Министерство науки и высшего образования Российской Федерации
Ярославский государственный университет им. П. Г. Демидова

А. Н. Максименко, Ю. В. Богомолов

Теория вероятностей

Учебное пособие

2-е издание, переработанное и дополненное

Ярославль
ЯрГУ
2019

УДК 519.21(075.8)
ББК В171я73
М17

Рекомендовано
Редакционно-издательским советом университета
в качестве учебного издания. План 2019 года

Рецензенты:

Е. И. Смирнов, д-р пед. наук, канд. физ.-мат. наук, проф.;
кафедра высшей математики РГАТУ им. П. А. Соловьева

Максименко, Александр Николаевич.
М17 Теория вероятностей : учебное пособие / А. Н. Максименко, Ю. В. Богомолов ; Яросл. гос. ун-т им. П. Г. Демидова. – 2-е изд., перераб. и доп. – Ярославль : ЯрГУ, 2019. – 128 с.

ISBN 978-5-8397-1171-6

Пособие написано на основе курсов лекций, читаемых авторами на факультете ИВТ. Обсуждение ключевых понятий дополняется большим количеством примеров и задач. Особое внимание уделяется задачам, связанным с испытаниями Бернулли. Им посвящена отдельная глава.

Предназначено для студентов, обучающихся по дисциплинам «Теория вероятностей» и «Теория вероятностей и математическая статистика».

УДК 519.21(075.8)
ББК В171я73

ISBN 978-5-8397-1171-6

Оглавление

Введение **5**

1. Случайные события **6**
1.1. Вероятностные пространства 6
1.1.1. Классическое определение вероятности 11
1.2. Комбинаторный метод вычисления вероятностей 12
1.2.1. Размещения с повторениями 13
1.2.2. Размещения без повторений 14
1.2.3. Сочетания 15
1.2.4. Разбиение множества на несколько подмножеств 16
1.2.5. Сочетания с повторениями 17
1.2.6. Гипергеометрическое распределение 18
1.3. Геометрические вероятности 19
1.4. Аксиоматика теории вероятностей 22
1.4.1. Основные свойства вероятности 24
1.5. Условная вероятность и независимость 28
1.5.1. Условная вероятность 28
1.5.2. Независимость 31
1.6. Формула полной вероятности и формула Байеса 34
1.6.1. Формула полной вероятности 34
1.6.2. Формула Байеса 36

2. Испытания Бернулли **39**
2.1. Формула Бернулли 39
2.2. Приближение Пуассона 42
2.3. Локальная теорема Муавра–Лапласа 46

2.4. Интегральная теорема Муавра–Лапласа 53
2.4.1. Вероятность отклонения относительной частоты от вероятности успеха 56

3. Случайные величины 58
3.1. Дискретные случайные величины 59
3.1.1. Примеры дискретных случайных величин 59
3.2. Совместные распределения 61
3.3. Математическое ожидание 64
3.3.1. Основные свойства математического ожидания . 67
3.3.2. Примеры вычисления математических ожиданий 70
3.4. Дисперсия дискретной случайной величины 71
3.4.1. Свойства дисперсии 73
3.4.2. Примеры вычисления дисперсии 75
3.5. Ковариация и коэффициент корреляции 77
3.5.1. Ковариация 77
3.5.2. Коэффициент корреляции 80
3.6. Функция распределения 83
3.7. Плотность распределения 90
3.7.1. Локальный смысл плотности 91
3.7.2. Примеры абсолютно непрерывных случайных величин 93
3.8. $\mathsf{M}(X)$ и $\mathsf{D}(X)$ для непрерывной случайной величины . 97
3.9. Преобразование случайных величин 101
3.9.1. Закон распределения зависимой случайной величины 101
3.9.2. Суммирование независимых случайных величин 106
3.9.3. Генерирование случайных величин 108
3.10. Закон больших чисел 113
3.10.1. Неравенства Чебышёва 113
3.10.2. Закон больших чисел 118

Литература 124

Приложения 125

Введение

Ключевое отличие теории вероятностей от большинства других математических дисциплин состоит в том, что она занимается изучением недетерминированных, случайных явлений. Здесь следует сразу сделать оговорку о том, что разделение явлений на детерминированные и случайные весьма условно. Например, состояние погоды в любой точке нашей планеты, судя по тем прогнозам метеорологов, которые мы наблюдаем каждый день, носит случайный характер. Тем не менее в среднем предсказания метеорологов сбываются, что говорит о детерминированности погоды. Проблема лишь в том, что множество факторов, определяющих погоду, слишком велико, чтобы можно было скрупулезно учесть их все. Поэтому теория вероятностей оказывается полезной везде, где по тем или иным причинам не удается учесть все множество факторов, оказывающих влияние на изучаемый процесс или явление.

Для успешного применения методов теории вероятностей (как, впрочем, и любых других математических методов) особенно важен правильный выбор математической модели случайного явления, опирающийся, прежде всего, на опыт проведения ряда однотипных экспериментов. Как правило, теория вероятностей оказывается бесполезной для предсказания результата уникального эксперимента.

Известно, что задачи по теории вероятностей одни из самых коварных. Даже знаменитые математики прошлых веков, прославившиеся своими достижениями в других областях, легко совершали ошибки при решении элементарных задач по теории вероятностей. Похоже, что единственный способ преуспеть в изучении теории вероятностей — это приобрести значительный опыт в решении задач.

Глава 1

Случайные события

1.1. Вероятностные пространства

Начнем с примеров.

Опыт показывает, что при («хорошем») бросании монеты она с равной вероятностью падает вверх цифрой (решкой) или гербом. Опыт также показывает, что в некоторых, весьма редких, случаях монета встает на ребро. Далее вероятность того, что монета встанет на ребро, будем полагать равной нулю. Во-первых, потому, что это — редкое событие. А во-вторых, для упрощения соответствующей математической модели.

Другой, очень удобный для иллюстраций, пример — игральная кость (кубик с пронумерованными от 1 до 6 гранями). Далее мы будем рассматривать идеальную игральную кость, при бросании которой с равной вероятностью появляется одна из шести цифр.

Одним из довольно простых, но интересных экспериментов является опыт, в котором монета бросается до тех пор, пока не выпадет герб. Теоретически число бросаний может оказаться любым. Практически же довольно трудно добиться даже 10-кратного появления решки. Если же такая ситуация все-таки происходит, то обычно это говорит о «неслучайности» бросаний.

Стрельба по мишени, в отличие от предыдущих примеров, является более сложным экспериментом. Во-первых, вероятности попадания в разные кольца мишени могут довольно сильно отличаться друг от друга в зависимости от условий эксперимента. Во-вторых, при проведении этого эксперимента можно отказаться от традиционного способа подсчета очков, заменив его более точным методом — измерением расстояния от центра мишени до места попадания.

Этот список каждый из читателей может легко дополнить своими примерами.

Обратим внимание, что каждый случайный эксперимент обладает некоторым заранее фиксированным набором (множеством) исходов.

Определение 1.1. *Пространством элементарных исходов* называется множество Ω, содержащее все возможные исходы данного случайного эксперимента, из которых в эксперименте происходит ровно один. Сами исходы являются элементами этого множества и обычно обозначаются греческими буквами ω или же ω_i.

Понятие элементарного исхода (или события) является относительным. Выбор того, что считать элементарным, а что — не элементарным, зависит не только от условий эксперимента и постановки задачи, но и от предпочтений того, кто будет эту задачу решать. Проиллюстрируем сказанное примерами.

Пример 1. При бросании игральной кости возможны, например, следующие представления Ω:

1. $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$.
2. $\Omega = \{\omega_1, \omega_2\}$, где $\omega_1 = \{$выпало четное число очков$\}$, $\omega_2 = \{$выпало нечетное число очков$\}$.

Заметим, что второе представление явно бесполезно, например, при рассмотрении события $A = \{$выпало не более двух очков$\}$.

Пример 2. При стрельбе по круглой мишени (для удобства будем рассматривать только случаи попадания в мишень) элементарные исходы можно различать по:

1) числу выбитых очков, тогда $\Omega = \{\omega_0, \omega_1, \ldots, \omega_n\}$, где $\omega_i = \{$выбито i очков$\}$, $i = 0, 1, \ldots, n$;

2) расстоянию от точки попадания до центра мишени, тогда Ω — это отрезок $[0, R]$, где R — радиус мишени;

3) координатам точки пробоины, тогда Ω — множество всех точек мишени.

Пример 3. Рассмотрим эксперимент, состоящий в подбрасывании монеты до тех пор, пока не выпадет герб. Исходы эксперимента будем различать по числу понадобившихся бросаний монеты. То есть $\omega_i = \{$при первых $i-1$ бросках появилась цифра, а при i-м бросании — герб$\}$, где $i \in \mathbb{N}$. Тогда множество Ω равномощно множеству всех натуральных чисел.

Определение 1.2. Пространство элементарных исходов называется *дискретным*, если оно конечно либо счетно.

В ближайшее время мы ограничимся рассмотрением дискретных пространств, более удобных для первого знакомства с основами теории вероятностей.

Определение 1.3. Подмножества пространства элементарных исходов называются *событиями*. Говорят, что событие произошло, если произошел один из исходов, входящих в это событие.

Во многих случаях корректность решения задачи напрямую зависит от способа выбора пространства элементарных исходов.

Задача 4 (парадокс Кардано[1]). При бросании двух игральных костей рассмотрим два события: $A = \{$сумма выпавших очков равна $9\}$, $B = \{$сумма выпавших очков равна $10\}$. Событие A состоит из двух исходов: $4+5$ и $3+6$. Событие B также состоит из двух исходов: $5+5$ и $4+6$. Почему при реальном проведении эксперимента событие A появляется чаще, чем B? Почему при бросании трех игральных костей событие B появляется чаще, чем A, хотя и то и другое состоят из шести исходов?

В качестве события может быть также рассмотрено все пространство элементарных исходов Ω или же его подмножество, не включающее в себя ни одного исхода (пустое множество).

[1] Джероламо Кардано (1501–1576) — итальянский математик, инженер, философ, медик и астролог.

Определение 1.4. *Достоверным* называется событие, включающее в себя все исходы из Ω (т. е. оно произойдет при любом исходе). *Невозможным* называется событие, не содержащее ни одного элементарного исхода (т. е. оно не произойдет никогда).

Отношения между событиями соответствуют отношениям между множествами содержащихся в этих событиях исходов. Эта связь отражена в таблице 1.1.

Таблица 1.1

Подмножества и события

Символ	Терия множеств	Теория вероятностей
Ω	(универсальное) множество	пространство элементарных исходов, или достоверное событие
$\omega \in \Omega$	элемент	элементарный исход
$A \subseteq \Omega$	множество	событие
$\varnothing$	пустое множество	невозможное событие
$A \subseteq B$	A — подмножество B	событие A влечет B
$\overline{A} = \Omega \setminus A$	дополнение множества	противоположное к A событие
$A \cap B$	пересечение множеств	произведение событий: AB
$A \cap B = \varnothing$	множества не пересекаются	события несовместны
$A \cup B$	объединение множеств	сумма событий: $A + B$

Определение 1.5. События A_1, A_2, ..., A_n *образуют полную группу*, если в результате проведения эксперимента должно произойти хотя бы одно из них: $A_1 + A_2 + \ldots + A_n = \Omega$.

Определение 1.6. События A_1, A_2, ..., A_n *образуют разбиение* пространства Ω, если они образуют полную группу и являются попарно несовместными, т. е. $A_iA_j = \varnothing$, $1 \leqslant i < j \leqslant n$.

Пример 5. Рассмотрим два события, которые могут произойти при бросании двух игральных костей: $A = \{$сумма выпавших очков не превосходит $10\}$, $B = \{$сумма очков не меньше $10\}$.

Тогда $A+B=\Omega$, то есть события A и B образуют полную группу. Но A и B не образуют разбиение, так как они совместны. А именно событие $C = AB = \{\text{сумма очков равна } 10\}$ не является невозможным. С другой стороны, события $\overline{B}$, C и $\overline{A}$ образуют разбиение.

Теперь, когда у нас есть четкое определение события, можно ввести понятие вероятности. По существу, вероятность — это некоторая числовая характеристика события, мера возможности его появления. В математике вероятность принято измерять долями от единицы, т. е. от вероятности достоверного события.

Определение 1.7 (вероятность в дискретном пространстве событий)**.** Пусть каждому элементарному исходу $\omega \in \Omega$ дискретного пространства Ω поставлено в соответствие число $\mathsf{P}(\omega) \in \mathbb{R}$ так, что для всех таких чисел выполняются следующие свойства:

1) $\mathsf{P}(w) \geqslant 0$;

2) $\sum\limits_{\omega\in\Omega} \mathsf{P}(\omega) = 1$.

Тогда число $\mathsf{P}(\omega)$ называется *вероятностью элементарного исхода ω*. *Вероятностью* произвольного *события* $A \subseteq \Omega$ называется число

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

Соответственно, $\mathsf{P}(\varnothing) = 0$.

Непосредственно из определения выводятся следующие **основные свойства вероятности:**

1. Для любого события A справедливо $0 \leqslant \mathsf{P}(A) \leqslant 1$, $\mathsf{P}(\Omega) = 1$;

2. Если $A \subseteq B$, то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$;

3. Для любых A и B верно $\mathsf{P}(A+B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$. Если же A и B несовместны, то $\mathsf{P}(A+B) = \mathsf{P}(A) + \mathsf{P}(B)$;

4. Для любого события A верно $\mathsf{P}(A) = 1 - \mathsf{P}(\overline{A})$.

Особо следует выделить последнее свойство, позволяющее вычислять вероятность события посредством рассмотрения противоположного к нему события. Обычно осознание ценности этого свойства приходит только после самостоятельного прорешивания нескольких десятков задач. В то же время есть и более простой путь. Достаточно заметить, что теоретически примерно в половине задач нахождение вероятности события $\overline{A}$ должно оказаться проще нахождения вероятности события A. Более того, в некоторых случаях переход к рассмотрению противоположного события на порядок сокращает трудоемкость решения задачи.

Пример 6. Найти вероятность события A из примера 5.

Решение. В зависимости от чисел очков, выпадающих на первой и второй костях, все пространство исходов Ω данного эксперимента может быть разбито на 36 элементарных исходов:

$$\begin{aligned} \Omega = \{&\text{⚀⚀}, \text{⚀⚁}, \ldots, \text{⚀⚅}, \\ &\text{⚁⚀}, \text{⚁⚁}, \ldots, \text{⚁⚅}, \\ &\ldots\ \ \ldots\ \ \ldots\ \ \ldots \\ &\text{⚅⚀}, \text{⚅⚁}, \ldots, \text{⚅⚅}\}. \end{aligned}$$

Так как все исходы равновозможны, то $P(\omega) = 1/36$ для каждого $\omega \in \Omega$.

Рассмотрим событие, противоположное к A (сумма очков больше 10). Оно состоит из трех исходов: $\overline{A} = \{\text{⚄⚅}, \text{⚅⚄}, \text{⚅⚅}\}$. Следовательно,

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{3}{36} = \frac{11}{12}. \qquad \square$$

1.1.1. Классическое определение вероятности

Рассмотрим частный случай дискретного пространства элементарных исходов, состоящего из конечного числа N элементарных исходов: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. Предположим, что все эти исходы равновозможны: $\mathsf{P}(\omega_1) = \mathsf{P}(\omega_2) = \ldots = \mathsf{P}(\omega_N) = 1/N$. Тогда вероятность любого события $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_k}\}$, состоящего из k исходов, вычисляется по формуле

$$\mathsf{P}(A) = \sum_{j=1}^{k} \mathsf{P}(\omega_{i_j}) = \sum_{j=1}^{k} \frac{1}{N} = \frac{k}{N} = \\ = \frac{\text{число благоприятных для } A \text{ исходов}}{\text{число всех исходов}}.$$

Определение 1.8. Говорят, что эксперимент удовлетворяет *классическому определению вероятности*, если пространство элементарных исходов Ω состоит из конечного числа равновозможных исходов.

В таком эксперименте вычисление вероятности события $A \subset \Omega$ сводится к подсчету числа благоприятных исходов.

Во многих задачах подсчет числа исходов выполняется с использованием комбинаторных методов, о которых пойдет речь в следующем разделе.

1.2. Комбинаторный метод вычисления вероятностей

Рассмотрим опыт, состоящий в случайном выборе k элементов (предметов) из n. Этот опыт допускает два способа выбора:

1. Один и тот же элемент может быть выбран несколько раз (*выбор с повторениями*);

2. Каждый элемент может быть выбран не более одного раза (*выбор без повторений*).

Кроме того, в зависимости от поставленной задачи может быть важно (или же, наоборот, не важно) учитывать порядок выбора элементов. С учетом указанных альтернатив можно выделить четыре различных схемы выбора:

I. Выбор с повторениями и упорядочиванием (*размещения с повторением*);

II. Выбор с упорядочиванием, но без повторений (*размещения без повторений*);

III. Выбор без повторений и без упорядочивания (*сочетания*);

IV. Выбор без упорядочивания, но с повторениями (*сочетания с повторением*).

Далее в этом разделе мы займемся подсчетом числа способов выбора k элементов из n для каждой из четырех схем.

1.2.1. Размещения с повторениями

Найдем число размещений с повторениями из n элементов по k. Каждый из k выбираемых элементов может быть выбран n способами независимо от выбора остальных. Следовательно, общее число таких размещений равно

$$\underbrace{n \cdot n \cdot \ldots \cdot n}_{k \text{ раз}} = n^k.$$

Пример 7. Найдем вероятность того, что при четырех бросаниях игральной кости хотя бы раз выпадет '1'.

Решение. Результат такого эксперимента, по существу, состоит в выборе шести граней игрального кубика с учетом порядка (номера броска) и, возможно, с повторениями, т. е. общее число исходов этого эксперимента равно 6^4. Причем все эти исходы равновозможны (это подтверждается многовековой историей игры в кости). Таким образом, эксперимент удовлетворяет классическому определению вероятности, и нам остается подсчитать число благоприятных для события

$$A = \{\text{'1' появилась хотя бы раз}\}$$

исходов. Решение данной задачи существенно упрощается, если перейти к рассмотрению противоположного события

$$\overline{A} = \{\text{'1' не появилась ни разу}\},$$

т. е. теперь нас интерсуют только те случаи, когда при бросании кости могли появляться только пять граней из шести. Число таких исходов равно 5^4. Следовательно,

$$\mathsf{P}(\overline{A}) = \frac{5^4}{6^4}, \qquad \mathsf{P}(A) = 1 - \mathsf{P}(\overline{A}) = 1 - \left(\frac{5}{6}\right)^4. \qquad \square$$

Задача 8 (Парадокс де Мере [3])**.** При 4-кратном бросании одной игральной кости вероятность того, что единица появится хотя бы раз, больше $1/2$. В то же время при 24-кратном бросании двух игральных костей вероятность того, что две единицы появятся одновременно хотя бы раз, меньше $1/2$. Парадокс состоит в том, что вероятность выпадения единицы в первом опыте ровно в шесть раз больше вероятности появления двух единиц во втором, а 24 ровно в 6 раз больше четырех.

1.2.2. Размещения без повторений

Отличие этой схемы выбора от предыдущей лишь в том, что каждый элемент может быть выбран не более одного раза. Используя похожие рассуждения, получаем формулу для числа размещений из n элементов по k без повторений (обычно словосочетание «без повторений» опускается):

$$A_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

Особый случай, когда нужно выбрать все n элементов из n, называется *перестановками.* Число перестановок n элементов равно $A_n^n = n!$

Пример 9. Эксперимент состоит в случайном выборе k элементов из n с учетом порядка и, возможно, с повторениями ($k \leqslant n$). Требуется найти вероятность события $A = \{$ни один элемент не появится дважды$\}$.

Решение. Общее число исходов равно числу размещений с повторениями из n по k, а число благоприятных для события A исходов равно числу размещений без повторений. Таким образом,

$$\mathsf{P}(A) = \frac{n(n-1)\dots(n-k+1)}{n^k}. \qquad \square$$

Пример 10. Найти вероятность того, что при 6 бросаниях игральной кости каждая грань появится ровно один раз.

Решение. Эта задача является частным случаем предыдущего примера для $n = 6$ и $k = 6$. Искомая вероятность равна

$$\frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 \cdot 6} = \frac{5}{324}. \qquad \square$$

Задача 11. Найти вероятность того, что среди k человек найдутся хотя бы двое с одинаковыми днями рождения (без учета года рождения). *Примечание:* для упрощения предполагаем, что среди выбранных k человек нет родившихся 29 февраля.

Интересно отметить, что уже для 23 человек вероятность совпадения дней рождения хотя бы у двоих примерно равна 0,5, а для 55 человек эта вероятность уже равна 0,99.

1.2.3. Сочетания

Рассмотрим схему выбора k элементов из n, при которой никакой элемент не может быть выбран дважды и порядок выбора не важен (т. е. k выбранных элементов рассматриваются не как упорядоченный набор, а как множество). Получившееся таким образом k-элементное подмножество называется *сочетанием из n элементов по k*.

Рассмотрим теперь одно такое множество из k элементов. Произвольная нумерация (упорядочивание) элементов превращает его в упорядоченную выборку (размещение без повторений). И наоборот, любая упорядоченная выборка размера k может быть получена соответствующей нумерацией одного из k-элементных подмножеств (сочетаний). Так как k элементов могут быть упорядочены $k!$ способами, то число C_n^k всех сочетаний из n по k может быть найдено по формуле[2]

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}.$$

В частности, так как $0! = 1$, то $C_n^0 = 1$.

[2]Для числа сочетаний C_n^k часто используют обозначение $\binom{n}{k}$.

Некоторые **свойства числа сочетаний**:

1) $C_n^k = C_n^{n-k}$;

2) $C_n^0 + C_n^1 + \ldots + C_n^n = 2^n$;

3) $C_n^k + C_n^{k+1} = C_{n+1}^{k+1}$.

Пример 12. Из колоды в 36 карт случайно (и без возвращения) выбирают 6 карт. Найти вероятность события

$$A = \{\text{среди выбранных карт есть хотя бы один туз}\}.$$

Решение. Так как порядок выбора карт не важен, число всех исходов равно C_{36}^6. При непосредственном подсчете числа благоприятных для события A исходов их следовало бы разбить на четыре случая — по числу тузов среди выбранных шести карт. Удобнее рассмотреть противоположное к A событие — не был выбран ни один туз. Число таких исходов равно числу сочетаний из 32 «не тузов» по 6. Следовательно,

$$\mathsf{P}(A) = 1 - \frac{C_{32}^6}{C_{36}^6} = \frac{100}{187}. \qquad \square$$

Задача 13. Подбрасываются четыре игральные кости. Найти вероятность того, что выпадут ровно две «тройки».

Задача 14. Из колоды в 52 карты случайным образом выбирают 5. Найти вероятность того, что все они окажутся разного достоинства.

1.2.4. Разбиение множества на несколько подмножеств

Предыдущий раздел был посвящен подсчету числа k-элементных подмножеств. Обратим теперь внимание на то, что, по существу, это число различных разбиений множества на два подмножества, состоящие из k и $n-k$ элементов. В этом разделе мы займемся подсчетом числа разбиений множества (состоящего из n элементов) на s подмножеств ($s > 2$), из которых первое содержит k_1 элементов, второе — k_2 элементов, ..., s-е подмножество — k_s элементов. Соответственно, $k_1 + k_2 + \cdots + k_s = n$.

Число способов выбора k_1 элементов в первое подмножество равно $C_n^{k_1}$. Число способов выбора k_2 элементов (из оставшихся $n - k_1$ элементов) во второе подмножество равно $C_{n-k_1}^{k_2}$. И т. д. Общее число разбиений равно произведению этих чисел:

$$C_n^{k_1} C_{n-k_1}^{k_2} \dots C_{k_s}^{k_s} = \frac{n!}{k_1! k_2! \dots k_s!}.$$

Пример 15. Найти число шестизначных телефонных номеров с тремя цифрами '3', двумя '2' и одной единицей.

Решение. Эта задача эквивалентна подсчету числа разбиений шести позиций цифр в телефонном номере на три подмножества, состоящих из трех, двух и одной цифры соответственно:

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

1.2.5. Сочетания с повторениями

Процесс выбора k элементов из n с повторениями, но без упорядочивания можно представить следующим образом. По n ящикам раскладываются k шаров. Причем в один ящик можно класть сколько угодно шаров и их порядок внутри ящика не важен. Результат одного такого раскладывания можно представить, например, так:

$$\underbrace{\underline{|\circ\circ\circ|}}_{\text{1-й ящик}}\quad \underbrace{\underline{|\,|}}_{\text{2-й ящик}}\quad \underbrace{\underline{|\circ|}}_{\text{3-й ящик}}\quad \dots\quad \underline{\ \ |}$$

Или так:

$$|\circ\circ\circ||||\circ|\dots|$$

А убрав «лишние» стенки ящиков (оставив только по одной перегородке между соседними ящиками), вот так:

$$\circ\circ\circ||\circ|\dots$$

Ясно, что каждое размещение k шаров по n ящикам может быть единственным образом записано в виде последовательности из k кружочков и $n-1$ черточек. И наоборот, каждая такая последовательность однозначно определяет размещение шаров по ящикам. Таким образом, задача подсчета числа сочетаний с повторениями из n элементов по k сводится к подсчету числа различных последовательностей, состоящих из k кружочков и $n-1$ черточек. Это число, очевидно, равно

$$C_{n+k-1}^{k} = C_{n+k-1}^{n-1}.$$

Задача 16. Пользуясь формулой числа сочетаний с повторениями, найти число костей домино.

