить стратегию биржевых игр и, следовательно, больше выиграть. При этом разработки аналитических отделов финансовых компаний представляют собой секреты, хранимые в большей тайне от конкурентов, нежели знаменитые секреты атомной бомбы. Наблюдается интересная ситуация: “официально”, для того чтобы по возможности ввести конкурентов в заблуждение, признается принцип нормальности, в то время как на самом деле финансовые расчеты, связанные с биржевыми играми, ведутся на совсем других основаниях.

*Центральная предельная теорема* — это собирательный термин, применимый к любому конкретному утверждению об асимптотической нормальности сумм случайных величин. Каждое из таких утверждений применяется к конкретной структурной модели. Например, в качестве математических моделей некоторых объектов могут выступать суммы как неслучайного, так и случайного числа слагаемых. Предельное поведение

ние сумм случайного числа слагаемых намного многообразнее, чем то, которое описывается приведенной выше центральной предельной теоремой. В частности, вместо нормального, случайные суммы одинаково распределенных независимых слагаемых с конечными дисперсиями могут в качестве предельного иметь распределение Коши. Фактически суммы неслучайного и случайного числа слагаемых являются *разными* структурными моделями.

Более того, каждое из утверждений типа центральной предельной теоремы справедливо при определенных условиях. Например, по стандартным курсам теории вероятностей наиболее хорошо известен вариант центральной предельной теоремы, в котором рассматриваются *стохастически независимые одинаково распределенные слагаемые с конечным вторым моментом*. Именно этот вариант был сформулирован выше. Если основываться на таком утверждении, то по крайней мере надо допускать, что в реальной ситуации слагаемые действительно независимы, действительно одинаково распределены и действительно имеют конечную дисперсию. Если же достаточных оснований для таких допущений или возможности для проверки выполнения этих условий нет, то увы, ссылки на центральную предельную теорему могут быть несостоятельными.

С нормальным распределением связано так называемое *правило трех сигм*. Это “правило” заключается в том, что если случайная величина  $X$  имеет нормальное распределение с математическим ожиданием  $a$  и дисперсией  $\sigma^2$ , то

$$P(|X - a| \leq 3\sigma) \approx 0,997$$

(другими словами, нормально распределенная случайная величина может отклониться от своего среднего значения более чем на три среднеквадратических отклонения с вероятностью примерно равной 0,003). В справедливости этого утверждения легко убедиться, заметив, что

$$P(|X - a| \leq 3\sigma) = 2\Phi(3) - 1,$$

и отыскивая значение  $\Phi(3)$  в специальных таблицах. Реально “правило трех сигм” означает, что практически всегда  $a - 3\sigma \leq X \leq a + 3\sigma$ . К сожалению, “правилу трех сигм” довольно часто придают некий универсальный смысл, забывая о том, что оно справедливо *только для нормально распределенных* случайных величин. Если же нет оснований однозначно считать распределение случайной величины  $X$  нормальным, то все, что



можно сказать, это то, что

$$P(|X - a| \leq 3\sigma) \geq \frac{8}{9} \approx 0,889$$

(в этом легко убедиться с помощью неравенства Чебышева). Другими словами, интервал  $[a - 3\sigma, a + 3\sigma]$  может не накрыть значение случайной величины  $X$  в среднем менее чем примерно в одиннадцати случаях из ста.

## 1.10. Равномерное распределение. Закон Бенфорда

Равномерное распределение выступает в качестве математической модели максимально неопределенной стохастической ситуации.

Рассмотрим стохастическую ситуацию с  $n$  возможными элементарными исходами  $\omega_1, \dots, \omega_n$ . Предположим, что эти исходы равновероятны:

$$P(\{\omega_1\}) = \dots = P(\{\omega_n\}) = \frac{1}{n}.$$

В таком случае говорят, что задано *дискретное равномерное распределение вероятностей*.

*Непрерывное равномерное распределение* вероятностей на некотором отрезке  $[a, b]$  задается с помощью плотности, имеющей вид

$$f(x) = \frac{\mathbf{1}_{[a,b]}(x)}{b-a} = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a, b], \\ 0, & \text{если } x \notin [a, b]. \end{cases}$$

Из этого определения, в частности, видно, что равномерное распределение не может быть задано на бесконечном интервале.

Если случайная величина  $X$  имеет равномерное распределение на отрезке  $[a, b]$ , то ее функция распределения  $F_X(x)$  имеет вид

$$F_X(x) = \begin{cases} 0, & \text{если } x < a, \\ \frac{x-a}{b-a}, & \text{если } a \leq x \leq b, \\ 1, & \text{если } x > b. \end{cases}$$

При этом

$$EX = \int_a^b \frac{x dx}{b-a} = \frac{1}{b-a} \left( \frac{a+b}{2} \cdot (b-a) \right) = \frac{a+b}{2},$$

$$DX = \frac{(b-a)^2}{12}.$$

**ПРИМЕР 10.1.1.** Пусть  $X$  – случайная величина, функция распределения  $F_X(x)$  которой строго монотонна (напомним, что функция  $F(x)$  называется строго монотонной, если для любых  $x$  и  $y$  таких, что  $x < y$ , выполнено неравенство  $F(x) < F(y)$ ). Построим новую случайную величину  $Y$ , положив

$$Y = F_X(X).$$

Найдем функцию распределения случайной величины  $Y$ . Во-первых, заметим, что  $0 \leq Y \leq 1$ , поскольку любая функция распределения принимает значения между нулем и единицей. Во-вторых, заметим, что поскольку по условию функция  $F_X(x)$  строго монотонна, определена обратная функция  $F_X^{-1}(y)$ , ставящая каждому числу  $y \in (0, 1)$  в соответствие решение уравнения

$$F_X(x) = y.$$

Более того, опять-таки, в силу строгой монотонности функции  $F_X(x)$  события  $\{\omega: F_X(X(\omega)) < x\}$  и  $\{\omega: X(\omega) < F_X^{-1}(x)\}$  оказываются эквивалентными. Поэтому при  $x \in [0, 1]$

$$F_Y(x) = P(Y < x) = P(F_X(X) < x) = P(X < F_X^{-1}(x)) \\ = F_X(F_X^{-1}(x)) \equiv x.$$

Таким образом,

$$F_X(x) = \begin{cases} 0, & \text{если } x < 0, \\ x, & \text{если } 0 \leq x \leq 1, \\ 1, & \text{если } x > 1, \end{cases}$$

что соответствует равномерному распределению на отрезке  $[0, 1]$  с плотностью

$$f_X(x) = \mathbf{1}_{[0,1]}(x).$$

**ПРИМЕР 10.1.2.** Пусть случайная величина  $X$  имеет равномерное распределение на отрезке  $[0, 1]$ ,  $F(x)$  – произвольная строго монотонная функция распределения. Как мы видели в предыдущем примере, в таком случае определена обратная функция  $F^{-1}(x)$ . Построим новую случайную величину  $Y$ , положив

$$Y = F^{-1}(X).$$

Найдем функцию распределения  $F_Y(x)$  случайной величины  $Y$ . Имеем

$$\begin{aligned} F_Y(x) &= P(Y < x) = P(F^{-1}(X) < x) \\ &= P(X < F(x)) = F_X(F(x)) \equiv F(x). \end{aligned}$$

Как известно, в любой современной системе программирования имеются стандартные процедуры, генерирующие так называемые псевдослучайные числа. Такие псевдослучайные числа можно считать независимыми реализациями случайных величин, имеющих равномерное распределение на отрезке  $[0, 1]$ . В то же время может возникнуть потребность в имитировании не равномерно распределенных случайных величин, а случайных величин, имеющих заданную функцию распределения. Приведенные выше рассуждения показывают, как это можно сделать. В частности, случайная величина  $Y$ , построенная из равномерно на  $[0, 1]$  распределенной случайной величины  $X$  с помощью преобразования

$$Y = -\frac{1}{\lambda} \log X,$$

где  $\lambda$  – некоторое положительное число, будет иметь показательное распределение с плотностью  $f_Y(x) = \lambda e^{-\lambda x}$  и функцией распределения  $F_Y(x) = 1 - e^{-\lambda x}$  ( $x \geq 0$ ).

**ПРИМЕР 10.1.3. Равномерное распределение на окружности.** В примере 5.2.1 мы уже встречались с дискретным равномерным распределением, сосредоточенным в равноотстоящих точках окружности. Рассмотрим непрерывный аналог такого распределения. Свернем отрезок  $[0, 1]$  в окружность единичной длины. Будем говорить, что случайно выбираемая точка на окружности имеет равномерное распределение, если вероятность ее попадания в любую дугу окружности равна длине этой дуги. Такое распределение должно иметь место при игре в рулетку. Однако чтобы это

было так, нужно, чтобы было выполнено некоторое дополнительное условие. А именно, рассмотрим угол, на который повернулось колесо рулетки до полной остановки, как непрерывную случайную величину  $X$  с плотностью  $f_X(x)$  (мы допускаем углы поворота, превосходящие  $360^\circ$ ). Введем новую единицу измерения угла, которую назовем, скажем, *оборот*. Тогда случайная величина  $X$  может быть представлена в виде

$$X = [X] + \{X\},$$

где  $[X]$  – число полных оборотов колеса рулетки, повернувшегося на угол  $X$ , а  $\{X\} = X - [X]$ . Так как длина окружности считается равной единице, то, очевидно,  $0 \leq \{X\} < 1$ . Величина  $\{X\}$  называется *дробной частью* величины  $X$ , а  $[X]$  называется *целой частью*.

Можно показать (см., например (Феллер, 1984), т. 2, с. 79), что для  $x \in [0, 1]$  плотность  $\tilde{f}(x)$  случайной величины  $\{X\}$  имеет вид

$$\tilde{f}(x) = \sum_{n=-\infty}^{+\infty} f_X(x+n) \left( = \lim_{N \rightarrow \infty} \sum_{n=-N}^N f_X(x+n) \right) \quad (10.1.1)$$

и  $\tilde{f}(x) = 0$  для  $x \notin [0, 1]$ . Оказывается, что для того, чтобы плотность  $\tilde{f}(x)$  была “примерно равномерной”, нужно, чтобы плотность  $f_X(x)$  исходной случайной величины  $X$  имела малый максимум, то есть чтобы распределение случайной величины  $X$  было “сильно размазано” по положительной полуоси. А именно, пусть максимум плотности  $f_X(x)$  равен, скажем,  $m$  и достигается, скажем, в точке  $x = a$ , причем при  $x < a$  функция  $f_X(x)$  возрастает, а при  $x > a$  – убывает. Тогда можно показать, что

$$|\tilde{f}(x) - 1| \leq m. \quad (10.1.2)$$

Действительно, так как  $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ , то в силу (10.1.1) плотность  $\tilde{f}(x)$  может быть представлена в виде

$$\tilde{f}(x) = 1 + \sum_{n=-\infty}^{+\infty} f_X(x+n) - \int_{-\infty}^{+\infty} f_X(x) dx. \quad (10.1.3)$$

Зафиксируем  $x$ . Для каждого целого  $k$  существует единственное  $x_k$  такое, что, во-первых,  $x_k = x + n$  при некотором целом  $n$  и, во-вторых,  $a + k \leq$

$x_k < a + k + 1$ . Тогда соотношение (10.1.3) можно переписать в виде

$$\begin{aligned} \tilde{f}(x) - 1 &= \sum_{k=-\infty}^{\infty} \int_{x_k}^{x_{k+1}} (f_X(x_k) - f_X(y)) dy \\ &= \sum_{k=-\infty}^{-2} \int_{x_k}^{x_{k+1}} (f_X(x_k) - f(y)) dy + \int_{x_{-1}}^a (f_X(x_k) - f_X(y)) dy \\ &\quad + \int_a^{x_0} (f_X(x_k) - f_X(y)) dy + \sum_{k=0}^{\infty} \int_{x_k}^{x_{k+1}} (f_X(x_k) - f_X(y)) dy \\ &\equiv -p + q, \end{aligned}$$

где

$$0 \leq p = \sum_{k=-\infty}^{-2} \int_{x_k}^{x_{k+1}} (f_X(y) - f_X(x_k)) dy + \int_{x_{-1}}^a (f_X(y) - f_X(x_k)) dy,$$

$$0 \leq q = \int_a^{x_0} (f_X(x_k) - f_X(y)) dy + \sum_{k=0}^{\infty} \int_{x_k}^{x_{k+1}} (f_X(x_k) - f_X(y)) dy.$$

Так как  $f_X(y)$  возрастает при  $y \leq a$ , то

$$\begin{aligned} p &\leq \sum_{k=-\infty}^{-2} \int_{x_k}^{x_{k+1}} (f_X(x_{k+1}) - f_X(x_k)) dy + \int_{x_{-1}}^a (f_X(a) - f_X(x_k)) dy \\ &= \sum_{k=-\infty}^{-2} (f_X(x_{k+1}) - f_X(x_k)) \cdot 1 + (f_X(a) - f_X(x_{-1})) \\ &= f_X(x_{-1}) + (f_X(a) - f_X(x_{-1}))(a - x_{-1}) \\ &\leq f_X(x_{-1}) + f_X(a) - f_X(x_{-1}) = f_X(a), \end{aligned}$$

поскольку  $a - x_{-1} \leq 1$ . Аналогично можно показать, что  $q \leq f_X(a)$ . Таким образом, мы имеем  $\tilde{f}(x) - 1 = q - p$ , где  $0 \leq q \leq f_X(a)$  и  $0 \leq p \leq f_X(a)$ . Значит,  $|\tilde{f}(x) - 1| \leq f_X(a) = m$ , что и требовалось установить. (Заметим, что в (Феллер, 1984), т. 2, с. 80 доказано неравенство, аналогичное (10.1.2), но с вдвое большей правой частью.)

Неравенство (10.1.2) означает, что чем меньше  $t$  (то есть чем сильнее размазано распределение случайной величины  $X$ ), тем ближе плотность  $f(x)$  к равномерной, то есть тем “случайнее” результат вращения колеса рулетки.

**ПРИМЕР 10.1.4. Закон Бенфорда.** Пусть  $X$  – случайная величина, равная первой значащей (ненулевой) цифре наугад выбранного числа из произвольно взятых таблиц (логарифмов, степеней, нормального распределения, длин рек, площади озер, численности населения и т. п.). Казалось бы, в условиях полной случайности выбора таблиц и числа в таблицах случайная величина  $X$  должна иметь дискретное равномерное распределение на множестве чисел  $\{1, 2, \dots, 9\}$ :

$$P(X = k) = \frac{1}{9} \approx 0,111, \quad k = 1, 2, \dots, 9.$$

Однако на практике эта закономерность отнюдь не наблюдается. Напротив, почти всегда наблюдается другая закономерность:

$$P(X = k) \approx p_k = \log_{10}(k + 1) - \log_{10} k, \quad k = 1, 2, \dots, 9.$$

Для наглядности приведем конкретные значения чисел  $p_k$ :

$$\begin{array}{lll} p_1 = 0,3010 & p_2 = 0,1761 & p_3 = 0,1249 \\ p_4 = 0,0969 & p_5 = 0,0792 & p_6 = 0,0669 \\ p_7 = 0,0580 & p_8 = 0,0512 & p_9 = 0,0458 \end{array}$$

При этом вероятность выпадения одной из цифр 1, 2, 3 или 4 совсем не равна ожидаемому числу  $\frac{4}{9} \approx 0,4444$ , а оказывается почти равной 0,7 (0,6992 — для любителей точности). Это обстоятельство может быть использовано при заключении пари подобно тому, как это с успехом делал некий известный прикладной математик, о чем упоминается в (Феллер, 1984), т. 2, с. 80 со ссылкой на статью Ф. Бенфорда, опубликованную в 1938 г. С тех пор описанная закономерность носит название *закона Бенфорда*. Эта закономерность настолько устойчива, что известен случай, когда на основании результатов ревизии прокурор Нью-Йорка возбудил уголовное дело в отношении некоей фирмы, в финансовом отчете которой данные не удовлетворяли закону Бенфорда.

Известны многие попытки теоретического обоснования этой эмпирической закономерности. В частности, такое обоснование может быть дано с помощью рассуждений, приведенных в предыдущем примере. А именно, наудачу выбранное число из наугад взятых таблиц может рассматриваться как положительная случайная величина с некоторым (неизвестным) распределением. Первая значащая цифра  $X$  такой случайной величины  $Y$  равна  $k$  в том и только в том случае, когда при некотором целом  $n$  выполнено соотношение

$$k \cdot 10^n \leq Y < (k + 1) \cdot 10^n. \quad (10.1.4)$$

Введем случайную величину  $Z = \log_{10} Y$ . Тогда соотношение (10.1.4) эквивалентно неравенствам

$$n + \log_{10} k \leq Z < n + \log_{10}(k + 1). \quad (10.1.5)$$

Как показано в предыдущем примере, если распределение случайной величины  $Y$  очень сильно размазано по положительной полуоси, то распределение случайной величины  $\{Z\}$  (дробной части величины  $Z$ ) является приближенно равномерным. Но тогда для каждого  $k = 1, 2, \dots, 9$  вероятность неравенств (10.1.5) оказывается приближенно равной  $p_k = \log_{10}(k + 1) - \log_{10} k$ , как и должно быть в соответствии с законом Бенфорда.

## 1.11. Многомерные случайные величины (случайные векторы)

Пусть  $n \geq 1$  — некоторое целое число. В математике набор чисел  $(x_1, \dots, x_n)$  принято называть ( $n$ -мерным) вектором. По аналогии, в данном разделе мы определим понятие *случайного* вектора.

В разделе 1.5 мы уже приводили пример стохастической ситуации (опрос общественного мнения), в которой каждому элементарному исходу  $\omega$  соответствует не одно, а несколько чисел  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ , где  $n \geq 1$ . В качестве других примеров таких ситуаций можно привести следующие.

**ПРИМЕР 11.1.1.** При клиническом анализе крови у каждого пациента (отождествляемого с элементарным исходом  $\omega$ ) определяются несколько характеристик: количество эритроцитов, лейкоцитов, тромбоцитов, свертываемость и так далее.

**ПРИМЕР 11.1.2.** При социологическом обследовании в каждом населенном пункте (отождествляемом с элементарным исходом  $\omega$ ) определяются такие показатели, как численность населения, занятость, количество тех или иных предприятий, кинотеатров и так далее.

В этих примерах номер (индекс)  $i$  у числа  $X_i(\omega)$  определяет название одной и той же конкретной характеристики для разных элементарных исходов  $\omega$ . Вспоминая определение случайной величины как функции элементарного исхода (см. раздел 1.3), мы приходим к заключению, что при каждом фиксированном индексе  $i$  функция  $X_i(\omega)$  элементарного исхода  $\omega$  является случайной величиной. Условимся называть набор  $\mathbf{X} = \mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))$  случайных величин *многомерной* (точнее, *n-мерной*) *случайной величиной*, или *n-мерным случайным вектором*. По отношению к случайному вектору  $\mathbf{X}$  составляющие его случайные величины  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  называются *компонентами*. В дальнейшем для краткости мы будем опускать аргумент  $\omega$  у компонент случайного вектора.

В разделе 1.5 мы познакомились с такими случайными векторами, компоненты которых стохастически независимы и рассмотрели простые числовые характеристики — ковариацию и коэффициент корреляции, — описывающие тесноту попарной линейной зависимости компонент. Чтобы более полно охарактеризовать совместные статистические свойства компонент случайного вектора  $\mathbf{X}$ , каждому возможному набору  $n$  чисел  $\mathbf{x} = (x_1, \dots, x_n)$  поставим в соответствие число  $F_{X_1, \dots, X_n}(x_1, \dots, x_n)$  по правилу

$$\begin{aligned} F_{X_1, \dots, X_n}(x_1, \dots, x_n) &= P(\{\omega: X_1(\omega) < x_1\} \cap \dots \cap \{\omega: X_n(\omega) < x_n\}) \\ &\equiv P(X_1 < x_1, \dots, X_n < x_n). \end{aligned}$$

Указанное соответствие определяет числовую функцию

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1, \dots, X_n}(x_1, \dots, x_n)$$



на множестве всевозможных наборов  $\mathbf{x} = (x_1, \dots, x_n)$ . Эта функция называется *совместной функцией распределения* случайного вектора  $\mathbf{X}$ .