Задача 17. Найти число различных наборов из трех чисел, возникающих при бросании трех внешне неразличимых игральных костей.

1.2.6. Гипергеометрическое распределение

Многие вероятностные задачи могут быть сведены к следующей модели.

В исходном множестве из N элементов имеется K помеченных. Из этого множества случайным образом и без повторений выбираются n элементов. Требуется найти вероятность $\mathsf{P}(k)$ того, что среди выбранных окажется ровно k помеченных. Так как порядок выбора элементов не важен и повторения не допускаются, то общее число исходов такого эксперимента можно взять равным C_N^n. При подсчете благоприятных для указанного события исходов требуется выбирать k помеченных элементов и $n-k$ непомеченных, что приводит к следующей формуле:

$$\mathsf{P}(k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}.$$

Набор чисел $\mathsf{P}(k)$ (при заранее фиксированных N, K, n) для всех допустимых[3] значений k называется *гипергеометрическим распределением.* Сумма всех чисел из этого набора, очевидно, должна быть равна 1.

Пример 18. Из колоды в 36 карт случайным образом выбираются 6 карт. Найти вероятности следующих событий: $A = \{$среди выбранных ровно два туза$\}$, $B = \{$среди выбранных ровно одна бубновой масти$\}$.

[3] Число k должно удовлетворять неравенствам $0 \leqslant k \leqslant K$ и $0 \leqslant n-k \leqslant N-K$.

Решение. Число всех элементарных исходов, очевидно, равно C_{36}^6. Для вычисления вероятности $\mathsf{P}(A)$ целесообразно «пометить» тузы, тогда

$$\mathsf{P}(A) = \frac{C_4^2 C_{32}^4}{C_{36}^6} = \frac{145}{1309}.$$

Для вычисления вероятности $\mathsf{P}(B)$ целесообразно «пометить» карты бубновой масти, тогда

$$\mathsf{P}(B) = \frac{C_9^1 C_{27}^5}{C_{36}^6} = \frac{121095}{324632} \approx 0{,}373. \qquad \square$$

1.3. Геометрические вероятности

Во многих задачах пространство элементарных исходов Ω не является дискретным. В наиболее простом варианте Ω представляет собой некоторую область в $\mathbb{R}^n$ и удовлетворяет следующему «естественному» условию.

Определение 1.9. Говорят, что *точка равномерно распределена в области* Ω, если вероятность попадания ее в любую область $A \subseteq \Omega$ прямо пропорциональна мере области A и не зависит от расположения A внутри Ω:

$$\mathsf{P}(A) = \frac{\mu(A)}{\mu(\Omega)}. \tag{1.1}$$

(Под мерой $\mu(\ )$ в случае $\Omega \subset \mathbb{R}^1$ обычно понимается длина, в случае $\Omega \subset \mathbb{R}^2$ — площадь, а в случае $\Omega \subset \mathbb{R}^n$, $n \geqslant 3$ — объем.) Обычно, когда речь идет о случайном выборе точки из некоторого множества (и никакие дополнительные условия выбора не оговариваются), подразумевается, что она равномерно распределена в этом множестве.

Заметим, что все элементарные исходы в Ω равновозможны, причем их бесконечно много. Поэтому вероятность попасть в некоторую фиксированную точку в таком эксперименте равна нулю. В то же время такое событие не является невозможным.

Рассмотрим несколько примеров.

Пример 19 (задача Бюффона[4])**.** На плоскость, расчерченную параллельными прямыми с одинаковыми расстояниями $2h$ между соседними линиями, падает игла длины $2l$, $l \leqslant h$. Найти вероятность того, что игла пересечет одну из прямых (событие A).

Решение. На рис. 1.1 а) изображены две параллельные прямые s_1 и s_2, между которыми попала середина C иглы. Не уменьшая общности, предположим, что C ближе к s_1, чем к s_2. (Другой случай является зеркальным отражением.) Опустим перпендикуляр CB на прямую s_1. По условию предполагается, что центр иглы равномерно распределен на плоскости. Тогда расстояние $y = CB$ от центра иглы до ближайшей прямой равномерно распределено на отрезке $[0, h]$ (где h — половина расстояния между прямыми s_1 и s_2). По той же причине угол $\varphi = \angle BCE$ равномерно (и независимо от y) распределен на отрезке $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Соответственно, каждый исход эксперимента можно представить точкой (φ, y), равномерно распределенной в прямоугольнике $\Omega = [-\frac{\pi}{2}, \frac{\pi}{2}] \times [0, h]$, изображенном на рис. 1.1 б).

Очевидно, игла пересекает прямую s_1 тогда и только тогда, когда проекция половины иглы CE на прямую CB не меньше длины CB:

$$l \cos\varphi \geqslant y.$$

[4]Жорж-Луи Леклерк, граф де Бюффон (1707–1788) — французский натуралист, биолог, математик, естествоиспытатель и писатель.

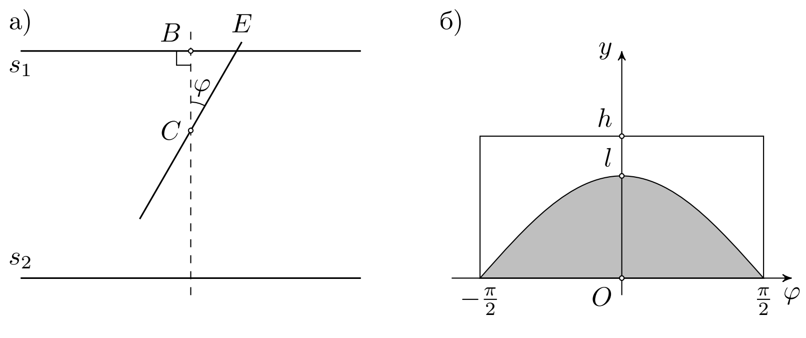

Рис. 1.1. Задача Бюффона об игле

Множество всех точек (φ, y), удовлетворяющих этому условию, закрашено на рис. 1.1 б). Вероятность события A равна отношению площади окрашенной части к площади всего прямоугольника Ω. Площадь закрашенной фигуры равна

$$\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} l\cos\varphi\, d\varphi = 2l.$$

Тогда

$$\mathsf{P}(A) = \frac{2l}{h\pi}. \qquad \square$$

Задача 20. Решить задачу Бюффона для случая $l > h$.

Пример 21. Коэффициенты p и q уравнения $x^2 + px + q = 0$ выбираются случайно и независимо друг от друга из отрезка $[0,1]$. Найти вероятность того, что корни уравнения действительны.

Решение. Результат эксперимента можно представить вектором (p, q), равномерно распределенным в квадрате $[0,1] \times [0,1]$. Корни уравнения действительны, если дискриминант $D = p^2 - 4q$ неотрицателен, т. е. при условии $q \leqslant \frac{p^2}{4}$. Соответствующая область закрашена на рис. 1.2. Следовательно, вероятность того, что корни окажутся действительными, равна

$$\int\limits_0^1 \frac{p^2}{4}\, dp = \frac{1}{12}. \qquad \square$$

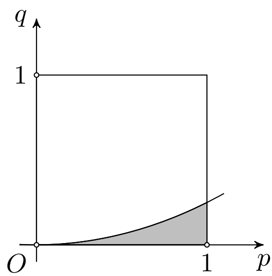

Рис. 1.2. Случаи действительных корней

Наличие подводных камней при таком подходе демонстрируется следующим парадоксом.

Задача 22 (парадокс Бертрана[5])**.** Случайным образом выбирается хорда окружности. Найти вероятность того, что длина хорды окажется больше длины стороны вписанного в эту окружность правильного треугольника. Рассмотреть три разных способа выбора хорды

[5] Жозеф Луи Франсуа Бертран (1822–1900) — французский математик.

(для удобства предполагаем, что окружность лежит в координатной плоскости с ортогональными осями Ox и Oy и центром в начале координат):

1. Случайно выбирается расстояние l от центра окружности до середины хорды и угол γ между нормалью хорды и осью Ox;
2. На окружности случайно выбираются две точки, служащие концами хорды;
3. Внутри круга случайно выбирается точка, которая будет серединой хорды.

Суть парадокса Бертрана состоит в том, что все три способа выбора хорды дают разные вероятности. При внимательном рассмотрении можно увидеть сходство этого парадокса с парадоксом Кардано (задача 4 на с. 8). В обоих случаях проблема заключается в выборе «правильного» пространства элементарных исходов, в котором вероятность была бы распределена между исходами равномерно.

1.4. Аксиоматика теории вероятностей

До настоящего момента рассматривались вероятности в дискретных пространствах и геометрические вероятности. Есть и другие, принципиально отличающиеся от этих, теоретико-вероятностные модели. В частности, в случае геометрических вероятностей случайно выбираемая точка может оказаться неравномерно распределенной[6]. Кроме того, существуют примеры сложных моделей, частично обладающих дискретными, а частично — непрерывными свойствами. С целью унификации всех этих моделей вводится некоторый универсальный набор правил — аксиом теории вероятностей[7]. Для его исчерпывающего описания нам потребуются понятия алгебры и σ-алгебры событий.

[6] Например, в парадоксе Бертрана (задача 22) длина хорды оказывается неравномерно распределенной при любом из трех предложенных способов выбора.

[7] Далее будет изложено строгое аксиоматическое построение основных понятий теории вероятностей, предложенное одним из основателей современной теории вероятностей Андреем Николаевичем Колмогоровым (1903–1987, Россия).

Определение 1.10. Множество $\mathcal{S} \subseteq 2^{\Omega}$, элементами которого являются (не обязательно все) подмножества множества Ω, называется *алгеброй событий*, если выполняются следующие условия:

1. $\Omega \in \mathcal{S}$;

2. $A \in \mathcal{S} \Rightarrow \overline{A} \in \mathcal{S}$;

3. $A, B \in \mathcal{S} \Rightarrow A \cup B \in \mathcal{S}$.

Заметим, что из условий 1 и 2 следует, что $\varnothing \in \mathcal{S}$, так как $\varnothing = \overline{\Omega}$. Из условия 3 следует справедливость его аналога для любого конечного набора множеств: если каждое из множеств A_1, A_2, …, A_n принадлежит $\mathcal{S}$, то их объединение $A_1 \cup A_2 \cup \dots \cup A_n$ также принадлежит $\mathcal{S}$. А из правила де Моргана $A \cap B = \overline{\overline{A} \cup \overline{B}}$ и условий 2 и 3 следует, что вместе с множествами A и B в $\mathcal{S}$ входит $A \cap B$.

В частности, если $\mathcal{S}$ содержит в себе все подмножества множества Ω, то $\mathcal{S}$ является алгеброй событий. Еще одним примером алгебры событий может служить множество $\mathcal{S} = \{\Omega, \varnothing, A, \overline{A}\}$, где A — произвольно выбранное подмножество множества Ω.

Понятия алгебры событий вполне достаточно для определения вероятности для конечного Ω. Но для более сложных случаев часто требуется выполнение более жесткого варианта условия 3:

3'. $A_1, A_2, \ldots \in \mathcal{S} \Rightarrow A_1 \cup A_2 \cup \ldots \in \mathcal{S}$.

Существенное отличие — в бесконечности (счетности) набора событий $A_1, A_2, \ldots \in \mathcal{S}$. Замена условия 3 на 3' приводит к понятию *σ-алгебры событий*. Более подробная информация о σ-алгебрах в теории вероятностей содержится в [6].

Определение 1.11 (Вероятностное пространство общего вида)**.** Пусть Ω — пространство элементарных исходов, $\mathcal{S}$ — некоторое множество его подмножеств, которые далее называются *событиями*, и $\mathcal{S}$ является σ-алгеброй. *Вероятностью* на $(\Omega, \mathcal{S})$ называется функция $\mathsf{P}\colon \mathcal{S} \to \mathbb{R}$, обладающая свойствами:

P1 (неотрицательность). Для каждого $A \in \mathcal{S}$ выполнено $\mathsf{P}(A) \geqslant 0$;

P2 (нормировка). $\mathsf{P}(\Omega) = 1$;

P3 (счетная аддитивность). Для любого счетного набора попарно несовместных событий $A_1,\ A_2,\ \ldots,\ A_n,\ \ldots \in \mathcal{S}$ выполнено

$$\mathsf{P}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i).$$

Тройка $(\Omega, \mathcal{S}, P)$ называется *вероятностным пространством.*

Те свойства вероятности, о которых было сказано ранее, в разделе 1.1 (на с. 10), оказываются справедливы и для вероятностного пространства общего вида.

1.4.1. Основные свойства вероятности

Свойство 1. $\mathsf{P}(\varnothing) = 0$.

$\triangleright$ Рассмотрим набор событий $A_i = \varnothing$, $i \in \mathbb{N}$. Все они попарно несовместны, и, следовательно, для них выполняется свойство P3:

$$\mathsf{P}(\varnothing) = \mathsf{P}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i) = \sum_{i=1}^{\infty} \mathsf{P}(\varnothing).$$

Очевидно, что это равенство возможно, только если $\mathsf{P}(\varnothing) = 0$. $\square$

Свойство 2 (конечная аддитивность). Для любого набора попарно несовместных событий $A_1, A_2, \ldots, A_n$ выполнено

$$\mathsf{P}\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i).$$

$\triangleright$ Достаточно в свойстве P3 положить $A_i = \varnothing$ для $i > n$ и воспользоваться тем, что $\mathsf{P}(\varnothing) = 0$. $\square$

Свойство 3. $\mathsf{P}(A) = 1 - \mathsf{P}(\overline{A})$.

$\triangleright$ Так как $A + \overline{A} = \Omega$ и события A и $\overline{A}$ несовместны, то

$$\mathsf{P}(\Omega) = \mathsf{P}(A + \overline{A}) = \mathsf{P}(A) + \mathsf{P}(\overline{A}).$$

Учитывая свойство P2, получаем $1 = \mathsf{P}(A) + \mathsf{P}(\overline{A})$. $\square$

Свойство 4. Из $A \subseteq B$ следует $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.

$\triangleright$ Разложим событие B на сумму двух несовместных событий:

$$B = AB + \overline{A}B.$$

Из $A \subseteq B$ следует $AB = A$. Таким образом,

$$\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(\overline{A}B) \geqslant \mathsf{P}(A). \qquad \square$$

Свойство 5. Для любого события A выполнено $0 \leqslant \mathsf{P}(A) \leqslant 1$.

$\triangleright$ Неравенство $0 \leqslant \mathsf{P}(A)$ декларируется свойством P1. А неравенство $\mathsf{P}(A) \leqslant 1$ следует из свойства 3:

$$\mathsf{P}(A) = 1 - \mathsf{P}(\overline{A}) \leqslant 1. \qquad \square$$

Свойство 6. $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$.

$\triangleright$ Заметим, что A представимо в виде суммы несовместных событий AB и $A\overline{B}$. Следовательно,

$$\mathsf{P}(A) = \mathsf{P}(AB) + \mathsf{P}(A\overline{B}).$$

Таким образом,

$$\mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB) = \mathsf{P}(A\overline{B}) + \mathsf{P}(B).$$

С другой стороны, $A + B$ тоже можно представить в виде суммы двух несовместных событий: B и $A\overline{B}$. А значит,

$$\mathsf{P}(A + B) = \mathsf{P}(B) + \mathsf{P}(A\overline{B}) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB). \qquad \square$$

Следствие 1 (неравенство треугольника).

$$\mathsf{P}(A + B) \leqslant \mathsf{P}(A) + \mathsf{P}(B).$$

Следствие 2 (формула включений-исключений).

$$\mathsf{P}\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum_{1 \leqslant i < j \leqslant n} \mathsf{P}(A_i A_j) + \\ + \sum_{i<j<k} \mathsf{P}(A_i A_j A_k) - \ldots + (-1)^{n+1} \mathsf{P}\left(\prod_{i=1}^{n} A_i\right).$$

Задача 23. Доказать формулу включений-исключений.

Свойство непрерывности. Пусть последовательность событий A_1, A_2, A_3, ... образует «матрёшку» $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$, причем $A = \prod\limits_{n=1}^{\infty} A_n$. Тогда

$$\lim_{n\to\infty} \mathsf{P}(A_n) = \mathsf{P}\left(A\right).$$

▷ Обозначим $B_i = A_i \setminus A_{i+1} = A_i\overline{A_{i+1}}$, $i \in \mathbb{N}$. Тогда события A, B_1, B_2, B_3, ... попарно несовместны. Воспользуемся для них представлением

$$A_n = A + B_n + B_{n+1} + \ldots$$

и свойством P3:

$$\mathsf{P}(A_1) = \mathsf{P}\left(A + \sum_{i=1}^{\infty} B_i\right) = \mathsf{P}(A) + \sum_{i=1}^{\infty} \mathsf{P}(B_i).$$

Так как вероятности всех событий неотрицательны и, кроме того, $\mathsf{P}(A_1) \leqslant 1$, то числовой ряд в правой части сходится абсолютно. Значит, его частичные суммы $\sum\limits_{i=n}^{\infty} \mathsf{P}(B_i)$ стремятся к нулю при n стремящемся к бесконечности. Еще раз воспользуемся свойством P3:

$$\mathsf{P}(A_n) = \mathsf{P}\left(A + \sum_{i=n}^{\infty} B_i\right) = \mathsf{P}(A) + \sum_{i=n}^{\infty} \mathsf{P}(B_i).$$

Переходя к пределу, получаем

$$\lim_{n\to\infty} \mathsf{P}(A_n) = \mathsf{P}(A) + \lim_{n\to\infty} \sum_{i=n}^{\infty} \mathsf{P}(B_i) = \mathsf{P}(A). \qquad \square$$

В завершение раздела разберем два примера использования формулы включений-исключений.

Пример 24. Бросаются три игральные кости. Найти вероятность того, что хотя бы на одной из костей появится ‘6’.

Решение. Разумеется, эту задачу можно решить тем же способом, что и пример 7. Но есть и альтернативное решение — с помощью формулы включений-исключений.

Пусть $A_i = \{$на i-й игральной кости появится ‘6’$\}$, где $i = 1, 2, 3$. По условию нужно найти вероятность события $A = A_1 + A_2 + A_3$. Заметим, что $\mathsf{P}(A_i) = \frac{1}{6}$, а $\mathsf{P}(A_1A_2) = \mathsf{P}(A_1A_3) = \mathsf{P}(A_2A_3) = \frac{1}{6^2}$. Аналогично $\mathsf{P}(A_1A_2A_3) = \frac{1}{6^3}$. Тогда, согласно формуле включений-исключений,

$$\mathsf{P}(A) = \mathsf{P}(A_1 + A_2 + A_3) =$$
$$= \sum_{i=1}^{3} \mathsf{P}(A_i) - \sum_{1 \leqslant i < j \leqslant 3} \mathsf{P}(A_iA_j) + \mathsf{P}(A_1A_2A_3) = \frac{91}{216}. \qquad \square$$

Пример 25 (задача о рассеянной секретарше)**.** Задача секретарши — отправить n различных по содержанию писем по n адресам. Уже готовы n подписанных конвертов. Секретарша раскладывает письма по конвертам случайным образом. Найти вероятность того, что хотя бы одно письмо попадет в «свой» конверт.

Решение. Наиболее простой способ решения этой задачи опирается на формулу включений-исключений. Представим интересующее нас событие в виде суммы событий $A_i = \{i$-е письмо попало в «свой» конверт$\}$, где $i = 1, \ldots, n$. Очевидно, $\mathsf{P}(A_i) = \frac{1}{n}$. Найдем вероятность события A_iA_j, где $i \neq j$:

$$\mathsf{P}(A_iA_j) = \frac{1}{n(n-1)}.$$

Аналогично

$$\mathsf{P}(A_iA_jA_k) = \frac{1}{n(n-1)(n-2)},$$
$$\ldots,$$
$$\mathsf{P}(A_1A_2\ldots A_n) = \frac{1}{n!}.$$

Подставляя в формулу включений-исключений, получаем

$$\mathsf{P}(A_1 + \ldots + A_n) = C_n^1 \cdot \frac{1}{n} - C_n^2 \cdot \frac{1}{n(n-1)} + \\ + C_n^3 \cdot \frac{1}{n(n-1)(n-2)} - \ldots + (-1)^{n+1}\frac{1}{n!} = \\ = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n+1}\frac{1}{n!}.$$

Учитывая известное разложение

$$e^{-1} = \frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!} + \ldots,$$

получаем

$$\mathsf{P}(A_1 + \ldots + A_n) \approx 1 - e^{-1} = 0{,}63212\ldots$$

с точностью до $\frac{1}{(n+1)!}$ (т. е. уже для $n = 6$ верны первые 3 знака после запятой). $\square$

1.5. Условная вероятность и независимость

1.5.1. Условная вероятность

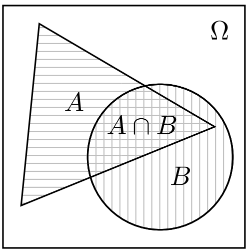

Рис. 1.3. Условная вероятность

Предположим, что точка равномерно распределена внутри квадрата Ω. Рассматриваются два события (см. рис. 1.3): $A = \{$точка попала в треугольник$\}$, $B = \{$точка попала в круг$\}$. Требуется найти вероятность события A при условии, что событие B произошло. Такая вероятность называется условной и обозначается $\mathsf{P}(A \mid B)$.

Так как событие B произошло, то мы можем предполагать, что точка выбирается случайно не из квадрата, а из круга. Тогда искомая вероятность равна отношению площади пересечения $A \cap B$ к площади круга B:

$$\mathsf{P}(A \mid B) = \frac{S(A \cap B)}{S(B)} = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}.$$

Определение 1.12 (условная вероятность)**.** Пусть B — событие положительной вероятности. Тогда вероятность события A *при условии, что B произошло*, определяется следующей формулой:

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}. \tag{1.2}$$

Альтернативные обозначения: $\mathsf{P}_B(A)$, $\mathsf{P}(A \mathbin{/} B)$.

Пример 26. В ящике лежат a белых и b черных шаров. Из ящика случайным образом вынимают один шар и, не глядя, откладывают в сторону. Затем вынимают еще один шар. Он оказался белым. Какова вероятность того, что 1-й вынутый шар тоже белый?

Решение. Введем обозначения: $A = \{$первый вынутый шар — белый$\}$, $B = \{$второй вынутый шар — белый$\}$. Если нам неизвестно, какой шар был вынут первым, то $\mathsf{P}(B) = \frac{a}{a+b}$. Вероятность того, что оба вынутых шара — белые:

$$\mathsf{P}(AB) = \frac{a(a-1)}{(a+b)(a+b-1)}.$$

Таким образом,

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)} = \frac{a-1}{a+b-1}. \qquad \square$$

Наиболее часто формула (1.2) используется как теорема об умножении вероятностей:

Теорема 1.13. $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B \mid A) = \mathsf{P}(B)\mathsf{P}(A \mid B)$.

Обратим теперь внимание на тот факт, что любая безусловная вероятность может быть представлена как условная:

$$\mathsf{P}(A) = \mathsf{P}_\Omega(A) = \mathsf{P}(A \mid \Omega).$$

В этом смысле все формулы, выведенные ранее для безусловных вероятностей, справедливы также и для условных вероятностей. Например, теорема умножения допускает следующую формулировку (в случае $\mathsf{P}(C) > 0$):

$$\mathsf{P}_C(AB) = \mathsf{P}_C(A)\mathsf{P}_C(B \mid A) = \mathsf{P}_C(B)\mathsf{P}_C(A \mid B),$$

или, что то же самое,

$$\mathsf{P}(AB \mid C) = \mathsf{P}(A \mid C)\mathsf{P}(B \mid AC) = \mathsf{P}(B \mid C)\mathsf{P}(A \mid BC).$$

Пользуясь последней формулой, можно вывести формулу для вероятности произведения трех событий:

$$\mathsf{P}(CBA) = \mathsf{P}(C(AB)) = \mathsf{P}(C)\mathsf{P}(AB \mid C) = \mathsf{P}(C)\mathsf{P}(B \mid C)\mathsf{P}(A \mid BC).$$

Таким образом, теорему умножения можно обобщить на случай произведения нескольких событий.

Теорема 1.14 (теорема умножения для нескольких событий)**.**

$$\mathsf{P}(A_1A_2\ldots A_n) = \mathsf{P}(A_1) \cdot \mathsf{P}(A_2 \mid A_1) \cdot \ldots \cdot \mathsf{P}(A_n \mid A_1 \ldots A_{n-1}).$$

Разумеется, эта теорема оказывается полезной в тех случаях, когда соответствующие условные вероятности вычисляются непосредственно (т. е. без использования формулы (1.2)).

Пример 27. Из колоды в 36 карт вынимаются три карты. Какова вероятность того, что все три бубновой масти?

Решение. По условию эксперимента порядок выбора карт не важен. Тем не менее мы можем пронумеровать выбранные карты. Это не изменит вероятность искомого события и позволит рассмотреть события $A_i = \{i\text{-я вынутая карта — бубновой масти}\}$, $i = 1, 2, 3$. Тогда искомая вероятность равна

$$\mathsf{P}(A_1A_2A_3) = \mathsf{P}(A_1) \cdot \mathsf{P}(A_2 \mid A_1) \cdot \mathsf{P}(A_3 \mid A_1A_2).$$

Очевидно, $\mathsf{P}(A_1) = \frac{9}{36}$. Для вычисления $\mathsf{P}(A_2 \mid A_1)$ предположим, что событие A_1 уже произошло. Следовательно, в колоде осталось 35 карт и среди них 8 бубновых. Тогда $\mathsf{P}(A_2 \mid A_1) = \frac{8}{35}$. Аналогично, предполагая, что A_1 и A_2 произошли, получаем $\mathsf{P}(A_3 \mid A_1A_2) = \frac{7}{34}$. В итоге

$$\mathsf{P}(A_1A_2A_3) = \frac{9 \cdot 8 \cdot 7}{36 \cdot 35 \cdot 34} = \frac{1}{85}. \qquad \square$$

1.5.2. Независимость

Обычно под независимостью каких-либо явлений понимается отсутствие причинно-следственной связи. Что же это означает на языке теории вероятностей?

Естественно считать (в рамках теории вероятностей), что событие A *не зависит от события* B, если вероятность события A не меняется при наступлении события B:

$$\mathsf{P}(A \mid B) = \mathsf{P}(A).$$

Предположим, что $\mathsf{P}(A) > 0$ и A не зависит от B. Тогда, в силу теоремы умножения,

$$\mathsf{P}(A)\mathsf{P}(B \mid A) = \mathsf{P}(B)\mathsf{P}(A \mid B) = \mathsf{P}(B)\mathsf{P}(A) \Rightarrow \mathsf{P}(B \mid A) = \mathsf{P}(B),$$

т. е. событие B тоже не зависит от A и свойство независимости взаимно. Это дает основания ввести следующее определение независимости событий.

Определение 1.15. События A и B называются *независимыми*, если выполнено равенство

$$\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B). \tag{1.3}$$

Теорема 1.16 (о независимости отрицаний событий)**.** *Если A и B независимы, то независимыми будут также следующие пары событий: A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.*

Доказательство. Заметим, что

$$\mathsf{P}(A) = \mathsf{P}(A\Omega) = \mathsf{P}(A(B + \overline{B})) = \mathsf{P}(AB) + \mathsf{P}(A\overline{B}).$$

Таким образом, $\mathsf{P}(A\overline{B}) = \mathsf{P}(A) - \mathsf{P}(AB)$. Если A и B независимы, то $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B)$. Получаем

$$\mathsf{P}(A\overline{B}) = \mathsf{P}(A) - \mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(A)(1 - \mathsf{P}(B)) = \mathsf{P}(A)\mathsf{P}(\overline{B}).$$

Следовательно, A и $\overline{B}$ тоже независимы.

Завершение доказательства предлагается читателю в качестве упражнения. □

Замечание. Теорема 1.16 допускает естественное обобщение на случай произвольного числа независимых событий.

Замечание. При решении реальных задач никто не проверяет независимость событий с помощью формулы (1.3). По существу, это равенство является необходимым (но не всегда достаточным) условием для отсутствия причинно-следственной связи. Проиллюстрируем эту мысль примерами.

Пример 28. Эксперимент состоит в бросании двух игральных костей. События $A = \{$‘1’ на первой кости$\}$ и $B = \{$четное число на второй кости$\}$, очевидно, никак не связаны друг с другом. Поэтому в данной ситуации можно пользоваться формулой $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B)$.

Пример 29. Точка (x, y) равномерно распределена в единичном квадрате $\Omega = [0,1] \times [0,1]$. Пусть a и b — некоторые числа из интервала $(0,1)$. Рассмотрим события $A = \{x \leqslant a\}$ и $B = \{y \leqslant b\}$. Очевидно, $\mathsf{P}(A) = a$, $\mathsf{P}(B) = b$, $\mathsf{P}(AB) = ab$ и события A и B независимы. Если же в качестве Ω рассмотреть единичный квадрат с дополнительным ограничением $x + y \geqslant \varepsilon$, где $\varepsilon > 0$ достаточно мало (см. рис. 1.4), то условие независимости нарушится (проверьте!).

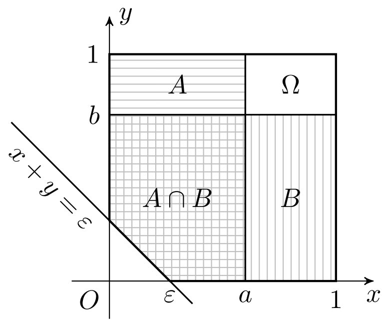

Рис. 1.4. Пример зависимых событий

Пример 30. Монету бросают два раза. Рассматриваются события $A = \{$при первом бросании появился герб$\}$ и $B = \{$в результате двух бросаний герб появился ровно один раз}. Проверка условия (1.3) показывает, что эти события независимы. Если же условия опыта слегка изменить, чуть-чуть погнув монету (точнее, изменив вероятность появления герба так, чтобы она немного отличалась от $1/2$), то условие независимости нарушится.

Замечание. Пусть A и B несовместны: $AB = \varnothing$. Тогда независимыми они будут, только если вероятность хотя бы одного из них равна нулю. Заметим также, что свойство несовместности событий не связано с их вероятностями.

Интуиция подсказывает, что если каждые два из трех событий A, B и C независимы, то и все три события должны быть независимы, т. е. для них должен выполняться аналог формулы (1.3):

$$\mathsf{P}(ABC) = \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C).$$

Оказывается, это не всегда верно.

Пример 31 (пример Бернштейна[8])**.** Монету бросают два раза. Рассматриваются следующие события:

$A = \{$при первом бросании появился герб$\}$,
$B = \{$в результате двух бросаний герб появился ровно один раз$\}$,
$C = \{$при втором бросании появился герб$\}$.

Нетрудно убедиться, что

$$\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = 1/2,$$
$$\mathsf{P}(AB) = \mathsf{P}(AC) = \mathsf{P}(BC) = 1/4,$$

т. е. события A, B и C попарно независимы. С другой стороны, событие ABC невозможно. Следовательно, $\mathsf{P}(ABC) \neq \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C)$.

Пример Бернштейна говорит о том, что из попарной независимости не всегда следует независимость в совокупности. Более того, выполнение условия $\mathsf{P}(ABC) = \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C)$ не означает, что события окажутся попарно независимыми.

[8]Сергей Натанович Бернштейн (1880–1968) — российский, советский математик.

Задача 32. Привести пример трех событий A, B и C, для которых условие $\mathsf{P}(ABC) = \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C)$ выполнено, но любая пара событий зависима.

Определение 1.17. События A_1, A_2, …, A_n называются *независимыми в совокупности*, если для любого $I \subseteq \{1, \ldots, n\}$, $I \neq \varnothing$, выполнено

$$\mathsf{P}\left(\prod_{i\in I} A_i\right) = \prod_{i\in I} \mathsf{P}(A_i). \tag{1.4}$$

Замечание. Как правило, определением независимости событий пользуются в обратную сторону. А именно, если между событиями A_1, A_2, …, A_n отсутствует причинно-следственная связь (это обычно явно следует из условий эксперимента), то для них верна каждая из $2^n - n - 1$ формул (1.4).

1.6. Формула полной вероятности и формула Байеса[9]

1.6.1. Формула полной вероятности

Рассмотрим полную группу несовместных событий H_1, H_2, …, H_n, $\mathsf{P}(H_i) > 0$, $i = 1, 2, \ldots, n$. Такой набор еще называют *разбиением* пространства Ω. Сами события H_1, H_2, …, H_n в этом разделе будут называться *гипотезами*.

Так как гипотезы H_1, …, H_n образуют полную группу, то

$$A = A \cdot \Omega = A(H_1 + \ldots + H_n) = AH_1 + \ldots + AH_n.$$

Кроме того, гипотезы H_1, …, H_n попарно несовместны. Значит, события AH_1, …, AH_n тоже попарно несовместны и справедливо

$$\mathsf{P}(A) = \mathsf{P}(AH_1) + \mathsf{P}(AH_2) + \ldots + \mathsf{P}(AH_n).$$

Воспользовавшись формулой умножения

$$\mathsf{P}(AH_i) = \mathsf{P}(H_i)\mathsf{P}(A \mid H_i), \quad i = 1, 2, \ldots, n,$$

[9] Томас Байес (1702–1761) — английский математик и пресвитерианский священник.

получаем *формулу полной вероятности*:

$$\mathsf{P}(A) = \mathsf{P}\left(\sum_{i=1}^{n} AH_i\right) = \sum_{i=1}^{n} \mathsf{P}(AH_i) = \sum_{i=1}^{n} \mathsf{P}(H_i)\mathsf{P}(A \mid H_i).$$

Эта формула полезна тем, что вычисление вероятностей $\mathsf{P}(H_i)$ и $\mathsf{P}(A \mid H_i)$ часто оказывается проще, чем непосредственное нахождение $\mathsf{P}(A)$.

Пример 33. Из 28 костей домино[10] случайным образом выбирается одна. Какова вероятность того, что вторую случайно выбранную кость можно будет приставить к первой?

Решение. Пусть $A = \{$вторую кость можно приставить к первой$\}$. Разобьем все исходы эксперимента на две гипотезы:

$H_1 = \{$первая выбранная кость оказалась дублем$\}$,

$H_2 = \overline{H_1} = \{$первая выбранная кость оказалась не дублем$\}$.

Так как всего имеется 7 дублей, то $\mathsf{P}(H_1) = \frac{7}{28} = \frac{1}{4}$, $\mathsf{P}(H_2) = \frac{3}{4}$.

Предположим, что гипотеза H_1 осуществилась, т. е. первая выбранная кость имеет вид $\{x, x\}$. К ней можно приставить только кости вида $\{x, y\}$. Всего таких костей осталось 6. Следовательно,

$$\mathsf{P}(A \mid H_1) = \frac{6}{27} = \frac{2}{9}.$$

Предположим, что осуществилась гипотеза H_2, т. е. первая выбранная кость имеет вид $\{x, y\}$. К этой кости можно приставить любую из 12 костей вида $\{x, z\}$ или $\{y, z\}$. А значит,

$$\mathsf{P}(A \mid H_2) = \frac{12}{27} = \frac{4}{9}.$$

И наконец, по формуле полной вероятности,

$$\mathsf{P}(A) = \mathsf{P}(H_1)\mathsf{P}(A \mid H_1) + \mathsf{P}(H_2)\mathsf{P}(A \mid H_2) = \frac{1}{4}\cdot\frac{2}{9} + \frac{3}{4}\cdot\frac{4}{9} = \frac{7}{18}. \qquad \square$$

[10] Каждая кость домино представляет собой неупорядоченную пару $\{x, y\}$ чисел $x, y \subset \{0,1,\ldots,6\}$. Кость $\{x, y\}$ называется дублем, если $x = y$. Кость $\{z, t\}$ можно приставить к $\{x, y\}$, если хотя бы одно из чисел z и t совпадает хотя бы с одним из x и y.

Пример 34. В первом ящике 9 белых и 1 черный шар, а во втором — 3 черных и 1 белый. Из первого ящика наудачу выбирают один шар и, не глядя, перекладывают во второй ящик. После этого из второго ящика наудачу выбирают один шар. С какой вероятностью он окажется белым?

Решение. Обозначим $A = \{$вынутый из второго ящика шар оказался белым$\}$. Относительно выбора шара из первого ящика можно сделать два предположения:

$H_1 = \{$из первого ящика вынут белый шар$\}$,

$H_2 = \{$из первого ящика вынут черный шар$\}$.

Очевидно, $\mathsf{P}(H_1) = \frac{9}{10}$, $\mathsf{P}(H_2) = \frac{1}{10}$. Предполагая, что произошло H_1 или, соответственно, H_2, находим:

$$\mathsf{P}(A \mid H_1) = \frac{2}{5}, \qquad \mathsf{P}(A \mid H_2) = \frac{1}{5}.$$

В итоге

$$\mathsf{P}(A) = \frac{9}{10} \cdot \frac{2}{5} + \frac{1}{10} \cdot \frac{1}{5} = \frac{19}{50}. \qquad \square$$

Задача 35 (проблема раздела ставки)**.** Два равносильных игрока должны были играть до тех пор, пока один из них не выиграет 6 раз (необязательно подряд). Игра неожиданно (по техническим причинам) закончилась со счетом 4 : 3 в пользу первого игрока. Как следует справедливо поделить призовой фонд, учитывая шансы игроков на победу?

1.6.2. Формула Байеса

Пусть H_1, H_2, …, H_n — разбиение пространства Ω. И пусть нам известны вероятности $\mathsf{P}(H_i)$ и вероятности события A при условии осуществления H_i, $i = 1, 2, \ldots, n$. Предположим, что в результате проведения эксперимента событие A произошло. Эта информация изменит наше представление о вероятностях осуществления гипотез. Новую (апостериорную) вероятность гипотезы H_k можно найти по *формуле Байеса*:

$$\mathsf{P}(H_k \mid A) = \frac{\mathsf{P}(H_k A)}{\mathsf{P}(A)} = \frac{\mathsf{P}(H_k)\mathsf{P}(A \mid H_k)}{\sum\limits_{i=1}^{n} \mathsf{P}(H_i)\mathsf{P}(A \mid H_i)}.$$

Пример 36. Имеются три игральные кости. Из них две — обычные. А одна — со смещенным центром тяжести так, что $\mathsf{P}(‘6’) = \frac{1}{4}$, $\mathsf{P}(‘2’) = \mathsf{P}(‘3’) = \mathsf{P}(‘4’) = \mathsf{P}(‘5’) = \frac{1}{6}$, $\mathsf{P}(‘1’) = \frac{1}{12}$. Случайно выбрав одну кость из трех, ее бросают три раза. Все три раза выпадает ‘6’. Найти вероятность того, что была выбрана кость со смещенным центром тяжести.

Решение. Через A обозначим событие {при бросании случайно выбранной кости три раза выпало ‘6’}.

При выборе одной из трех костей возможны две взаимоисключающие ситуации:

$H_1 =$ {выбрана кость со смещенным центром тяжести},

$H_2 =$ {выбрана обычная кость}.

В задаче требуется найти условную вероятность $\mathsf{P}(H_1 \mid A)$.

Предположим, что нам неизвестны результаты описаного эксперимента (т. е. событие A еще не произошло). Тогда

$$\mathsf{P}(H_1) = \frac{1}{3}, \qquad \mathsf{P}(H_2) = \frac{2}{3}.$$

В предположении, что была выбрана кость со смещенным центром тяжести, находим

$$\mathsf{P}(A \mid H_1) = \left(\frac{1}{4}\right)^3 = \frac{1}{64}.$$

Аналогично

$$\mathsf{P}(A \mid H_2) = \left(\frac{1}{6}\right)^3 = \frac{1}{216}.$$

По формуле Байеса, искомая вероятность равна

$$\mathsf{P}(H_1 \mid A) = \frac{\mathsf{P}(H_1)\mathsf{P}(A \mid H_1)}{\mathsf{P}(H_1)\mathsf{P}(A \mid H_1) + \mathsf{P}(H_2)\mathsf{P}(A \mid H_2)} =$$

$$= \frac{\frac{1}{3} \cdot \frac{1}{64}}{\frac{1}{3} \cdot \frac{1}{64} + \frac{2}{3} \cdot \frac{1}{216}} = \frac{27}{27 + 16} = \frac{27}{43}.$$

Как и следовало ожидать, информация о том, что произошло событие A, существенно увеличивает вероятность того, что была выбрана необычная кость. □

Пример 37. Агентство по страхованию автомобилей разделяет водителей по трем группам: H_1 — водит осторожно, H_2 — обычный водитель, H_3 — высока вероятность аварий. Вероятность попасть хотя бы в одну аварию в течение года для водителя из группы H_1 равна 0,1, для водителя из 2-й группы эта вероятность равна 0,2 и для водителя из 3-й группы — 0,3. По данным компании известно, что доля первой группы составляет 30 % от всех водителей, доля второй группы — 50 % и доля третьей группы — 20 %. Некий водитель страхует в агентстве свой автомобиль и в течение года попадает в аварию. Учитывая эту информацию, оценить апостериорные вероятности принадлежности его к каждой из трех групп.

Решение. Пусть $A = \{\text{водитель в течение года попал в аварию}\}$. По условию задачи известно:

$$\mathsf{P}(H_1) = 0{,}3, \qquad \mathsf{P}(H_2) = 0{,}5, \qquad \mathsf{P}(H_3) = 0{,}2.$$

Также известны условные вероятности:

$$\mathsf{P}(A \mid H_1) = 0{,}1, \qquad \mathsf{P}(A \mid H_2) = 0{,}2, \qquad \mathsf{P}(A \mid H_3) = 0{,}3.$$

Пользуясь формулой Байеса, находим апостериорные вероятности гипотез:

$$\mathsf{P}(H_1 \mid A) = \frac{0{,}3 \cdot 0{,}1}{0{,}3 \cdot 0{,}1 + 0{,}5 \cdot 0{,}2 + 0{,}2 \cdot 0{,}3} = \frac{3}{19},$$
$$\mathsf{P}(H_2 \mid A) = \frac{0{,}5 \cdot 0{,}2}{0{,}3 \cdot 0{,}1 + 0{,}5 \cdot 0{,}2 + 0{,}2 \cdot 0{,}3} = \frac{10}{19},$$
$$\mathsf{P}(H_3 \mid A) = \frac{0{,}2 \cdot 0{,}3}{0{,}3 \cdot 0{,}1 + 0{,}5 \cdot 0{,}2 + 0{,}2 \cdot 0{,}3} = \frac{6}{19}.$$

Откуда следует весьма логичный (с точки зрения страховой компании) вывод:

$$\mathsf{P}(H_1 \mid A) < \mathsf{P}(H_1), \qquad \mathsf{P}(H_2 \mid A) \approx \mathsf{P}(H_2), \qquad \mathsf{P}(H_3 \mid A) > \mathsf{P}(H_3).$$

□

Глава 2

Испытания Бернулли

2.1. Формула Бернулли[1]

Испытаниями Бернулли называется последовательность независимых в совокупности испытаний, в каждом из которых возможны лишь два исхода — «успех» и «неудача». Вероятность успеха в одном испытании обозначается буквой p, а вероятность неудачи — q, $q = 1 - p$. Эти вероятности остаются неизменными на протяжении всего эксперимента.

Разнообразие задач, в которых эта модель оказывается полезной, довольно велико. Наиболее известным примером испытаний Бернулли является многократное бросание монеты, где под успехом понимается появление герба (решки) в отдельном бросании. Другим примером может служить проверка серии однотипных изделий на соответствие стандарту. В таком случае отдельное испытание представляет собой проверку одного экземпляра. Соответственно, под успехом можно понимать соответствие стандарту. (Наличие брака в некотором смысле тоже можно принять за успех, если с точки зрения решения задачи это оказывается удобнее.) Можно также интересоваться у случайных людей, не является ли настоящий день их днем рождения. Это тоже пример испытаний Бернулли, в которых вероятность успеха приблизительно равна $\frac{1}{365}$ (при условии, что настоящий

[1] Якоб Бернулли (1654–1705) — швейцарский математик.

день — не 29 февраля).

Для одного испытания Бернулли пространство $\Omega = \{\text{У}, \text{Н}\}$ состоит из двух исходов. Для n испытаний $\Omega = \{\text{У}, \text{Н}\}^n$. Каждый элементарный исход испытаний Бернулли можно представить последовательностью успехов и неудач. В силу независимости испытаний вероятность такого исхода равна произведению вероятностей соответствующих успехов и неудач. Например:

$$\mathsf{P}(\text{УННУУ}) = \mathsf{P}(\text{У})\mathsf{P}(\text{Н})\mathsf{P}(\text{Н})\mathsf{P}(\text{У})\mathsf{P}(\text{У}) = p^3q^2.$$

Обозначим через S_n число успехов в n испытаниях Бернулли. Во многих задачах важно уметь находить вероятность $\mathsf{P}(S_n = k)$ того, что n испытаний Бернулли закончатся ровно k успехами и $n-k$ неудачами. Заметим, что событие $S_n = k$ содержит столько элементарных исходов, сколько существует последовательностей, состоящих из k букв У и $n-k$ букв Н, то есть из C_n^k элементарных исходов. Причем каждый элементарный исход имеет вероятность p^kq^{n-k}. Результатом этих рассуждений является следующая формула.

Теорема 2.1 (формула Бернулли)**.**

$$\mathsf{P}(S_n = k) = C_n^k p^k q^{n-k}, \quad \textit{где } q = 1-p, \quad k = 0, 1, \ldots, n.$$

Заметим, что сумма всех чисел $\mathsf{P}(S_n = n)$ при фиксированном n равна единице, согласно формуле бинома Ньютона:

$$1 = (p+q)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k}.$$

По этой причине набор чисел $\mathsf{P}(S_n = n)$, $k = 0, 1, \ldots, n$, называется *биномиальным распределением.*

Пример 38. Тестовое задание состоит из шести вопросов. Для каждого вопроса предлагается 3 варианта ответов, только один из которых правильный. Найти вероятность того, что человек, не знающий верных ответов и выбирающий их наудачу, ответит правильно не менее чем на 3 вопроса.

Решение. Переведем условие задачи на язык испытаний Бернулли.

Проводятся 6 испытаний с вероятностью $p = \frac{1}{3}$ успеха в каждом. Требуется найти $\mathsf{P}(S_6 \geqslant 3)$. По формуле Бернулли:

$$\mathsf{P}(S_6 \geqslant 3) = \sum_{k=3}^{6} \mathsf{P}\big(S_6 = k\big) = \sum_{k=3}^{6} \frac{6!}{k!(6-k)!} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{6-k} = \\ = \frac{20 \cdot 2^3 + 15 \cdot 2^2 + 6 \cdot 2 + 1}{3^6} = \frac{233}{729}. \qquad \square$$

Пример 39 (задача о спичечных коробках Банаха[2])**.** Некий математик имел привычку носить в каждом из двух карманов пальто по коробку спичек. Всякий раз, когда ему хотелось закурить трубку, он выбирал наугад один из коробков и доставал из него спичку. Первоначально в каждом коробке было по N спичек. Найти вероятность того, что в момент, когда математик впервые вытащит пустой коробок, в другом останется ровно k спичек.

Решение. Интересующее нас событие (обозначим его A) естественным образом разбивается на два:

$A_1 = \{$пустая коробка была извлечена из левого кармана$\}$,

$A_2 = \{$пустая коробка была извлечена из правого кармана$\}$.

Очевидно, $\mathsf{P}(A_1) = \mathsf{P}(A_2)$. Поэтому далее сосредоточимся на поиске вероятности события A_1.

При выборе одного из карманов под успехом будем понимать выбор левого кармана. Тогда $p = q = \frac{1}{2}$. Согласно условию задачи, до извлечения пустой коробки должно быть вынуто N спичек из левой коробки и $N-k$ спичек — из правой, т. е. $2N-k$ испытаний должны закончиться ровно N успехами. После этого с вероятностью $p = \frac{1}{2}$ будет вынута левая коробка. Итак,

$$\mathsf{P}(A_1) = \frac{1}{2}\mathsf{P}\big(S_{2N-k} = N\big) = \frac{1}{2} C_{2N-k}^{N} \left(\frac{1}{2}\right)^{2N-k}.$$

Следовательно, искомая вероятность равна

$$\mathsf{P}(A) = 2\mathsf{P}(A_1) = \frac{C_{2N-k}^{N}}{2^{2N-k}}. \qquad \square$$

[2]Эта шутливая задача о привычках известного польского математика Стефана Банаха была придумана Гуго Дионисием Штейнгаузом.

Утверждение 2.2 (наиболее вероятное число успехов)**.** *Пусть проводятся n испытаний Бернулли с вероятностью успеха p. Тогда вероятность $\mathsf{P}(S_n = k)$ принимает наибольшее значение при всяком целом k, принадлежащем отрезку $[np - q, np + p]$.*

Доказательство. Рассмотрим соотношение

$$\frac{\mathsf{P}(S_n = k)}{\mathsf{P}(S_n = k-1)} = \frac{C_n^k p^k q^{n-k}}{C_n^{k-1} p^{k-1} q^{n-k+1}} = \frac{(n-k+1)p}{kq}, \tag{2.1}$$

из которого следует, что неравенство $P_n(k) > P_n(k-1)$ эквивалентно неравенству $(n-k+1)p > kq$, то есть $(n+1)p > k(p+q) = k$. Следовательно, с ростом k вероятность $P_n(k)$ растет, пока выполняется $k < (n+1)p$, и убывает, если $k > (n+1)p$. Таким образом, максимум $P_n(k)$ достигается в наибольшем целом k, удовлетворяющем условию $k \leqslant (n+1)p$. Если $(n+1)p$ — целое, то имеется еще одна точка максимума $k = (n+1)p - 1 = np - q$. □

2.2. Приближение Пуассона[3]

Пример 40. Для того чтобы получить суперприз в игре «5 из 36», необходимо угадать 5 чисел из 36. Найти вероятность того, что хотя бы два из 750 тыс. билетов, участвующих в розыгрыше, выиграют суперприз.

Решение. Перейдем к рассмотрению противоположного события: $\overline{A} = \{$выигрышным окажется только один или ни одного билета$\}$.

По существу, в данной задаче мы имеем дело с $n = 750\,000$ испытаниями Бернулли. Вероятность выиграть суперприз для одного билета равна

$$p = \frac{1}{C_{36}^5} = \frac{31!5!}{36!} = \frac{1}{376\,992}.$$

По формуле Бернулли,

$$\mathsf{P}(\overline{A}) = \mathsf{P}(S_n = 0) + \mathsf{P}(S_n = 1) = (1-p)^n + np(1-p)^{n-1}.$$

[3] Симеон Дени Пуассон (1781–1840) — французский математик и физик.