Совместная функция распределения обладает следующими свойствами.

1. Для любого  $i = 1, 2, \dots, n$ , если  $x'_i \leq x''_i$ , то

$$F_{X_1, \dots, X_i, \dots, X_n}(x_1, \dots, x'_i, \dots, x_n) \leq F_{X_1, \dots, X_i, \dots, X_n}(x_1, \dots, x''_i, \dots, x_n);$$

2. Для любого  $i = 1, 2, \dots, n$

$$\lim_{x_i \rightarrow -\infty} F_{X_1, \dots, X_i, \dots, X_n}(x_1, \dots, x_i, \dots, x_n) = 0;$$

3. Для любого  $i = 1, 2, \dots, n$

$$\begin{aligned} \lim_{x_i \rightarrow +\infty} F_{X_1, \dots, X_{i-1}, X_i, X_{i+1}, \dots, X_n}(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) \\ = F_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n}(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n). \end{aligned}$$

Если все аргументы совместной функции распределения одинаковы то есть  $x_1 = \dots = x_n \equiv x$ , то

$$\begin{aligned} F_{X_1, \dots, X_n}(x, \dots, x) &= P(X_1 < x, \dots, X_n < x) \\ &= P(\max\{X_1, \dots, X_n\} < x). \end{aligned}$$

Пусть  $\Delta$  – малое положительное число. Если на множестве всевозможных наборов  $\mathbf{x} = (x_1, \dots, x_n)$  определена неотрицательная функция  $f(\mathbf{x}) = f(x_1, \dots, x_n)$ , такая, что для любого набора  $\mathbf{x} = (x_1, \dots, x_n)$

$$\lim_{\Delta \rightarrow 0} \frac{P(x_1 \leq X_1 < x_1 + \Delta, \dots, x_n \leq X_n < x_n + \Delta)}{\Delta^n} = f(x_1, \dots, x_n), \quad (11.1.)$$

то случайный вектор  $\mathbf{X} = (X_1, \dots, X_n)$  называется *непрерывным*, а функция  $f(x_1, \dots, x_n)$  называется *совместной плотностью* (распределения вероятностей) компонент случайного вектора  $\mathbf{X}$ .

Чтобы проиллюстрировать смысл понятий совместной функции распределения и совместной плотности, рассмотрим ситуацию, когда случайный вектор  $\mathbf{X}$  состоит из двух компонент:  $\mathbf{X} = (X_1, X_2)$ , причем обе компоненты —  $X_1$  и  $X_2$  — неотрицательны. В таком случае областью возможных значений случайного вектора  $\mathbf{X} = (X_1, X_2)$  является первый квадрант координатной плоскости. Пусть  $x_1$  и  $x_2$  — два произвольных положительных числа. Тогда значение  $F_{\mathbf{X}}(x_1, x_2)$  совместной функции распределения в точке с координатами  $(x_1, x_2)$  равно вероятности того, что случайный вектор  $\mathbf{X} = (X_1, X_2)$  попал в прямоугольник, ограниченный отрезками, последовательно соединяющими точки  $(0, 0)$ ,  $(x_1, 0)$ ,  $(x_1, x_2)$  и  $(0, x_2)$ .

Для некоторого  $\Delta > 0$  найдем вероятность попадания определенного выше случайного вектора  $\mathbf{X} = (X_1, X_2)$  в квадрат с вершинами  $(x_1, x_2)$ ,  $(x_1 + \Delta, x_2)$ ,  $(x_1 + \Delta, x_2 + \Delta)$ ,  $(x_1, x_2 + \Delta)$ . Обозначим этот квадрат через  $S$ , прямоугольник с вершинами  $(0, 0)$ ,  $(x_1, 0)$ ,  $(x_1, x_2)$ ,  $(0, x_2)$  — через  $S_1$ , прямоугольник с вершинами  $(0, x_2)$ ,  $(x_1, x_2)$ ,  $(x_1, x_2 + \Delta)$ ,  $(0, x_2 + \Delta)$  — через  $S_2$ , прямоугольник с вершинами  $(x_1, 0)$ ,  $(x_1 + \Delta, 0)$ ,  $(x_1 + \Delta, x_2)$ ,  $(x_1, x_2)$  — через  $S_3$ . Тогда

$$S = (S \cup S_1 \cup S_2 \cup S_3) \setminus ((S_1 \cup S_2) \cup (S_1 \cup S_3)).$$

Но так как

$$F_{\mathbf{X}}(x_1 + \Delta, x_2 + \Delta) = P(\mathbf{X} \in S \cup S_1 \cup S_2 \cup S_3),$$

$$F_{\mathbf{X}}(x_1, x_2 + \Delta) = P(\mathbf{X} \in S_1 \cup S_2),$$

$$F_{\mathbf{X}}(x_1 + \Delta, x_2) = P(\mathbf{X} \in S_1 \cup S_3),$$

$$F_{\mathbf{X}}(x_1, x_2) = P(\mathbf{X} \in S_1) = P(\mathbf{X} \in ((S_1 \cup S_2) \cap (S_1 \cup S_3))),$$

то с учетом формулы  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  мы получаем, что искомая вероятность  $P(\mathbf{X} \in S)$  равна

$$\begin{aligned} P(\mathbf{X} \in S) &= P(x_1 \leq X_1 < x_1 + \Delta, x_2 \leq X_2 < x_2 + \Delta) \\ &= F_{\mathbf{X}}(x_1 + \Delta, x_2 + \Delta) - F_{\mathbf{X}}(x_1, x_2 + \Delta) \\ &\quad - F_{\mathbf{X}}(x_1 + \Delta, x_2) + F_{\mathbf{X}}(x_1, x_2). \end{aligned}$$

Из соотношения (11.1.1) вытекает, что если случайный вектор  $\mathbf{X} =$

$= (X_1, X_2)$  непрерывен, то

$$\begin{aligned} P(\mathbf{X} \in S) &= P(x_1 \leq X_1 < x_1 + \Delta, x_2 \leq X_2 < x_2 + \Delta) \\ &= \Delta^2 f(x_1, x_2) + \delta(\Delta^2), \end{aligned}$$

где  $\delta(x)$  – некоторая функция, обладающая свойством

$$\lim_{x \rightarrow 0} \frac{\delta(x)}{x} = 0.$$

Таким образом, при малых  $\Delta$  можно записать

$$P(\mathbf{X} \in S) \approx \Delta^2 f(x_1, x_2). \quad (11.1.2)$$

Но правая часть соотношения (11.1.2) равна объему параллелепипеда, основанием которого является квадрат  $S$ , а высота равна  $f(x_1, x_2)$ . В трехмерном пространстве рассмотрим множество всех точек  $(x_1, x_2, x_3)$ , удовлетворяющих соотношению

$$x_3 = f(x_1, x_2).$$

Такое множество задает поверхность, которая по аналогии с одномерным случаем называется *графиком* функции  $f(x_1, x_2)$ . Пусть теперь  $A$  – некоторое множество на плоскости  $x_1 O x_2$ . Предположим, что множество  $A$  можно представить в виде объединения одинаковых квадратов  $S^{(i)}$  стороны которых параллельны координатным осям и равны малому числу  $\Delta$ :

$$A = \bigcup_i S^{(i)},$$

причем площадь фигуры, получающейся при пересечении любых из этих квадратов, равна нулю. В таком случае множество  $A$  называется *квადрируемым*. Пусть  $(x_1^{(i)}, x_2^{(i)})$  – координаты одной из вершин квадрата  $S^{(i)}$ . Так как пересечение квадратов  $S^{(i)}$  имеет нулевую площадь, то в соответствии с формулой (11.1.2)

$$P(\mathbf{X} \in A) = \sum_i P(\mathbf{X} \in S^{(i)}) \approx \sum_i \Delta^2 f(x_1^{(i)}, x_2^{(i)}).$$

Но при стремлении  $\Delta$  к нулю последняя сумма стремится к объему тела, ограниченного “снизу” множеством  $A$  на плоскости  $x_1 O x_2$ , “сверху” — графиком функции  $f(x_1, x_2)$ , а “с боков” — поверхностью, содержащей все прямые, перпендикулярные плоскости  $x_1 O x_2$  и проведенные через все точки границы множества  $A$ . По аналогии с одномерной ситуацией объем такого тела называется *интегралом функции  $f(x_1, x_2)$  по множеству  $A$*  и обозначается символом

$$\iint_A f(x_1, x_2) dx_1 dx_2.$$

Таким образом, для непрерывного случайного вектора  $X = (X_1, X_2)$  мы можем записать

$$P(X \in A) = \iint_A f(x_1, x_2) dx_1 dx_2.$$

Пусть  $R$  — некоторое положительное число и

$$A = A(R) = \{(x_1, x_2) : |x_1| \leq R, |x_2| \leq R\}.$$

Под интегралом функции  $f(x_1, x_2)$  по всей плоскости  $\mathbb{R}^2 = x_1 O x_2$  мы будем понимать предел

$$\iint_{\mathbb{R}^2} f(x_1, x_2) dx_1 dx_2 = \lim_{R \rightarrow \infty} \iint_{A(R)} f(x_1, x_2) dx_1 dx_2.$$

При этом, очевидно,

$$\iint_{\mathbb{R}^2} f(x_1, x_2) dx_1 dx_2 = \lim_{R \rightarrow \infty} P(|X_1| \leq R, |X_2| \leq R) = P(X \in \mathbb{R}^2) = 1.$$

Если зафиксировать какое-либо значение  $x_1$ , а переменной  $x_2$  при этом разрешить пробегать все свои возможные значения, то из функции двух переменных  $f(x_1, x_2)$  мы получим функцию одной переменной, которую мы обозначим  $g_{x_1}(x_2)$ ,  $g_{x_1}(x_2) = f(x_1, x_2)$ . Аргументом этой функции

является  $x_2$ , а  $x_1$  при этом можно считать параметром. Функция  $g_{x_1}(x_2)$  неотрицательна, стало быть, мы можем рассмотреть ее (одномерный) интеграл (значение которого, очевидно, зависит от  $x_1$ ). Другими словами, интеграл функции  $g_{x_1}(x_2)$  по  $dx_2$  сам является функцией от  $x_1$ :

$$\int_{-\infty}^{\infty} g_{x_1}(x_2) dx_2 = \lim_{R \rightarrow \infty} \int_{-R}^R g_{x_1}(x_2) dx_2 \equiv h_1(x_1).$$

Функция  $h_1(x_1)$ , очевидно, неотрицательна, и мы, в свою очередь, можем рассмотреть ее интеграл. При этом, в силу того, что  $f(x_1, x_2)$  – плотность, принимая во внимание равенство  $g_{x_1}(x_2) = f(x_1, x_2)$ , для любого  $A > 0$  мы получим

$$\begin{aligned} \int_{-A}^A h_1(x_1) dx_1 &= \int_{-A}^A \left( \int_{-\infty}^{\infty} g_{x_1}(x_2) dx_2 \right) dx_1 = \int_{-A}^A \left( \int_{-R}^R g_{x_1}(x_2) dx_2 \right) dx_1 \\ &= \lim_{R \rightarrow \infty} \left[ \int_{-A}^A \left( \lim_{R \rightarrow \infty} \int_{-R}^R f(x_1, x_2) dx_2 \right) dx_1 \right]. \end{aligned}$$

Для произвольных  $A > 0$  и  $R > 0$  обозначим

$$S(A, R) = \{(x_1, x_2) : |x_1| \leq A, |x_2| \leq R\}.$$

Тогда, исходя из геометрического определения двумерного интеграла как объема некоего тела, продолжая начатую выше цепочку соотношений, мы получим

$$\begin{aligned} \int_{-A}^A h_1(x_1) dx_1 &= \lim_{R \rightarrow \infty} \left[ \int_{-A}^A \left( \int_{-R}^R f(x_1, x_2) dx_2 \right) dx_1 \right] \\ &= \lim_{R \rightarrow \infty} \iint_{S(A, R)} f(x_1, x_2) dx_1 dx_2 \\ &= \lim_{R \rightarrow \infty} P(X \in S(A, R)) \\ &= \lim_{R \rightarrow \infty} P(|X_1| \leq A, |X_2| \leq R) \\ &= P(|X_1| \leq A, -\infty < X_2 < \infty) \\ &= P(|X_1| \leq A). \end{aligned}$$

Таким образом, функция  $h_1(x_1)$  является плотностью вероятностей случайной величины  $X_1$ .

Аналогично можно убедиться, что функция

$$h_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$$

является плотностью вероятностей случайной величины  $X_2$ .

Таким образом, если известна совместная функция распределения  $F_{X_1, X_2}(x_1, x_2)$  случайного вектора  $\mathbf{X} = (X_1, X_2)$ , то для того, чтобы найти функцию распределения одной из компонент этого вектора, скажем,  $X_1$ , следует аргумент, соответствующий другой компоненте (в данном случае это  $x_2$ ), положить равным  $+\infty$ :

$$F_{X_1}(x_1) = F_{X_1, X_2}(x_1, +\infty).$$

Аналогично,

$$F_{X_2}(x_2) = F_{X_1, X_2}(+\infty, x_2).$$

Если же известна совместная плотность вероятностей  $f_{X_1, X_2}(x_1, x_2)$  случайного вектора  $\mathbf{X} = (X_1, X_2)$ , то для того, чтобы найти плотность распределения одной из компонент этого вектора, скажем,  $X_1$ , следует проинтегрировать совместную плотность по аргументу, соответствующему другой компоненте (в данном случае это  $x_2$ ), от  $-\infty$  до  $+\infty$ :

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_2.$$

Аналогично,

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_1.$$

По отношению к двумерному распределению непрерывного случайного вектора  $(X_1, X_2)$  обе плотности  $f_{X_1}(x)$  и  $f_{X_2}(x)$  называются *маргинальными плотностями*. Они определяют *маргинальные распределения* компонент этого вектора: плотность  $f_{X_1}(x)$  задает распределение компоненты  $X_1$  в то время как плотность  $f_{X_2}(x)$  задает распределение компоненты  $X_2$ .

Вектор, составленный из математических ожиданий случайных величин  $X_1, X_2, \dots, X_n$ , называется *математическим ожиданием случайного вектора*  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ :

$$E\mathbf{X} = (EX_1, EX_2, \dots, EX_n).$$

Для случайного вектора  $\mathbf{X} = (X_1, \dots, X_n)$  составим таблицу из  $n$  строк и  $n$  столбцов по следующему правилу. В клетку этой таблицы, стоящую на пересечении  $i$ -й строки и  $j$ -го столбца, поставим число  $c_{i,j}$ , равное ковариации случайных величин  $X_i$  и  $X_j$ :

$$c_{i,j} = \text{cov}(X_i, X_j) = E(X_i X_j) - EX_i EX_j.$$

Так как, очевидно,  $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$ , то  $c_{i,j} = c_{j,i}$ . При этом  $c_{i,i} = EX_i^2 - (EX_i)^2 = DX_i$ . Такая таблица называется *ковариационной матрицей* случайного вектора  $\mathbf{X} = (X_1, \dots, X_n)$ .

Рассмотрим поподробнее ситуацию, в которой компоненты случайного вектора  $\mathbf{X} = (X_1, \dots, X_n)$  независимы в совокупности. В таком случае все элементы ковариационной матрицы, стоящие вне диагонали указанной выше таблицы, проходящей из левого верхнего угла в правый нижний ее угол (эта диагональ называется *главной*), равны нулю. В таком случае ковариационная матрица называется *диагональной*. Эту же структуру ковариационная матрица, очевидно, сохраняет и в более общих ситуациях, когда компоненты случайного вектора  $\mathbf{X} = (X_1, \dots, X_n)$  попарно независимы или просто некоррелированы.

Если компоненты случайного вектора  $\mathbf{X} = (X_1, \dots, X_n)$  независимы в совокупности, то совместная функция распределения такого случайного вектора равна произведению одномерных функций распределения его компонент:

$$\begin{aligned} F_{X_1, \dots, X_n}(x_1, \dots, x_n) &= P(X_1 < x_1, \dots, X_n < x_n) \\ &= P(X_1 < x_1) \cdot \dots \cdot P(X_n < x_n) \\ &= F_{X_1}(x_1) \cdot \dots \cdot F_{X_n}(x_n). \end{aligned}$$

Аналогично, в такой ситуации совместная плотность случайного вектора  $\mathbf{X} = (X_1, \dots, X_n)$  равна произведению одномерных плотностей его компонент:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n).$$

## 1.12. Формулы свертки

Рассмотрим случайный вектор  $\mathbf{X} = (X, Y)$ . Нашей целью в данном разделе будет изучение распределения суммы  $X + Y$  его компонент.

Предположим вначале, что случайные величины  $X$  и  $Y$  дискретны и принимают соответственно значения  $x_0, x_1, x_2, \dots$  и  $y_0, y_1, y_2, \dots$ , причем

$$P(X = x_i, Y = y_j) = p_{i,j}, \quad i, j = 0, 1, 2, \dots$$

При этом так как события  $\{X = x_i, Y = y_j\}$ ,  $i, j = 0, 1, 2, \dots$  не пересекаются, то

$$\begin{aligned} P(X = x_i) &= P\left(\bigcup_j \{X = x_i, Y = y_j\}\right) \\ &= \sum_j P(X = x_i, Y = y_j) = \sum_j p_{i,j} \equiv p_{i..} \end{aligned}$$

Аналогично,

$$\begin{aligned} P(Y = y_j) &= P\left(\bigcup_i \{X = x_i, Y = y_j\}\right) \\ &= \sum_i P(X = x_i, Y = y_j) = \sum_i p_{i,j} \equiv p_{.j}. \end{aligned}$$

Так как

$$\sum_{i,j} p_{i,j} = 1,$$

то

$$\sum_i p_{i..} = \sum_j p_{.j} = 1,$$

то есть как набор чисел  $\{p_{i..}\}_{i \geq 1}$ , так и набор чисел  $\{p_{.j}\}_{j \geq 1}$  являются дискретными распределениями вероятностей. По отношению к двумерному распределению дискретного случайного вектора  $(X, Y)$  оба указанных выше набора чисел называются *маргинальными распределениями*. Они определяют распределения компонент этого вектора: набор  $\{p_{i..}\}_{i \geq 1}$



— это распределение компоненты  $X$  в то время как набор  $\{p_j\}_{j \geq 1}$  — это распределение компоненты  $Y$ .

Для произвольного числа  $x$  мы имеем

$$\begin{aligned} P(X + Y < x) &= P\left(\bigcup_{i,j: x_i+y_j < x} \{X = x_i, Y = y_j\}\right) \\ &= \sum_{i,j: x_i+y_j < x} P(X = x_i, Y = y_j). \end{aligned}$$

Пусть далее,  $x_i = i$ ,  $y_j = j$  при всех  $i, j = 0, 1, 2, \dots$ . Тогда возможные значения  $x_i + y_j$  суммы  $X + Y$  составляют множество целых неотрицательных чисел, причем для любого целого  $n \geq 0$

$$\begin{aligned} P(X + Y = n) &= \sum_{i,j: i+j=n} P(X = i, Y = j) \\ &= \sum_{i=0}^{\infty} P(X = i, Y = n - i) = \sum_{i=0}^{\infty} p_{i,n-i}. \end{aligned} \quad (12.1.1)$$

Если при этом случайные величины  $X$  и  $Y$  независимы, то для любых  $i$  и  $j$

$$p_{i,j} = p_i \cdot p_j.$$

Поэтому в ситуации, когда случайные величины  $X$  и  $Y$  независимы, формула (12.1.1) принимает вид

$$P(X + Y = n) = \sum_{i=0}^{\infty} p_i \cdot p_{(n-i)}. \quad (12.1.2)$$

Соотношение (12.1.2) называется *формулой свертки* последовательностей  $\{p_i\}_{i \geq 0}$  и  $\{p_j\}_{j \geq 0}$ .