Учитывая, что p близко к нулю, воспользуемся известной формулой $\ln(1+p) \sim p$:

$$(1-p)^n = e^{n\ln(1-p)} \approx e^{-np}.$$

Так как p — мало, то $(1-p)^{n-1} \approx (1-p)^n$. Введем обозначение $\lambda = np = \frac{750\,000}{376\,992}$. Тогда

$$\mathsf{P}(A) \approx 1 - e^{-\lambda} - \lambda e^{-\lambda} = 0{,}591\ldots \qquad \square$$

Рассуждения, используемые в примере с игрой «5 из 36» для приближенного вычисления вероятностей, можно применять во всех задачах, где число испытаний n велико, а вероятность успеха p близка к нулю. Обобщение этой идеи носит название *закона редких событий* или *закона малых чисел* и имеет многочисленные приложения в физике, теории массового обслуживания, теории надежности и т. д.

Теорема 2.3 (приближение Пуассона)**.** *Пусть в испытаниях Бернулли $n \to \infty$ и $p = p(n) \to 0$, причем $np = \lambda_n \to \lambda$. Тогда для любого неотрицательного целого k выполняется*

$$\mathsf{P}(S_n = k) = C_n^k p^k q^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

Доказательство. В формуле Бернулли сделаем замену $p = \frac{\lambda_n}{n}$:

$$C_n^k p^k q^{n-k} = \frac{n(n-1)\ldots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} =$$
$$= \frac{\lambda_n^k}{k!} \cdot \frac{n(n-1)\ldots(n-k+1)}{n^k}\left(1 - \frac{\lambda_n}{n}\right)^{-k}\left(1 - \frac{\lambda_n}{n}\right)^{n}.$$

Заметим, что

$$\frac{n(n-1)\ldots(n-k+1)}{n^k} = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right) \to 1,$$

так как k фиксировано, а $n \to \infty$. Аналогично

$$\left(1 - \frac{\lambda_n}{n}\right)^{-k} \to 1, \quad \text{при } n \to \infty.$$

Пользуясь известной формулой $\ln(1+x) \sim x$ при $x \to 0$, получаем

$$\left(1 - \frac{\lambda_n}{n}\right)^n = e^{n \ln\left(1 - \frac{\lambda_n}{n}\right)} \sim e^{-\lambda_n} \to e^{-\lambda}.$$

Откуда и следует справедливость теоремы. $\square$

Результат этой теоремы обычно представляется в виде приближенной формулы Пуассона:

$$\mathsf{P}(S_n = k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{где } \lambda = np,$$

достаточно точной при больших значениях n и малых p. В тех приложениях, где эта формула оказывется полезной, λ интерпретируется как среднее (ожидаемое) число успехов.

Далее будем пользоваться обозначением

$$\Pi(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{где } \lambda = np.$$

Для точности приближения Пуассона имеется следующая оценка.

Теорема 2.4. *Для любого набора индексов $I \subset \{0, 1, 2, \ldots\}$*

$$\left| \sum_{k \in I} \Big(\mathsf{P}(S_n = k) - \Pi(k) \Big) \right| \leqslant \min\{np^2, p\}.$$

Реальные значения погрешности $\Delta(n,p) = \max\limits_k |\mathsf{P}(S_n = k) - \Pi(k)|$ для некоторых значений n и p приводятся в таблице 2.1. Наибольшее значение погрешности $|\mathsf{P}(S_n = k) - \Pi(k)|$ достигается на одном из целых k, ближайших к λ.

Замечание. Ранее было показано, что для биномиального распределения вероятностей справедлива формула (2.1):

$$\mathsf{P}(S_n = k) = \frac{(n-k+1)p}{kq} \cdot \mathsf{P}(S_n = k-1).$$

Аналогичная формула для распределения Пуассона:

$$\Pi(k) = \frac{\lambda}{k} \cdot \Pi(k-1).$$

Таблица 2.1

Погрешность $\boldsymbol{\Delta(n,p) = \max_k |\mathsf{P}(S_n = k) - \Pi(k)|}$ приближения Пуассона для некоторых значений $\boldsymbol{n}$ и $\boldsymbol{p}$

n	10	20	100	500
p	0,2	0,05	0,1	0,01
$\Delta(n,p)$	0,0313	0,0095	0,0068	0,0009

Ясно, что при больших n и малых p справедливо

$$\frac{(n-k+1)p}{kq} = \frac{np}{k} \cdot \frac{1 - \frac{k-1}{n}}{1-p} \approx \frac{\lambda}{k},$$

что дает еще один способ вывода приближенной формулы Пуассона для $\mathsf{P}(S_n = k)$, начиная с $\mathsf{P}(S_n = 0) \approx e^{-\lambda}$.

Пример 41. При изготовлении булочек на хлебозаводе тесто тщательно перемешивается с изюмом. Предположим, что за один цикл в чан с тестом закладывается n изюминок и изготавливается k булочек. Выберем случайно одну из свежеприготовленных булочек. При ее изготовлении каждая из n изюминок могла попасть в эту булочку с вероятностью

$$p = \frac{1}{\text{число всех булочек}} = \frac{1}{k}.$$

Следовательно, число изюминок в булочке подчиняется закону Пуассона с параметром $\lambda = np$. Причем параметр λ, очевидно, определяет среднее число изюминок в булочке.

Задача 42. В среднем левши составляют 10 % населения. Каким должен быть размер случайной выборки людей, чтобы с вероятностью 99 % в ней оказался хотя бы один левша?

2.3. Локальная теорема Муавра–Лапласа[4]

При сравнении графиков функции $f_{n,p}(k) = \mathsf{P}(S_n = k)$ для различных значений параметров n и p обнаруживается, что все они имеют одинаковую колоколообразную форму (см. рис. 2.1). Отличие, по существу, заключается лишь в смещении «колокола» по оси Ox и в коэффициентах растяжения графика по вертикали и горизонтали. Иными словами, можно подобрать некую, не зависящую от параметров n и p, функцию $\psi(x)$ такую, что

$$f_{n,p}(k) \approx a \cdot \psi\left(\frac{k-c}{b}\right), \tag{2.2}$$

где коэффициенты растяжения a и b и величина смещения c зависят только от параметров n и p.

Не уменьшая общности, можно считать, что вершина колоколообразной кривой $y = \psi(x)$ располагается в точке $(0,1)$. В таком случае, согласно формуле (2.2), вершина графика функции $y = f_{n,p}(x)$ будет лежать в точке (c, a). Откуда можно найти формулы для вычисления параметров c и a. В частности, из утверждения 2.2 следует, что $c \in [np - q, np + p]$. Оптимальный вариант:

$$c = np. \tag{2.3}$$

Теперь найдем $a \approx f_{n,p}(c)$. (Заметим, что функция $f_{n,p}$ определена только для целых значений аргумента. С другой стороны, целью данных рассуждений является приближенная формула. Поэтому противоречие не возникает.) Значение $f_{n,p}(c)$ найдем приближенно, воспользовавшись формулой Стирлинга $n! \approx \sqrt{2\pi n}\; n^n e^{-n}$:

---

[4] Абрахам де Муавр (1667–1754, английский математик французского происхождения) доказал частный случай теоремы для $p = 1/2$ в 1733 г. В общем виде теорема была доказана Лапласом в 1812 г.

Пьер-Симон, маркиз де Лаплас (1749–1827) — французский математик, механик, физик и астроном.

$$
\begin{aligned}
f_{n,p}(k) = \frac{n!\,p^k\,q^{n-k}}{k!(n-k)!} \approx& \\
\approx \frac{\left(\sqrt{2\pi n}\;n^n\,\cancel{e^{-n}}\right)p^k\,q^{n-k}}{\left(\sqrt{2\pi k}\;k^k\,\cancel{e^{-k}}\right)\left(\sqrt{2\pi(n-k)}\;(n-k)^{n-k}\,\cancel{e^{k-n}}\right)} =& \\
= \frac{1}{\sqrt{\frac{2\pi k(n-k)}{n}}}\cdot\frac{n^n\,p^k\,q^{n-k}}{k^k(n-k)^{n-k}} =& \qquad (2.4)\\
= \frac{\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}}{\sqrt{\frac{2\pi k(n-k)}{n}}}.&
\end{aligned}
$$

Подставляя $k = c = np$ (соответственно, $n-k = n-np = nq$), получаем

$$f_{n,p}(c) \approx \frac{1}{\sqrt{2\pi npq}}.$$

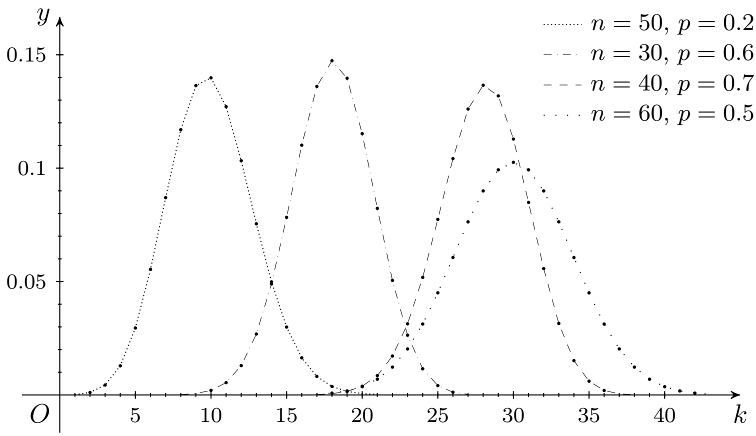

Рис. 2.1. Графики функций $y = f_{n,p}(k) = \mathsf{P}(S_n = k)$ для некоторых значений n и p

Это и есть значение параметра a:

$$a = \frac{1}{\sqrt{2\pi npq}}. \tag{2.5}$$

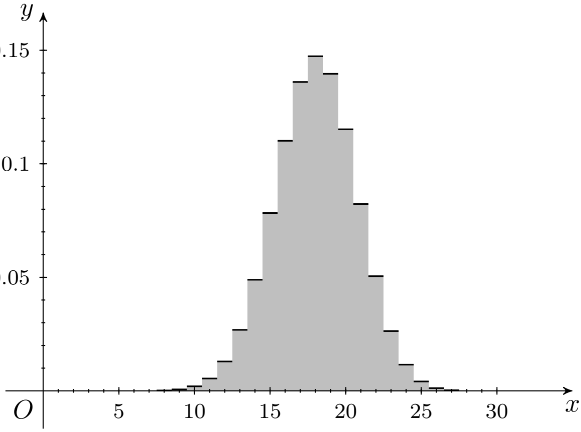

Рис. 2.2. Площадь под графиком функции
$\tilde{f}_{n,p}(x) = f_{n,p}\left(\lfloor x + 0.5 \rfloor\right)$, где $x \in [-0.5, n + 0.5)$,
всегда равна единице. (Изображен случай $n = 30$, $p = 0.6$.)

Для определения параметра b воспользуемся следующим рассуждением. Вместо функции $f_{n,p}(k)$ рассмотрим ее интегрируемый аналог — ступенчатую функцию

$$\tilde{f}_{n,p}(x) = f_{n,p}\left(\lfloor x + 0.5 \rfloor\right), \quad x \in \left[-0.5, n + 0.5\right).$$

Площадь под ее графиком (см. рис. 2.2) равна единице вне зависимости от выбора параметров n и p, так как

$$\sum_{k=0}^{n} f_{n,p}(k) = 1.$$

Следовательно, площадь под графиком функции $a \cdot \psi\left(\frac{k-c}{b}\right)$ тоже должна быть инвариантна относительно выбора параметров n и p.

Сдвиг $\psi(x-c)$ графика функции $\psi(x)$ по оси Ox, очевидно, не меняет площадь. Тогда как растяжения $a\cdot\psi(x)$ и $\psi(x/b)$ увеличивают площадь в a и в b раз соответственно. Следовательно, для сохранения площади требуется выполнение условия $ab = \mathrm{const}$. Полагая для удобства вычислений $ab = \frac{1}{\sqrt{2\pi}}$, получаем

$$b = \sqrt{npq}. \tag{2.6}$$

Остается подставить в (2.2) значения параметров из формул (2.3), (2.5) и (2.6):

$$f_{n,p}(k) \approx \frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{2\pi}}\cdot\psi\left(\frac{k-np}{\sqrt{npq}}\right).$$

Оказывается, такая универсальная функция $\psi(x)$ действительно существует (см. рис. 2.3). Ее точное представление было найдено Лапласом в 1812 году.

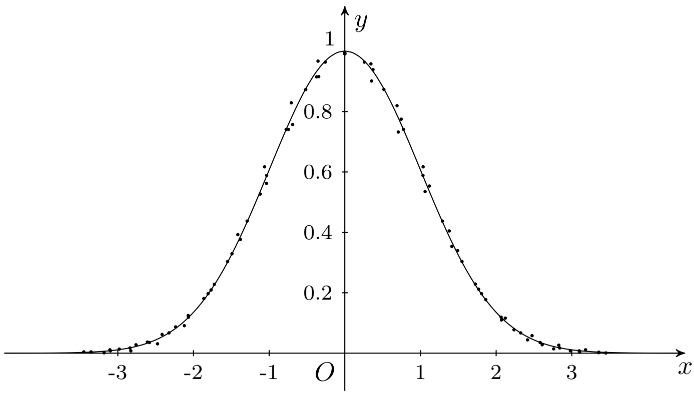

Рис. 2.3. Точки с рисунка 2.1, приведенные к виду $\left(\frac{k-c}{b}, \frac{f_{n,p}(k)}{a}\right)$, и график кривой Гаусса $y = \psi(x) = e^{-x^2/2}$

В настоящее время функция $\dfrac{1}{\sqrt{2\pi}}\cdot\psi(x)$ называется *плотностью нормального распределения* и обозначается $\varphi(x)$:

$$\varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}. \tag{2.7}$$

Таблица ее значений имеется в Приложении 1.

Некоторые свойства функции $\varphi(x)$**:**

1. $\varphi(-x) = \varphi(x)$ (четность);

2. $\int\limits_{-\infty}^{+\infty} \varphi(x)\,dx = 1$.

Теорема 2.5 (локальная теорема Муавра–Лапласа)**.**

$$\mathsf{P}(S_n = k) \approx \frac{1}{\sqrt{npq}}\,\varphi\left(\frac{k-np}{\sqrt{npq}}\right), \quad \textit{где } \varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}. \tag{2.8}$$

Доказательство. Точность формулы (2.8) пропорциональна коэффициенту $\frac{1}{\sqrt{npq}}$, стоящему перед функцией $\varphi(x)$ в правой части формулы. Поэтому далее предполагаем, что npq относительно велико. Обычно формулой Муавра–Лапласа рекомендуется пользоваться при $npq > 20$.

Заметим, что при k, значительно отличающихся от np, вероятность $\mathsf{P}(S_n = k)$ практически равна нулю. Это следует из соотношения (2.1) и иллюстрируется рисунками 2.1 и 2.3.

Правая часть формулы (2.8) также практически равна нулю уже при $\left|\frac{k-np}{\sqrt{npq}}\right| \geqslant 3$, так как $\varphi(x) \leqslant 0{,}0045$ при $|x| \geqslant 3$ (на рис. 2.3 изображен график функции $y = \sqrt{2\pi}\varphi(x) \approx 2{,}5\,\varphi(x)$). Поэтому далее будем рассматривать только те случаи, для которых разность

$$x_k = \frac{k}{n} - p \tag{2.9}$$

мала по абсолютной величине. Точнее, будем предполагать, что

$$|x_k| = \varepsilon\,\frac{\sqrt{npq}}{n}, \quad \text{где } 0 \leqslant \varepsilon < 3. \tag{2.10}$$

Это предположение эквивалентно неравенству $\left|\frac{k-np}{\sqrt{npq}}\right| < 3$.

Ниже понадобится следующее очевидное замечание:

$$\left|\frac{x_k}{p}\right| = \varepsilon\,\frac{\sqrt{npq}}{np} \leqslant \frac{\varepsilon}{\sqrt{npq}}, \quad \text{и} \quad \left|\frac{x_k}{q}\right| \leqslant \frac{\varepsilon}{\sqrt{npq}}. \tag{2.11}$$

Другими словами, левые части неравенств (2.11) относительно малы. Во всяком случае, они меньше единицы при $npq \geqslant 9$.

Воспользуемся выведенной выше приближенной формулой (2.4). Сделаем замену (2.9):

$$k = n(p + x_k), \qquad n - k = n - n(x_k + p) = n(q - x_k).$$

Получаем

$$\mathsf{P}(S_n = k) \approx \frac{\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}}{\sqrt{\frac{2\pi k(n-k)}{n}}} = \frac{\left(\frac{p}{p+x_k}\right)^{n(p+x_k)} \left(\frac{q}{q-x_k}\right)^{n(q-x_k)}}{\sqrt{2\pi n(p+x_k)(q-x_k)}}. \tag{2.12}$$

Прологарифмируем числитель:

$$n(p+x_k)\ln\left(\frac{p}{p+x_k}\right) + n(q-x_k)\ln\left(\frac{q}{q-x_k}\right) =$$
$$= -n\left\{(p+x_k)\ln\left(1+\frac{x_k}{p}\right) + (q-x_k)\ln\left(1-\frac{x_k}{q}\right)\right\}.$$

Воспользуемся разложением Тейлора для $\ln(1+t) = t - \frac{t^2}{2} + \frac{t^3}{6} - \ldots$ и тем фактом, что $\left|\frac{x_k}{p}\right|$ и $\left|\frac{x_k}{q}\right|$ относительно малы (см. оценки (2.11)). Тогда логарифм числителя равен

$$-n\left\{(p+x_k)\left(\frac{x_k}{p} - \frac{x_k^2}{2p^2} + \frac{x_k^3}{6p^3} - \ldots\right) + \right.$$
$$\left. +(q-x_k)\left(-\frac{x_k}{q} - \frac{x_k^2}{2q^2} - \frac{x_k^3}{6q^3} - \ldots\right)\right\} =$$
$$= -n\left\{\left(x_k + \frac{x_k^2}{2p} - \frac{x_k^3}{3p^2} + \ldots\right) + \left(-x_k + \frac{x_k^2}{2q} + \frac{x_k^3}{3q^2} + \ldots\right)\right\} =$$
$$= -n\left\{\frac{x_k^2(p+q)}{2pq} + \frac{x_k^3(p^2-q^2)}{3p^2q^2} + \ldots\right\} =$$
$$= \frac{-nx_k^2}{pq}\left\{\frac{1}{2} + \frac{x_k(p-q)}{3pq} + \ldots\right\}.$$

Согласно формуле (2.10), имеем $\frac{x_k(p-q)}{3pq} = \frac{\varepsilon(p-q)}{3\sqrt{npq}}$. Предполагая, что npq велико, отбросим второе и последующие слагаемые в фигурных скобках. Подставляя результат в (2.12) и делая замену, обратную (2.9), окончательно получаем

$$\mathsf{P}(S_n = k) \approx \frac{e^{\frac{-nx_k^2}{2pq}}}{\sqrt{2\pi n(p+x_k)(q-x_k)}} \approx \frac{e^{-\frac{(k-np)^2}{2npq}}}{\sqrt{2\pi npq}} = \frac{1}{\sqrt{npq}}\varphi\left(\frac{k-np}{\sqrt{npq}}\right).$$

□

Значения погрешности

$$\Delta(n,p) = \max_k \left|\mathsf{P}(S_n = k) - \frac{1}{\sqrt{npq}} \cdot \varphi\left(\frac{k-np}{\sqrt{npq}}\right)\right|$$

локальной формулы Муавра–Лапласа для некоторых значений n и p приводятся в таблице 2.2. Погрешность формулы пропорциональна $\frac{1}{\sqrt{npq}}$. А при фиксированном $\frac{1}{\sqrt{npq}}$ наибольшая точность достигается при значениях p близких к $\frac{1}{2}$.

Таблица 2.2

Погрешность $\boldsymbol{\Delta(n,p) = \max_k \left|\mathsf{P}(S_n = k) - \frac{1}{\sqrt{npq}} \cdot \varphi\left(\frac{k-np}{\sqrt{npq}}\right)\right|}$

n	80	100	125	230	450
p	0,5	0,3	0,2	0,1	0,05
npq	20	21	20	20,7	21,4
$\Delta(n{,}p)$	0,0003	0,0018	0,0028	0,0037	0,0041

Пример 43. Найдем вероятность того, что при 100 бросаниях симметричной монеты герб выпадет не менее 49 и не более 51 раза. Значение $npq = 100 \cdot 0{,}5 \cdot 0{,}5 = 25$ достаточно велико. Кроме того, $p = \frac{1}{2}$, что существенно повышает точность формулы Муавра–Лапласа.

$$\mathsf{P}(49 \leqslant S_n \leqslant 51) = \mathsf{P}(S_n = 49) + \mathsf{P}(S_n = 50) + \mathsf{P}(S_n = 51) \approx$$
$$\approx \frac{1}{\sqrt{25}}\left(\varphi\left(\frac{49 - 100 \cdot 0{,}5}{\sqrt{25}}\right) + \right.$$
$$\left. + \varphi\left(\frac{50-50}{5}\right) + \varphi\left(\frac{51-50}{5}\right)\right) =$$

$$= \frac{\varphi(-0{,}2) + \varphi(0) + \varphi(0{,}2)}{5} \approx$$
$$\approx \frac{0{,}391 + 0{,}3989 + 0{,}391}{5} \approx 0{,}236.$$

Точное значение в данном случае отличается менее чем на 0,001.

2.4. Интегральная теорема Муавра–Лапласа

Во многих задачах (см. пример 43), связанных с испытаниями Бернулли, требуется находить вероятность $\mathsf{P}(m_1 \leqslant S_n \leqslant m_2)$ того, что число успехов попадет в некоторый промежуток $[m_1, m_2]$. Если промежуток $[m_1, m_2]$ велик, то это потребует большого объема вычислений. Избежать этого помогает интегральная формула Муавра–Лапласа.

Определение 2.6. Функция $\Phi(x) = \int\limits_0^x \varphi(t)\,dt$, где $\varphi(t) = \dfrac{e^{-t^2/2}}{\sqrt{2\pi}}$, называется *функцией Лапласа.*

В Приложении 2 имеется таблица значений этой функции. А ее график изображен на рис. 2.4.

Некоторые свойства функции Лапласа:

1. $\Phi(-x) = -\Phi(x)$ (нечетность);

2. $\Phi(+\infty) = -\Phi(-\infty) = 1/2$.

Теорема 2.7 (интегральная теорема Муавра–Лапласа)**.** *Вероятность того, что в n испытаниях Бернулли число успехов S_n попадет в полуинтервал $[m_1; m_2)$, может быть приближенно найдена по формуле*

$$\mathsf{P}(m_1 \leqslant S_n < m_2) \approx \Phi(x_2) - \Phi(x_1), \quad \textit{где } x_i = \frac{m_i - np}{\sqrt{npq}},\ i = 1, 2.$$

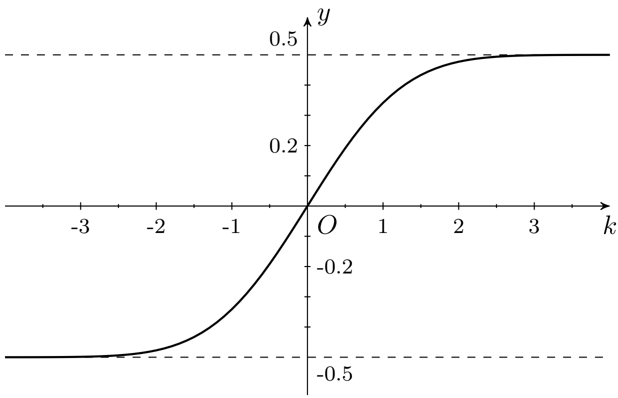

Рис. 2.4. Функция Лапласа $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-t^2/2}\, dt$

Доказательство. Введем обозначения:

$$t(k) = \frac{k - np}{\sqrt{npq}} \quad \text{и} \quad \Delta t = t(k+1) - t(k) = \frac{1}{\sqrt{npq}}.$$

Воспользуемся локальной формулой Муавра–Лапласа (2.8):

$$\mathsf{P}(S_n = k) \approx \frac{1}{\sqrt{npq}}\, \varphi(t(k)) = \varphi(t(k))\Delta t.$$

Выражение в правой части представляет собой площадь прямоугольника ширины Δt и высоты $\varphi(t(k))$ и приближенно равно площади под графиком функции $y = \varphi(t)$ на отрезке $t \in \left[t\left(k - \frac{1}{2}\right), t\left(k + \frac{1}{2}\right)\right]$ (см. рис. 2.5).

Другими словами,

$$\mathsf{P}(S_n = k) \approx \varphi(t(k))\Delta t \approx \int_{t\left(k-\frac{1}{2}\right)}^{t\left(k+\frac{1}{2}\right)} \varphi(t)\, dt.$$

Причем чем меньше «ширина» Δt, тем выше точность приближения.

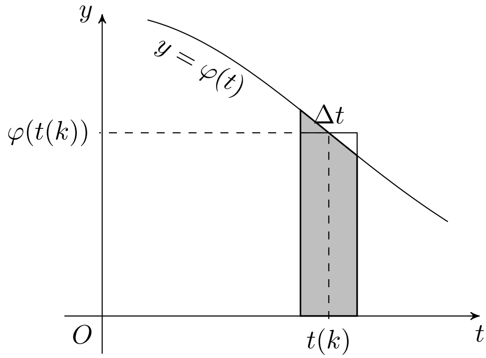

Рис. 2.5. Площадь прямоугольника ширины Δt и высоты $\varphi(t(k))$ приближенно равна площади криволинейной трапеции под графиком функции $\varphi(t)$, $t \in \left[t\left(k-\frac{1}{2}\right), t\left(k+\frac{1}{2}\right)\right]$

Реализуя эту замену для всех целых $k \in [m_1, m_2)$, получаем

$$\begin{aligned}
\mathsf{P}(m_1 \leqslant k < m_2) &= \sum_{k=m_1}^{m_2-1} \mathsf{P}(S-n=k) \approx \\
&\approx \sum_{k=m_1}^{m_2-1} \int_{t\left(k-\frac{1}{2}\right)}^{t\left(k+\frac{1}{2}\right)} \varphi(t)\,dt = \int_{t\left(m_1-\frac{1}{2}\right)}^{t\left(m_2-\frac{1}{2}\right)} \varphi(t)\,dt \approx \\
&\approx \int_{t(m_1)}^{t(m_2)} \varphi(t)\,dt = \Phi\left(\frac{m_2-np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1-np}{\sqrt{npq}}\right).
\end{aligned}$$

□

Оценить точность интегральной формулы Муавра–Лапласа позволяет следующее утверждение.

Теорема 2.8 (неравенство Бе́рри–Эссе́ена[5])**.**

$$\left|\mathsf{P}\left(x_1 < \frac{k-np}{\sqrt{npq}} < x_2\right) - \big(\Phi(x_2)-\Phi(x_1)\big)\right| \leqslant \frac{C(p^2+q^2)}{\sqrt{npq}}, \quad C \approx 0{,}4.$$

[5]Это неравенство было независимо выведено Эндрю Берри в 1941 г. и Карлом-Густавом Эссееном в 1942 г.

Замечание. Как правило, погрешность формулы Муавра–Лапласа значительно меньше правой части неравенства Берри–Эссеена.

Пример 44. Найдем вероятность того, что при 100 бросаниях монеты герб появится не более 35 раз.

Прежде всего $np = 100/2 = 50$, $npq = 25$ — определяет относительно хорошую точность формулы Муавра–Лапласа (особенно при $p = 1/2$). Подставляем исходные данные в формулу:

$$\mathsf{P}(0 \leqslant S_n < 36) \approx \Phi\left(\frac{36-50}{5}\right) - \Phi\left(\frac{0-50}{5}\right) = \\ = \Phi(-2{,}8) - \Phi(-10) = -\Phi(2{,}8) + \Phi(10).$$

По таблице (см. Приложение 2) находим $\Phi(2{,}8) \approx 0{,}4974$. Обратим внимание, что $\Phi(x)$ — возрастающая функция и

$$0{,}5 > \Phi(x) > \Phi(5) \approx 0{,}4999997, \quad \text{при } x > 5,$$

т. е. можно считать $\Phi(x) \approx 0{,}5$ при $x > 5$. Ошибка при этом окажется не более 0,0000003. (Поэтому для $x > 5$ значения $\Phi(x)$ в таблице не приводятся.) Итак, $\mathsf{P}(0 \leqslant S_n < 36) \approx 0{,}5 - 0{,}4974 = 0{,}0026$. Заметим, что в этом примере мы фактически нашли $\mathsf{P}(-\infty < S_n < 36)$.

2.4.1. Вероятность отклонения относительной частоты от вероятности успеха

Пусть проводятся n испытаний Бернулли. Соотношение $\frac{S_n}{n}$ числа всех появившихся успехов к числу n всех испытаний называется *относительной частотой* появления «успеха», или *статистической вероятностью.* Найдем вероятность того, что относительная частота $\frac{S_n}{n}$ будет отличаться от вероятности p «успеха» не более чем на ε:

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = \mathsf{P}(-\varepsilon n + np < S_n < \varepsilon n + np) \approx \\ \approx \Phi\left(\frac{\varepsilon n}{\sqrt{npq}}\right) - \Phi\left(\frac{-\varepsilon n}{\sqrt{npq}}\right) = 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \tag{2.13}$$

Пример 45. Проводится опрос, цель которого — выяснить процент сторонников партии X среди населения всей страны.

Чтобы получить точное значение, потребовалось бы опросить всех граждан. Предположим, что нас устраивает погрешность не более 1 % (число ε в формуле (2.13)), полученная с вероятностью не менее 95 % (доверительный уровень). Нужно найти наименьшее число опрашиваемых, достаточное для соблюдения этих условий.

Решение. Согласно формуле (2.13),

$$0{,}95 \leqslant \mathsf{P}\left(\left|\frac{S_n}{n} - p\right| < 0{,}01\right) \approx 2\Phi\left(0{,}01\sqrt{\frac{n}{pq}}\right). \tag{2.14}$$

По таблице значений функции Лапласа (Приложение 2) находим, что неравенство $\Phi(x) \geqslant 0{,}475$ выполняется при $x \geqslant 1{,}96$. Следовательно, неравенство (2.14) эквивалентно $0{,}01\sqrt{\frac{n}{pq}} \geqslant 1{,}96$, или, что то же самое,

$$\sqrt{n} \geqslant 196\sqrt{p(1-p)}.$$

Вероятность p нам неизвестна. Но $p \in [0,1]$, а значит, $0 \leqslant p(1-p) \leqslant \frac{1}{4}$. Причем наибольшее значение достигается при $p = \frac{1}{2}$. Следовательно, ориентируясь на «худший» случай, получаем

$$n \geqslant 98^2.$$

Обычно в таких ситуациях берут n не менее 10 000. □

Глава 3

Случайные величины

Во многих случайных экспериментах интерес в первую очередь представляют не сами исходы, а их некоторые числовые характеристики. Так, например, в испытаниях Бернулли важной характеристикой является число реализовавшихся успехов, при бросании игральной кости (или нескольких костей) результатом является число (или сумма чисел), при случайном выборе человека нам могут быть интересны его рост или вес и т. д.

Определение 3.1. *Случайной величиной* называется функция, заданная на пространстве элементарных исходов Ω и принимающая числовые значения:

$$X:\ \Omega \to \mathbb{R}.$$

Замена исходов эксперимента их числовыми характеристиками существенно упрощает соответствующую математическую модель и связанные с ней исследования. Абстрагируясь от исходного пространства Ω, вместо случайных величин обычно рассматривают их законы распределения.

Определение 3.2. Говорят, что *закон распределения случайной величины X известен*, если для любых чисел $a, b \in \mathbb{R}$, $a \leqslant b$ определена вероятность $\mathsf{P}(a \leqslant X \leqslant b)$.

Знакомство со случайными величинами и их свойствами удобнее начинать на примере дискретных случайных величин.

3.1. Дискретные случайные величины

Определение 3.3. Случайная величина называется *дискретной*, если множество ее значений конечно либо счетно.

Закон распределения дискретной случайной величины естественно представлять в виде *ряда (таблицы) распределения*, отражающего зависимость между возможными значениями x_i, $i = 1, 2, \ldots$, случайной величины X и вероятностями $p_i = \mathsf{P}(X = x_i)$:

X	x_1	x_2	x_3	$\ldots$
P	p_1	p_2	p_3	$\ldots$

Иногда, с целью экономии времени и места, далее вместо таблицы будет приводиться функция $p_i = \mathsf{P}(X = x_i)$, $i = 1, 2, \ldots$

Так как при любом исходе эксперимента случайная величина принимает ровно одно значение, то

$$\sum_i p_i = \sum_i \mathsf{P}(X = x_i) = 1.$$

3.1.1. Примеры дискретных случайных величин

Рассмотрим несколько примеров наиболее известных распределений.

1. *Вырожденная случайная величина* — (не)случайная величина с единственным возможным значением $x_1 = C$. Соответственно, таблица распределения для нее имеет вид:

C	C
P	1

2. Пусть A — некоторое событие. *Индикатором события* A называется случайная величина I_A, принимающая значение 1, если в ходе эксперимента событие A произошло, и значение 0, если событие A не произошло. Таблица распределения такой случайной величины имеет вид:

I_A	0	1
P	$1 - P(A)$	$P(A)$

3. Пусть проводится серия из n независимых однотипных испытаний с вероятностью успеха p в каждом из них (серия испытаний Бернулли), а случайная величина X — количество успехов в этой серии испытаний. В этом случае говорят, что случайная величина X распределена *по биномиальному закону с параметрами n и p.*

Множество значений такой случайной величины есть множество целых чисел $0, 1, \dots, n$, а вероятности $p_k = \mathsf{P}(X = k)$, $k = 0, 1, \dots, n$ определяются формулой Бернулли $p_k = C_n^k p^k q^{n-k}$ (здесь $q = 1 - p$ — вероятность «неудачи»). Таблица распределения:

X	0	1	$\dots$	k	$\dots$	n
P	q^n	$C_n^1 p q^{n-1}$	$\dots$	$C_n^k p^k q^{n-k}$	$\dots$	p^n

4. Предположим, что испытания Бернулли проводятся до появления первого успеха. Обозначим X — число проведенных испытаний. Говорят, что такая случайная величина распределена *по геометрическому закону с параметром p* (вероятность успеха в одном испытании). Согласно определению, такая величина X принимает значения $k \in \mathbb{N}$ с вероятностями $p_k = q^{k-1}p$.

X	1	2	3	$\dots$	k	$\dots$
P	p	qp	q^2p	$\dots$	$q^{k-1}p$	$\dots$

5. Случайная величина, распределенная *по закону Пуассона с параметром λ*, принимает значения $k = 0, 1, 2, \dots$ с вероятностями $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$.

X	0	1	2	$\dots$	k	$\dots$
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2!} e^{-\lambda}$	$\dots$	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\dots$

6. Рассмотрим множество, состоящее из N элементов, среди которых имеется K помеченных. Из этого множества случайным образом выбирают n элементов, причем никакой элемент не может быть выбран повторно. Пусть X — число помеченных элементов среди выбранных. Такая случайная величина называется *распределенной по гипергеометрическому закону.* Ранее в разделе 1.2.6 была выведена формула $P(X = k) = \dfrac{C_K^k C_{N-K}^{n-k}}{C_N^n}$, где $0 \leqslant k \leqslant K$, $0 \leqslant n - k \leqslant N - K$.

3.2. Совместные распределения

Рассмотрим две дискретные случайные величины X и Y, определенные на одном пространстве элементарных исходов. Тогда содержательная часть *таблицы совместного распределения* величин X и Y будет состоять из вероятностей $p_{ij} = \mathsf{P}(X = x_i,\ Y = y_j)$, где $\{x_1, x_2, \dots\}$ — множество значений величины X, а $\{y_1, y_2, \dots\}$ — множество значений величины Y.

X \ Y	y_1	y_2	y_3	$\cdots$
x_1	p_{11}	p_{12}	p_{13}	$\cdots$
x_2	p_{21}	p_{22}	p_{23}	$\cdots$
x_3	p_{31}	p_{32}	p_{33}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Действуя по аналогии, можно рассмотреть распределение для $n > 2$ случайных величин. Тогда соответствующая таблица будет n-мерной.

Пример 46. Монету бросают три раза. Пусть X — число гербов, появившихся при первом бросании (0 или 1), а Y — общее число появлений герба при трех бросаниях. Требуется построить таблицу совместного распределения для X и Y.

Решение. События $\{X = 0\}$ и $\{Y = 0\}$ могут произойти одновременно только в одном случае, когда герб не появится ни разу. Следовательно, $\mathsf{P}(X = 0,\ Y = 0) = 1/8$. Событию $\{X = 0,\ Y = 1\}$ соответствуют два элементарных исхода: ‘РГР’ и ‘РРГ’. Значит, $\mathsf{P}(X = 0,\ Y = 1) = 2/8$. Событие $\{X = 1,\ Y = 0\}$ невозможно, так как общее число гербов не может быть меньше числа гербов при первом бросании. Таким образом, $\mathsf{P}(X = 1,\ Y = 0) = 0$.

X \ Y	0	1	2	3
0	1/8	2/8	1/8	0
1	0	1/8	2/8	1/8

Аналогично разбираются остальные пять случаев. $\square$

Заметим, что событие $\{X = x_i\}$ может быть представлено в виде суммы попарно несовместных событий:

$$\{X = x_i\} = \sum_j \{X = x_i,\ Y = y_j\}.$$

Следовательно,

$$p_i = \mathsf{P}(X = x_i) = \sum_j \mathsf{P}\left(X = x_i,\ Y = y_j\right) = \sum_j p_{ij}. \qquad (3.1)$$

То есть, просуммировав вероятности p_{ij} в строках таблицы, мы получим ряд распределения случайной величины X.

Аналогично, суммируя вероятности в столбцах, можно построить ряд распределения случайной величины Y:

$$\mathsf{P}(Y = y_j) = \sum_i \mathsf{P}\left(X = x_i,\ Y = y_j\right) = \sum_i p_{ij}. \qquad (3.2)$$

Две случайные величины X и Y называются *независимыми*, если независимы любые пары событий, непосредственно связанные одно с X, а другое — с Y. Нетрудно доказать, что для независимости дискретных случайных величин X и Y достаточно выполнения всех условий вида

$$\mathsf{P}(X = x_i,\ Y = y_j) = \mathsf{P}(X = x_i) \cdot \mathsf{P}(Y = y_j).$$

В частности, случайные величины в примере 46 зависимы, так как $\mathsf{P}(X = 1,\ Y = 0) = 0$, в то время как $\mathsf{P}(X = 1) = 1/2$ и $\mathsf{P}(Y = 0) = 1/8$ отличаются от нуля.

Пусть $y = f(x)$ — числовая функция, а X — случайная величина. Тогда $Y = f(X)$ — тоже случайная величина. Причем

$$\mathsf{P}\bigl(f(X) = c\bigr) = \sum_{i:\ f(x_i)=c} \mathsf{P}(X = x_i) = \sum_{i:\ f(x_i)=c} p_i.$$

Аналогично определяется функция от двух и более случайных величин. Соответственно,

$$\mathsf{P}\bigl(f(X,Y) = c\bigr) = \sum_{\substack{i,j:\\ f(x_i,y_j)=c}} \mathsf{P}(X = x_i,\ Y = y_j) = \sum_{\substack{i,j:\\ f(x_i,y_j)=c}} p_{ij}.$$

Пример 47. Составить ряды распределения случайных величин $(Y-1)^2$, $Y-X$ и $X\cdot(Y-1)$ по таблице совместного распределения, построенной в примере 46.

Решение. При первом знакомстве ряд распределения для функции от случайной величины рекомендуется строить в два этапа.

Согласно условию,

Y	0	1	2	3
P	1/8	3/8	3/8	1/8

Первый этап:

$(Y-1)^2$	$(0-1)^2$	$(1-1)^2$	$(2-1)^2$	$(3-1)^2$
P	1/8	3/8	3/8	1/8

Чтобы привести этот ряд к окончательному виду, нужно объединить столбцы с одинаковыми значениями для $(y_i-1)^2$, а соответствующие вероятности сложить:

$(Y-1)^2$	0	1	4
P	3/8	$1/8+3/8$	1/8

Построим ряд распределения случайной величины $Y-X$ по таблице распределения из примера 46. Первый этап (события с нулевыми вероятностями не рассматриваем):

$Y-X$	$0-0$	$1-0$	$2-0$	$1-1$	$2-1$	$3-1$
P	1/8	2/8	1/8	1/8	2/8	1/8

Окончательный результат:

$Y-X$	0	1	2
P	$1/8+1/8$	$2/8+2/8$	$1/8+1/8$

Построение ряда распределения для $X\cdot(Y-1)$ предлагается в качестве упражнения читателю. □

3.3. Математическое ожидание

При исследовании случайной величины или сравнении её с другими случайными величинами полезно ввести в рассмотрение некоторые суммарные характеристики, отражающие её типичные свойства. Одной из таких характеристик является среднее значение случайной величины, или математическое ожидание.

Пример 48. Рассмотрим три случайные величины X, Y и Z со следующими распределениями:

X	0	1
P	0,5	0,5

Y	0	1
P	0,99	0,01

Z	0	1
P	0,01	0,99

Все эти величины имеют одинаковый набор значений.

Предположим, что проведено большое количество испытаний (экспериментов), и в каждом из них случайные величины X, Y и Z принимали какие-то числовые значения (0 или 1). Попробуем оценить, какими будут средние арифметические значений каждой из этих случайных величин.

Случайная величина X равновозможно принимает значения 0 и 1, поэтому в серии из большого количества испытаний примерно в половине случаев X примет значение 1, а в остальных — значение 0. Тогда среднее арифметическое значений, которые принимала величина X в этой серии испытаний, будет примерно равно 0,5.

В то же время случайная величина Y с вероятностью 0,99 принимает значение 0, поэтому чаще всего в большой серии испытаний Y будет принимать значение, равное 0, и очень редко — значение 1. А значит, среднее арифметическое значений, которые принимала величина Y, будет близко к 0 (так как это среднее арифметическое большого количества нулей и нескольких единиц).

Аналогично среднее арифметическое значений случайной величины Z в этой серии испытаний будет близко к 1, так как такая случайная величина с большой вероятностью принимает значение 1 и очень редко нулевое значение.

Запишем более формально определение числовой характеристики, которая хорошо согласуется с описанным представлением о среднем значении случайной величины.

Определение 3.4. *Математическое ожидание* $\mathsf{M}(X)$ дискретной случайной величины X определяется формулой[1]

$$\mathsf{M}(X) = \sum_i x_i \cdot \mathsf{P}(X = x_i),$$

где x_i — возможные значения случайной величины X. Математическое ожидание $\mathsf{M}(X)$ существует, только если числовой ряд в правой части формулы сходится абсолютно[2].

Пример 49. Вычислить математическое ожидание случайных величин, описанных в примере 48.

Решение. Исходя из приведенного выше определения математического ожидания дискретной случайной величины, получаем:

$$\begin{aligned} \mathsf{M}(X) &= 0 \cdot 0{,}5 + 1 \cdot 0{,}5 = 0{,}5, \\ \mathsf{M}(Y) &= 0 \cdot 0{,}99 + 1 \cdot 0{,}01 = 0{,}99, \\ \mathsf{M}(Z) &= 0 \cdot 0{,}01 + 1 \cdot 0{,}99 = 0{,}01. \quad \square \end{aligned}$$

Пример 50. Игра состоит в бросании трех игральных костей. Перед началом игры игрок делает ставку (выигрыш зависит от размера ставки). Если ни на одной из трех игральных костей не появляется грань с шестью очками, то игрок теряет ставку. В противном случае игрок получает назад поставленные деньги и выигрыш, равный числу «шестерок», помноженному на размер ставки. Найти математическое ожидание выигрыша.

Решение. Обозначим через X выигрыш игрока, измеряемый в ставках. Тогда

X	-1	1	2	3
P	$\frac{5^3}{6^3}$	$\frac{3\cdot 5^2}{6^3}$	$\frac{3\cdot 5}{6^3}$	$\frac{1}{6^3}$

[1] В англоязычной литературе обычно используется обозначение $\mathsf{E}(X)$ от англ. *expected value* или *expectation*. Название появилось еще в XVII веке, когда теория вероятностей являлась частью теории азартных игр и игроков в первую очередь интересовало среднее значение ожидаемого выигрыша (проигрыша), вычисляемое математически.

[2] Абсолютная сходимость гарантирует, что сумма ряда не изменится при перестановке слагаемых.

Находим математическое ожидание выигрыша:

$$\mathsf{M}(X) = -1 \cdot \frac{5^3}{6^3} + 1 \cdot \frac{3 \cdot 5^2}{6^3} + 2 \cdot \frac{3 \cdot 5}{6^3} + 3 \cdot \frac{1}{6^3} = -\frac{17}{216}.$$

Отрицательное значение математического ожидания говорит о том, что эта игра (в среднем) не выгодна для игрока. Более того, значение $\mathsf{M}(X) = -\frac{17}{216}$ позволяет измерить величину «невыгодности». В частности, после проведения 216 игр с одинаковым размером ставки игрок потеряет в среднем 17 ставок. □

Пример 51 (Санкт-Петербургский парадокс[3])**.** Игра состоит в бросании монеты до тех пор, пока не выпадет решка. Если это произойдет при k-м бросании, то игрок получает 2^k рублей. Сколько игроку следует заплатить за участие в игре, чтобы игра была справедливой (чтобы среднее значение выигрыша оказалось равным 0)?

Парадокс заключается в том, что математическое ожидание выигрыша равно бесконечности:

$$2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{8} + \ldots = 1 + 1 + 1 + \ldots = \infty,$$

т. е. больше любого выигрыша.

Способы разрешения этого парадокса можно найти в [3] и [4].

Теорема 3.5. *Пусть $y = f(x)$ — функция, действующая из $\mathbb{R}$ в $\mathbb{R}$, а X — дискретная случайная величина с возможными значениями x_i и вероятностями $p_i = \mathsf{P}(X = x_i)$, $i = 1, 2, \ldots$ Тогда математическое ожидание случайной величины $f(X)$ может быть найдено по формуле*

$$\mathsf{M}(f(X)) = \sum_i f(x_i)p_i, \tag{3.3}$$

если числовой ряд в правой части сходится абсолютно.

Для функции от двух переменных справедлива аналогичная формула:

$$\mathsf{M}(f(X,Y)) = \sum_{i,j} f(x_i,y_j)\mathsf{P}(X = x_i,\ Y = y_j). \tag{3.4}$$

[3] Этот парадокс был опубликован Даниилом Бернулли в 1738 г. в Комментариях Санкт-Петербургской Академии наук.

▷ Докажем справедливость формулы (3.3). Предположим, что величина $Y = f(X)$ принимает значения y_j, $j = 1, 2, \ldots$ Согласно определению функции от случайной величины, $\mathsf{P}(Y = y_j) = \sum\limits_{i:\ f(x_i)=y_j} p_i$. Находим математическое ожидание:

$$\mathsf{M}(Y) = \sum_j y_j \mathsf{P}(Y = y_j) = \sum_j y_j \sum_{i:\ f(x_i)=y_j} p_i = \sum_j \sum_{i:\ f(x_i)=y_j} f(x_i) p_i. \qquad (3.5)$$

Так как f — однозначная функция, то сумма (3.5) совпадает с правой частью формулы (3.3).

Справедливость формулы (3.4) доказывается аналогично. □

3.3.1. Основные свойства математического ожидания

1. Если $\mathsf{P}(X = c) = 1$, то $\mathsf{M}(X) = c$.

2. $\mathsf{M}(X + c) = \mathsf{M}(X) + c$, если $\mathsf{M}(X)$ существует.

3. $\mathsf{M}(c \cdot X) = c \cdot \mathsf{M}(X)$, если $\mathsf{M}(X)$ существует.

4. $\mathsf{M}(X + Y) = \mathsf{M}(X) + \mathsf{M}(Y)$ при условии, что $\mathsf{M}(X)$ и $\mathsf{M}(Y)$ существуют.
Следствие: $\mathsf{M}(X_1 + X_2 + \cdots + X_n) = \mathsf{M}(X_1) + \mathsf{M}(X_2) + \cdots + \mathsf{M}(X_n)$ при условии, что все $\mathsf{M}(X_i)$ существуют.
Следствие (свойство линейности): если $\mathsf{M}(X)$ и $\mathsf{M}(Y)$ существуют, то $\mathsf{M}(a \cdot X + b \cdot Y + c) = a \cdot \mathsf{M}(X) + b \cdot \mathsf{M}(Y) + c$, где $a, b, c \in \mathbb{R}$.

5. Если X и Y независимы и их математические ожидания существуют, то $\mathsf{M}(X \cdot Y) = \mathsf{M}(X) \cdot \mathsf{M}(Y)$.

6. Если $\mathsf{M}(X)$ и $\mathsf{M}(Y)$ существуют и $X \leqslant Y$ при любом исходе эксперимента, то $\mathsf{M}(X) \leqslant \mathsf{M}(Y)$.
Следствие: если $a \leqslant X \leqslant b$, то $a \leqslant M(X) \leqslant b$.

Доказательство.

1. По определению математического ожидания: $\mathsf{M}(X) = c \cdot 1 = c$.

2. Согласно формуле (3.3),

$$\mathsf{M}(X + c) = \sum_i (x_i + c)\, p_i = \sum_i x_i\, p_i + c \sum_i p_i = \mathsf{M}(X) + c \cdot 1.$$

3. Согласно формуле (3.3),

$$\mathsf{M}(c\,X) = \sum_i c\,x_i\,p_i = c\cdot\sum_i x_i\,p_i = c\cdot\mathsf{M}(X).$$

4. Согласно формуле (3.4),

$$\mathsf{M}(X+Y) = \sum_{i,j}(x_i+y_j)\cdot\mathsf{P}(X=x_i,\ Y=y_j).$$

Введем обозначение $p_{ij} = \mathsf{P}(X=x_i,\ Y=y_j)$. Тогда

$$\mathsf{M}(X+Y) = \sum_{i,j} x_i p_{ij} + \sum_{i,j} y_j p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij}.$$

Воспользовавшись формулами (3.1) и (3.2), получаем

$$\mathsf{M}(X+Y) = \sum_i x_i\cdot\mathsf{P}(X=x_i) + \sum_j y_j\cdot\mathsf{P}(Y=y_j).$$

Первое следствие является обобщением свойства 4 на случай произвольного количества слагаемых. Доказывается методом математической индукции по n.

Справедливость второго следствия (линейность математического ожидания) вытекает из свойств 1, 2, 3 и 4.