Теперь предположим, что случайный вектор  $\mathbf{X} = (X, Y)$  непрерывен и имеет совместную плотность  $f_{X,Y}(x, y)$ . В таком случае функция распределения  $F_{X+Y}(z) = P(X + Y < z)$  суммы  $X + Y$  его компонент имеет вид

$$F_{X+Y}(z) = \iint_{\{(x,y): x+y < z\}} f_{X,Y}(x, y) dx dy.$$

Более того, поскольку

$$\{(x, y): x + y < z\} = \{(x, y): -\infty < x < \infty, y < z - x\},$$

для последнего интеграла справедливо равенство

$$\iint_{\{(x, y): x+y < z\}} f_{X,Y}(x, y) dx dy = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{z-x} f_{X,Y}(x, y) dy \right) dx.$$

Предположим теперь дополнительно, что случайные величины  $X$  и  $Y$  независимы. В таком случае  $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ . Поэтому приведенная выше формула примет вид

$$F_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) \left( \int_{-\infty}^{z-x} f_Y(y) dy \right) dx = \int_{-\infty}^{\infty} F_Y(z-x) f_X(x) dx. \quad (12.1.3)$$

Поскольку в приведенных выше рассуждениях случайные величины  $X$  и  $Y$  абсолютно равноправны, мы точно так же получаем и формулу

$$F_{X+Y}(z) = \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy. \quad (12.1.4)$$

Соотношения (12.1.3) и (12.1.4) называют *формулами свертки для функций распределения*.

Исходя из геометрических построений, легко убедиться, что для любого числа  $\Delta > 0$

$$\begin{aligned} & \left\{ x \leq X < x + \frac{\Delta}{2}, y - x \leq Y < y - x + \frac{\Delta}{2} \right\} \\ & \subseteq \{x \leq X < x + \Delta, y \leq X + Y < y + \Delta\} \\ & \subseteq \{x \leq X < x + 2\Delta, y - x - \Delta \leq Y < y - x + \Delta\}. \end{aligned}$$

Поэтому, с одной стороны, из определения (11.1.1) вытекает, что

$$\begin{aligned} & \lim_{\Delta \rightarrow 0} \frac{P(x \leq X < x + \Delta, y \leq X + Y < y + \Delta)}{\Delta^2} \\ & \geq \lim_{\Delta \rightarrow 0} \frac{4}{\Delta^2} P\left(x \leq X < x + \frac{\Delta}{2}, y - x \leq Y < y - x + \frac{\Delta}{2}\right) \\ & = f_{X,Y}(x, y - x), \end{aligned} \quad (12.1.5)$$

а с другой стороны, из того же соотношения (11.1.1) мы получим

$$\begin{aligned} & \lim_{\Delta \rightarrow 0} \frac{P(x \leq X < x + \Delta, y \leq X + Y < y + \Delta)}{\Delta^2} \\ & \leq \lim_{\Delta \rightarrow 0} \frac{P(x \leq X < x + 2\Delta, y - x - \Delta \leq Y < y - x + \Delta)}{4\Delta^2} \\ & = f_{X,Y}(x, y - x). \end{aligned} \quad (12.1.6)$$

Из неравенств (12.1.5) и (12.1.6) вытекает, что

$$\lim_{\Delta \rightarrow 0} \frac{P(x \leq X < x + \Delta, y \leq X + Y < y + \Delta)}{\Delta^2} = f_{X,Y}(x, y - x).$$

Но в силу определения (11.1.1) последнее соотношение означает, что совместная плотность случайных величин  $X$  и  $X + Y$  выражается через совместную плотность случайных величин  $X$  и  $Y$  как

$$f_{X, X+Y}(x, y) = f_{X,Y}(x, y - x). \quad (12.1.7)$$

Таким образом, чтобы теперь найти плотность суммы  $X + Y$ , нам надо проинтегрировать совместную плотность  $f_{X, X+Y}(x, y)$  по  $dx$ , откуда мы получаем

$$f_{X+Y}(y) = \int_{-\infty}^{\infty} f_{X, X+Y}(x, y) dx = \int_{-\infty}^{\infty} f_{X,Y}(x, y - x) dx.$$

Если при этом случайные величины  $X$  и  $Y$  независимы, то из последнего соотношения мы получаем формулу

$$f_{X+Y}(y) = \int_{-\infty}^{\infty} f_Y(y - x) f_X(x) dx. \quad (12.1.8)$$

Поскольку в приведенных выше рассуждениях случайные величины  $X$  и  $Y$  абсолютно равноправны, мы точно так же получаем и формулу

$$f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(x - y) f_Y(y) dy. \quad (12.1.9)$$

Соотношения (12.1.8) и (12.1.9) называются *формулами свертки для плотностей*.

ПРИМЕР 12.1.1. Пусть  $X$  и  $Y$  – независимые случайные величины, имеющие одинаковое показательное распределение с параметром  $\lambda$ . Тогда

$$f_Y(y) = \begin{cases} 0, & y < 0, \\ \lambda e^{-\lambda y}, & y \geq 0, \end{cases} \quad f_X(x-y) = \begin{cases} \lambda e^{-\lambda(x-y)}, & y \leq x, \\ 0, & y > x. \end{cases}$$

Поэтому при  $x \geq 0$  мы имеем

$$\begin{aligned} f_{X+Y}(x) &= \int_{-\infty}^{\infty} f_X(x-y) f_Y(y) dy = \int_0^x f_X(x-y) f_Y(y) dy \\ &= \int_0^x \lambda^2 e^{-\lambda(x-y)} e^{-\lambda y} dy = \lambda^2 e^{-\lambda x} \int_0^x dy = x \lambda^2 e^{-\lambda x}. \end{aligned}$$

Для отрицательных  $x$  функция  $f_{X+Y}(x)$  равна нулю.

ПРИМЕР 12.1.2. Пусть  $X$  и  $Y$  – независимые случайные величины, имеющие нормальные распределения с параметрами  $a_1, \sigma_1^2$  и  $a_2, \sigma_2^2$  соответственно. Тогда по формуле (12.1.9), используя обозначение  $\exp\{z\} = e^z$ , мы получим

$$\begin{aligned} f_{X+Y}(x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(y-a_1)^2}{2\sigma_1^2}\right\} \\ &\quad \times \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(x-y-a_2)^2}{2\sigma_2^2}\right\} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left\{-\frac{y^2}{2\sigma_1^2} + \frac{a_1 y}{\sigma_1^2} - \frac{a_1^2}{2\sigma_1^2} - \frac{x^2}{2\sigma_2^2} - \frac{y^2}{2\sigma_2^2} + \frac{xy}{2\sigma_2^2} - \frac{a_2^2}{2\sigma_2^2} + \frac{a_2 x}{\sigma_2^2} + \frac{a_2 y}{\sigma_2^2}\right\} dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{x^2}{2\sigma_2^2} - \frac{a_1^2}{2\sigma_1^2} - \frac{a_2^2}{2\sigma_2^2} + \frac{a_2 x}{\sigma_2^2}\right\} \\ &\quad \times \int_{-\infty}^{\infty} \exp\left\{-\left[y^2 \left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\right) - y \left(\frac{a_1}{\sigma_1^2} + \frac{a_2}{\sigma_2^2} + \frac{x}{2\sigma_2^2}\right)\right]\right\} dy \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma_1\sigma_2} \sqrt{\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}} \exp \left\{ -\frac{x^2}{2\sigma_2^2} - \frac{a_1^2}{2\sigma_1^2} - \frac{a_2^2}{2\sigma_2^2} + \frac{a_2x}{\sigma_2^2} \right. \\
&\quad \left. - \frac{a_2^2\sigma_2^4}{\sigma_1^2 + \sigma_2^2} + \frac{\sigma_1^4(2a_2 + x)^2}{2(\sigma_1^2 + \sigma_2^2)} + \frac{a_1\sigma_1^2\sigma_2^2(2a_2 + x)}{\sigma_1^2 + \sigma_2^2} \right\} \\
&\quad \times \int_{-\infty}^{\infty} \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2\pi\sigma_1^2\sigma_2^2}} \\
&\quad \times \exp \left\{ -\frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1^2\sigma_2^2} \left( y - \frac{2a_1\sigma_2^2 + \sigma_1^2(2a_2 + x)}{2(\sigma_1^2 + \sigma_2^2)} \right)^2 \right\} dy \\
&= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp \left\{ -\frac{(x - (a_1 + a_2))^2}{2(\sigma_1^2 + \sigma_2^2)} \right\},
\end{aligned}$$

поскольку интеграл в предпоследней строке, как несложно видеть, равен единице, так как подынтегральная функция является плотностью нормального распределения с параметрами

$$a = \frac{2a_1\sigma_2^2 + \sigma_1^2(2a_2 + x)}{2(\sigma_1^2 + \sigma_2^2)} \text{ и } \sigma^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

Таким образом, распределение суммы независимых нормально распределенных случайных величин снова оказывается нормальным, причем параметры нормального распределения суммы оказываются равными суммам соответствующих параметров нормальных распределений слагаемых.

## 1.13. Условные распределения

В этом разделе мы снова будем иметь дело с двумерными случайными векторами  $X = (X, Y)$ .

### 1.13.1. Дискретные условные распределения

Сначала предположим, что и компонента  $X$ , и компонента  $Y$  дискретны. Возможные значения компоненты  $X$  обозначим  $x_1, x_2, \dots$ . Возможные

значения компоненты  $Y$  обозначим  $y_1, y_2, \dots$ . Также обозначим

$$p_{i,j} = P(X = x_i, Y = y_j), \quad i, j = 1, 2, \dots$$

Здесь  $p_{i,j}$  — это безусловная (априорная) вероятность того, что одновременно будут наблюдаться значения  $x_i$  компоненты  $X$  и  $y_j$  компоненты  $Y$ .

Найдем условную (апостериорную) вероятность того, что компонента  $X$  примет значение  $x_i$ , если известно, что значение компоненты  $Y$  равно  $y_j$ . По определению условной вероятности мы имеем

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{i,j}}{p_{\cdot,j}} \equiv q_i^{(j)}$$

(мы используем введенное в предыдущем разделе обозначение

$$p_{\cdot,j} = \sum_i p_{i,j} = P(Y = y_j)).$$

Просуммировав вероятности  $q_i^{(j)}$  по всем  $i$ , мы получим

$$\sum_i q_i^{(j)} = \sum_i \frac{p_{i,j}}{p_{\cdot,j}} = \frac{1}{p_{\cdot,j}} \sum_i p_{i,j} = \frac{p_{\cdot,j}}{p_{\cdot,j}} = 1.$$

Таким образом, при каждом  $j \geq 1$ , набор чисел

$$q_1^{(j)}, q_2^{(j)}, \dots$$

задает дискретное распределение вероятностей. Это распределение называется *условным распределением случайной величины  $X$  при условии  $Y = y_j$* .

### 1.13.2. Дискретные условные математические ожидания

Число

$$E(X | Y = y_j) = \sum_i x_i q_i^{(j)}$$

называется *условным математическим ожиданием случайной величины  $X$  при условии  $Y = y_j$* . Если множество значений случайной величины  $X$  бесконечно, то для существования условного математического ожидания этой случайной величины при условии  $Y = y_j$  необходимо и достаточно, чтобы

$$\sum_i |x_i| q_i^{(j)} < \infty.$$

Если указанное условие выполнено, то определена дискретная случайная величина  $E(X | Y)$ , которая принимает значения

$$E(X | Y = y_1), E(X | Y = y_2) \dots$$

соответственно с вероятностями  $p_{.1} = P(Y = y_1)$ ,  $p_{.2} = P(Y = y_2)$ , ...:

$$P(E(X | Y) = E(X | Y = y_j)) = P(Y = y_j) \equiv p_{.j}, \quad j \geq 1.$$

Определенная таким образом случайная величина  $E(X | Y)$  называется *условным математическим ожиданием случайной величины  $X$  относительно случайной величины  $Y$* .

Как и обычное математическое ожидание, условное математическое ожидание обладает свойством линейности: если  $X = \alpha X_1 + \beta X_2$ , где  $\alpha$  и  $\beta$  — числа, а  $X_1$  и  $X_2$  — случайные величины, то

$$E(X | Y) = \alpha E(X_1 | Y) + \beta E(X_2 | Y).$$

Математическое ожидание условного математического ожидания равно "безусловному" математическому ожиданию:

$$\begin{aligned} EE(X | Y) &= \sum_j p_{.j} \left( \sum_i x_i q_i^{(j)} \right) = \sum_i x_i \sum_j p_{.j} q_i^{(j)} \\ &= \sum_i x_i \sum_j p_{.j} \frac{p_{i,j}}{p_{.j}} = \sum_i x_i p_{i.} = \sum_i x_i P(X = x_i) = EX. \end{aligned}$$

**ПРИМЕР 13.2.1.** В разделе 1.7 мы рассматривали задачу о риске лейкемии в двух городах. При ее решении мы установили, что если  $X_1$  и  $X_2$  — две независимые случайные величины, имеющие распределение Пуассона с параметрами соответственно  $\lambda_1$  и  $\lambda_2$ , то условное распределение

случайной величины  $X_1$  при условии  $X_1 + X_2 = n$  является биномиальным с параметрами  $n$  и  $p = \lambda_1 / (\lambda_1 + \lambda_2)$ . Следовательно, условное математическое ожидание случайной величины  $X_1$  при условии  $X_1 + X_2 = n$  имеет вид

$$E(X_1 | X_1 + X_2 = n) = \frac{n\lambda_1}{\lambda_1 + \lambda_2}.$$

Но, как мы выяснили при решении той же задачи, случайная величина  $Y = X_1 + X_2$  имеет распределение Пуассона с параметром  $\lambda_1 + \lambda_2$ . Поэтому условное математическое ожидание  $E(X_1 | X_1 + X_2)$  случайной величины  $X_1$  относительно случайной величины  $X_1 + X_2$  — это случайная величина, распределение которой задается соотношением

$$P\left(E(X_1 | X_1 + X_2) = \frac{n\lambda_1}{\lambda_1 + \lambda_2}\right) = \exp\{\lambda_1 + \lambda_2\} \frac{(\lambda_1 + \lambda_2)^n}{n!},$$

$n = 0, 1, 2, \dots$

Отсюда легко видеть, что в рассматриваемом случае

$$E(X_1 | X_1 + X_2) = \frac{\lambda_1(X_1 + X_2)}{\lambda_1 + \lambda_2}.$$

### 1.13.3. Условные плотности

Теперь перейдем к рассмотрению ситуации, в которой случайный вектор  $X = (X, Y)$  имеет непрерывное распределение, задаваемое совместной плотностью  $f_{X,Y}(x, y)$ . В этой ситуации, как мы убедились выше, плотность  $f_Y(y)$  случайной величины  $Y$  можно выразить в виде интеграла от совместной плотности:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx.$$

К сожалению, в таком случае нельзя непосредственно выписать условную вероятность некоторого события  $X < x$  при условии  $Y = y$ , какими бы ни были  $x$  и  $y$ , поскольку при формальном подходе, основанном на определении условной вероятности, мы получим неопределенность вида

$$P(X < x | Y = y) = \frac{P(\{X < x\} \cap \{Y = y\})}{P(Y = y)} = \frac{0}{0}.$$



Чтобы обойти это препятствие, зафиксируем произвольное положительное число  $\varepsilon$  и рассмотрим условную вероятность события  $\{X < x\}$  при условии  $y \leq Y < y + \varepsilon$ . Мы получим

$$\begin{aligned} P(X < x \mid y \leq Y < y + \varepsilon) &= \frac{P(X < x, y \leq Y < y + \varepsilon)}{P(y \leq Y < y + \varepsilon)} \\ &= \frac{\iint_{S_\varepsilon(x, y)} f_{X, Y}(x_1, x_2) dx_1 dx_2}{\int_y^{y+\varepsilon} f_Y(z) dz}, \end{aligned} \quad (13.3.1)$$

где  $S_\varepsilon(x, y) = \{(x_1, x_2): -\infty < x_1 < x, y \leq x_2 < y + \varepsilon\}$ . При этом из определения плотности вытекает, что

$$\int_y^{y+\varepsilon} f_Y(z) dz = \varepsilon f_Y(y) + \delta(\varepsilon), \quad (13.3.2)$$

где  $\delta(\varepsilon)$  — некоторая функция, обладающая свойством

$$\lim_{\varepsilon \rightarrow 0} \frac{\delta(\varepsilon)}{\varepsilon} = 0 \quad (13.3.3)$$

(в этом легко убедиться, отталкиваясь от геометрической интерпретации интеграла как площади некоей криволинейной трапеции). Аналогично,

$$\iint_{S_\varepsilon(x, y)} f_{X, Y}(x_1, x_2) dx_1 dx_2 = \varepsilon \int_{-\infty}^x f_{X, Y}(x_1, y) dx_1 + \delta^*(\varepsilon), \quad (13.3.4)$$

где  $\delta^*(\varepsilon)$  — некоторая функция, обладающая свойством (13.3.3).

Таким образом, если мы, желая избавиться от описанной выше неопределенности, под условной вероятностью  $P(X < x \mid Y = y)$  будем понимать предел

$$P(X < x \mid Y = y) = \lim_{\varepsilon \rightarrow 0} P(X < x \mid y \leq Y < y + \varepsilon),$$

то на основании соотношений (13.3.1)–(13.3.4) мы получим

$$\begin{aligned}
 P(X < x | Y = y) &= \lim_{\varepsilon \rightarrow 0} P(X < x | y \leq Y < y + \varepsilon) \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{\iint_{S_\varepsilon(x, y)} f_{X, Y}(x_1, x_2) dx_1 dx_2}{\int_y^{y+\varepsilon} f_Y(z) dz} \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon \int_{-\infty}^x f_{X, Y}(z, y) dz + \delta^*(\varepsilon)}{\varepsilon f_Y(y) + \delta(\varepsilon)} \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{\int_{-\infty}^x f_{X, Y}(z, y) dz + \frac{\delta^*(\varepsilon)}{\varepsilon}}{f_Y(y) + \frac{\delta(\varepsilon)}{\varepsilon}} \\
 &= \frac{1}{f_Y(y)} \int_{-\infty}^x f_{X, Y}(z, y) dz = \int_{-\infty}^x \frac{f_{X, Y}(z, y)}{f_Y(y)} dz.
 \end{aligned}$$

Другими словами, мы получили формулу

$$P(X < x | Y = y) = \int_{-\infty}^x \frac{f_{X, Y}(z, y)}{f_Y(y)} dz, \quad (13.3.5)$$

справедливую для любых  $x$  и  $y$  при единственном условии  $f_Y(y) \neq 0$ . Поэтому по аналогии с определением “безусловной” плотности  $f_Z(x)$  случайной величины  $Z$  как неотрицательной функции, для которой при каждом  $x$  справедливо соотношение

$$P(Z < x) = \int_{-\infty}^x f_Z(z) dz,$$

соотношение (13.3.5) позволяет нам назвать функцию

$$g_Y(x) = \frac{f_{X, Y}(x, y)}{f_Y(y)} \quad (13.3.6)$$

условной плотностью непрерывной случайной величины  $X$  при условии  $Y = y$ .

### 1.13.4. Условные математические ожидания непрерывных случайных величин

Определенная в предыдущем подразделе функция  $g_y(x)$  при каждом значении  $y$ , для которого  $f_Y(y) \neq 0$ , обладает всеми свойствами плотности вероятностей. Поэтому по аналогии с “безусловным” математическим ожиданием непрерывной случайной величины мы можем определить *условное математическое ожидание* непрерывной случайной величины  $X$  при условии  $Y = y$  (которое, как и в дискретном случае, мы будем обозначать  $E(X | Y = y)$ ) как

$$E(X | Y = y) = \int_{-\infty}^{\infty} x g_y(x) dx = \int_{-\infty}^{\infty} \frac{x f_{X,Y}(x, y)}{f_Y(y)} dx.$$

Очевидно, что введенное таким образом условное математическое ожидание непрерывной случайной величины  $X$  при условии  $Y = y$  является (неслучайной) функцией аргумента  $y$ . Эта функция называется *регрессией  $X$  на  $Y$* .