5. Для независимых случайных величин выполняются равенства $\mathsf{P}(X=x_i,\ Y=y_j) = \mathsf{P}(X=x_i)\cdot\mathsf{P}(Y=y_j)$. Таким образом,

$$\begin{aligned}\mathsf{M}(XY) &= \sum_{i,j} x_i y_j\cdot\mathsf{P}(X=x_i,\ Y=y_j) = \\ &= \sum_i x_i\cdot\mathsf{P}(X=x_i)\sum_j y_j\cdot\mathsf{P}(Y=y_j) = \mathsf{M}(X)\cdot\mathsf{M}(Y).\end{aligned}$$

6. Рассмотрим вспомогательную случайную величину $Z = Y - X$. Так как $X \leqslant Y$, то $Z \geqslant 0$. Кроме того, в силу свойства линейности, $\mathsf{M}(Z) = \mathsf{M}(Y-X) = \mathsf{M}(Y) - \mathsf{M}(X)$. Таким образом, достаточно доказать, что из $Z \geqslant 0$ следует $\mathsf{M}(Z) \geqslant 0$. Но это очевидно, так как из условия $Z \geqslant 0$ следует неотрицательность всех слагаемых суммы $\mathsf{M}(Z) = \sum\limits_i z_i\mathsf{P}(Z=z_i)$.

Если рассмотреть $Y = b$ (вырожденную случайную величину), то из $X \leqslant b$ следует $\mathsf{M}(X) \leqslant \mathsf{M}(b) = b$. Аналогично, из $a \leqslant X$ следует $a \leqslant \mathsf{M}(X)$. □

Задача 52. Одно из альтернативных определений выпуклой функции следующее. Функция $y = f(x)$ называется *выпуклой вниз*, если через любую точку $(x_0, f(x_0))$ её графика можно провести прямую таким образом, что весь график функции f окажется не ниже этой прямой[4]. Другими словами, для каждого x_0 можно подобрать коэффициент $k = k(x_0)$ так, что неравенство $f(x) \geqslant f(x_0) + k \cdot (x - x_0)$ выполнено для всех x из области определения функции.

Пользуясь этим определением, доказать неравенство Йенсена[5].

$$f\big(\mathsf{M}(X)\big) \leqslant \mathsf{M}\big(f(X)\big),$$

где функция f выпукла вниз.

Определение 3.6. Будем говорить, что X *больше* Y *по вероятности*, если $\mathsf{P}(X > Y) > 1/2$.

Можно доказать, что если X больше Y по вероятности, то не обязательно $\mathsf{M}(X) > \mathsf{M}(Y)$.

Пример 53 (парадокс транзитивности)**.** Занумеруем грани трех игральных костей числами 1, 2, ..., 18 следующим образом [3]:

Грани первой кости: 5, 7, 8, 9, 10, 18;
Грани второй кости: 2, 3, 4, 15, 16, 17;
Грани третьей кости: 1, 6, 11, 12, 13, 14.

Пусть X_1 — число, выпавшее при бросании первой кости, X_2 — второй, X_3 — третьей.

Легко проверяется, что $\mathsf{M}(X_1) = \mathsf{M}(X_2) = \mathsf{M}(X_3)$. Если же сравнивать эти случайные величины по вероятности (оставляем это в качестве упражнения читателю), то оказывается, что X_1 больше X_2, X_2 больше X_3, но X_3 больше X_1. Т. е. такой способ сравнения не обладает свойством транзитивности.

[4] Такая прямая называется опорной.

[5] Иоган Людвиг Виллиам Вальдемар Йенсен (1859–1925) — датский математик и инженер.

3.3.2. Примеры вычисления математических ожиданий

Пример 54. Пусть X распределена по биномиальному закону с параметрами n и p, т. е. X — число успехов в n испытаниях Бернулли.

Рассмотрим вспомогательные случайные величины X_k — число успехов в k-м испытании, $k = 1, 2, \ldots, n$. Таким образом,

X_k	0	1
P	$1-p$	p

Заметим, что общее число успехов X можно представить в виде суммы

$$X = X_1 + X_2 + \ldots + X_n.$$

По свойству математического ожидания суммы,

$$\mathsf{M}(X) = \mathsf{M}(X_1 + \ldots + X_n) = \mathsf{M}(X_1) + \ldots + \mathsf{M}(X_n).$$

Так как $\mathsf{M}(X_k) = 0 \cdot (1-p) + 1 \cdot p = p$, то $\mathsf{M}(X) = np$.

Пример 55. Пусть Y распределена по закону Пуассона с параметром λ. Тогда

$$\mathsf{M}(Y) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \lambda \cdot \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} =$$

$$= \left| \begin{array}{c} \text{Замена:} \\ i = k-1,\ k = i+1 \end{array} \right| = \lambda \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda}.$$

Сумма $\sum\limits_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda}$ есть не что иное, как сумма вероятностей для распределения Пуассона, а значит, она равна единице. Следовательно, $\mathsf{M}(Y) = \lambda$.

Пример 56. Пусть Z распределена по геометрическому закону с параметром p. Тогда

$$\mathsf{M}(Z) = \sum_{k=1}^{\infty} k q^{k-1} p = p \sum_{k=1}^{\infty} k q^{k-1}. \tag{3.6}$$

Найдем $\sum\limits_{k=1}^{\infty} kq^{k-1}$. Для этого проинтегрируем степенной ряд $\sum\limits_{k=1}^{\infty} kx^{k-1}$. (В базовом курсе математического анализа доказывается, что внутри интервала сходимости[6] знаки интегрирования и суммирования можно менять местами.)

$$\int_0^q \sum_{k=1}^{\infty} kx^{k-1}\,dx = \sum_{k=1}^{\infty} \int_0^q kx^{k-1}\,dx = \sum_{k=1}^{\infty} q^k = \frac{q}{1-q}.$$

Тогда

$$\sum_{k=1}^{\infty} kq^{k-1} = \left(\frac{q}{1-q}\right)'_q = \frac{1}{(1-q)^2} = \frac{1}{p^2}.$$

Подставляя в (3.6), получаем

$$\mathsf{M}(Z) = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

3.4. Дисперсия дискретной случайной величины

Кроме среднего значения, часто полезно уметь оценивать ширину разброса случайной величины. Рассмотрим в качестве примера случайные величины X и Y:

X	2	3	4
P	1/6	2/3	1/6

Y	-3	0	3
P	1/4	1/2	1/4

Их *многоугольники распределения* изображены на рис. 3.1. Видно, что разброс случайной величины Y явно больше, чем у X.

Заметим, что ширина разброса случайной величины никак не зависит от ее математического ожидания. В частности, математическое ожидание случайной величины Y, изображенной на рис. 3.1, равно 0, что не мешает ей быть более разбросанной, чем X.

[6] Здесь мы пользуемся тем, что ряд $\sum\limits_{k=1}^{\infty} kx^{k-1}$ сходится при $x \in (-1, 1)$.

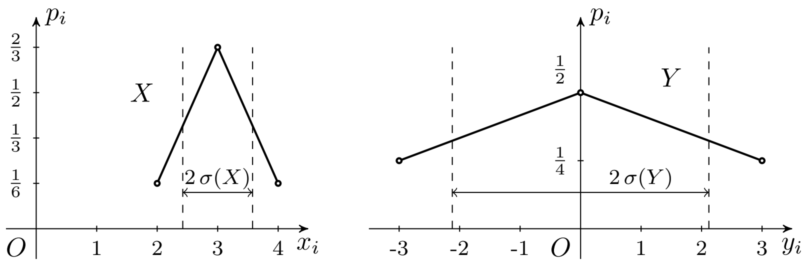

Рис. 3.1. Ширина разброса случайной величины X значительно меньше, чем у Y

Определение 3.7. Случайная величина $X - \mathsf{M}\,X$ называется *отклонением величины X от своего математического ожидания*, или *центрированной случайной величиной.*

Из определения следует, что математическое ожидание центрированной случайной величины равно 0.

Естественной мерой разброса случайной величины является среднее значение абсолютной величины отклонения: $\mathsf{M}\big|X - \mathsf{M}\,X\big|$. Оно называется *средним отклонением.* Но по ряду причин такой способ измерения разброса оказывается неудобным. Альтернативой ему служат дисперсия и среднеквадратическое отклонение.

Определение 3.8. *Дисперсия* $\mathsf{D}(X)$ случайной величины X вычисляется по формуле[7]

$$\mathsf{D}(X) = \mathsf{M}\left(\big(X - \mathsf{M}\,X\big)^2\right). \tag{3.7}$$

Определение 3.9. *Среднеквадратическое отклонение*[8] $\sigma(X)$ случайной величины X равно корню из её дисперсии $\mathsf{D}(X)$.

Пример 57. Найти среднеквадратическое отклонение числа очков, выпавших при однократном бросании игральной кости.

[7]В англоязычной литературе используется обозначение $\mathsf{Var}(X)$ от англ. *variance.*

[8]Используются также названия *стандартное отклонение* и *стандартный разброс.*

Решение. Обозначим через X число очков, выпавших на игральной кости. Тогда

$$\mathsf{M}\,X = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + \ldots + 6\cdot\frac{1}{6} = 3{,}5.$$

Находим дисперсию:

$$\mathsf{D}\,X = \mathsf{M}\big(X - \mathsf{M}\,X\big)^2 = (1-3{,}5)^2\cdot\frac{1}{6} + (2-3{,}5)^2\cdot\frac{1}{6} + \ldots + (6-3{,}5)^2\cdot\frac{1}{6} = \frac{35}{12}.$$

Среднеквадратическое отклонение $\sigma(X) = \sqrt{\mathsf{D}\,X} \approx 1{,}71$. □

Задача 58. Привести пример случайной величины X, математическое ожидание которой существует и конечно, а дисперсия не существует.

Задача 59. Для каждого натурального n привести пример случайной величины X_n такой, что $\sigma(X_n) \geqslant n\cdot\mathsf{M}\big|X_n - \mathsf{M}\,X_n\big| > 0$ (среднее отклонение в n раз меньше среднеквадратического).

3.4.1. Свойства дисперсии

Во многих случаях дисперсию удобнее вычислять так.

Формула для дисперсии.

$$\mathsf{D}\,X = \mathsf{M}\left(X^2\right) - (\mathsf{M}\,X)^2\,. \tag{3.8}$$

▷ Положим для удобства $c = \mathsf{M}\,X$ и воспользуемся свойством линейности математического ожидания:

$$\begin{aligned}\mathsf{D}\,X = \mathsf{M}(X-c)^2 = \mathsf{M}\left(X^2 - 2\,c\,X + c^2\right) = \\ = \mathsf{M}\left(X^2\right) - 2\,c\,\mathsf{M}\,X + c^2 = \mathsf{M}\left(X^2\right) - c^2.\end{aligned}$$ □

Перечислим **основные свойства дисперсии**:

1. $\mathsf{D}\,X \geqslant 0$ всегда;

2. Дисперсия константы (вырожденной случайной величины) равна 0. Верно и обратное: $\mathsf{D}\,X = 0$ только если X вырождена, т. е. $\mathsf{P}(X = c) = 1$ для некоторого постоянного $c \in \mathbb{R}$;

3. $\mathsf{D}(c \cdot X) = c^2 \cdot \mathsf{D}\, X$;

4. Введем обозначение

$$\operatorname{cov}(X, Y) = \mathsf{M}(X \cdot Y) - \mathsf{M}\, X \cdot \mathsf{M}\, Y.$$

Тогда

$$\mathsf{D}(X + Y) = \mathsf{D}(X) + \mathsf{D}(Y) + 2\operatorname{cov}(X, Y); \tag{3.9}$$

5. Если X и Y независимы, то $\mathsf{D}(X + Y) = \mathsf{D}(X) + \mathsf{D}(Y)$.

Доказательство.

1. Справедливость неравенства $\mathsf{D}\, X \geqslant 0$ следует из того, что дисперсия есть математическое ожидание неотрицательной случайной величины $Y = (X - \mathsf{M}\, X)^2$.

2. Пусть c — константа, тогда $\mathsf{D}(c) = \mathsf{M}(c - \mathsf{M}\, c)^2 = \mathsf{M}(0) = 0$.

Предположим теперь, что $\mathsf{D}\, X = 0$. То есть математическое ожидание неотрицательной случайной величины $Y = (X - \mathsf{M}\, X)^2$ равно нулю. Но это возможно, только если $\mathsf{P}(Y > 0) = 0$ (иначе $\mathsf{M}\, Y > 0$, так как одно из слагаемых в $\mathsf{M}\, Y$ больше нуля). Следовательно, $\mathsf{P}(X = \mathsf{M}\, X) = 1$.

3. $\mathsf{D}(cX) = \mathsf{M}(cX - \mathsf{M}(cX))^2 = \mathsf{M}\big(c^2(X - \mathsf{M}\, X)^2\big) = c^2\, \mathsf{M}(X - \mathsf{M}\, X)^2$.

4. Воспользуемся формулой (3.8):

$$\begin{aligned}\mathsf{D}\, X &= \mathsf{M}(X + Y)^2 - (\mathsf{M}(X + Y))^2 = \\ &= \mathsf{M}\left(X^2 + 2XY + Y^2\right) - (\mathsf{M}\, X + \mathsf{M}\, Y)^2 = \\ &= \mathsf{M}\, X^2 + 2\ \mathsf{M}(XY) + \mathsf{M}\, Y^2 - (\mathsf{M}\, X)^2 - 2\ \mathsf{M}\, X\ \mathsf{M}\, Y - (\mathsf{M}\, Y)^2.\end{aligned}$$

Остается заметить, что $\mathsf{M}\, X^2 - (\mathsf{M}\, X)^2 = \mathsf{D}\, X$ и $\mathsf{M}\, Y^2 - (\mathsf{M}\, Y)^2 = \mathsf{D}\, Y$.

5. Согласно формуле (3.9) достаточно показать, что $\operatorname{cov}(X, Y) = 0$ для независимых X и Y. Действительно, если X и Y независимы, тогда, по свойству 5 математического ожидания, $\mathsf{M}(XY) = \mathsf{M}\, X\ \mathsf{M}\, Y$. $\square$

Следствие 1. $\mathsf{D}(X + C) = \mathsf{D}\, X$.

Следствие 2. $\mathsf{D}(X - Y) = \mathsf{D}\, X + \mathsf{D}\, Y$ для независимых X и Y.

Следствие 3. Для независимых в совокупности случайных величин X_1, X_2, …, X_n верна формула

$$\mathsf{D}(X_1 + X_2 + \ldots + X_n) = \mathsf{D}\,X_1 + \mathsf{D}\,X_2 + \ldots + \mathsf{D}\,X_n. \tag{3.10}$$

Задача 60. Доказать, что формула (3.10) справедлива для попарно независимых случайных величин.

Из свойств дисперсии вытекают **свойства среднекватратического отклонения**:

1. $\sigma(c) = 0$;

2. $\sigma(X + c) = \sigma(X)$;

3. $\sigma(c \cdot X) = |c| \cdot \sigma(X)$.

3.4.2. Примеры вычисления дисперсии

Пример 61. Найдем дисперсию случайной величины X, распределенной по биномиальному закону с параметрами n и p.

Действуя по аналогии с примером 54, представим X в виде суммы $X_1 + X_2 + \ldots + X_n$, где X_i равно числу успехов в i-м испытании. Так как случайные величины X_1, …, X_n независимы, то

$$\mathsf{D}\,X = \mathsf{D}\,X_1 + \mathsf{D}\,X_2 + \ldots + \mathsf{D}\,X_n.$$

Так как $\mathsf{D}\,X_i = \mathsf{M}\,X_i^2 - (\mathsf{M}\,X_i)^2 = p - p^2$, то $\mathsf{D}\,X = np(1-p) = npq$.

Пример 62. Пусть Y распределена по закону Пуассона с параметром λ.

Найдем $\mathsf{D}\,Y$, воспользовавшись рассуждениями из примера 55 и формулой (3.8):

$$\mathsf{M}(Y^2) = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} (k(k-1) + k) \frac{\lambda^k}{k!} e^{-\lambda} =$$

$$= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda}.$$

Последняя сумма есть не что иное, как математическое ожидание $\mathsf{M}\,Y = \lambda$. Таким образом,

$$\mathsf{M}(Y^2) = \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} + \lambda = \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!}e^{-\lambda} + \lambda =$$
$$= \left| \begin{array}{c} \text{Замена:} \\ i = k-2,\ k = i+2 \end{array} \right| = \lambda^2 \sum_{i=0}^{\infty} \frac{\lambda^i}{i!}e^{-\lambda} + \lambda = \lambda^2 + \lambda.$$

Следовательно,

$$\mathsf{D}\,Y = \mathsf{M}(Y^2) - (\mathsf{M}\,Y)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Пример 63. Пусть Z распределена по геометрическому закону с параметром p.

Для нахождения дисперсии Z воспользуемся формулой (3.8) и примером 56.

$$\mathsf{M}(Z^2) = \sum_{k=1}^{\infty} k^2 q^{k-1} p = p \sum_{k=1}^{\infty} k^2 q^{k-1}.$$

Проинтегрируем степенной ряд $\sum\limits_{k=1}^{\infty} k^2 x^{k-1}$:

$$\int_0^q \sum_{k=1}^{\infty} k^2 x^{k-1}\,dx = \sum_{k=1}^{\infty} \int_0^q k^2 x^{k-1}\,dx =$$
$$= \sum_{k=1}^{\infty} k q^k = q \sum_{k=1}^{\infty} k q^{k-1}.$$

В примере 56 была получена формула $\sum\limits_{k=1}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2}$. Следовательно,

$$\int_0^q \sum_{k=1}^{\infty} k^2 x^{k-1}\,dx = q\frac{1}{(1-q)^2}.$$

Тогда

$$\sum_{k=1}^{\infty} k^2 q^{k-1} = \left(\frac{q}{(1-q)^2}\right)'_q = \frac{1+q}{(1-q)^3} = \frac{1+q}{p^3}.$$

Подведем итог:

$$\mathsf{D}\,Z = \mathsf{M}(Z^2) - (\mathsf{M}\,Z)^2 = p\frac{1+q}{p^3} - \frac{1}{p^2} = \frac{q}{p^2}.$$

3.5. Ковариация и коэффициент корреляции

Рассмотренные ранее математическое ожидание и дисперсия являются числовыми характеристиками отдельной случайной величины и дают общую информацию о характере распределения. В данном разделе описаны числовые характеристики, которые при некоторых условиях могут использоваться как индикаторы зависимости или показатели силы связи между случайными величинами.

3.5.1. Ковариация

Выше, при нахождении дисперсии суммы (см. свойство 4), было введено обозначение

$$\operatorname{cov}(X, Y) = \mathsf{M}(XY) - \mathsf{M}\,X \cdot \mathsf{M}\,Y. \tag{3.11}$$

Эта числовая функция называется *ковариацией* случайных величин X и Y и характеризует степень зависимости X и Y.

Отметим также, что для независимых случайных величин X и Y математическое ожидание их произведения равно произведению математических ожиданий (свойство 5): $\mathsf{M}(XY) = \mathsf{M}(X) \cdot \mathsf{M}(Y)$, откуда следует $\operatorname{cov}(X, Y) = 0$ (то есть независимые случайные величины обладают нулевой ковариацией).

Ковариацию случайных величин можно определить и иначе.

Определение 3.10. Ковариацией случайных величин X и Y называется числовая характеристика, определяемая следующим образом:

$$\operatorname{cov}(X, Y) = \mathsf{M}\big((X - \mathsf{M}\,X)(Y - \mathsf{M}\,Y)\big). \tag{3.12}$$

Иначе говоря, ковариация — это математическое ожидание произведения центрированных случайных величин.

Теорема 3.11. *Определения* (3.11) *и* (3.12) *эквивалентны.*

Доказательство. Рассмотрим формулу (3.12) и равносильными преобразованиями выведем из нее формулу (3.11). Для удобства промежуточных вычислений введем обозначения: $\mathsf{M}\,X = a$, $\mathsf{M}\,Y = b$.

$$\begin{aligned}\operatorname{cov}(X,Y) &= \mathsf{M}\big((X-\mathsf{M}\,X)(Y-\mathsf{M}\,Y)\big) = \mathsf{M}\big((X-a)(Y-b)\big) = \\ &= \mathsf{M}(XY - aY - bX + ab) = \\ &= \mathsf{M}(XY) - a\,\mathsf{M}\,Y - b\,\mathsf{M}\,X + ab = \\ &= \mathsf{M}(XY) - ab - ba + ab = \mathsf{M}(XY) - \mathsf{M}\,X\ \mathsf{M}\,Y.\end{aligned}$$

□

Непосредственно из свойств математического ожидания и дисперсии выводятся **свойства ковариации**:

1. Если X и Y независимы, то $\operatorname{cov}(X,Y) = 0$;

2. $\operatorname{cov}(X,Y) = \operatorname{cov}(Y,X)$;

3. $\operatorname{cov}(X,X) = DX$;

4. $\operatorname{cov}(aX,bY) = ab\operatorname{cov}(X,Y)$;

5. $\operatorname{cov}(X+a,Y+b) = \operatorname{cov}(X,Y)$;

6. $\operatorname{cov}(c,Y) = 0$;

7. $\operatorname{cov}(X+Z,Y) = \operatorname{cov}(X,Y) + \operatorname{cov}(Z,Y)$.

Для доказательства данных свойств будем использовать доказанные ранее свойства математического ожидания.

Доказательство.

1. Так как математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий, то

$$\operatorname{cov}(X,Y) = \mathsf{M}(XY) - \mathsf{M}\,X\ \mathsf{M}\,Y = \mathsf{M}\,X\ \mathsf{M}\,Y - \mathsf{M}\,X\ \mathsf{M}\,Y = 0.$$

2. Следует из определения ковариации:

$$\operatorname{cov}(X,Y) = \mathsf{M}(XY) - \mathsf{M}\,X\ \mathsf{M}\,Y = \mathsf{M}(YX) - \mathsf{M}\,Y\ \mathsf{M}\,X = \operatorname{cov}(Y,X).$$

3. $\operatorname{cov}(X, X) = \mathsf{M}(X \cdot X) - \mathsf{M}\,X\,\mathsf{M}\,X = \mathsf{M}\,X^2 - (\mathsf{M}\,X)^2 = \mathsf{D}\,X$.

4. Воспользуемся тем, что при умножении случайной величины на константу её математическое ожидание тоже умножается на эту константу:

$$\begin{aligned}\operatorname{cov}(aX, bY) &= \mathsf{M}((aX)(bY)) - \mathsf{M}(aX)\ \mathsf{M}(bY) = \\ &= \mathsf{M}(abXY) - (a\ \mathsf{M}\,X)(b\ \mathsf{M}\,Y) = \\ &= ab\ \mathsf{M}(XY) - ab\ \mathsf{M}\,X\ \mathsf{M}\,Y = ab\ \operatorname{cov}(X, Y).\end{aligned}$$

5. Здесь будем исходить из второго определения ковариации:

$$\begin{aligned}\operatorname{cov}(X + a, Y + b) &= \mathsf{M}\Big(\big(X + a - \mathsf{M}(X + a)\big)\big(Y + b - \mathsf{M}(Y + b)\big)\Big) = \\ &= \mathsf{M}\big((X + a - \mathsf{M}\,X - a)(Y + b - \mathsf{M}\,Y - b)\big) = \\ &= \mathsf{M}(X - \mathsf{M}\,X)(Y - \mathsf{M}\,Y) = \operatorname{cov}(X, Y).\end{aligned}$$

6. $\operatorname{cov}(c, Y) = \mathsf{M}(cY) - \mathsf{M}\,c\,\mathsf{M}\,Y = c\,\mathsf{M}\,Y - c\,\mathsf{M}\,Y = 0$.

7. Так как математическое ожидание суммы случайных величин равно сумме их математических ожиданий, получим:

$$\begin{aligned}\operatorname{cov}(X + Z, Y) &= \mathsf{M}((X + Z)Y) - \mathsf{M}(X + Z)\,\mathsf{M}\,Y = \\ &= \mathsf{M}(XY + ZY) - (\mathsf{M}\,X + \mathsf{M}\,Z)\,\mathsf{M}\,Y = \\ &= \mathsf{M}(XY) + \mathsf{M}(ZY) - \mathsf{M}\,X\,\mathsf{M}\,Y - \mathsf{M}\,Z\,\mathsf{M}\,Y = \\ &= (\mathsf{M}(XY) - \mathsf{M}\,X\,\mathsf{M}\,Y) + (\mathsf{M}(ZY) - \mathsf{M}\,Z\,\mathsf{M}\,Y) = \\ &= \operatorname{cov}(X, Y) + \operatorname{cov}(Z, Y).\end{aligned}$$

Отметим также, что данное свойство можно обобщить на случай произвольного количества слагаемых:

$$\operatorname{cov}(X_1 + X_2 + \dots + X_m, Y_1 + Y_2 + \dots + Y_n) = \sum_{i,j} \operatorname{cov}(X_i, Y_j).$$

Доказательство данного обобщенного свойства проводится аналогично тому, как это было сделано в случае двух слагаемых. □

Задача 64. Вывести формулу для дисперсии суммы нескольких случайных величин:

$$\mathsf{D}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{cov}(X_i, X_j) = \sum_{i=1}^{n} \mathsf{D}\,X_i + 2\sum_{1 \leqslant i < j \leqslant n} \operatorname{cov}(X_i, X_j).$$

Как было сказано ранее, ковариация характеризует степень зависимости X и Y. Более того, было показано, что некоторые преобразования случайных величин не влияют на значение этой числовой характеристики: например, из свойства $\operatorname{cov}(X+a, Y+b) = \operatorname{cov}(X, Y)$ следует, что при добавлении констант к случайным величинам ковариация не меняется (что довольно разумно, ведь степень зависимости случайных величин $X + 5$ и $Y - 3$ должна быть такой же, что и степень зависимости величин X и Y).