Если мы теперь вместо  $y$  подставим в эту функцию случайную величину  $Y$ , то мы, естественно, получим случайную величину (обратим внимание, что формальная подстановка  $Y$  вместо  $y$  в обозначение  $E(X | Y = y)$  приведет к конфузу).

Под *условным математическим ожиданием непрерывной случайной величины  $X$  относительно непрерывной случайной величины  $Y$*  мы будем понимать случайную величину

$$E(X | Y) = \int_{-\infty}^{\infty} x g_Y(x) dx = \int_{-\infty}^{\infty} \frac{x f_{X,Y}(x, Y)}{f_Y(Y)} dx.$$

Как и в дискретном случае, условное математическое ожидание непрерывной случайной величины  $X$  относительно непрерывной случайной величины  $Y$  обладает свойством линейности: если  $X = \alpha X_1 + \beta X_2$ , где  $\alpha$  и  $\beta$  – числа, а  $X_1$  и  $X_2$  – случайные величины, то

$$E(X | Y) = \alpha E(X_1 | Y) + \beta E(X_2 | Y).$$

Как и в дискретном случае, “безусловное” математическое ожидание условного математического ожидания непрерывной случайной величины

$X$  относительно непрерывной случайной величины  $Y$  равно “безусловно-математическому ожиданию случайной величины  $X$ ”:

$$\begin{aligned} EE(X | Y) &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \frac{xf_{X,Y}(x, y)}{f_Y(y)} dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \frac{xf_{X,Y}(x, y)}{f_Y(y)} f_Y(y) dx \right) dy \\ &= \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} xf_{X,Y}(x, y) dx \right) dy \\ &= \int_{-\infty}^{\infty} x \left( \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \right) dx = \int_{-\infty}^{\infty} xf_X(x) dx = EX. \end{aligned}$$

**ПРИМЕР 13.4.1.** Пусть  $X$  и  $Y$  — независимые случайные величины, имеющие одинаковое показательное распределение с параметром  $\lambda$ . В силу независимости случайных величин  $X$  и  $Y$  их совместная плотность имеет вид

$$f_{X,Y}(x, y) = \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda y} = \lambda^2 e^{-\lambda(x+y)}$$

при  $x, y \geq 0$ . В силу соотношения (12.1.7) совместная плотность случайных величин  $X$  и  $X + Y$  равна

$$f_{X, X+Y}(x, y) = f_{X,Y}(x, y - x) = \lambda^2 e^{-\lambda(x+y-x)} = \lambda^2 e^{-\lambda y}$$

при  $0 \leq x \leq y$  и равна нулю вне интервала  $[0, y]$ . Обратим внимание, что найденная совместная плотность случайных величин  $X$  и  $X + Y$  постоянна по  $x$  в указанном интервале. Далее, в примере 1 мы установили, что плотность суммы  $X + Y$  имеет вид

$$f_{X+Y}(z) = z\lambda^2 e^{-\lambda z}$$

при неотрицательных  $z$  и эта плотность равна нулю для отрицательных  $z$ . Найдем теперь условную плотность случайной величины  $X$  при условии  $X + Y = s$  для некоторого положительного  $s$ . По определению (13.3.6) эта плотность имеет вид

$$g_s(x) = \frac{f_{X, X+Y}(x, s)}{f_{X+Y}(s)} = \frac{\lambda^2 e^{-\lambda s}}{s\lambda^2 e^{-\lambda s}} = \frac{1}{s}$$

для  $x \in [0, s]$  и  $g_s(x) = 0$  для  $x \notin [0, s]$ . Таким образом, мы установили, что условная плотность случайной величины  $X$  при условии  $X + Y = s$  совпадает с плотностью равномерного распределения на отрезке  $[0, s]$ . При этом легко видеть, что

$$E(X | X + Y = s) = \int_0^s \frac{dx}{s} = \frac{s}{2}$$

и соответственно

$$E(X | X + Y) = \frac{X + Y}{2}.$$

# Основные понятия прикладной статистики

## 2.1. Задачи математической и прикладной статистики

Основной целью теории вероятностей является построение математических моделей стохастических ситуаций, выявление закономерностей в проявлении случайных событий, позволяющих выразить вероятности более сложных событий через вероятности более простых.

Однако, как правило, эти закономерности записываются не в абсолютно точном виде, а нуждаются в последующем уточнении с учетом реально наблюдаемых значений тех или иных случайных величин. Поясним сказанное простым примером.

**ПРИМЕР 1.1.1.** Пусть  $X_1, X_2, \dots, X_n$  — результаты прямого измерения некоторого параметра  $a$ . На каждое измерение влияют некоторые факторы, из-за которых каждое измерение содержит случайную погрешность:  $X_j = a + \varepsilon_j, j = 1, \dots, n$ . Предполагая, что систематическая погрешность отсутствует, на основании закона больших чисел мы делаем заключение о том, что в качестве итоговой оценки параметра  $a$  следует взять среднее арифметическое наблюдений:

$$a \approx \frac{1}{n} \sum_{j=1}^n X_j. \quad (1.1.1)$$

Более того, на основании центральной предельной теоремы мы можем заключить, что если все измерения  $X_1, \dots, X_n$  равноточны с одной и той же дисперсией  $DX_j = \sigma^2$ , то погрешность в (1.1.1) имеет приближенно нормальное распределение с параметрами 0 и  $\sigma^2/n$ . Как видим, чтобы дать исчерпывающий ответ о значении параметра  $a$ , который должен содержать и информацию о точности нашей оценки, мы должны по наблюдениям  $X_1, \dots, X_n$  каким-то образом оценить неизвестный параметр  $\sigma^2$ .

Получение выводов о неизвестных параметрах вероятностных моделей на основе реальных данных является основной задачей *математической статистики*, к описанию основных процедур которой мы и приступаем.

Однако при применении статистических процедур, обоснованность которых вытекает из теорем математической статистики, мы сталкиваемся с одной важной особенностью. В любой математической дисциплине, в том числе и в математической статистике, имеются фундаментальное и прикладное направления. Красивые математически строгие утверждения фундаментальной статистики, являясь теоремами, начинаются со слов “предположим, что ...”, за которыми следует набор математических условий, при которых справедливо утверждение теоремы. К сожалению, как правило, эти условия на практике либо очень трудно, либо невозможно проверить без каких-либо дополнительных предположений или допущений. Но, как мы уже говорили выше, специалистам-прикладникам приходится решать не те задачи, которые *можно* решить, а те задачи, которые *нужно* решать. Поэтому неизбежно возникает необходимость изучения закономерностей в применении статистических методов к практическим задачам. Ведь любой метод можно применить правильно и неправильно. Как и столярные или слесарные инструменты, каждый метод математической статистики предназначен для своей конкретной ситуации и, будучи применен не по назначению, может принести существенный вред. *Прикладная статистика* как раз и занимается практическим применением методов математической статистики, выработкой практических рекомендаций, какие методы, как и в каких ситуациях применять можно, а какие — нельзя. Так как при этом часто приходится руководствоваться нематематическими, трудно формализуемыми соображениями, то прикладную

статистику можно считать как *наукой*, так и *искусством*. Мы не будем подробно говорить о всех направлениях этого искусства, ведь для серьезного разговора на эту тему надо иметь полное представление о всех методах математической статистики. Мы ограничимся лишь тем, что затронем некоторые особенности практического применения некоторых из статистических процедур, которые будут описаны во второй главе.

## 2.2. Выборочные характеристики

### 2.2.1. Выборка. Вариационный ряд.

#### Порядковые статистики

Выводы о значениях параметров вероятностной модели делаются на основе наблюдаемых значений случайных величин, связанных с изучаемой стохастической ситуацией. Как правило, наблюдаемые значения можно интерпретировать как значения одной и той же случайной величины, полученные в результате нескольких независимых воспроизведений стохастической ситуации. Но это то же самое, что наблюдаемые значения нескольких независимых и одинаково распределенных случайных величин. Поскольку мы будем обсуждать свойства *процедур*, или алгоритмов статистических выводов, которые не должны зависеть от конкретных значений начальных данных, то мы будем считать, что данные представляют собой выборку, понимаемую в следующем смысле.

**ОПРЕДЕЛЕНИЕ 2.1.1.** *Выборкой* называются независимые одинаково распределенные случайные величины  $X_1, X_2, \dots, X_n$ . Количество  $n$  наблюдений в выборке называется *объемом* выборки.

Иногда, чтобы подчеркнуть независимость наблюдений в выборке, выборку, определенную выше, называют *независимой*. Чтобы подчеркнуть совпадение распределений элементов выборки, выборку, определенную выше, иногда называют *однородной*.

Пусть  $F(x)$  — функция распределения элемента выборки, то есть  $F(x) = P(X_1 < x)$ . Иногда в связи со свойством воспроизводимости стохастической ситуации говорят, что  $X_1, X_2, \dots, X_n$  — это выборка из *генеральной совокупности* с распределением  $F(x)$ . Чисто формально в дальнейшем мы будем отождествлять понятие генеральной совокупнос-



ти с функцией распределения  $F(x)$  элемента выборки. Основная задача математической статистики заключается в описании функции распределения  $F(x)$ . Если эта функция распределения известна с точностью до некоторого параметра:  $F(x) = F(x; \theta)$ , то задача описания функции распределения  $F(x; \theta)$  сводится к отысканию параметра  $\theta$ . В таком случае говорят о задаче *параметрической статистики*. Если о функции распределения  $F(x)$  заранее ничего не известно, то о задаче ее описания по имеющимся наблюдениям  $X_1, X_2, \dots, X_n$  говорят как о задаче *непараметрической статистики*.

**ОПРЕДЕЛЕНИЕ 2.1.2.** Пусть  $X_1, X_2, \dots, X_n$  — выборка, то есть *независимые* случайные величины, занумерованные, скажем, в хронологическом порядке, то есть в порядке их регистраций. Расположим элементы выборки в порядке их возрастания, то есть перенумеруем их так, чтобы первым стал наименьший элемент выборки, а последним — наибольший. Мы получим набор  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  такой, что  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ . Упорядоченная таким образом выборка называется *вариационным рядом*. Элементы вариационного ряда называются *порядковыми статистиками*.

Отметим, что порядковые статистики уже не являются независимыми между собой. Найдем распределения экстремальных порядковых статистик

$$X_{(1)} = \min\{X_1, X_2, \dots, X_n\} \quad \text{и} \quad X_{(n)} = \max\{X_1, X_2, \dots, X_n\}.$$

Поскольку элементы выборки независимы между собой, мы имеем

$$\begin{aligned} P(X_{(n)} < x) &= P(\max\{X_1, X_2, \dots, X_n\} < x) \\ &= P(X_1 < x, X_2 < x, \dots, X_n < x) \\ &= P(X_1 < x) \cdot P(X_2 < x) \cdot \dots \cdot P(X_n < x) = (F(x))^n, \\ P(X_{(1)} < x) &= 1 - P(X_{(1)} \geq x) = 1 - P(\min\{X_1, X_2, \dots, X_n\} \geq x) \\ &= 1 - P(X_1 \geq x, X_2 \geq x, \dots, X_n \geq x) \\ &= 1 - P(X_1 \geq x) \cdot P(X_2 \geq x) \cdot \dots \cdot P(X_n \geq x) \\ &= 1 - (1 - F(x))^n. \end{aligned}$$

### 2.2.2. Статистические выводы о параметрах положения, разброса и формы распределения

Мы уже видели, что центр распределения, то есть такое значение, вокруг которого группируются значения выборки, можно описать многими способами, используя с этой целью, например, такие характеристики, как математическое ожидание, медиану или моду. В данном разделе мы опишем выборочные аналоги этих характеристик.

Эмпирическим (выборочным) аналогом математического ожидания естественно считать среднее арифметическое элементов выборки. Для него в математической статистике используется специальное обозначение  $\bar{X}_n$ :

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Отметим некоторые свойства величины  $\bar{X}_n$ . Если  $EX_1 = a$ , то в силу закона больших чисел с вероятностью единица

$$\bar{X}_n \rightarrow a$$

при неограниченном увеличении объема выборки. Из этого свойства вытекает естественность использования величины  $\bar{X}_n$  в качестве приближенного значения (в математической статистике говорят: в качестве *оценки*) математического ожидания  $a$ .

Рассмотрим стандартные рекомендации об использовании центральной предельной теоремы для оценивания точности приближения математического ожидания с помощью выборочного среднего.

Если  $0 < DX_1 = \sigma^2 < \infty$ , то в силу центральной предельной теоремы величина  $\bar{X}_n$  асимптотически нормальна: для любого числа  $x$

$$\lim_{n \rightarrow \infty} P \left( \frac{\sqrt{n}}{\sigma} \cdot (\bar{X}_n - a) < x \right) = \Phi(x).$$

Отсюда вытекает приближенная формула

$$P(\bar{X}_n < x) \approx \Phi \left( \sqrt{n} \cdot \frac{x - a}{\sigma} \right), \quad (2.2.1)$$

которую можно использовать при исследовании точности приближения математического ожидания величиной  $\bar{X}_n$ . А именно, пусть  $\varepsilon$  – положительное число. С помощью формулы (2.2.1) мы легко получаем соотношение

$$\begin{aligned} P(\bar{X}_n - \varepsilon \leq a \leq \bar{X}_n + \varepsilon) &\approx \Phi(\varepsilon\sqrt{n}/\sigma) - \Phi(-\varepsilon\sqrt{n}/\sigma) \\ &= 2\Phi(\varepsilon\sqrt{n}/\sigma) - 1 \end{aligned} \quad (2.2.2)$$

в силу того обстоятельства, что  $1 - \Phi(x) = \Phi(-x)$ , каково бы ни было положительное число  $x$ . В соотношении (2.2.2) величина  $\varepsilon$  характеризует точность приближенной формулы (2.2.1). Потребуем, чтобы вероятность (2.2.2) была равна заданному числу  $\gamma$ :

$$2\Phi(\varepsilon\sqrt{n}/\sigma) - 1 = \gamma. \quad (2.2.3)$$

В этом соотношении число  $\gamma$  характеризует степень нашей уверенности в том, что интервал со случайными концами  $[\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon]$  накроет (вообще говоря, неизвестное) математическое ожидание  $a$ , другими словами, число  $\gamma$  характеризует надежность нашего вывода.

Найдем соотношение между точностью  $\varepsilon$  и надежностью  $\gamma$  нашего вывода о значении  $a$ . В главе 1 для произвольного фиксированного  $\alpha \in (0, 1)$  мы определили  $\alpha$ -квантиль  $u_\alpha$  стандартного нормального распределения как решение уравнения  $\Phi(u_\alpha) = \alpha$ . Напомним, что значения  $\alpha$ -квантилей стандартного нормального распределения при различных  $\alpha$  можно найти в специальных таблицах. Из (2.2.3) мы приходим к уравнению

$$\Phi(\varepsilon\sqrt{n}/\sigma) = \frac{1 + \gamma}{2},$$

откуда по определению  $\alpha$ -квантили мы заключаем, что

$$\frac{\varepsilon\sqrt{n}}{\sigma} = u_{\frac{1+\gamma}{2}},$$

то есть

$$\varepsilon = \frac{\sigma}{\sqrt{n}} \cdot u_{\frac{1+\gamma}{2}}. \quad (2.2.4)$$

Соотношение (2.2.4) показывает, как точность зависит от объема выборки  $n$  и надежности  $\gamma$ . Обратим внимание на то обстоятельство, что рассуждения о точности статистических процедур обязательно зависят от

требований к их надежности. Выбор конкретного значения  $\gamma$  зависит от нематематических обстоятельств.

Мы описали стандартные рекомендации, включенные во многие учебные пособия. Однако эти рекомендации, к сожалению, никак не учитывают погрешность формулы (2.2.1). Для того, чтобы оценить эту погрешность, предположим, что существует  $\mu = E|X_1 - a|^3$  и обозначим  $L_3 = 0,7655\mu/(\sigma^3\sqrt{n})$ . Как мы видели в разделе 1.9,

$$\Phi\left(\sqrt{n} \cdot \frac{x-a}{\sigma}\right) - L_3 \leq P(\bar{X}_n < x) \leq \Phi\left(\sqrt{n} \cdot \frac{x-a}{\sigma}\right) + L_3.$$

Следовательно, рассуждая более аккуратно, вместо (2.2.2) мы должны записать

$$2\Phi(\varepsilon\sqrt{n}/\sigma) - 1 - 2L_3 \leq P(\bar{X}_n - \varepsilon \leq a \leq \bar{X}_n + \varepsilon) \leq 2\Phi(\varepsilon\sqrt{n}/\sigma) - 1 + 2L_3.$$

Разумно потребовать, чтобы вероятность того, что интервал со случайными концами  $[\bar{X}_n - \varepsilon, \bar{X}_n + \varepsilon]$  накроет неизвестное математическое ожидание  $a$ , была бы гарантированно не меньше, чем  $\gamma$ . Поэтому вместо (2.2.3) следует руководствоваться условием

$$2\Phi(\varepsilon\sqrt{n}/\sigma) - 1 - 2L_3 \geq \gamma,$$

очевидно, эквивалентным условию

$$\Phi(\varepsilon\sqrt{n}/\sigma) \geq \frac{1+\gamma}{2} + L_3,$$

откуда мы получаем, что точность  $\varepsilon$  вместо (2.2.4) на самом деле удовлетворяет более аккуратному неравенству

$$\varepsilon \geq \frac{\sigma}{\sqrt{n}} \cdot u_{\frac{1+\gamma}{2}+L_3}. \quad (2.2.5)$$

Как мы видим, соотношение (2.2.4) фактически дает *неоправданно более оптимистичную* оценку точности приближения математического ожидания с помощью среднего арифметического по сравнению с худшей (*но верной!*) оценкой (2.2.5). При этом можно показать, что всегда

$L_3 \geq 0,4/\sqrt{n}$ . Поэтому, если  $\frac{1+\gamma}{2} \geq 1 - 0,4/\sqrt{n}$ , то есть  $\gamma \geq 1 - 0,8/\sqrt{n}$  или  $n \leq 0,64/(1 - \gamma)^2$ , то, вообще говоря, нельзя сделать абсолютно уверенного вывода о том, какова точность аппроксимации (2.2.1).

Мы убедились, что если зафиксировать объем выборки  $n$  и требуемую надежность  $\gamma$ , то с помощью приведенных выше рассуждений можно получить оценки точности приближения математического ожидания с помощью выборочного среднего. В то же время, с помощью тех же соотношений мы можем получить оценки для необходимого объема выборки при фиксированных требованиях к точности  $\varepsilon$  и надежности  $\gamma$ . Такие оценки можно использовать для ответа на вопрос, репрезентативна ли выборка для обоснованного вывода о значении математического ожидания наблюдаемых величин. К подобной задаче мы вернемся в разделе 2.2.4. Наконец, с помощью тех же соотношений мы можем оценить надежность нашего вывода при фиксированном объеме выборки и заданных требованиях к его точности.

В разделе 1.4 для каждого  $q \in (0, 1)$  мы определили  $q$ -квантиль произвольной случайной величины. В соответствии с этим определением  $q$ -квантилью случайной величины  $X_1$  называется число  $x_q$  такое, что одновременно

$$P(X_1 \leq x_q) \geq q \quad \text{и} \quad P(X_1 \geq x_q) \geq 1 - q.$$

**ОПРЕДЕЛЕНИЕ 2.2.1** Если  $X_{(1)}, \dots, X_{(n)}$  — вариационный ряд, построенный по выборке  $X_1, \dots, X_n$ , то *эмпирической (или выборочной)  $q$ -квантилью* называется порядковая статистика с номером  $[nq]$ , где символ  $[x]$  обозначает целое число, ближайшее к числу  $x$  из всех целых чисел, не превосходящих  $x$ . В соответствии с определением медианы, *эмпирической (или выборочной) медианой* называется величина  $\bar{m}_n$ , определяемая как

$$\bar{m}_n = \begin{cases} X_{([n/2]+1)}, & \text{если } n \text{ нечетное,} \\ \frac{1}{2}(X_{([n/2])} + X_{([n/2]+1)}), & \text{если } n \text{ четное.} \end{cases}$$

Другими словами, выборочная медиана — это средний элемент вариационного ряда.

Вопросу о том, какая функция от выборки лучше оценивает неизвестный параметр (в данном случае в качестве неизвестного параметра

рассматривается “центр” распределения), можно придать строгий смысл. Прежде всего следует определить критерий качества оценки. В качестве такого критерия разумно взять величину, характеризующую точность оценки, например, ожидаемое значение квадрата разности между оцениваемым (неизвестным) параметром и его статистической оценкой: если оцениваемый параметр обозначить  $\theta$ , а его статистическую оценку обозначить  $\theta_n$  ( $\theta_n = \theta_n(X_1, \dots, X_n)$ ), то рассматриваемый критерий имеет вид

$$E(\theta_n - \theta)^2.$$

При этом, если функция от выборки  $\theta_n(X_1, \dots, X_n)$  оценивает параметр  $\theta$  без систематического смещения, то есть

$$E\theta_n = \theta$$

(как мы уже отмечали выше, в таком случае говорят, что оценка  $\theta_n$  является *несмещенной* оценкой параметра  $\theta$ ), то по определению дисперсии мы получаем

$$E(\theta_n - \theta)^2 = D\theta_n.$$

Для унификации единиц измерения в качестве критерия качества несмещенных оценок можно рассматривать их среднеквадратическое отклонение  $\sqrt{D\theta_n}$ .

**ПРИМЕР 2.2.1.** Предположим, что в выборке  $X_1, \dots, X_n$  все элементы независимы и имеют одну и ту же непрерывную плотность распределения  $f(x)$ . Обозначим  $m = \text{med}X_1$ . Предположим, что  $f(m) > 0$ . Пусть  $\bar{m}_n$  – выборочная медиана, построенная по выборке  $X_1, \dots, X_n$ . Еще в 1931 г. А. Н. Колмогоров<sup>1</sup> показал, что при  $n \rightarrow \infty$

$$P(\sqrt{n}(\bar{m}_n - m) < x) \rightarrow \Phi(2f(m)x),$$

где, как обычно,  $\Phi(y)$  – стандартная нормальная функция распределения, так что  $\Phi(2f(m)x)$  – функция распределения нормально распределенной

<sup>1</sup>А. Н. Колмогоров. Метод медианы в теории ошибок. *Матем. сборник*, 1931, т. 38, № 3/4, с. 47-50. Также см. А. Н. Колмогоров. *Теория вероятностей и математическая статистика*. Сборник статей. “Наука”, Москва, 1986, с. 111-114.

случайной величины с дисперсией  $(2f(m))^{-2}$ . Таким образом, для рассмотренного выше критерия качества выборочной медианы при больших  $n$  мы имеем

$$E(\bar{m}_n - m)^2 = \frac{1}{n} E[\sqrt{n}(\bar{m}_n - m)]^2 \approx \frac{1}{4n(f(m))^2}.$$

Если мы дополнительно обозначим  $a = EX_1$ , то согласно центральной предельной теореме

$$P(\sqrt{n}(\bar{X}_n - a) < x) \rightarrow \Phi(x/\sqrt{DX_1}),$$

то есть при больших  $n$

$$E(\bar{X}_n - a)^2 = \frac{1}{n} E[\sqrt{n}(\bar{X}_n - a)]^2 \approx \frac{DX_1}{n}.$$

Таким образом, ответ на вопрос о том, какая из оценок – выборочное среднее или выборочная медиана – лучше, можно получить, скажем, вычислив отношение

$$\frac{E(\bar{X}_n - a)^2}{E(\bar{m}_n - m)^2} \approx 4(f(m))^2 DX_1$$

(это отношение называется *относительной эффективностью* оценок  $\bar{X}_n$  и  $\bar{m}_n$ ). В частности, если

$$f(x) = \frac{1}{\sqrt{2\pi DX_1}} \exp \left\{ -\frac{(x-a)^2}{2DX_1} \right\}$$

– плотность нормального распределения со средним  $a$  и дисперсией  $DX_1$ , то, во-первых,  $a = m$  и, во-вторых,  $f(m) = 1/\sqrt{2\pi DX_1}$ , так что

$$\frac{E(\bar{X}_n - a)^2}{E(\bar{m}_n - m)^2} \approx \frac{2}{\pi}.$$

Мы видим, что, если в нормальном случае для оценивания параметра положения мы будем использовать выборочную медиану, то для того, чтобы достичь той же точности, что при использовании выборочного

среднего, нам понадобится в  $\pi/2 \approx 1,57$  раз больше наблюдений, то есть в таком случае выборочная медиана более чем в полтора раза менее эффективна, нежели выборочное среднее.

При использовании выборочного среднего и выборочной медианы в качестве статистических оценок параметра, характеризующего “центр” распределения, следует заметить, что выборочная медиана обладает большей устойчивостью к присутствию в выборке так называемых “загрязняющих” наблюдений. Действительно, если выборка  $X_1, \dots, X_n$  в некотором смысле не является однородной, то есть наряду с наблюдениями, имеющими функцию распределения  $F(x)$ , в ней присутствуют наблюдения с какой-то другой функцией распределения, то в выборочное среднее наряду с “правильными” наблюдениями войдут значения “загрязняющих” наблюдений. При этом, если значения “загрязняющих” наблюдений велики, то их присутствие, естественно, сильно смажет итоговую картину. В то же время, отклонения выборочной медианы от ее “правильного” значения зависит не столько от значений “загрязняющих” наблюдений, сколько от их числа. С этим свойством медианы мы познакомились, разбирая Задачу 4.1.3. Такое свойство выборочной медианы называется *робастностью*.

**ПРИМЕР 2.2.2.** Вышеупомянутое свойство робастности выборочной медианы хорошо иллюстрируется на примере следующей ситуации. Предположим, что в выборке  $X_1, \dots, X_n$  все элементы независимы и имеют плотность распределения

$$f(x) = \frac{1-p}{\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2}\right\} + \frac{p}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\},$$

где  $0 < p < 1$  и  $\sigma^2 > 1$ . Эту ситуацию можно интерпретировать как наличие в выборке примерно  $(1-p) \cdot 100\%$  наблюдений с нормальным распределением, имеющим параметры  $a$  и  $1$ , и примерно  $p \cdot 100\%$  наблюдений с нормальным распределением, имеющим параметры  $a$  и  $\sigma^2 > 1$ , то есть изучаемая популяция (генеральная совокупность) является *смесью* двух популяций: нормально распределенной с параметрами  $a$  и  $1$  и нормально распределенной с параметрами  $a$  и  $\sigma^2 > 1$ , причем доли этих двух суб-популяций (*компонент смеси*) составляют соответственно  $(1-p) \cdot 100\%$  и  $p \cdot 100\%$ . Если при этом  $p$  мало, то говорят, что выборка



из первой суб-популяции загрязнена объектами (наблюдениями) из второй суб-популяции. Заметим, что параметры "центра" у обеих компонент смеси одинаковы. Легко видеть, что  $a = m$  и

$$f(m) = \frac{1}{\sqrt{2\pi}} \left(1 - p + \frac{p}{\sigma}\right).$$

Далее,

$$\begin{aligned} DX_1 &= \int_{-\infty}^{\infty} (x - a)^2 f(x) dx \\ &= \frac{1-p}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - a)^2 \exp\left\{-\frac{(x-a)^2}{2}\right\} dx \\ &\quad + \frac{p}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - a)^2 \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\} dx \\ &= 1 + p(\sigma^2 - 1). \end{aligned}$$

Вычислим относительную эффективность выборочного среднего и выборочной медианы:

$$\frac{E(\bar{X}_n - a)^2}{E(\bar{m}_n - m)^2} \approx 4(f(m))^2 DX_1 = \frac{2}{\pi} \left(1 - p + \frac{p}{\sigma}\right)^2 (1 + p(\sigma^2 - 1)).$$

Несложно видеть, что при каждом фиксированном значении  $p$  правая часть последнего соотношения неограниченно возрастает при увеличении  $\sigma^2$ . К примеру, если  $p = 0,01$ , то выборочная медиана эффективнее выборочного среднего для  $\sigma^2 > 61$ . Если же  $p = 0,05$ , то выборочная медиана эффективнее выборочного среднего для  $\sigma^2 > 14$ . Наконец, если доля "загрязняющих" наблюдений составляет 10%, то выборочная медиана эффективнее выборочного среднего уже для  $\sigma^2 > 9,1$ .

Перейдем к описанию выборочных аналогов характеристик разброса. При определении величины  $\bar{X}_n$ , являющейся выборочным аналогом математического ожидания, мы по сути заменили символ математического ожидания  $E$  символом арифметического усреднения  $\frac{1}{n} \sum_{k=1}^n$ . Это наводит нас на мысль о том, что в качестве выборочного аналога дисперсии

можно взять величину

$$\tilde{S}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Однако при аппроксимации дисперсии с помощью величины  $\tilde{S}_n^2$  мы обнаруживаем систематическое смещение. Дело в том, что если  $DX_1 = \sigma^2$ , то

$$E\tilde{S}_n^2 = \frac{n-1}{n} \cdot \sigma^2 = \sigma^2 - \frac{\sigma^2}{n}.$$

При больших объемах выборки смещение  $\sigma^2/n$  незначительно, но при умеренных  $n$  это смещение может заметно исказить результат. Это смещение легко устранить, взяв в качестве *несмещенной выборочной дисперсии* величину

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Легко убедиться, что  $E S_n^2 = \sigma^2$ .

В качестве эмпирического аналога интерквартильного размаха естественно взять величину

$$X_{((3n/4))} - X_{((n/4))}.$$

С помощью описанной выше замены символа математического ожидания символом арифметического усреднения легко получить выборочные аналоги коэффициентов асимметрии и эксцесса, определенных в разделе 4.3. А именно, если  $\bar{X}_n$  и  $\tilde{S}_n^2$  — соответственно выборочные среднее и дисперсия, то в качестве *выборочного коэффициента асимметрии* естественно взять величину

$$\bar{x}_3 = \frac{\sum_{j=1}^n (X_j - \bar{X}_n)^3}{n\tilde{S}_n^3} = \frac{\sqrt{n} \sum_{j=1}^n (X_j - \bar{X}_n)^3}{\left[ \sum_{j=1}^n (X_j - \bar{X}_n)^2 \right]^{3/2}},$$

а в качестве *выборочного коэффициента эксцесса* — величину

$$\bar{x}_4 = \frac{\sum_{j=1}^n (X_j - \bar{X}_n)^4}{n\tilde{S}_n^4} = \frac{n \sum_{j=1}^n (X_j - \bar{X}_n)^4}{\left[ \sum_{j=1}^n (X_j - \bar{X}_n)^2 \right]^2}.$$

### 2.2.3. Непараметрическое оценивание распределения генеральной совокупности

**Эмпирическая функция распределения.** Пусть  $X_1, X_2, \dots, X_n$  – независимая выборка из генеральной совокупности с распределением  $F(x)$  (то есть  $F(x) = P(X_1 < x)$ ), то в качестве приближения для  $F(x)$  можно взять эмпирическую функцию распределения  $F_n(x)$ , определяемую следующим образом. Для произвольного  $x$  пусть  $\nu(x)$  – число тех элементов выборки  $X_1, X_2, \dots, X_n$ , которые меньше  $x$ ,  $\nu(x) = \sum_{j=1}^n \mathbf{1}_{(-\infty, x)}(X_j)$ , где, как и ранее, мы используем обозначение  $\mathbf{1}_A(y)$  для индикаторной функции множества  $A$ :

$$\mathbf{1}_A(y) = \begin{cases} 1, & \text{если } y \in A; \\ 0, & \text{если } y \notin A. \end{cases}$$

Тогда эмпирическая функция распределения  $F_n(x)$  определяется как

$$F_n(x) = \frac{\nu(x)}{n} = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(-\infty, x)}(X_j), \quad -\infty < x < \infty.$$

При каждом фиксированном  $x$  эмпирическая функция распределения  $F_n(x)$  является случайной величиной. Несложно убедиться, что при этом

$$E F_n(x) = F(x), \quad D F_n(x) = \frac{1}{n} F(x)(1 - F(x)).$$

Отсюда видно, что при неограниченном возрастании объема выборки эмпирическая функция распределения все более и более сближается с теоретической функцией распределения  $F(x)$ . Например, в силу неравенства Чебышева для любого сколь угодно малого положительного числа  $\varepsilon$  мы имеем

$$P(|F_n(x) - F(x)| > \varepsilon) \leq \frac{D F_n(x)}{\varepsilon^2} = \frac{F(x)(1 - F(x))}{n\varepsilon^2} \rightarrow 0 \quad (n \rightarrow \infty).$$

Более того, на самом деле имеет место существенно более сильное утверждение: с вероятностью единица предел при  $n \rightarrow \infty$  наибольшего (по  $x$ ) возможного значения величины  $|F_n(x) - F(x)|$  равен нулю. Это утверждение носит название *теоремы Гливенко*.

**Гистограмма и полигон.** Если выборка  $X_1, X_2, \dots, X_n$  представляет собой результаты  $n$  независимых наблюдений (реализаций) дискретной случайной величины  $X$ , принимающей значения  $x_1, x_2, \dots$  (то есть каждое из наблюдений  $X_1, X_2, \dots, X_n$  представляет собой одно из чисел  $x_1, x_2, \dots$ ) с вероятностями соответственно  $p_1, p_2, \dots$  ( $p_i > 0$ ,  $i = 1, 2, \dots$ ,  $p_1 + p_2 + \dots = 1$ ), то *выборочной частотой* значения  $x_i$  называется величина

$$\tilde{p}_i = \frac{v_i}{n},$$

где  $v_i$  – количество элементов выборки  $X_1, X_2, \dots, X_n$ , равных  $x_i$ . Несложно видеть, что выборочные частоты удовлетворяют соотношению  $\tilde{p}_1 + \tilde{p}_2 + \dots = 1$ . В силу закона больших чисел для каждого  $i$  с вероятностью единица предел при  $n \rightarrow \infty$  выборочной частоты  $\tilde{p}_i$  совпадает с  $p_i$ .

Если выборка  $X_1, X_2, \dots, X_n$  представляет собой результаты  $n$  независимых наблюдений (реализаций) непрерывной случайной величины  $X$ , распределение которой имеет функцию плотности  $p(x)$ , то можно построить выборочные аналоги функции плотности – так называемые *гистограмму* и *полигон*. С этой целью задают целое положительное число  $k$  и разбивают интервал  $[X_{(1)}, X_{(n)}]$  (напомним, что  $X_{(1)}$  – наименьший элемент выборки, а  $X_{(n)}$  – наибольший) на  $k$  равных непересекающихся частей. Обозначим полученные подынтервалы символами  $\Delta_j$ ,  $j = 1, \dots, k$  (в формальной записи  $\Delta_j = [X_{(1)} + (j-1)\delta, X_{(1)} + j\delta]$ ,  $j = 1, \dots, k$ , где  $\delta = (X_{(n)} - X_{(1)})/k$ ). Пусть  $v_j$  – число тех элементов выборки  $X_1, X_2, \dots, X_n$ , которые попали в интервал  $\Delta_j$ . Для  $x \in \Delta_j$  определим функцию  $\tilde{p}_n(x)$  равной  $v_j/n$ ,  $j = 1, \dots, k$ . Так определенная функция  $\tilde{p}_n(x)$  является ступенчатой и называется *гистограммой*. Желая сделать наш аналог плотности более гладким, можно последовательно соединить отрезками средние точки ступенек гистограммы, а для крайних ступенек – провести прямые линии через средние точки верхней и боковой грани ступенек до пересечения с осью абсцисс. Построенные отрезки образуют фигуру, называемую *полигоном*.

Ясно, что вид гистограммы и полигона существенно зависит от выбора числа  $k$ . На практике можно выбирать  $k$  равным ближайшему целому к  $\sqrt{n}$ , если  $n \leq 300$ , и равным ближайшему целому к  $C \log_2 n$ , если  $n > 300$ , где  $C = \sqrt{300}/\log_2 300$ .

**Гладкие оценки функции распределения и плотности.** При всей простоте, гистограмма и полигон имеют несколько существенных недостатков. Во-первых, гистограмма и полигон не являются в достаточной степени гладкими функциями. Во-вторых, и гистограмма, и полигон строятся по сгруппированным данным, а стало быть, происходит потеря информации при группировании, когда наблюдения, попавшие в один интервал  $\Delta_j$ , фактически заменяются их средним значением.

Идея построения более совершенных оценок плотности заключается в следующем. Если наблюдаемыми значениями выборки  $\mathbf{X} = (X_1, \dots, X_n)$  является набор  $\mathbf{x} = (x_1, \dots, x_n)$ , то соответствующая реализация эмпирической функции распределения

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{(-\infty, x)}(x_j) = \frac{1}{n} \sum_{j=1}^n Q_j(x)$$

является средним арифметическим функций

$$Q_j(x) = \begin{cases} 0, & \text{если } x \leq x_j, \\ 1, & \text{если } x > x_j. \end{cases}$$

Каждая функция  $Q_j(x)$  представляет собой вырожденную функцию распределения, соответствующую случайной величине, с вероятностью единица принимающей значение  $x_j$ . Теперь ясно, что если вместо функций  $Q_j(x)$  взять какие-нибудь гладкие (непрерывные) функции распределения  $G_j(x)$ , то соответствующая оценка для функции распределения  $F(x)$  также станет гладкой. На практике в качестве  $G_j(x)$  берут функции вида  $G_j(x) = G(x - X_j/a_n)$ , где  $G(x)$  – некоторая фиксированная функция распределения, а  $a_n > 0$  – так называемый параметр гладкости, выбор которого является прерогативой исследователя, так что получается приближенная формула

$$F(x) \approx \frac{1}{n} \sum_{j=1}^n G\left(\frac{x - X_j}{a_n}\right). \quad (2.3.1)$$

Легко убедиться, что если при этом функции распределения  $G(x)$  соответствует плотность  $g(x)$ , то есть  $G(x) = \int_{-\infty}^x g(x) dx$ , то функции

распределения, стоящей в правой части формулы (2.3.1) соответствует плотность

$$f_n(x) = \frac{1}{na_n} \sum_{j=1}^n g\left(\frac{x - X_j}{a_n}\right). \quad (2.3.2)$$

Функция  $f_n(x)$  представляет собой оценку для неизвестной плотности  $p(x)$ . Оценки типа (2.3.2) называются *ядерными*, а соответствующая плотность  $g(x)$  называется *ядром*.