В то же время из формулы $\operatorname{cov}(a\cdot X, b\cdot Y) = ab\cdot\operatorname{cov}(X, Y)$ следует, что величина ковариации зависит от коэффициентов a и b. Но с точки зрения здравого смысла степень зависимости, например, случайных величин $2X$ и $3Y$ должна быть такой же, как у X и Y. Другими словами, выбор единиц измерения для величин X и Y не должен влиять на степень их зависимости. Наиболее естественный способ решения этой проблемы рассматривается в следующем параграфе.

3.5.2. Коэффициент корреляции

Начнем со вспомогательного определения.

Определение 3.12. Случайная величина

$$X^* = \frac{X - \mathsf{M}\,X}{\sigma(X)}$$

называется *нормированной.*

Задача 65. Докажите, что $\mathsf{M}\,X^* = 0$, а $\sigma(X^*) = \mathsf{D}\,X^* = 1$.

Определение 3.13. *Коэффициент корреляции* случайных величин X и Y определяется формулой

$$\rho(X, Y) = \operatorname{cov}(X^*, Y^*) = \frac{\operatorname{cov}(X, Y)}{\sigma(X)\,\sigma(Y)} = \frac{\mathsf{M}(XY) - \mathsf{M}\,X\ \mathsf{M}\,Y}{\sqrt{\mathsf{D}\,X\,\mathsf{D}\,Y}}.$$

Свойства коэффициента корреляции:

1. Если X и Y независимы, то $\rho(X, Y) = 0$;

2. $\rho(X + a, Y + b) = \rho(X, Y)$;

3. $\rho(aX, bY) = \frac{ab}{|ab|}\rho(X, Y) = \operatorname{sgn}(ab) \cdot \rho(X, Y)$ при $a, b \neq 0$;

4. $\rho(X, X) = 1$.

Доказательство.

1. Ранее было доказано, что для независимых случайных величин X и Y выполняется следующее свойство: $\operatorname{cov}(X, Y) = 0$. Тогда

$$\rho(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sigma(X)\,\sigma(Y)} = \frac{0}{\sigma(X)\,\sigma(Y)} = 0.$$

2. С учетом свойств $\operatorname{cov}(X+a, Y+b) = \operatorname{cov}(X, Y)$ и $\sigma(X+c) = \sigma X$, получаем:

$$\rho(X + a, Y + b) = \frac{\operatorname{cov}(X + a, Y + b)}{\sigma(X + a)\,\sigma(Y + b)} = \frac{\operatorname{cov}(X, Y)}{\sigma(X)\,\sigma(Y)} = \rho(X, Y).$$

3. Аналогично, с использованием ранее доказанных свойств ковариации и среднеквадратического отклонения, получаем:

$$\rho(aX, bY) = \frac{\operatorname{cov}(aX, bY)}{\sigma(aX)\,\sigma(bY)} = \frac{ab\operatorname{cov}(X, Y)}{|a|\sigma(X)\,|b|\sigma(Y)} = \frac{ab}{|ab|}\rho(X, Y).$$

Напомним, что функция $\operatorname{sgn}(x)$ принимает значение 1 при $x > 0$, значение -1 при $x < 0$ и значение 0 при $x = 0$. Если a и b отличны от 0, то $\frac{ab}{|ab|} = \operatorname{sgn}(ab)$. Таким образом, если умножить случайные величины X и Y на константы одного знака, то их коэффициент корреляции не изменится, а если умножить на константы разных знаков, то коэффициент корреляции поменяет знак. Абсолютная величина коэффициента корреляции при этом останется неизменной.

4. Следует из свойства ковариации и определения среднеквадратического отклонения:

$$\rho(X, X) = \frac{\operatorname{cov}(X, X)}{\sigma(X)\,\sigma(X)} = \frac{\mathsf{D}\,X}{\mathsf{D}\,X} = 1. \qquad \square$$

Пример 66 ([4]). Пусть с. в. X принимает значения -2, -1, 1 и 2 с одинаковыми вероятностями $1/4$. Пусть $Y = X^2$ (т. е. X и Y зависимы). Покажем, что $\rho(X, Y) = 0$.

Легко вычисляется: $\mathsf{M}\,X = 0$ и $\mathsf{M}\,Y = 5/2$. В то же время

$$\mathsf{M}(XY) = \mathsf{M}(X^3) = (-2)^3 \cdot \frac{1}{4} + (-1)^3 \cdot \frac{1}{4} + 1^3 \cdot \frac{1}{4} + 2^3 \cdot \frac{1}{4} = 0.$$

Следовательно, $\operatorname{cov}(X,Y) = \mathsf{M}(XY) - \mathsf{M}\,X\ \mathsf{M}\,Y = 0$ и $\rho(X,Y) = 0$.

Теорема 3.14 (о коэффициенте корреляции)**.** *Для любых случайных величин X и Y верно*

$$|\rho(X,Y)| \leqslant 1.$$

Причем $|\rho(X,Y)| = 1$ тогда и только тогда, когда найдутся числа $a \neq 0$ и b такие, что $Y = aX + b$.

Доказательство. Пусть X^* и Y^* — нормированные случайные величины (а значит, $\mathsf{D}\,X^* = \mathsf{D}\,Y^* = 1$). Тогда, согласно формуле (3.9),

$$\begin{aligned} \mathsf{D}(X^* \pm Y^*) &= \mathsf{D}(X^*) + \mathsf{D}(Y^*) \pm 2\operatorname{cov}(X^*,Y^*) = \\ &= 2 \pm 2\operatorname{cov}(X^*,Y^*) = 2(1 \pm \rho(X,Y)). \end{aligned} \tag{3.13}$$

Так как $\mathsf{D}(X^* \pm Y^*) \geqslant 0$ всегда, то справедливы неравенства

$$1 + \rho(X,Y) \geqslant 0 \qquad \text{и} \qquad 1 - \rho(X,Y) \geqslant 0.$$

Откуда

$$-1 \leqslant \rho(X,Y) \leqslant 1,$$

и первое утверждение теоремы доказано.

Если $Y = aX + b$, то, согласно свойствам 2–4 коэффициента корреляции,

$$\rho(X,Y) = \rho(X, aX+b) = \rho(X, aX) = \frac{a}{|a|}\rho(X,X) = \pm 1.$$

Предположим теперь, что $\rho(X,Y) = 1$. Тогда из формулы (3.13) следует $\mathsf{D}(X^* - Y^*) = 0$. А значит, разность $X^* - Y^*$ равна[9] некоторой константе c. Другими словами,

$$\frac{X - \mathsf{M}\,X}{\sigma X} - \frac{Y - \mathsf{M}\,Y}{\sigma Y} = c.$$

[9]Строго говоря, с вероятностью 1.

Откуда получаем $Y = aX + b$, где

$$a = \frac{\sigma Y}{\sigma X}, \qquad b = \mathsf{M}\,Y - \left(c + \frac{\mathsf{M}\,X}{\sigma X}\right)\sigma Y.$$

Отметим, что в этом случае $a > 0$.

Аналогично доказывается, что если $\rho(X, Y) = -1$, то $Y = aX + b$ для некоторых a и b (причем $a < 0$). □

Пример 67. Пусть X и Y — независимые, одинаково распределенные случайные величины, и пусть $Z = X + Y$. Найти $\rho(X, Z)$.

Решение. Обозначим для удобства $b = \mathsf{D}\,X = \mathsf{D}\,Y$ (X и Y одинаково распределены, а значит, их дисперсии совпадают). Воспользуемся тем, что $\operatorname{cov}(X, Y) = 0$ и $\mathsf{D}(X + Y) = \mathsf{D}\,X + \mathsf{D}\,Y$ для независимых X и Y:

$$\rho(X, Z) = \frac{\operatorname{cov}(X, X + Y)}{\sqrt{\mathsf{D}\,X\ \mathsf{D}(X + Y)}} = \frac{\operatorname{cov}(X, X) + \operatorname{cov}(X, Y)}{\sqrt{b(b + b)}} = \frac{b + 0}{b\sqrt{2}} = \frac{1}{\sqrt{2}}.$$

□

3.6. Функция распределения

До настоящего момента речь шла только о дискретных случайных величинах, множество значений которых конечно либо счетно. Наиболее естественным способом описания их законов распределения являются ряды распределения. Далее мы будем рассматривать случайные величины других типов и для описания их распределений потребуются другие понятия[10]. В частности, универсальным способом описания закона распределения случайной величины любого типа является функция распределения.

Определение 3.15. *Функция F_X распределения* случайной величины X определяется соотношением

$$F_X(x) = \mathsf{P}(X < x), \quad \text{где } x \in \mathbb{R}.$$

[10]Множество значений недискретных случайных величин несчетно и поэтому не может быть записано в ряд.

Далее, если из контекста ясно, о какой случайной величине идет речь, или, если речь идет об общих свойствах случайных величин, мы будем пользоваться обозначением F вместо F_X.

Рассмотрим несколько примеров.

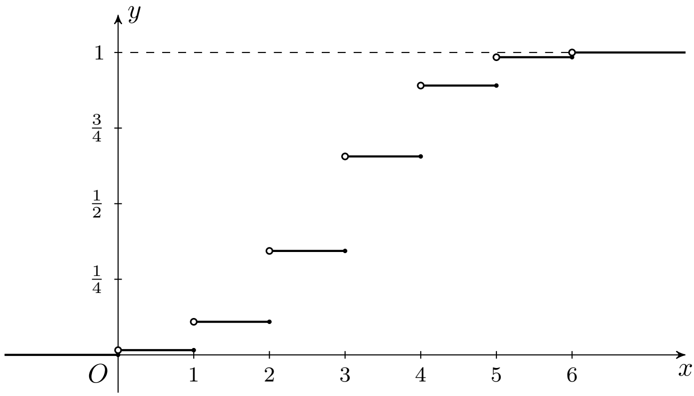

Рис. 3.2. Функция распределения $y = F_X(x)$ для с. в. X, распределенной по биномиальному закону с параметрами $n = 6$ и $p = 1/2$

Предположим, что X распределена по биномиальному закону с параметрами n и p. Тогда, очевидно, $F_X(x) = P(X < x) = 0$ для всех $x \leqslant 0$. Если же $x > n$, то $F_X(x) = P(X < x) = 1$. И, наконец,

$$F_X(x) = \sum_{k<x} P(X = k), \quad \text{где } k = 0, 1, \ldots, n.$$

В целом график функции $F_X(x)$ представляет собой ступенчатую линию, где высота k-й ступеньки (относительно предыдущей ступеньки) равна $P(X = k - 1) = C_n^k p^k q^{n-k}$. В частности, ряд распределения X для $n = 6$ и $p = 1/2$ имеет вид

X	0	1	2	3	4	5	6
P	$\frac{1}{64}$	$\frac{6}{64}$	$\frac{15}{64}$	$\frac{20}{64}$	$\frac{15}{64}$	$\frac{6}{64}$	$\frac{1}{64}$

А график соответствующей функции распределения приведен на рис. 3.2. На рис. 3.3 изображены еще два примера графиков функций распределения.

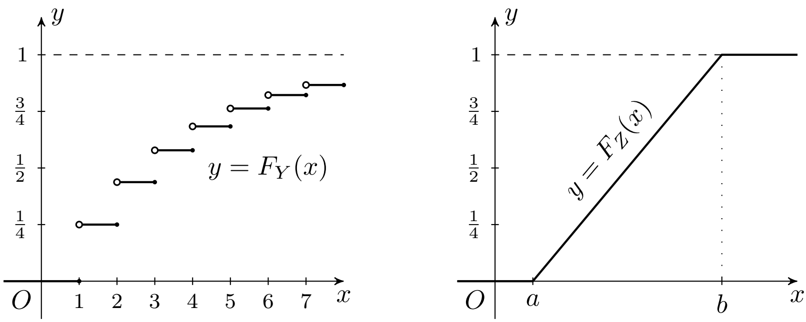

Рис. 3.3. Функции распределения для с. в. Y, распределенной по геометрическому закону с параметром $p = 1/4$, и для с. в. Z, равномерно распределенной на отрезке $[a, b]$

Непосредственно из определения 3.15 и из свойств вероятности выводятся **свойства функции распределения:**

1. $0 \leqslant F(x) \leqslant 1$ для любого $x \in \mathbb{R}$;

2. $P(x_1 \leqslant X < x_2) = F_X(x_2) - F_X(x_1)$, $x_1 < x_2$. В частности, отсюда следует, что F_X — неубывающая функция;

3. $F(+\infty) = 1$, $F(-\infty) = 0$;

4. F непрерывна слева: $\lim\limits_{x \to x_0 - 0} F(x) = F(x_0)$.

Доказательство.

1. Свойство вытекает из определения функции распределения как вероятности некоторого события:

$$F_X(x) = \mathsf{P}(X < x) \in [0, 1].$$

2. Пусть $x_1 < x_2$. Тогда событие $\{X < x_2\}$ можно представить в виде суммы двух несовместных событий:

$$\{X < x_2\} = \{X < x_1\} + \{x_1 \leqslant X < x_2\}.$$

Соответственно, для вероятностей этих событий выполнено соотношение:

$$F_X(x_2) = \mathsf{P}(X < x_2) = \mathsf{P}(\{X < x_1\} + \{x_1 \leqslant X < x_2\}) = \\ = \mathsf{P}(X < x_1) + \mathsf{P}(x_1 \leqslant X < x_2) = F_X(x_1) + \mathsf{P}(x_1 \leqslant X < x_2).$$

Отсюда непосредственно получаем, что

$$\mathsf{P}(x_1 \leqslant X < x_2) = F_X(x_2) - F_X(x_1).$$

Отметим, что так как $F_X(x_2) - F_X(x_1) = \mathsf{P}(x_1 \leqslant X < x_2) \geqslant 0$, то $F_X(x_2) \geqslant F_X(x_1)$. Следовательно, функция F_X — неубывающая (нестрого возрастающая).

3. Из монотонности (свойство 2) и ограниченности сверху и снизу (свойство 1) функции F_X следует, что у неё существуют пределы при $x \to +\infty$ и $x \to -\infty$ (по теореме Вейерштрасса). Осталось доказать, что эти пределы равны 1 и 0 соответственно. Для этого будем использовать следующий подход: рассмотрим подходящую последовательность $x_n \to +\infty$ (при $n \to \infty$) и докажем, что $F_X(x_n) \to 1$ (при $n \to \infty$), аналогично для предела на $-\infty$.

Пусть $x_n = -n \to -\infty$ (при $n \to \infty$). Рассмотрим последовательность попарно несовместных событий $A_k = \{-k-1 \leqslant X < -k\}$, их вероятности обозначим через $p_k = \mathsf{P}(-k-1 \leqslant X < -k)$. Заметим, что событие $X < -1$ раскладывается в сумму: $\{X < -1\} = \sum\limits_{k=1}^{\infty} A_k$, поэтому

$$\mathsf{P}(X < -1) = \mathsf{P}\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathsf{P}(A_k) = \sum_{k=1}^{\infty} p_k.$$

Получили, что $\sum\limits_{k=1}^{\infty} p_k$ — это сходящийся (к $\mathsf{P}(X < -1)$) ряд, поэтому его остаток $\sum\limits_{k=n}^{\infty} p_k$ сходится к 0. Заметим, что именно это и требовалось:

$$F_X(x_n) = \mathsf{P}(X < x_n) = \mathsf{P}(X < -n) = \sum_{k=n}^{\infty} p_k \to 0.$$

Аналогично для доказательства $F_X(+\infty) = 1$ рассмотрим последовательность $x_n = n \to +\infty$ (при $n \to \infty$) и покажем, что $F_X(x_n) = F_X(n) \to 1$. Рассмотрим такую последовательность попарно несовместных событий $A_k = \{k \leqslant X < k+1\}$, их вероятности обозначим через $p_k = \mathsf{P}(k \leqslant X < k+1)$, также заметим, что $\{X \geqslant 1\} = \sum\limits_{k=1}^{\infty} A_k$. Отсюда

$$\mathsf{P}(X \geqslant 1) = \mathsf{P}\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathsf{P}(A_k) = \sum_{k=1}^{\infty} p_k.$$

Снова получаем, что $\sum\limits_{k=1}^{\infty} p_k$ — это сходящийся (к $\mathsf{P}(X \geqslant 1)$) ряд, поэтому его остаток $\sum\limits_{k=n}^{\infty} p_k$ сходится к 0. Тогда

$$\mathsf{P}(X \geqslant n) = \sum_{k=n}^{\infty} p_k \to 0.$$

Соответственно,

$$F_X(x_n) = F_X(n) = \mathsf{P}(X < n) = 1 - \mathsf{P}(X \geqslant n) \to 1 - 0 = 1$$

при $n \to +\infty$, что и требовалось доказать.

4. Существование предела $F_X(x_0 - 0) = \lim\limits_{x \to x_0 - 0} F_X(x)$ следует из монотонности и ограниченности функции F_X, поэтому достаточно (как и при доказательстве свойства 3) рассмотреть произвольную подходящую последовательность $x_n \to x_0 - 0$ (при $n \to \infty$) и убедиться, что $F_X(x_n) \to F_X(x_0)$.

Пусть $x_n = x_0 - \frac{1}{n}$. Рассмотрим последовательность попарно несовместных событий $A_k = \left\{x_0 - \dfrac{1}{k+1} \leqslant X < x_0 - \dfrac{1}{k}\right\} = \{x_{k+1} \leqslant X < x_k\}$. Их вероятности $p_k = \mathsf{P}(x_{k+1} \leqslant X < x_k)$, $\{x_0 - 1 \leqslant X < x_0\} = \sum\limits_{k=1}^{\infty} A_k$. Поэтому

$$\mathsf{P}(x_0 - 1 \leqslant X < x_0) = \mathsf{P}\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathsf{P}(A_k) = \sum_{k=1}^{\infty} p_k.$$

Это сходящийся ряд, поэтому его остаток $\sum\limits_{k=n}^{\infty} p_k$ сходится к 0. Тогда

$$F_X(x_0) - F_X(x_n) = \mathsf{P}(x_n \leqslant X < x_0) = \sum_{k=n}^{\infty} p_k \to 0,$$

откуда следует, что $F_X(x_n) \to F_X(x_0)$ при $n \to \infty$, что и требовалось доказать. □

Определение 3.16. Случайная величина называется *непрерывной*, если непрерывна ее функция распределения.

Например, на рис. 3.3 изображен график функции распределения непрерывной случайной величины Z, а также график функции распределения случайной величины Y, не являющейся непрерывной.

5. Для любой непрерывной случайной величины X и любого $x \in \mathbb{R}$ выполнено $\mathsf{P}(X = x) = 0$.

Последнее свойство означает, что любое свое значение непрерывная случайная величина принимает с нулевой вероятностью. Этот факт является еще одной причиной, почему для некоторых случайных величин неприемлем способ описания закона распределения в виде таблицы распределения (т. е. в форме перечисления значений и указания вероятностей, с которыми эти значения достигаются). Также для непрерывных случайных величин в свойстве 2 границы промежутков можно произвольно включать или не включать в соответствующий интервал:

$$\begin{aligned} F_X(x_2) - F_X(x_1) = \mathsf{P}(x_1 \leqslant X < x_2) = \mathsf{P}(x_1 \leqslant X \leqslant x_2) = \\ = \mathsf{P}(x_1 < X \leqslant x_2) = \mathsf{P}(x_1 < X < x_2). \end{aligned}$$

Доказательство свойства 5. Покажем, что для любой случайной величины X справедливо более общее свойство:

$$\lim_{x \to x_0 + 0} F_X(x) = F_X(x_0) + \mathsf{P}(X = x_0), \quad \forall x_0 \in \mathbb{R}.$$

Для этого достаточно доказать, что

$$\lim_{x \to x_0 + 0} \mathsf{P}(x_0 \leqslant X < x) = \lim_{x \to x_0 + 0} (F_X(x) - F_X(x_0)) = \mathsf{P}(X = x_0).$$

Существование предела следует из монотонности и ограниченности функции F_X. Поэтому, как и при доказательстве свойств 3 и 4, достаточно рассмотреть подходящую последовательность $x_n \to x_0+0$ и доказать, что $\mathsf{P}(x_0 \leqslant X < x_n) \to \mathsf{P}(X = x_0)$.

Выберем $x_n = x_0 + \frac{1}{n}$ и рассмотрим попарно несовместные события $A_k = \left\{x_0 + \dfrac{1}{k+1} \leqslant X < x_0 + \dfrac{1}{k}\right\} = \{x_{k+1} \leqslant X < x_k\}$, их вероятности традиционно обозначим как $p_k = \mathsf{P}(x_{k+1} \leqslant X < x_k)$. Заметим, что $\{x_0 \leqslant X < x_0 + 1\} = \{X = x_0\} + \sum\limits_{k=1}^{\infty} A_k$, поэтому

$$\mathsf{P}(x_0 \leqslant X < x_0 + 1) = \mathsf{P}\left(\{X = x_0\} + \sum_{k=1}^{\infty} A_k\right) = \\ = \mathsf{P}(X = x_0) + \sum_{k=1}^{\infty} \mathsf{P}(A_k) = \mathsf{P}(X = x_0) + \sum_{k=1}^{\infty} p_k.$$

Следовательно, $\sum\limits_{k=1}^{\infty} p_k$ — это сходящийся ряд, поэтому его остаток $\sum\limits_{k=n}^{\infty} p_k$ сходится к 0. Теперь осталось завершить доказательство:

$$\mathsf{P}(x_0 \leqslant X < x_n) = \mathsf{P}\left(\{X = x_0\} + \sum_{k=n}^{\infty}\{x_{k+1} \leqslant X < x_k\}\right) = \\ = \mathsf{P}(X = x_0) + \sum_{k=n}^{\infty} \mathsf{P}(x_{k+1} \leqslant X < x_k) = \\ = \mathsf{P}(X = x_0) + \sum_{k=n}^{\infty} p_k \to \mathsf{P}(X = x_0) + 0 = \mathsf{P}(X = x_0).$$

Для доказательства исходного свойства 5 достаточно отметить, что если X — непрерывная случайная величина, то по определению F_X — непрерывная функция, поэтому $\lim\limits_{x \to x_0+0} F_X(x) = F_X(x_0)$, а тогда

$$\mathsf{P}(X = x_0) = \lim_{x \to x_0+0} F_X(x) - F_X(x_0) = F_X(x_0) - F_X(x_0) = 0. \qquad \square$$

Оказывается, свойств 3, 4 и свойства монотонности достаточно для того, чтобы соответствующая функция оказалась функцией распределения некоторой случайной величины.

Теорема 3.17. *Функция $F : \mathbb{R} \to \mathbb{R}$ является функцией распределения некоторой случайной величины тогда и только тогда, когда выполняются следующие свойства:*

1. *F неубывающая на всей числовой оси;*

2. *$F(+\infty) = 1$, $F(-\infty) = 0$;*

3. *F непрерывна слева.*

Задача 68. Обратим внимание на то, что график дискретной случайной величины представляет собой ступенчатую линию (см. рис. 3.2 и 3.3), где длина ступени равна расстоянию между соседними значениями случайной величины. При этом крайние ступени на рис. 3.2 и 3.3 имеют бесконечную длину. Попробуйте привести пример дискретной случайной величины, все ступени которой имеют длину 1.

3.7. Плотность распределения

Определение 3.18. Случайная величина X называется *абсолютно непрерывной*, если найдется функция f_X такая, что

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(t)\,dt, \qquad \forall x \in \mathbb{R}.$$

Эта функция f_X называется *плотностью распределения* случайной величины X.

Из определения следует, что для абсолютно непрерывной случайной величины X ее функция распределения F_X является кусочно дифференцируемой и

$$f_X(x) = F'_X(x) \qquad \text{почти всюду}^{11}.$$

Далее будем пользоваться обозначением f вместо f_X, если из контекста ясно, о какой случайной величине идет речь, или выбор случайной величины не имеет значения.

Из определения плотности и из свойств функции распределения выводятся **свойства плотности распределения:**

1. $f(x) \geqslant 0$ для всех $x \in R$;

2. $\mathsf{P}(x_1 \leqslant X \leqslant x_2) = \int\limits_{x_1}^{x_2} f_X(x)\,dx$;

3. $\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1$.

3.7.1. Локальный смысл плотности

Приведённое выше свойство 2 плотности распределения случайной величины даёт нам возможность вычислять вероятность попадания абсолютно непрерывной случайной величины в заданный интервал. Для этого достаточно взять интеграл от плотности распределения по этому интервалу (или, что то же самое, вычислить площадь под графиком плотности распределения на заданном интервале). В то же время, само значение плотности распределения в конкретной точке для нас пока не несёт никакой смысловой нагрузки (в самом деле, из соотношений $f_X(x_0) = 10$ или $f_X(x_0) = 0{,}4$ мы пока не можем сделать никаких выводов). Ответ на вопрос, какой же математический смысл несёт значение плотности распределения, даёт следующая теорема о локальном смысле плотности.

Теорема 3.19. *Пусть X — абсолютно непрерывная случайная величина, $f_X(x)$ — её плотность распределения, непрерывная в точ-*

[11] Термин «почти всюду» подразумевает, что данное равенство выполняется для всех x, кроме, быть может, множества точек меры 0. Например, допустимо, чтобы функция распределения была недифференцируемой в нескольких точках или даже в счетном множестве точек.