При использовании ядерных оценок плотности главными проблемами являются выбор ядра и выбор параметра гладкости. Как правило, используются ядра, удовлетворяющие условиям

$$\int_{-\infty}^{\infty} g(x) dx = 1, \quad \int_{-\infty}^{\infty} xg(x) dx = 0, \quad \int_{-\infty}^{\infty} x^2g(x) dx = 1.$$

Первое из этих условий вытекает из требования, чтобы функция  $g(x)$  была плотностью распределения, второе условие означает, что случайная величина с плотностью распределения  $g(x)$  имеет нулевое математическое ожидание, а третье условие означает, что дисперсия этой случайной величины равна единице. Чаще всего в качестве  $g(x)$  используются равномерная плотность  $g(x) = 1_{[-\sqrt{3}, \sqrt{3}]}(x)$  (в этом случае получается непрерывная оценка для функции распределения  $F(x)$ , но ступенчатая оценка для плотности  $f(x)$ ) или стандартная нормальная плотность  $g(x) = \varphi(x)$ . Некоторые исследователи отмечают, что хорошие, наглядные результаты дает применение квадратичного ядра

$$g(x) = \begin{cases} 0 & \text{при } x < -2,5, \\ -\frac{576x^2}{390625} + \frac{144}{15625} & \text{при } -2,5 \leq x \leq 2,5, \\ 0 & \text{при } x > 2,5. \end{cases}$$

При малых значениях параметра гладкости  $a_n$  ядерная оценка имеет много довольно часто расположенных острых зубцов. При увеличении параметра  $a_n$  ядерная оценка становится все более и более гладкой. При этом в качестве окончательного значения выбирается то, при котором вид ядерной оценки плотности в наибольшей степени устраивает исследователя. Другими словами, выбор параметра сглаживания на практике — это в большей степени искусство или шаманство, нежели математика.

### 2.2.4. Репрезентативность выборки

Часто возникает вопрос о том, достаточно ли имеющихся статистических данных для того, чтобы выводы, сделанные на их основе, были точными и надежными, другими словами, репрезентативна ли имеющаяся выборка. Эта довольно общая проблема в некоторых случаях может быть сформулирована более конкретно. Рассмотрим следующую задачу.

**Задача 2.4.1.** Предположим, что с целью определения рейтинга некоего политического деятеля опрошено  $n$  человек. Какова точность оценки рейтинга по итогам этого опроса?

Сразу заметим, что формулировка вопроса нуждается в уточнении. Прежде всего, необходимо уяснить, что такое рейтинг, и построить математическую модель рейтинга. Предположим, что каждому респонденту задается один и тот же вопрос: "Поддерживаете ли Вы данного политического деятеля?" На такой вопрос можно дать лишь один из двух возможных ответов: "да" или "нет". Предположим также, что респонденты отвечают на этот вопрос независимо друг от друга. Тогда результаты опроса представляют собой  $n$  независимых случайных величин  $X_1, X_2, \dots, X_n$ , каждая из которых принимает одно из двух возможных значений: 0 (что соответствует ответу "нет") и 1 (что соответствует ответу "да"), причем

$$P(X_j = 1) = p = 1 - P(X_j = 0), \quad j = 1, \dots, n.$$

В этой ситуации рейтингом данного политического деятеля разумно считать вероятность  $p$  того, что наугад выбранный респондент поддерживает его. Таким образом, с формальной точки зрения мы имеем дело с последовательностью испытаний Бернулли. В качестве *эмпирического рейтинга* естественно взять эмпирическую частоту, которая в рассматриваемой ситуации может быть записана как среднее арифметическое значение величин  $X_1, X_2, \dots, X_n$ :

$$p_n^* = \frac{1}{n} \sum_{j=1}^n X_j = \bar{X}_n.$$

Рассмотрим вопрос о том, какова точность приближения

$$p_n^* \approx p.$$

К сожалению, в отличие от детерминированных (неслучайных) схем, при анализе случайных данных *одного* параметра, характеризующего точность, недостаточно, так как событие

$$|p_n^* - p| \leq \varepsilon$$

является случайным, каково бы ни было число  $\varepsilon \in (0, 1)$ , поскольку для одной выборки  $X_1, X_2, \dots, X_n$  это событие может осуществиться, а для какой-либо другой – нет. Поэтому наряду с параметром  $\varepsilon$ , характеризующим точность, зададим еще один параметр  $\gamma \in (0, 1)$  и потребуем, чтобы вероятность указанного события была бы не меньше  $\gamma$  или, что то же самое,

$$P(|p_n^* - p| \geq \varepsilon) \leq 1 - \gamma. \quad (2.4.1)$$

При этом ясно, что  $\varepsilon$  должно быть близко к нулю, а  $\gamma$  должно быть близко к единице, характеризуя нашу уверенность в правильности вывода. Другими словами, параметр  $\gamma$  характеризует надежность статистического вывода.

**Решение, основанное на неравенстве Чебышева.** Несложно видеть, что  $E p_n^* = p$ , а  $D p_n^* = p(1-p)/n$ . Тогда по неравенству Чебышева

$$P(|p_n^* - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}. \quad (2.4.2)$$

Из соотношений (2.4.1) и (2.4.2) мы получим неравенство

$$\frac{p(1-p)}{n\varepsilon^2} \leq 1 - \gamma,$$

откуда

$$n \geq \frac{p(1-p)}{\varepsilon^2(1-\gamma)}. \quad (2.4.3)$$

К сожалению, правая часть этого неравенства зависит от неизвестного параметра  $p$ . Однако это препятствие можно обойти. Известно, что наибольшее значение величины  $p(1-p)$  при  $0 \leq p \leq 1$  равно  $\frac{1}{4}$ . Поэтому с целью получить *гарантированную* оценку для объема выборки мы в неравенстве (2.4.3) выражение  $p(1-p)$  заменим его наибольшим значением и окончательно получим неравенство

$$n \geq \frac{1}{4\varepsilon^2(1-\gamma)}. \quad (2.4.4)$$



В частности, если  $\gamma = 0,95$ ,  $\varepsilon = 0,001$ , то  $n \geq 5000000$ .

Соотношение (2.4.4) можно использовать для решения обратной задачи. Предположим, что опрошено  $n = 1500$  человек (что типично для опросов, результаты которых публикуются в средствах массовой информации). Какова точность соответствующей оценки?

Из соотношения (2.4.4) вытекает неравенство

$$\varepsilon \geq \frac{1}{2\sqrt{n(1-\gamma)}}.$$

Поэтому для  $n = 1500$  и, скажем,  $\gamma = 0,95$  мы получаем  $\varepsilon \geq 5,77\%$ , то есть при таком объеме выборки и таких требованиях к надежности *гарантированная* точность приближения  $p_n^* \approx p$  составляет  $\pm 5,77\%$ . Другими словами, мы видим, что иной раз погрешность может превышать сам рейтинг.

**Решение, основанное на теореме Муавра–Лапласа.** Если воспользоваться теоремой Муавра–Лапласа, то оценки для необходимого объема выборки и соответствующей точности можно улучшить. А именно, поскольку случайная величина  $X_1 + \dots + X_n$  имеет биномиальное распределение с параметрами  $n$  и  $p$ , по теореме Муавра–Лапласа мы имеем

$$\begin{aligned} P(|p_n^* - p| \geq \varepsilon) &= P\left(\left|\frac{\sum_{j=1}^n X_j - np}{\sqrt{np(1-p)}}\right| \geq \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \\ &\approx 2\left[1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)\right] \end{aligned} \quad (2.4.5)$$

(см. задачу 7.2.1). Учитывая требования к надежности нашего вывода, потребуем, чтобы вероятность (2.4.5) была бы не больше  $1 - \gamma$ :

$$2\left[1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)\right] \leq 1 - \gamma, \quad (2.4.5')$$

откуда с учетом определения  $\alpha$ -квантили  $u_\alpha$  стандартного нормального закона (см. раздел 2.2) мы приходим к неравенству

$$\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \geq u_{\frac{1+\gamma}{2}},$$

откуда

$$n \geq u_{\frac{1+\gamma}{2}}^2 \cdot \frac{p(1-p)}{\varepsilon^2}.$$

Избавляясь от неизвестного параметра  $p$  точно так же, как в соотношении (2.4.4), мы окончательно получаем гарантированную оценку

$$n \geq \frac{u_{\frac{1+\gamma}{2}}^2}{4\varepsilon^2}. \quad (2.4.6)$$

В частности, если  $\gamma = 0,95$ ,  $\varepsilon = 0,001$ , то из этой оценки вытекает, что  $n \geq 960400$  (из таблиц мы находим, что  $u_{0,975} = 1,96$ ).

Используем неравенство (2.4.6) для решения обратной задачи, а именно, найдем точность  $\varepsilon$ , если известно, что  $n = 1500$  и  $\gamma = 0,95$ . Из (2.4.6) мы получаем неравенство

$$\varepsilon \geq u_{\frac{1+\gamma}{2}} \cdot \frac{1}{2\sqrt{n}} \geq 0,02530,$$

то есть погрешность примерно равна  $\pm 2,53\%$ .

Более того, из неравенства (2.4.5') вытекает, что при  $n = 1500$ ,  $\varepsilon = 0,001$  и  $0,01 \leq p \leq 0,5$  надежность выводов колеблется от 0,1% (для  $p = 0,5$ ) до 29,6% (для  $p = 0,01$ ).

Мы видим, что оценка для числа респондентов (то есть объема выборки), полученная с использованием теоремы Муавра–Лапласа, более чем в 5 раз “лучше” оценки, полученной с использованием неравенства Чебышева. Оценка точности по теореме Муавра–Лапласа примерно в 2,3 раза “лучше”, чем по неравенству Чебышева. Однако оптимизм от кажущегося преимущества решения, основанного на теореме Муавра–Лапласа, не должен нас слишком расслаблять. Дело в том, что неравенство Чебышева дает хоть и более грубые, но абсолютно корректные, гарантированные оценки для объема выборки и точности. В то же время, привлекая теорему Муавра–Лапласа, мы используем приближенное равенство (2.4.5), которое само вносит погрешность в наш вывод. Поэтому, рассуждая так же, как при исследовании точности приближения математического ожидания с помощью выборочного среднего, мы должны заметить, что на самом деле вместо (2.4.6) нужно использовать более аккуратную оценку

$$n \geq \frac{u_{\frac{1+\gamma}{2}+L_3}^2}{4\varepsilon^2},$$

решение которого отнюдь не представляет собой простую задачу, поскольку величина  $L_3$  зависит от  $n$  (и, кстати, от неизвестного параметра  $p$ ). Однако с учетом неравенства  $L_3 \geq 0,4/\sqrt{n}$  мы можем сказать, что на самом деле при  $n = 1500$  и  $\gamma = 0,95$  точность заведомо хуже, чем

$$\varepsilon \geq u_{\frac{1+\gamma}{2}+L_3} \cdot \frac{1}{2\sqrt{n}} \geq u_{0,985} \cdot \frac{1}{2\sqrt{1500}} = 0,0281 = 2,81\%.$$

Таким образом, при выборке объемом 1500 респондентов организации, проводящие опрос общественного мнения и/или средства массовой информации могут допускать любые, в том числе умышленные отклонения обнаруживаемого рейтинга от его "истинного" значения в пределах почти шестипроцентного интервала, так как такие отклонения можно объяснить чисто случайной погрешностью.

Более того, при  $n \leq 0,64/(1-\gamma)^2$  теореме Муавра–Лапласа вообще нельзя использовать для гарантированного вывода о точности результатов опроса. Отсюда вытекает, что, не пренебрегая точностью, теоремой Муавра–Лапласа можно пользоваться, только если выполняется соотношение  $\gamma \leq 1 - 0,8/\sqrt{n}$ . В частности, при  $n = 1500$  максимально возможная надежность вывода меньше 0,98.

В только что рассмотренной задаче цель статистического исследования была сформулирована конкретно: оценить рейтинг, что в формальной постановке сводится к оцениванию вероятности успеха в испытаниях Бернулли. Далекое не всегда цель исследования можно заранее столь же конкретно сформулировать. Очень часто эта цель подвергается корректировке в ходе самого исследования. В силу подобных причин иногда приходится сталкиваться со следующей, казалось бы, совсем общей и потому неразрешимой задачей.

**ЗАДАЧА 2.4.2.** Имеется независимая однородная выборка  $X_1, \dots, X_n$  объема  $n$ . Репрезентативна ли она?

Кажется, что задача сформулирована так, что при поиске разумного ответа не за что зацепиться, и потому эта задача не имеет разумного решения. Однако мы сейчас покажем, что существует подход, который приводит к вполне обоснованному ответу.

Итак, поскольку никакой дополнительной информации нет, мы можем заключить, что целью исследования является описание неизвестного

распределения генеральной совокупности  $F(x) = P(X_1 < x)$ . Мы уже знаем, что в качестве оценки для  $F(x)$  мы можем использовать эмпирическую функцию распределения  $F_n(x)$ , определенную в предыдущем разделе. Оказывается, что распределение случайной величины

$$D_n = \max_x |F_n(x) - F(x)|$$

одинаково для всех непрерывных функций распределения  $F(x)$ . Более того, существует функция распределения  $K(x)$  такая, что

$$\lim_{n \rightarrow \infty} P(\sqrt{n}D_n \leq x) = K(x). \quad (2.4.7)$$

Функция  $K(x)$  называется *функцией распределения Колмогорова*, а соотношение (2.4.7) составляет суть *теоремы Колмогорова*. Используя приведенные выше рассуждения о точности и надежности статистического вывода, потребуем, чтобы объем выборки  $n$  гарантировал нужную точность приближения  $F(x)$  посредством эмпирической функции распределения  $F_n(x)$  с заданной надежностью. Более формально, зададим два числа  $\varepsilon > 0$  и  $\gamma > 0$  и потребуем, чтобы

$$P(\max_x |F_n(x) - F(x)| \leq \varepsilon) \geq \gamma. \quad (2.4.8)$$

С учетом (2.4.7) мы получаем

$$P(\max_x |F_n(x) - F(x)| \leq \varepsilon) = P(\sqrt{n}D_n \leq \sqrt{n}\varepsilon) \approx K(\sqrt{n}\varepsilon). \quad (2.4.9)$$

Пусть теперь  $\alpha \in (0, 1)$  и  $k(\alpha)$  – решение уравнения

$$K(k(\alpha)) = \alpha$$

( $k(\alpha)$  – это  $\alpha$ -квантиль распределения Колмогорова). Значения  $k(\alpha)$  для разных значений  $\alpha$  можно найти в специальных таблицах. Тогда из (2.4.8) и (2.4.9) мы получаем

$$\sqrt{n}\varepsilon \approx k(\gamma), \quad (2.4.10)$$

откуда

$$n \approx \frac{(k(\gamma))^2}{\varepsilon^2}. \quad (2.4.11)$$

Из (2.4.10) мы получаем оценку для точности:

$$\varepsilon \approx \frac{k(\gamma)}{\sqrt{n}}. \quad (2.4.12)$$

В частности, если  $\gamma = 0,95$  и  $\varepsilon = 0,001$ , то из (2.4.11) следует, что  $n \approx 1849600$ .

Если же, например,  $n = 1500$  и  $\gamma = 0,95$ , то из (2.4.12) следует, что  $\varepsilon \approx 0,0351$  ( $\varepsilon \approx 3,5\%$ ).

Особо следует отметить, что в данном случае возможная потеря точности из-за использования приближенной формулы (2.4.10) не является критичной. Дело в том, что распределение величины  $D_n$  известно (и табулировано) для каждого  $n$ . Поэтому вместо приближения (2.4.10) (которое, кстати, имеет приемлемую точность уже при  $n \geq 20$ ) всегда можно использовать точную формулу для квантилей соответствующего распределения. Критичным фактором в этой задаче является непрерывность функции распределения  $F(x)$ .

Таким образом, ответ на поставленный вопрос о том, достаточно ли  $n$  наблюдений, мы можем дать, сопоставив полученные по приведенным выше формулам значения точности и надежности на основе данного объема выборки с нашими представлениями о том, какими должны быть эти параметры.

## 2.3. Статистический анализ нормальных выборок

Этот раздел посвящен статистическим выводам о параметрах генеральной совокупности с нормальным распределением.

### 2.3.1. Распределения вероятностей, связанные с нормальным законом. Распределения хи-квадрат, Стьюдента, Фишера–Снедекора

Пусть  $X_1, X_2, \dots, X_n$  — независимые случайные величины с одним и тем же стандартным нормальным распределением, соответствующим плот-

НОСТИ

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

ОПРЕДЕЛЕНИЕ 3.1.1. Распределение случайной величины  $X = X_1^2 + X_2^2 + \dots + X_n^2$  называется *распределением хи-квадрат с  $n$  степенями свободы*.

Можно показать, что распределению хи-квадрат с  $n$  степенями свободы соответствует плотность  $f_n(x)$ , равная нулю при отрицательных  $x$ , а при  $x > 0$

$$f_n(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)},$$

где

$$\Gamma(z) = \int_0^{\infty} e^{-x} x^{z-1} dx$$

— так называемая *гамма-функция Эйлера*. Если  $z$  — целое неотрицательное число, то  $\Gamma(z+1) = z!$ .

ОПРЕДЕЛЕНИЕ 3.1.2. Пусть  $X$  и  $Y$  — независимые случайные величины, причем  $X$  имеет стандартное нормальное распределение, а  $Y$  имеет распределение хи-квадрат с  $n$  степенями свободы. Распределение случайной величины

$$Z = \frac{X}{\sqrt{Y/n}}$$

называется *распределением Стьюдента с  $n$  степенями свободы*.

Можно показать, что распределению Стьюдента с  $n$  степенями свободы соответствует плотность

$$t_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n} \Gamma(n/2) (1 + x^2/n)^{(n+1)/2}}.$$

ОПРЕДЕЛЕНИЕ 3.1.3. Пусть  $X$  и  $Y$  — независимые случайные величины, причем  $X$  имеет распределение хи-квадрат с  $n$  степенями свободы, а  $Y$  имеет распределение хи-квадрат с  $m$  степенями свободы. Распределение случайной величины

$$Z = \frac{mX}{nY}$$

называется *распределением Фишера–Снедекора с  $n$  и  $m$  степенями свободы*.

Можно показать, что распределению Фишера–Снедекора с  $n$  и  $m$  степенями свободы соответствует плотность  $p_{n,m}(x)$ , равная нулю при отрицательных  $x$ , а при  $x > 0$

$$p_{n,m}(x) = \left(\frac{n}{m}\right)^{n/2} \frac{\Gamma((n+m)/2)x^{n/2-1}}{\Gamma(n/2)\Gamma(m/2)(1+nx/m)^{(n+m)/2}}.$$

Распределения хи-квадрат, Стьюдента и Фишера–Снедекора табулированы для всевозможных значений их параметров (степеней свободы).

### 2.3.2. Статистические выводы о параметрах нормального распределения

Предположим, что имеется независимая однородная выборка  $X_1, \dots, X_n$  из генеральной совокупности, распределенной по нормальному закону с параметрами  $a$  и  $\sigma^2$ , то есть  $P(X_1 < x) = \Phi((x-a)/\sigma)$ . При этом параметры  $a$  и  $\sigma^2$  неизвестны и целью статистического анализа выборки  $X_1, X_2, \dots, X_n$  являются выводы о значениях параметров  $a$  и  $\sigma^2$ . К подобной статистической задаче, например, может привести схема прямых измерений

$$X_j = a + \varepsilon_j$$

параметра  $a$ . Как уже говорилось, в силу центральной предельной теоремы можно считать, что погрешность  $\varepsilon_j$  имеет нормальное распределение с параметрами 0 (систематическая погрешность отсутствует) и  $\sigma^2$ . При этом параметр  $\sigma^2$  (дисперсия погрешности) характеризует точность измерений.

Как и ранее, выборочное среднее и несмещенную выборочную дисперсию будем обозначать  $\bar{X}_n$  и  $S_n^2$ ,

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j, \quad S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2.$$

Мы будем использовать следующий замечательный результат, известный как *теорема Фишера*.

1. Выборочное среднее  $\bar{X}_n$  имеет нормальное распределение с параметрами  $a$  и  $\sigma^2/n$ ;
2. Нормированная несмещенная выборочная дисперсия  $(n-1)S_n^2/\sigma^2$  имеет распределение хи-квадрат с  $n-1$  степенями свободы;
3. Случайные величины  $\bar{X}_n$  и  $S_n^2$  стохастически независимы.

При статистическом анализе нормальных выборок мы будем различать три ситуации, которые мы рассмотрим по порядку.

**I. Параметр  $a$  (среднее значение) известен, а параметр  $\sigma^2$  (дисперсия) неизвестен.** В схеме прямых измерений такая ситуация возникает, когда речь идет об определении точности измерительного прибора. В этом случае параметр  $\sigma^2$  может быть оценен с помощью величины

$$\bar{S}^2 = \frac{1}{n} \sum_{j=1}^n (X_j - a)^2.$$

Чтобы оценить погрешность приближения

$$\sigma^2 \approx \bar{S}^2,$$

заметим, что каждая из независимых величин  $(X_i - a)/\sigma$ ,  $i = 1, \dots, n$ , имеет стандартное нормальное распределение, а стало быть, согласно Определению 3.1.1, случайная величина

$$X = \frac{1}{\sigma^2} \sum_{j=1}^n (X_j - a)^2 = \frac{n\bar{S}^2}{\sigma^2}$$

имеет распределение хи-квадрат с  $n$  степенями свободы. Поэтому если  $\alpha$ -квантиль распределения хи-квадрат с  $n$  степенями свободы ( $0 < \alpha < 1$ ) обозначить  $\chi_n^2(\alpha)$ , то по определению  $\alpha$ -квантили для любого  $\alpha \in (0, 1/2)$  мы будем иметь

$$P \left( \chi_n^2 \left( \frac{\alpha}{2} \right) \leq \frac{n\bar{S}^2}{\sigma^2} \leq \chi_n^2 \left( 1 - \frac{\alpha}{2} \right) \right) = 1 - \alpha. \quad (3.2.1)$$



Поэтому, зафиксировав коэффициент доверия (надежность)  $\gamma (= 1 - \alpha)$ , из соотношения (3.2.1) мы заключаем, что с уверенностью  $\gamma \cdot 100\%$  можно утверждать, что

$$\frac{n\bar{S}^2}{\chi_n^2\left(\frac{1+\gamma}{2}\right)} \leq \sigma^2 \leq \frac{n\bar{S}^2}{\chi_n^2\left(\frac{1-\gamma}{2}\right)}.$$

II. Параметр  $\sigma^2$  известен, а параметр  $a$  неизвестен. В схеме прямых измерений такая ситуация возникает, когда известна точность измерительного прибора, но измеряемый параметр неизвестен. В этой ситуации наилучшим приближением для параметра  $a$ , очевидно, будет

$$a \approx \bar{X}_n. \quad (3.2.2)$$

Чтобы оценить точность приближения (3.2.2), заметим, что из первого пункта теоремы Фишера вытекает, что в рассматриваемой ситуации случайная величина

$$\frac{\sqrt{n}(\bar{X}_n - a)}{\sigma}$$

имеет стандартное нормальное распределение. Поэтому, рассуждая так же, как в разделе 11.2, мы получаем, что для любого  $\gamma \in (0, 1)$

$$P\left(\left|\frac{\sqrt{n}(\bar{X}_n - a)}{\sigma}\right| \leq u_{\frac{1+\gamma}{2}}\right) = \gamma,$$

где  $u_{\frac{1+\gamma}{2}} - \frac{1+\gamma}{2}$ -квантиль стандартного нормального распределения. Поэтому с уверенностью  $\gamma \cdot 100\%$  мы можем утверждать, что

$$\bar{X}_n - u_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq a \leq \bar{X}_n + u_{\frac{1+\gamma}{2}} \cdot \frac{\sigma}{\sqrt{n}}.$$

III. Оба параметра  $a$  и  $\sigma^2$  неизвестны. В схеме прямых измерений такая ситуация возникает, когда неизвестны ни измеряемый параметр, ни точность измерительного прибора. В этой ситуации наилучшими приближениями для параметров  $a$  и  $\sigma^2$  будут

$$a \approx \bar{X}_n, \quad \sigma^2 \approx S_n^2. \quad (3.2.3)$$

Чтобы сделать вывод о точности этих приближенных равенств, мы заметим, что из теоремы Фишера и Определения 3.1.2 вытекает, что случайная величина

$$T_n = \frac{\sqrt{n}(\bar{X}_n - a)}{S}$$

имеет распределение Стьюдента с  $n - 1$  степенями свободы. Если для  $\alpha \in (0, 1)$  мы обозначим  $\alpha$ -квантиль такого распределения через  $t_{n-1}(\alpha)$ , то для любого  $\gamma \in (0, 1)$  мы будем иметь

$$P \left( \left| \frac{\sqrt{n}(\bar{X}_n - a)}{S} \right| \leq t_{n-1} \left( \frac{1 + \gamma}{2} \right) \right) = \gamma.$$

Поэтому с уверенностью  $\gamma \cdot 100\%$  мы можем утверждать, что

$$\bar{X}_n - t_{n-1} \left( \frac{1 + \gamma}{2} \right) \cdot \frac{S}{\sqrt{n}} \leq a \leq \bar{X}_n + t_{n-1} \left( \frac{1 + \gamma}{2} \right) \cdot \frac{S}{\sqrt{n}}.$$

Обратим внимание, что мы оценили точность приближения  $a \approx \bar{X}_n$ , не привлекая никакой информации о неизвестном параметре  $\sigma^2$ . Чтобы оценить точность приближения  $\sigma^2 \approx S^2$ , воспользуемся вторым пунктом теоремы Фишера, из которого, рассуждая так же, как в ситуации II, мы приходим к следующему выводу: зафиксировав произвольный уровень доверия  $\gamma \in (0, 1)$ , с уверенностью  $\gamma \cdot 100\%$  мы можем утверждать, что

$$\frac{(n-1)S^2}{\chi_{n-1}^2 \left( \frac{1+\gamma}{2} \right)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1}^2 \left( \frac{1-\gamma}{2} \right)}.$$

При анализе нормальных выборок часто возникает задача об объединении однородных выборок. Если имеются две нормальные выборки  $X_1, \dots, X_n$  и  $X_{n+1}, \dots, X_{n+m}$ , то статистические выводы на основе объединенной выборки  $X_1, \dots, X_{n+m}$  должны быть более точны, нежели выводы на основе каждой из выборок объемов  $n$  и  $m$  в отдельности. Единственным неизменяемым условием возможности объединения выборок является их однородность: параметры обеих нормальных выборок  $X_1, \dots, X_n$  и  $X_{n+1}, \dots, X_{n+m}$  должны совпадать. Проверить, совпадают ли эти параметры, можно статистически. Рассмотрим следующую задачу.

Пусть имеются две независимые выборки:  $X_1, \dots, X_n$  и  $Y_1, \dots, Y_m$ , причем выборка  $X_1, \dots, X_n$  взята из нормальной генеральной совокупности с параметрами  $a_1$  и  $\sigma_1^2$ , а выборка  $Y_1, \dots, Y_m$  взята из нормальной генеральной совокупности с параметрами  $a_2$  и  $\sigma_2^2$ . Требуется проверить, что параметры совокупностей совпадают:  $a_1 = a_2$  и  $\sigma_1^2 = \sigma_2^2$ .

Обозначим

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j, \quad \bar{Y}_m = \frac{1}{m} \sum_{j=1}^m Y_j,$$

$$S_X^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X}_n)^2, \quad S_Y^2 = \frac{1}{m-1} \sum_{j=1}^m (Y_j - \bar{Y}_m)^2.$$

Согласно второму утверждению теоремы Фишера, величина  $(n-1)S_X^2/\sigma_1^2$  имеет распределение хи-квадрат с  $n-1$  степенями свободы, а величина  $(m-1)S_Y^2/\sigma_2^2$  имеет распределение хи-квадрат с  $m-1$  степенями свободы. Поэтому, если  $\sigma_1^2 = \sigma_2^2$ , то согласно Определению 3.1.3, величина

$$Z_{n,m} = \frac{(n-1)S_X^2}{(m-1)S_Y^2}$$

должна иметь распределение Фишера–Снедекора с  $n-1$  и  $m-1$  степенями свободы. Поэтому, чтобы проверить гипотезу о совпадении дисперсий  $\sigma_1^2 = \sigma_2^2$ , следует зафиксировать малое положительное число  $\alpha$ , интерпретируемое как вероятность практически невозможного события (следует отметить, что в реальных исследованиях принято выбирать значения  $\alpha = 0,01$  или  $\alpha = 0,05$ ), и проверить, попадает ли наблюдаемое значение величины  $Z_{n,m}$  в интервал  $[z_{n-1,m-1}(\frac{\alpha}{2}), z_{n-1,m-1}(1 - \frac{\alpha}{2})]$ , где  $z_{n-1,m-1}(\frac{\alpha}{2})$  и  $z_{n-1,m-1}(1 - \frac{\alpha}{2})$  — соответственно  $\frac{\alpha}{2}$ - и  $(1 - \frac{\alpha}{2})$ -квантили распределения Фишера–Снедекора с  $n-1$  и  $m-1$  степенями свободы. Если величина  $Z_{n-1,m-1}$  попала в указанный интервал, то гипотезу о совпадении дисперсий следует принять. В противном случае — отвергнуть. При этом вероятность ошибочного отказа от гипотезы равенства дисперсий в случае, когда эта гипотеза на самом деле верна, будет равна  $\alpha$ .

Предположим, что гипотеза о совпадении дисперсий рассматриваемых нормальных генеральных совокупностей принята, то есть  $\sigma_1^2 = \sigma_2^2 \equiv \sigma^2$ .

Тогда можно убедиться, что случайная величина  $\bar{X}_n - \bar{Y}_m$  имеет нормальное распределение с параметрами  $a_1 - a_2$  и  $(\frac{1}{n} + \frac{1}{m})\sigma^2$ . В то же время, по Определению 3.1.1 и второму пункту теоремы Фишера, случайная величина  $[(n-1)S_X^2 + (m-1)S_Y^2]/\sigma^2$  имеет распределение хи-квадрат с  $n+m-2$  степенями свободы. Таким образом, если  $a_1 = a_2$ , то по Определению 3.1.2, случайная величина

$$T_{n,m} = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \cdot \sqrt{\frac{n+m}{nm(n+m-2)}}$$

имеет распределение Стьюдента с  $n+m-2$  степенями свободы. Поэтому, чтобы проверить гипотезу о совпадении средних значений  $a_1 = a_2$ , следует зафиксировать малое положительное число  $\alpha$ , интерпретируемое как вероятность практически невозможного события, и проверить, попадает ли наблюдаемое значение величины  $T_{n,m}$  в интервал  $[-t_{n+m-2}(1-\frac{\alpha}{2}), t_{n+m-2}(1-\frac{\alpha}{2})]$ , где  $t_{n+m-2}(1-\frac{\alpha}{2}) - (1-\frac{\alpha}{2})$ -квантиль распределения Стьюдента с  $m+n-2$  степенями свободы. Если величина  $T_{n,m}$  попала в указанный интервал, то гипотезу о совпадении средних значений следует принять. В противном случае — отвергнуть. При этом вероятность ошибочного отказа от гипотезы равенства средних значений в случае, когда эта гипотеза на самом деле верна, будет равна  $\alpha$ .

Следует особо обсудить *формулировку* вывода о результатах проверки статистических гипотез во всех упомянутых выше ситуациях. Решение о том, что гипотеза отвергается, формально должно иметь вид: *имеющиеся данные противоречат гипотезе*. В то же время, в противоположной ситуации было бы ошибкой заключить, что *гипотеза верна*. Ведь фактически у нас просто нет достаточных оснований отвергнуть ее. Поэтому если мы не отвергаем гипотезу, то это *не означает, что мы соглашаемся с ней*. Наш вывод должен иметь вид: *имеющиеся данные не противоречат гипотезе*. Другими словами, рассмотренные выше критерии имеют характер необходимых, но не достаточных условий.

В заключение следует остановиться на некоторых стереотипах, связанных с нормальным законом. Нормальное распределение является аналитически удобной идеальной абстракцией, обладающей многими благоприятными свойствами. К числу таких свойств, несомненно, относится стохастическая независимость выборочного среднего и выборочной дис-

персии в нормальных выборках. Кстати, заметим, что это свойство является характеристическим, то есть нормальное распределение является единственным законом, обладающим таким свойством.

В разделе 1.9 мы уже говорили о том, что нормальное распределение во многих практических ситуациях довольно далеко от того, чтобы быть адекватной моделью происходящего. Использование нормального распределения во многих случаях как бы реализует "принцип пьяницы", который ищет монетку не там, где он ее потерял, а там, где стоит фонарь, потому что там светлее.

## 2.4. Методы построения оценок неизвестных параметров

### 2.4.1. Метод наибольшего правдоподобия

При статистическом анализе стохастических ситуаций очень полезен и эффективен принцип наибольшего правдоподобия, который в самом общем виде может быть сформулирован так: при поиске причин тех или иных событий, если нет никаких других соображений, следует исходить из наиболее правдоподобных объяснений.

Этот общий принцип реализуется в задачах математической статистики как *метод максимального правдоподобия*, математическая формализация которого выглядит следующим образом. Пусть в распоряжении исследователя имеется независимая однородная выборка  $X_1, \dots, X_n$ . Предположим, что распределение генеральной совокупности  $P(X_1 < x)$  зависит от некоторого неизвестного параметра  $\theta$ :  $P(X_1 < x) = F(x; \theta)$ . *Функция правдоподобия*  $L(\theta; X_1, \dots, X_n)$  определяется по-разному в зависимости от того, дискретно или непрерывно распределение  $F(x; \theta)$ . Если функция распределения  $F(x; \theta)$  дискретна, то есть каждая из случайных величин  $X_1, \dots, X_n$  может принимать значения  $x_1, x_2, \dots$  с вероятностями соот-

ответственно  $p(x_1; \theta)$ ,  $p(x_2; \theta)$ , ..., то определим функцию  $p(x; \theta)$  как

$$p(x; \theta) = \begin{cases} p(x_1; \theta), & \text{если } x = x_1, \\ \dots & \\ p(x_n; \theta), & \text{если } x = x_n, \\ \dots & \\ 0, & \text{если } x \notin \{x_1, x_2, \dots\}. \end{cases}$$

Если же функция распределения  $F(x; \theta)$  непрерывна, то определим функцию  $p(x; \theta)$  как плотность распределения  $F(x; \theta)$ : для любого  $y$

$$F(y; \theta) = \int_{-\infty}^y p(x; \theta) dx.$$

Для обоих этих случаев функция правдоподобия определяется как

$$L(\theta; X_1, \dots, X_n) = \prod_{j=1}^n p(X_j; \theta).$$

Функция правдоподобия является функцией аргумента  $\theta$  и имеет смысл вероятности наблюдать то, что фактически наблюдается. Поскольку, как правило, происходят события, вероятность которых велика, то принцип наибольшего правдоподобия сводится к следующим рассуждениям. Если мы наблюдаем выборку  $X_1, \dots, X_n$ , то событие, результатом которого является выборка, имеет большую вероятность. Стало быть, наиболее правдоподобным значением параметра  $\theta$  является то, при котором эта вероятность максимальна. Таким образом, искать приближение  $\hat{\theta}$  к неизвестному "истинному" значению параметра  $\theta$  следует, исходя из условия

$$L(\hat{\theta}; X_1, \dots, X_n) \geq L(\theta; X_1, \dots, X_n)$$

для всех возможных значений параметра  $\theta$ .

**ЗАДАЧА 4.1.1.** Пусть имеется выборка  $X_1, \dots, X_n$  из совокупности с нормальным распределением с параметрами  $m$  и единица, то есть

$$P(X_1 < x) = \Phi(x - m), \quad -\infty < x < \infty,$$

причем неизвестный параметр  $m$  может принимать только целые значения. Найти наиболее правдоподобное значение параметра  $m$ .

Функция правдоподобия для этой задачи имеет вид

$$\begin{aligned} L(m; X_1, \dots, X_n) &= \prod_{j=1}^n \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(X_j - m)^2}{2} \right\} \\ &= \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^n (X_j - m)^2 \right\}. \end{aligned}$$

Наибольшее значение функции правдоподобия будем искать, сравнивая отношение ее значений при соседних значениях  $m$  с единицей. Имеем

$$\begin{aligned} &\frac{L(m; X_1, \dots, X_n)}{L(m+1; X_1, \dots, X_n)} \\ &= \frac{\exp \left\{ -\frac{1}{2} \sum_{j=1}^n (X_j - m)^2 \right\}}{\exp \left\{ -\frac{1}{2} \sum_{j=1}^n (X_j - (m+1))^2 \right\}} \\ &= \exp \left\{ -\frac{1}{2} \left[ \sum_{j=1}^n (X_j - m)^2 - \sum_{j=1}^n (X_j - (m+1))^2 \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[ \sum_{j=1}^n X_j^2 - 2m \sum_{j=1}^n X_j + nm^2 - \sum_{j=1}^n X_j^2 \right. \right. \\ &\quad \left. \left. - 2(m+1) \sum_{j=1}^n X_j - nm^2 - 2nm - n \right] \right\} \\ &= \exp \left\{ n \left( m + \frac{1}{2} \right) - \sum_{j=1}^n X_j \right\}. \end{aligned}$$

Последнее выражение не меньше единицы тогда и только тогда, когда

$$n \left( m + \frac{1}{2} \right) - \sum_{j=1}^n X_j \geq 0$$

или, что то же самое,

$$m \geq \bar{X}_n - \frac{1}{2}.$$

Таким образом, наиболее правдоподобное значение  $\hat{m}$  параметра  $m$  равно  $[\bar{X}_n - \frac{1}{2}]$ , если  $L([\bar{X}_n - \frac{1}{2}]; X_1, \dots, X_n) \geq L([\bar{X}_n - \frac{1}{2}] + 1; X_1, \dots, X_n)$ , а если  $L([\bar{X}_n - \frac{1}{2}]; X_1, \dots, X_n) < L([\bar{X}_n - \frac{1}{2}] + 1; X_1, \dots, X_n)$ , то оно равно  $[\bar{X}_n - \frac{1}{2}] + 1$ . Здесь символом  $[x]$  мы обозначили наибольшее из всех целых чисел, не превосходящих  $x$ .

## 2.4.2. Метод моментов

Как и в предыдущем разделе, предположим, что в распоряжении исследователя имеется независимая однородная выборка  $X_1, \dots, X_n$ . Предположим, что распределение генеральной совокупности  $P(X_1 < x)$  зависит от некоторых неизвестных параметров  $\theta_1, \dots, \theta_r$ :  $P(X_1 < x) = F(x; \theta_1, \dots, \theta_r)$ . Если случайная величина  $X_1$  дискретна и принимает значения  $x_1, x_2, \dots$ , то от неизвестных параметров  $\theta_1, \dots, \theta_r$  зависят вероятности  $P(X_1 = x_i) = p_i(\theta_1, \dots, \theta_r)$ ,  $i = 1, 2, \dots$ . Если случайная величина  $X_1$  непрерывна, то от неизвестных параметров  $\theta_1, \dots, \theta_r$  зависит ее плотность  $p(x) = p(x; \theta_1, \dots, \theta_r)$ . Ясно, что в обеих ситуациях числовые характеристики случайной величины  $X_1$  также зависят от неизвестных параметров  $\theta_1, \dots, \theta_r$ . Идея метода моментов заключается в приравнивании теоретических числовых характеристик случайной величины  $X_1$  их эмпирическим (выборочным) аналогам и решении получаемых при этом (систем) уравнений относительно параметров  $\theta_1, \dots, \theta_r$ .

Чаще всего в качестве числовых характеристик генеральной совокупности (то есть распределения случайной величины  $X_1$ ) рассматриваются ее моменты  $EX_1^k = \mu_k(\theta_1, \dots, \theta_r)$ , где

$$\mu_k(\theta_1, \dots, \theta_r) = \begin{cases} \sum_{i=1}^{\infty} x_i^k p_i(\theta_1, \dots, \theta_r), & \text{если распределение случайной величины } X_1 \\ & \text{дискретно,} \\ \int_{-\infty}^{\infty} x^k p(x; \theta_1, \dots, \theta_r) dx, & \text{если распределение случайной величины } X_1 \\ & \text{непрерывно.} \end{cases}$$



В таком случае уравнения метода моментов принимают вид

$$\mu_k(\theta_1, \dots, \theta_r) = \frac{1}{n} \sum_{j=1}^n X_j^k.$$

При этом, как правило, берутся первые  $r$  уравнений такого вида:  $k = 1, \dots, r$ . Однако такой выбор необязателен. Выбор уравнений диктуется вычислительной эффективностью.

**ПРИМЕР 4.2.1.** Пусть выборка  $X_1, \dots, X_n$  взята из нормальной генеральной совокупности,  $P(X_1 < x) = \Phi((x - a)/\sigma)$ . Параметры  $a$  и  $\sigma^2$  неизвестны. Тогда

$$EX_1 = a, \quad EX_1^2 = a^2 + \sigma^2,$$

и система уравнений метода моментов принимает вид

$$\begin{cases} a = \frac{1}{n} \sum_{j=1}^n X_j \equiv \bar{X}_n, \\ a^2 + \sigma^2 = \frac{1}{n} \sum_{j=1}^n X_j^2. \end{cases}$$

Решение этой системы имеет вид

$$\begin{cases} \tilde{a} = \bar{X}_n, \\ \tilde{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2 \equiv \tilde{S}_n^2. \end{cases}$$

В тех случаях, когда моменты генеральной совокупности не определены (не существуют), в методе моментов можно брать другие числовые характеристики, что иллюстрирует следующий пример.

**ПРИМЕР 4.2.2.** Пусть выборка  $X_1, \dots, X_n$  взята из генеральной совокупности, распределенной по закону Коши с плотностью

$$p(x) = \frac{1}{\pi(1 + (x - a)^2)}.$$

Параметр  $a$  неизвестен. Как мы видели, у такого распределения математическое ожидание (первый момент) не определено, а стало быть, нет и моментов старших порядков. Однако легко видеть, что  $\text{med } X_1 = a$ . Поэтому соотношение

$$a = \bar{m}_n,$$

где  $\bar{m}_n$  – выборочная медиана, можно рассматривать как реализацию метода моментов, сводящуюся к оценке  $\tilde{a} = \bar{m}_n$ .

## 2.5. Проверка согласия экспериментальных данных с теоретической моделью распределения генеральной совокупности

Степень адекватности математической модели, описывающей ту или иную стохастическую ситуацию, можно проверить с помощью так называемых критериев согласия. В данном разделе мы рассмотрим два таких критерия – критерий согласия хи-квадрат и критерий согласия Колмогорова.

### 2.5.1. Критерий согласия хи-квадрат

Критерий согласия хи-квадрат использует сгруппированные данные подобно тому, как это было сделано при рассмотрении гистограммы в разделе 2.2.3.

Пусть имеется независимая однородная выборка  $X_1, \dots, X_n$  из генеральной совокупности с неизвестным распределением  $F(x) = P(X_1 < x)$ . Предположим, что для описания вида распределения  $F(x)$  сформулирована модель  $F_0(x)$ . Проверка адекватности этой модели по выборке  $X_1, \dots, X_n$  эквивалентна проверке гипотезы о том, что  $F(x) \equiv F_0(x)$ . Критерий согласия хи-квадрат как раз и предназначен для проверки этой гипотезы. Заключение о справедливости указанной выше гипотезы делается на основе сравнения *статистики хи-квадрат* с соответствующим пороговым значением. Опишем эту процедуру подробнее.

Пусть  $a$  и  $b$  – числа, удовлетворяющие неравенствам  $a < X_{(1)}$ ,  $b > X_{(n)}$  (напомним, что  $X_{(1)}$  – наименьший элемент выборки, а  $X_{(n)}$

— наибольший). Зададим целое положительное число  $k$  и разобьем интервал  $[a, b]$  на  $k$  равных непересекающихся частей. Обозначим полученные подынтервалы символами  $\Delta_j$ ,  $j = 1, \dots, k$  (в формальной записи  $\Delta_j = [a + (j-1)\delta, a + j\delta)$ ,  $j = 1, \dots, k$ , где  $\delta = (b-a)/k$ ). Пусть  $v_j$  — число тех элементов выборки  $X_1, X_2, \dots, X_n$ , которые попали в интервал  $\Delta_j$ . С помощью модельной (гипотетической) функции распределения  $F_0(x)$  определим числа  $p_j^{(0)}$ , положив  $p_j^{(0)} = F_0(j\delta) - F_0((j-1)\delta)$ ,  $j = 1, \dots, k$  (другими словами,  $p_j^{(0)}$  — это вероятность того, что случайно взятый элемент генеральной совокупности попадает в интервал  $\Delta_j$ , вычисленная в предположении о том, что  $F(x) \equiv F_0(x)$ ). *Статистикой хи-квадрат* называется величина

$$X^2 = \sum_{j=1}^k \frac{(v_j - np_j^{(0)})^2}{np_j^{(0)}}.$$

В терминах выборочных частот  $\tilde{p}_j = v_j/n$ , введенных в разделе 11.3, статистика хи-квадрат может быть записана в виде

$$X^2 = n \sum_{j=1}^k \frac{(\tilde{p}_j - p_j^{(0)})^2}{p_j^{(0)}}.$$

Статистика хи-квадрат характеризует суммарное отклонение выборочных (наблюдаемых) частот от теоретических (гипотетических). По тому, насколько велика эта статистика, можно сделать вывод о неадекватности или адекватности (согласии) теоретического распределения с экспериментальными данными. Чем эта статистика больше, тем менее адекватна теоретическая модель. А именно, можно показать, что верна так называемая *теорема Пирсона*, устанавливающая, что, если гипотеза  $F(x) \equiv F_0(x)$  верна, то при неограниченно увеличивающемся объеме выборки ( $n \rightarrow \infty$ ) распределение случайной величины  $X^2$ , введенной выше, все больше и больше приближается с распределением хи-квадрат с  $k - 1$  степенями свободы (см. раздел 2.3.1).

Зафиксируем малое положительное число  $\alpha$  (на практике традиционно выбирается  $\alpha = 0,01$  или  $\alpha = 0,05$ ). Пусть, как и ранее,  $\chi_{k-1}^2(1 - \alpha)$  —  $(1 - \alpha)$ -квантиль распределения хи-квадрат с  $k - 1$  степенями свободы.

Процедура проверки указанной гипотезы с помощью критерия хи-квадрат заключается в следующем. Значение статистики хи-квадрат  $X^2$  сравнивается с порогом  $\chi_{k-1}^2(1-\alpha)$ . Если  $X^2 > \chi_{k-1}^2(1-\alpha)$ , то гипотеза о том, что  $F(x) \equiv F_0(x)$  отвергается. Если же  $X^2 \leq \chi_{k-1}^2(1-\alpha)$ , то делается вывод о том, что экспериментальные данные не противоречат выдвинутой гипотезе, то есть согласуются с ней. При этом вероятность ошибочного отклонения гипотезы  $F(x) \equiv F_0(x)$ , если она на самом деле верна, равна  $\alpha$ .

На практике критерий согласия хи-квадрат можно применять, если наименьшая из величин  $np_1^{(0)}, \dots, np_k^{(0)}$  не меньше пяти.

Критерий согласия хи-квадрат можно применять и тогда, когда сформулированная гипотеза описывает распределение генеральной совокупности не однозначно, а с точностью до некоторых неизвестных параметров:  $F(x) \equiv F_0(x; \theta_1, \dots, \theta_r)$ . В этом случае необходимо предварительно оценить неизвестные параметры и вычислить значения  $p_j^{(0)}$  как  $p_j^{(0)} = F_0(j\delta; \hat{\theta}_1, \dots, \hat{\theta}_r) - F_0((j-1)\delta; \hat{\theta}_1, \dots, \hat{\theta}_r)$ ,  $j = 1, \dots, k$ . При этом, однако, предельным распределением случайной величины  $X^2$  будет распределение хи-квадрат с  $k - r - 1$  степенями свободы, и стало быть, величину  $X^2$  надо сравнивать с  $(1-\alpha)$ -квантилью именно этого распределения.

При использовании критерия согласия хи-квадрат надо, однако, принимать во внимание следующие обстоятельства.

- а) Критерий хи-квадрат имеет асимптотический характер: только при "бесконечно большом" объеме выборки распределение статистики  $X^2$  совпадает с распределением хи-квадрат. Точность же приближения истинного (допредельного) распределения этой статистики предельным распределением хи-квадрат, вообще говоря, неизвестна. Поэтому истинная вероятность ошибки, совершаемой при отказе от верной гипотезы, не совпадает с  $\alpha$ .
- б) Более того, если проверяемая гипотеза неоднозначно задает распределение генеральной совокупности, то предельное распределение статистики  $X^2$  будет совпадать с распределением хи-квадрат (с соответствующим числом степеней свободы), только если неизвестные параметры оцениваются с помощью так называемого по-

линомиального метода максимального правдоподобия. По крайней мере, сходимость распределения статистики  $X^2$  к распределению хи-квадрат доказана только для такого случая.

- с) Поскольку базой для вычисления статистики критерия согласия хи-квадрат являются сгруппированные данные типа гистограммы, конкретное значение этой статистики существенно зависит от того, как сгруппированы данные, то есть от числа  $k$  интервалов и выбора точек  $a$  и  $b$ .
- д) Критерий согласия хи-квадрат позволяет сделать вывод о том, что данные *не согласуются* с той или иной гипотезой. Однако с его помощью нельзя сделать вывода о том, что данные *согласуются* с конкретной гипотезой. Можно лишь сделать вывод о том, что данные ей *не противоречат*.
- е) Чрезмерно малые (близкие к нулю) значения статистики  $X^2$ , на основании которых формально надо делать вывод о том, что данные не противоречат проверяемой гипотезе, свидетельствуют о нарушении условий независимости или однородности наблюдений, как если бы при многократном воспроизведении серий, скажем, по четыре испытания Бернулли с вероятностью успеха в одном испытании, скажем, равной  $\frac{1}{4}$ , каждый раз наблюдался бы ровно один успех.

### 2.5.2. Критерий согласия Колмогорова

Если теоретическая (гипотетическая) функция распределения генеральной совокупности непрерывна, то адекватность выбранной модели можно проверять с помощью критерия согласия Колмогорова. Он основан на сравнении *статистики Колмогорова*, аналогичной случайной величине  $D_n$ , которая была введена в разделе 2.2.4 (Задача 2.4.2), с соответствующим пороговым значением. Опишем эту процедуру подробнее.

Пусть  $F_n(x)$  — эмпирическая функция распределения, построенная по выборке  $X_1, \dots, X_n$  так, как это было описано в разделе 2.2.3. Пусть в отношении (неизвестного) распределения генеральной совокупности  $F(x)$  выдвинута гипотеза  $F(x) \equiv F_0(x)$ . Определим *статистику Колмогорова*

$D_n^{(0)}$  как

$$D_n^{(0)} = \max_x |F_n(x) - F_0(x)|.$$

Значение этой статистики, как несложно видеть, можно вычислить по формуле

$$D_n^{(0)} = \max_{j=1, \dots, n} |F_n(X_{(j)}) - F_0(X_{(j)})|.$$

Статистика Колмогорова характеризует отклонение выборочной (эмпирической) функции распределения от теоретической (гипотетической). По тому, насколько велика эта статистика, можно сделать вывод о неадекватности или адекватности теоретического распределения (его согласии с экспериментальными данными). Чем эта статистика больше, тем менее адекватна теоретическая модель. А именно, можно показать, что в силу соотношения (2.4.7), если верна гипотеза  $F(x) \equiv F_0(x)$ , то при неограниченном увеличении объема выборки ( $n \rightarrow \infty$ ) распределение величины  $\sqrt{n}D_n^{(0)}$  все больше и больше сближается с функцией распределения Колмогорова  $K(x)$ .

Поэтому, если мы зафиксируем произвольное малое положительное число  $\alpha$  и, как и ранее,  $(1 - \alpha)$ -квантиль распределения Колмогорова обозначим через  $k(1 - \alpha)$ , то указанная гипотеза отклоняется, если  $\sqrt{n}D_n^{(0)} > k(1 - \alpha)$ . Если же  $\sqrt{n}D_n^{(0)} \leq k(1 - \alpha)$ , то делается вывод о том, что экспериментальные данные не противоречат выдвинутой гипотезе, то есть согласуются с ней. При этом вероятность ошибочного отклонения гипотезы  $F(x) \equiv F_0(x)$ , если она на самом деле верна, равна  $\alpha$ .

Критерий согласия Колмогорова можно применять только тогда, когда выдвинутая гипотеза однозначно описывает непрерывное распределение генеральной совокупности, то есть не содержит никаких неизвестных параметров. Например, с его помощью *нельзя* проверять гипотезу "распределение генеральной совокупности нормально", поскольку нормальных распределений бесконечно много и каждое из них определяется парой параметров, но *можно* проверить гипотезу "распределение генеральной совокупности нормально с параметрами 0 и 1". При подстановке оценок параметров, построенных по выборке, вместо неизвестных параметров гипотетической функции распределения в статистику Колмогорова  $D_n^{(0)}$  *изменяется ее предельное распределение*, которое становится зависящим от конкретного вида гипотетической функции распределения и спосо-

ба получения оценок. А это означает, что истинная вероятность ошибки будет отличаться от требуемого значения (оставаясь, вообще говоря, неизвестной).

## 2.6. Исследование стохастических зависимостей

### 2.6.1. Выборочный коэффициент корреляции

Предположим, что изучаются  $n$  объектов, причем для каждого из них независимо от других объектов одновременно измеряются два параметра (например, для каждого из  $n$  человек измеряются рост и вес). Результаты измерения представляют собой набор  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ , где  $X_i$  — значение первого параметра, назовем его, скажем,  $X$ , зафиксированное для  $i$ -го объекта,  $Y_i$  — значение второго параметра, назовем его, скажем,  $Y$ , у  $i$ -го объекта. В этом случае говорят, что наблюдения (измерения) представляют собой *двумерную* выборку. Степень взаимозависимости наблюдаемых параметров может характеризовать коэффициент корреляции (см. раздел 1.5.2). В качестве *выборочного коэффициента корреляции* признаков  $X$  и  $Y$  естественно взять величину

$$r_{XY} = \frac{\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y})}{n\tilde{S}_X\tilde{S}_Y} = \frac{\sum_{j=1}^n (X_j - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2}}$$

$$= \frac{\sum_{j=1}^n X_j Y_j - n\bar{X}\bar{Y}}{\sqrt{\sum_{j=1}^n (X_j - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2}},$$

где

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j, \quad \bar{Y} = \frac{1}{n} \sum_{j=1}^n Y_j,$$

$$\tilde{S}_X^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2, \quad \tilde{S}_Y^2 = \frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})^2.$$

## 2.6.2. Линейная регрессия. Метод наименьших квадратов

При исследовании зависимостей часто сталкиваются со следующей задачей, являющейся развитием проблем, для решения которых был введен коэффициент корреляции. Пусть в отношении взаимосвязи между признаками (параметрами)  $X$  и  $Y$  сделано предположение, что их математические ожидания связаны линейной зависимостью:

$$EY = aEX + b. \quad (6.2.1)$$

Зависимость типа (6.2.1), связывающая не сами наблюдения  $(X_j, Y_j)$ , а их ожидаемые величины, называется *регрессионной*. При этом в случае (6.2.1) говорят о *линейной регрессии  $Y$  на  $X$* .

Соотношению (6.2.1) соответствует связь между наблюдениями  $(X_j, Y_j)$ , имеющая вид

$$Y_j = aX_j + b + \varepsilon_j, \quad (6.2.2)$$

где  $\varepsilon_j$ ,  $j = 1, \dots, n$  — случайные величины с нулевым математическим ожиданием:  $E\varepsilon_j = 0$ .

Параметры  $a$  и  $b$  моделей (6.2.1) и (6.2.2), как правило, неизвестны и подлежат определению (оцениванию) с помощью статистических процедур.

Эффективный способ оценивания параметров  $a$  и  $b$  заключается в минимизации суммы квадратов отклонений величин  $Y_j$  от  $aX_j + b$ , то есть в выборе в качестве оценок для  $a$  и  $b$  таких функций  $\hat{a}$  и  $\hat{b}$  от  $X_j$  и  $Y_j$ ,  $j = 1, \dots, n$ , для которых выполнено соотношение

$$\sum_{j=1}^n (Y_j - \hat{a}X_j - \hat{b})^2 \leq \sum_{j=1}^n (Y_j - aX_j - b)^2$$

для любых допустимых значений  $a$  и  $b$ . Такой подход приводит к следующим приближениям. Во введенных выше обозначениях мы имеем

$$a \approx \hat{a} = \frac{\sum_{j=1}^n X_j Y_j - \frac{1}{n} \sum_{j=1}^n X_j \sum_{j=1}^n Y_j}{\sum_{j=1}^n X_j^2 - \frac{1}{n} \left( \sum_{j=1}^n X_j \right)^2} = \frac{\frac{1}{n} \sum_{j=1}^n X_j Y_j - \bar{X} \cdot \bar{Y}}{\frac{1}{n} \sum_{j=1}^n X_j^2 - (\bar{X})^2},$$

$$b \approx \hat{b} = \bar{Y} - \hat{a}\bar{X}.$$



Несложно видеть, что

$$\hat{a} = r_{XY} \cdot \frac{\tilde{S}_Y}{\tilde{S}_X}.$$

## Рекомендуемая литература

1. С. А. Айвазян, И. С. Енюков и Л. Д. Мешалкин. *Прикладная статистика. Основы моделирования и первичная обработка данных*. "Финансы и статистика", Москва, 1983.
2. С. А. Айвазян, И. С. Енюков и Л. Д. Мешалкин. *Прикладная статистика. Исследование зависимостей*. "Финансы и статистика", Москва, 1985.
3. С. А. Айвазян, В. М. Бухштабер, И. С. Енюков и Л. Д. Мешалкин. *Прикладная статистика. Классификация и снижение размерности*. "Финансы и статистика", Москва, 1989.
4. С. А. Айвазян и В. С. Мхитарян. *Прикладная статистика и основы экономики*. Издательское объединение ЮНИТИ, Москва, 1998.
5. Б. В. Гнеденко и А. Я. Хинчин. *Элементарное введение в теорию вероятностей*. 9-е изд. "Наука", Москва, 1982.
6. Г. Кимбл. *Как правильно пользоваться статистикой*. "Финансы и статистика", Москва, 1982.
7. А. И. Китайгородский. *Невероятно – не факт*. "Молодая гвардия", Москва, 1978.
8. Б. А. Севастьянов. *Вероятностные модели*. "Наука", Москва, 1992.
9. В. Феллер. *Введение в теорию вероятностей и ее приложения*. Т. 1. "Мир", Москва, 1984.
10. В. Феллер. *Введение в теорию вероятностей и ее приложения*. Т. 2. "Мир", Москва, 1984.

ЗБС  
К-682

В.Ю. Королев

# ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УЧЕБНИК

Учебник содержит изложение основных разделов теории вероятностей и математической статистики. Основное внимание уделено особенностям практического применения описываемых моделей и методов теории вероятностей и математической статистики. Учебник ориентирован на формирование у читателя понимания возможностей теории вероятностей и математической статистики и ошибок, происходящих при неправильном применении вероятностных моделей и статистических процедур. Наряду с разделами, традиционно включаемыми в стандартные курсы теории вероятностей и математической статистики, содержатся некоторые специальные разделы, в частности, связанные со свойствами равномерного распределения и его ролью. Для понимания основных частей курса достаточно хорошего знания математики в рамках школьной программы.

Для студентов, обучающихся по экономическим, инженерным, психологическим и другим специальностям, использующим методы теории вероятностей и процедуры статистического анализа в качестве инструмента для решения конкретных практических задач.

Научная библиотека МГУ



62005083



• ПРОСПЕКТ •

Издательство «ПРОСПЕКТ»  
111020, Москва, ул. Боровая, д. 7, стр. 4  
(095) 967-1572  
e-mail: mail@prospekt.org  
www.prospekt.org



9 785482 002742