ке x_0. Пусть также $A_\delta(x_0) = \{x : |x - x_0| < \delta\}$ — интервал с центром в точке x_0 (ширина интервала равна $|A_\delta(x_0)| = 2\delta$). Тогда

$$\lim_{|A_\delta(x_0)|\to 0} \frac{\mathsf{P}(X \in A_\delta(x_0))}{|A_\delta(x_0)|} = f_X(x_0).$$

Несмотря на несколько пугающую формулировку, это утверждение может быть переформулировано более простыми словами: плотность распределения случайной величины в точке — это «приблизительно» отношение вероятности попадания случайной величины в малую окрестность данной точки к длине этой окрестности. Точно таким же образом определяется, например, локальная плотность неоднородного тела — как отношение массы небольшого фрагмента тела к объёму этого фрагмента. Поэтому слово «плотность» в таком способе определения закона распределения случайной величины неслучайно и имеет совершенно естественный физический смысл.

Доказательство. Первый способ. Для одномерной (скалярной) случайной величины можно воспользоваться таким простым рассуждением:

$$\frac{\mathsf{P}(X \in A_\delta(x_0))}{|A_\delta(x_0)|} = \frac{F_X(x_0+\delta) - F_X(x_0-\delta)}{(x_0+\delta)-(x_0-\delta)} = F'_X(c) = f_X(c),$$

где c — некоторая точка на интервале $A_\delta(x_0)$ (по теореме Лагранжа о промежуточном значении). Если теперь устремить ширину интервала к 0, то $f_X(c) \to f_X(x_0)$ (из непрерывности функции f_X в точке x_0), откуда и следует утверждение теоремы.

Второй способ. Воспользуемся определением непрерывности функции f_X. Выберем $\varepsilon > 0$, тогда существует такое $\delta > 0$, что для всех $x \in A_\delta(x_0)$: $|f_X(x) - f_X(x_0)| < \varepsilon$ (для удобства можем полагать, что $\delta \leqslant \varepsilon$: если это не так, то окрестность $A_\delta(x_0)$ можно сузить до нужной ширины — чтобы при $\varepsilon \to 0$ выполнялось $\delta \to 0$).

Тогда

$$\mathsf{P}(X \in A_\delta) = \int\limits_{A_\delta} f_X(x)dx = \int\limits_{A_\delta} f_X(x_0)dx + \int\limits_{A_\delta} (f_X(x) - f_X(x_0))dx =$$

$$= f_X(x_0)|A_\delta| + \int\limits_{A_\delta} (f_X(x) - f_X(x_0))dx.$$

Используя определение непрерывности f_X, оценим второй интеграл:

$$\left|\int\limits_{A_\delta}(f_X(x)-f_X(x_0))dx\right| \leqslant \int\limits_{A_\delta}|f_X(x)-f_X(x_0)|dx \leqslant \int\limits_{A_\delta}\varepsilon dx = \varepsilon|A_\delta|.$$

Тогда

$$|\mathsf{P}(X\in A_\delta) - f_X(x_0)|A_\delta|| \leqslant \varepsilon|A_\delta|,$$

что после деления на $|A_\delta|$ даёт следующее ограничение:

$$\left|\frac{|\mathsf{P}(X\in A_\delta)}{|A_\delta|} - f_X(x_0)\right| \leqslant \varepsilon.$$

Теперь устремим $\varepsilon \to 0$, при этом также $|A_\delta| \to 0$, откуда получаем, что подмодульное выражение слева стремится к 0, что и требовалось.

Отметим, что второй способ доказательства без изменений переносится на случай абсолютно непрерывных многомерных случайных величин (в окрестностях точек, где их плотность распределения непрерывна). $\square$

Доказанная теорема позволяет проводить грубую визуальную оценку особенностей распределения случайной величины по графику её плотности распределения. Утверждение теоремы даёт оценку вероятности попадания случайной величины в малый интервал как значение плотности распределения в точке из этого интервала, умноженное на длину интервала. Иначе говоря, если рассмотреть малый интервал на числовой прямой, то вероятность попадания случайной величины в такой интервал будет тем больше, чем больше значение плотности распределения.

3.7.2. Примеры абсолютно непрерывных случайных величин

Случайная величина X *распределена равномерно на отрезке* $[a,b]$, если ее плотность распределения постоянна на данном отрезке и равна нулю вне его. (Следовательно, X может принимать значения только из отрезка $[a,b]$, т. е. $a \leqslant X \leqslant b$ при любом исходе.)

Предположим, что X распределена равномерно на отрезке $[a, b]$ и ее плотность равна постоянной c на этом отрезке, тогда значение c однозначно определяется с помощью свойства 3. А именно, интеграл

$$\int\limits_{-\infty}^{+\infty} f_X(x)\,dx = \int\limits_{-\infty}^{a} 0\,dx + \int\limits_{a}^{b} c\,dx + \int\limits_{b}^{+\infty} 0\,dx = \int\limits_{a}^{b} c\,dx = c(b-a)$$

должен быть равен единице. Откуда $c = \frac{1}{b-a}$.

Таким образом, плотность случайной величины, равномерно распределенной на отрезке $[a, b]$, определяется формулой

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

Найдем общий вид для функции распределения случайной величины X, равномерно распределенной на отрезке $[a, b]$. Для этого разберем три случая:

1) если $x < a$, то $F_X(x) = P(X < x) = 0$, так как $X \geqslant a$ всегда;
2) если $x > b$, то $F_X(x) = P(X < x) = 1$, так как $X \leqslant b$ всегда;
3) если же $x \in [a, b]$, то

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(t)\,dt = \int\limits_{-\infty}^{a} f_X(t)\,dt + \int\limits_{a}^{x} f_X(t)\,dt =$$
$$= \int\limits_{-\infty}^{a} 0\,dt + \int\limits_{a}^{x} \frac{dt}{b-a} = \frac{x-a}{b-a}.$$

Таким образом,

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

Графики плотности и функции распределения для такой случайной величины изображены на рис. 3.4.

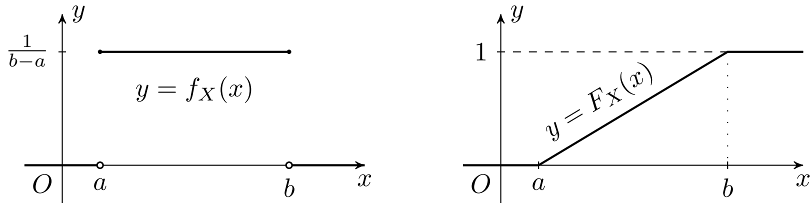

Рис. 3.4. Графики плотности $y = f_X(x)$
и функции $y = F_X(x)$ распределения для с. в. X,
равномерно распределенной на отрезке $[a,b]$

Приведем еще пару примеров абсолютно непрерывных случайных величин.

Плотность распределения случайной величины Y, *распределенной по показательному закону с параметром* $\alpha > 0$, выражается формулой

$$f_Y(x) = \begin{cases} \alpha e^{-\alpha x}, & \text{при } x \geqslant 0, \\ 0, & \text{иначе.} \end{cases}$$

(Т. е. $Y \geqslant 0$ всегда.) Соответственно, функция распределения $F_Y(x)$ равна нулю при $x < 0$. Если же $x \geqslant 0$, то

$$F_Y(x) = \int\limits_{-\infty}^{x} f_Y(t)\,dt = \int\limits_{-\infty}^{0} 0\,dt + \int\limits_{0}^{x} \alpha e^{-\alpha t}\,dt =$$
$$= -e^{-\alpha t}\Big|_0^x = 1 - e^{-\alpha x}.$$

Графики для $y = f_Y(x)$ и $y = F_Y(x)$ изображены на рис. 3.5.

Случайная величина Z называется *распределенной по нормальному закону с параметрами* a *и* σ^2, если ее плотность распределения задается формулой

$$f_Z(x) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x-a}{\sigma}\right), \quad x \in \mathbb{R},$$

где функция $\varphi(x)$ определяется формулой (2.7) на с. 49. Там же при-

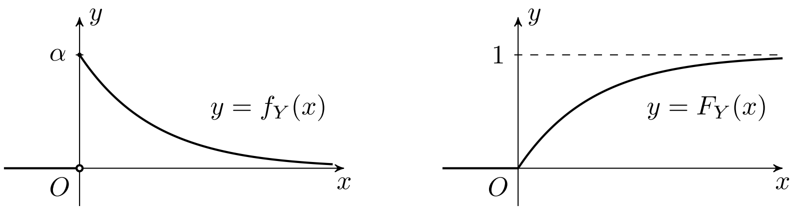

Рис. 3.5. Графики плотности и функции распределения для с. в. Y, распределенной по показательному закону с параметром α

водятся её основные свойства. Кроме того, согласно определению 2.6,

$$\Phi(x) = \int\limits_0^x \varphi(t)\,dt,$$

где $\Phi(x)$ — функция Лапласа.

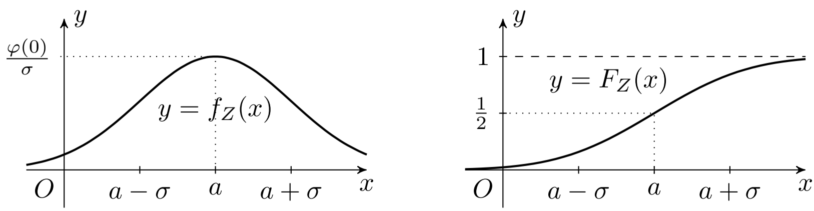

Рис. 3.6. Графики плотности и функции распределения для с. в. Z, распределенной по нормальному закону с параметрами a и σ^2

Соответственно, функция распределения с. в. Z равна

$$F_Z(x) = \int\limits_{-\infty}^x \frac{1}{\sigma}\,\varphi\left(\frac{t-a}{\sigma}\right)\,dt.$$

Сделаем замену $y = \frac{t-a}{\sigma}$ (следовательно, $dy = \frac{dt}{\sigma}$) и воспользуемся тем, что

$$\int_{-\infty}^{z} \varphi(y)\,dy = \int_{-\infty}^{0} \varphi(y)\,dy + \int_{0}^{z} \varphi(y)\,dy = 1/2 + \Phi(z),$$

где $\Phi(z)$ — функция Лапласа. Получаем

$$F_Z(x) = \int_{-\infty}^{\frac{x-a}{\sigma}} \varphi(y)\,dy = 1/2 + \Phi\left(\frac{x-a}{\sigma}\right).$$

3.8. Математическое ожидание и дисперсия непрерывных случайных величин

Определение 3.20. *Математическое ожидание* абсолютно непрерывной случайной величины X вычисляется по формуле

$$\mathsf{M}\,X = \int_{-\infty}^{+\infty} x f_X(x)\,dx.$$

Математическое ожидание существует, если интеграл в правой части сходится абсолютно.

Так, например, случайная величина X, равномерно распределенная на отрезке $[a,b]$, имеет плотность

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

А ее математическое ожидание равно

$$\mathsf{M}\,X = \int_{a}^{b} x \cdot \frac{1}{b-a}\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}.$$

Математическое ожидание непрерывной случайной величины обладает теми же свойствами (см. раздел 3.3.1 на с. 67), что и математическое ожидание в дискретном случае. В частности,

$$\mathsf{M}(g(X)) = \int\limits_{-\infty}^{+\infty} g(x) f_X(x)\, dx$$

при условии, что интеграл в правой части существует[12].

Дисперсия непрерывной случайной величины X вычисляется по формуле

$$\mathsf{D}\, X = \mathsf{M}(X - \mathsf{M}\, X)^2 = \int\limits_{-\infty}^{+\infty} (x - \mathsf{M}\, X)^2 f(x)\, dx,$$

или

$$\mathsf{D}\, X = \mathsf{M}(X^2) - (\mathsf{M}\, X)^2 = \int\limits_{-\infty}^{+\infty} x^2 f(x)\, dx - (\mathsf{M}\, X)^2.$$

В частности, дисперсия с. в. X, равномерно распределенной на отрезке $[a, b]$, равна

$$\mathsf{D}\, X = \int\limits_{a}^{b} \frac{x^2\, dx}{b-a} - \left(\frac{b+a}{2}\right)^2 = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

А среднеквадратическое отклонение

$$\sigma X = \sqrt{\mathsf{D}\, X} = \frac{b-a}{2\sqrt{3}}.$$

[12]Нахождение математического ожидания для функции от двух случайных величин, вообще говоря, предполагает вычисление двойного интеграла

$$\mathsf{M}(g(X,Y)) = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} g(x,y) f(x,y)\, dxdy,$$

где $f(x,y)$ — плотность распределения вектора (X,Y). Более подробно этот вопрос рассматривается, например, в [6] и [5].

Пример 69. Найти математическое ожидание и дисперсию для случайной величины X, распределенной по показательному закону с параметром $\alpha > 0$.

Решение. Плотность распределения вероятностей случайной величины X равна $\alpha e^{-\alpha x}$ при $x \geqslant 0$ и равна нулю при $x < 0$. Согласно определению,

$$\mathsf{M}\,X = \int\limits_0^{+\infty} x\,\alpha\,e^{-\alpha x}\,dx.$$

Воспользуемся методом интегрирования по частям:

$$\int x\,\alpha\,e^{-\alpha x}\,dx = \int x\,d(-e^{-\alpha x}) = -e^{-\alpha x}x + \int e^{-\alpha x}\,dx = \\ = -e^{-\alpha x}x + \frac{e^{-\alpha x}}{-\alpha} = -\frac{x+1/\alpha}{e^{\alpha x}}.$$

Следовательно,

$$\mathsf{M}\,X = -\frac{x+1/\alpha}{e^{\alpha x}}\bigg|_0^{+\infty} = \frac{0+1/\alpha}{e^0} - \lim_{x\to+\infty}\frac{x+1/\alpha}{e^{\alpha x}} = \frac{1}{\alpha}.$$

Действуя по аналогии, находим

$$\mathsf{M}\,X^2 = \int\limits_0^{+\infty} x^2\alpha e^{-\alpha x}\,dx = \int\limits_0^{+\infty} x^2\,d(-e^{-\alpha x}) = \\ = -e^{-\alpha x}x^2\bigg|_0^{+\infty} + \int\limits_0^{+\infty} 2xe^{-\alpha x}\,dx = 0 + \frac{2}{\alpha}\,\mathsf{M}\,X = \frac{2}{\alpha^2}.$$

Следовательно,

$$\mathsf{D}\,X = \mathsf{M}\,X^2 - (\mathsf{M}\,X)^2 = \frac{2}{\alpha^2} - \frac{1}{\alpha^2} = \frac{1}{\alpha^2}. \qquad \square$$

Пример 70. Найти математическое ожидание и дисперсию для случайной величины X, распределенной по нормальному закону с параметрами a и σ^2.

Решение. Плотность распределения такой случайной величины задается следующей формулой:

$$f_X(x) = \frac{1}{\sigma} \cdot \varphi\left(\frac{x-a}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}, \quad x \in \mathbb{R}.$$

Найдем математическое ожидание случайной величины X:

$$\mathsf{M}\,X = \int\limits_{-\infty}^{+\infty} \frac{x}{\sigma}\,\varphi\left(\frac{x-a}{\sigma}\right) dx = \left| \begin{array}{c} \text{Замена: } t = \frac{x-a}{\sigma} \\ x = \sigma\, t + a,\ dx = \sigma\, dt \end{array} \right| =$$

$$= \int\limits_{-\infty}^{+\infty} (\sigma\, t + a)\,\varphi(t)\,dt = \sigma \int\limits_{-\infty}^{+\infty} t\,\varphi(t)\,dt + a \int\limits_{-\infty}^{+\infty} \varphi(t)\,dt.$$

Воспользуемся свойствами функции $\varphi(t)$ (см. с. 49). Во-первых, эта функция четная, следовательно, функция $t\,\varphi(t)$ — нечетная. Тогда $\int\limits_{-\infty}^{+\infty} t\,\varphi(t)\,dt = 0$. Во-вторых, $\int\limits_{-\infty}^{+\infty} \varphi(x)\,dx = 1$. Таким образом,

$$\mathsf{M}\,X = \sigma \cdot 0 + a \cdot 1 = a.$$

Действуя по аналогии, найдем дисперсию:

$$\mathsf{D}\,X = \int\limits_{-\infty}^{+\infty} (x - \mathsf{M}\,X)^2 f_X(x)\,dx = \int\limits_{-\infty}^{+\infty} \frac{(x-a)^2}{\sigma}\,\varphi\left(\frac{x-a}{\sigma}\right) dx =$$

$$= \left| \begin{array}{c} \text{Замена: } t = \frac{x-a}{\sigma} \\ x = \sigma\, t + a,\ dx = \sigma\, dt \end{array} \right| =$$

$$= \int\limits_{-\infty}^{+\infty} \sigma^2\, t^2\, \varphi(t)\,dt = \frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} t^2 e^{-t^2/2}\,dt.$$

Проинтегрируем по частям:

$$\int\limits_{-\infty}^{+\infty} t^2 e^{-t^2/2}\,dt = \int\limits_{-\infty}^{+\infty} t \cdot t e^{-t^2/2}\,dt = \left| \begin{array}{c} u = t,\ dv = te^{-t^2/2}\,dt \\ du = dt,\ v = -e^{-t^2/2} \end{array} \right| =$$

$$= -te^{-t^2/2}\Big|_{-\infty}^{+\infty} + \int\limits_{-\infty}^{+\infty} e^{-t^2/2}\,dt = 0 + \int\limits_{-\infty}^{+\infty} e^{-t^2/2}\,dt.$$

Следовательно,

$$\mathsf{D}\,X = \frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-t^2/2}\,dt = \sigma^2 \int\limits_{-\infty}^{+\infty} \varphi(x)\,dx = \sigma^2.$$

Таким образом, если мы говорим о случайной величине, распределенной по нормальному закону с параметрами a и σ^2, то первый из этих параметров является математическим ожиданием, а второй — дисперсией этой случайной величины. □

3.9. Преобразование случайных величин

3.9.1. Закон распределения зависимой случайной величины

При рассмотрении формул для расчета математического ожидания было упомянуто, что если известен закон распределения случайной величины X и задана функциональная зависимость $Y = g(X)$, то для нахождения математического ожидания случайной величины Y существует достаточно удобная вычислительная формула. В этом разделе мы рассмотрим, каким образом не просто найти числовые характеристики зависимой случайной величины, но и описать закон ее распределения.

Начнем с рассмотрения важного частного случая — линейной зависимости между случайными величинами.

Теорема 3.21. *Пусть плотность распределения случайной величины X равна $f_X(x)$, а случайная величина Y линейно зависит от X: $Y = aX + b$ ($a \neq 0$). Тогда плотность распределения случайной величины Y равна*

$$f_Y(x) = f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right). \tag{3.14}$$

Доказательство. Обозначим функции распределения случайных величин X и Y через $F_X(x)$ и $F_Y(x)$ соответственно.

Рассмотрим два случая.

1. $a > 0$. Выразим функцию распределения $F_Y(x)$ через функцию распределения случайной величины X:

$$F_Y(x) = \mathsf{P}(Y < x) = \mathsf{P}(aX + b < x) = \mathsf{P}\left(X < \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right).$$

Для нахождения плотности распределения продифференцируем найденную функцию распределения как сложную функцию:

$$f_Y(x) = (F_Y(x))' = \left(F_X\left(\frac{x-b}{a}\right)\right)' = \\ = F_X'\left(\frac{x-b}{a}\right) \cdot \left(\frac{x-b}{a}\right)' = f_X\left(\frac{x-b}{a}\right) \cdot \frac{1}{a}.$$

2. $a < 0$. Выполним аналогичные преобразования с учетом того, что при делении на отрицательный коэффициент a знак неравенства изменится на противоположный:

$$F_Y(x) = \mathsf{P}(Y < x) = \mathsf{P}(aX + b < x) = \mathsf{P}\left(X > \frac{x-b}{a}\right) = \\ = 1 - \mathsf{P}\left(X \leqslant \frac{x-b}{a}\right) = 1 - F_X\left(\frac{x-b}{a}\right).$$

Здесь мы также воспользовались тем фактом, что для непрерывной случайной величины X нестрогие неравенства при некоторых преобразованиях можно заменить строгими: $\mathsf{P}(X \leqslant t) = \mathsf{P}(X < t) = F_X(t)$.

Теперь найдем плотность распределения случайной величины Y:

$$f_Y(x) = (F_Y(x))' = \left(1 - F_X\left(\frac{x-b}{a}\right)\right)' = \\ = -F_X'\left(\frac{x-b}{a}\right) \cdot \left(\frac{x-b}{a}\right)' = f_X\left(\frac{x-b}{a}\right) \cdot \frac{1}{-a}.$$

Заметим, что в обоих рассмотренных случаях итоговое выражение для $F_Y(x)$ можно записать единообразно:

$$f_Y(x) = f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right). \qquad \square$$

Пример 71. Пусть случайная величина X равномерно распределена на отрезке $[1, 3]$. Найдите закон распределения случайной величины $Y = 3X - 4$.

Решение. Плотность распределения исходной случайной величины:

$$f_X(x) = \begin{cases} \frac{1}{2}, & x \in [1, 3], \\ 0, & x \notin [1, 3]. \end{cases}$$

Так как Y линейно зависит от X, то для нахождения плотности $f_Y(x)$ можно воспользоваться формулой (3.14):

$$f_Y(x) = f_{3X-4}(x) = \frac{1}{3} \cdot f_X\left(\frac{x+4}{3}\right) = \begin{cases} \frac{1}{3} \cdot \frac{1}{2}, & 1 \leqslant \frac{x+4}{3} \leqslant 3, \\ 0, & \frac{x+4}{3} \notin [1, 3], \end{cases} =$$

$$= \begin{cases} \frac{1}{6}, & x \in [-1, 5], \\ 0, & x \notin [-1, 5]. \end{cases}$$

По найденной плотности распределения можно сделать вывод, что случайная величина Y равномерно распределена на отрезке $[-1, 5]$. □

Замечание. Результат, полученный в предыдущем примере, несложно обобщается на произвольные равномерно распределенные случайные величины: если случайная величина X равномерно распределена на некотором отрезке $[a, b]$, а случайная величина Y связана с ней линейной функциональной зависимостью $Y = kX + c$, $k \neq 0$, то Y также является равномерно распределенной случайной величиной (на отрезке $[ka + c, kb + c]$ при $k > 0$ или на отрезке $[kb + c, ka + c]$ при $k < 0$).

Задача 72. Пусть случайная величина X распределена по нормальному закону с параметрами 0 и 1 (*стандартное нормальное распределение*). Докажите, что случайная величина $Y = \sigma X + a$ распределена по нормальному закону с параметрами a и σ^2.

Задача 73. Пусть случайная величина X распределена по нормальному закону с параметрами a и σ^2. Докажите, что случайная величина $Y = \dfrac{X - a}{\sigma}$ распределена по нормальному закону с параметрами 0 и 1 (то есть имеет стандарное нормальное распределение).

Задача 74. Пусть случайная величина X распределена по нормальному закону с параметрами a и σ^2. Докажите, что случайная величина $Y = kX + b$ $(k \neq 0)$ тоже распределена по нормальному закону. Найдите параметры распределения величины Y. (Иначе говоря, покажите, что нормальное распределение остается нормальным при линейных преобразованиях, меняются лишь его параметры.)

Полученное соотношение для плотности распределения случайной величины, линейно зависящей от заданной величины X, может быть несложным образом обобщено и на случай нелинейных монотонных функциональных зависимостей.

Теорема 3.22. *Пусть плотность распределения случайной величины X равна $f_X(x)$, а функция $g : \mathbb{R} \to \mathbb{R}$ монотонна. Тогда плотность распределения случайной величины $Y = g(X)$ равна*

$$f_Y(x) = f_{g(X)}(x) = \left|(g^{-1}(x))'\right| f_X(g^{-1}(x)).$$

(Здесь $g^{-1}(x)$ — функция, обратная к $g(x)$.)

Доказательство данного утверждения в целом аналогично выводу соответствующего соотношения для случая линейной функциональной зависимости (то есть для случая $g(x) = ax + b$), при этом придется рассмотреть два случая: 1) g — монотонно возрастающая функция, 2) g — монотонно убывающая функция.

Для немонотонных функций g закон распределения случайной величины $Y = g(X)$ придется находить «вручную», сводя функцию распределения зависимой случайной величины к известной функции распределения заданной случайной величины X. Рассмотрим несколько примеров.

Пример 75. Пусть случайная величина X равномерно распределена на отрезке $[-1, 3]$. Найдите закон распределения случайной величины $Y = X^2$.

Решение. Будем выражать функцию распределения случайной величины Y через известную функцию распределения величины X:

$$F_X(x) = \begin{cases} 0, & x < -1, \\ \frac{x+1}{4}, & x \in [-1, 3], \\ 1, & x > 3. \end{cases}$$

По определению: $F_Y(x) = \mathsf{P}(Y < x) = \mathsf{P}(X^2 < x)$. Заметим, что случайная величина $Y = X^2$ может принимать только неотрицательные значения. Следовательно, $F_Y(x) = 0$ при $x \leqslant 0$.

Для $x > 0$ выполняем преобразования:

$$F_Y(x) = \mathsf{P}(-\sqrt{x} < X < \sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x}).$$

При $0 < x \leqslant 9$ верно неравенство $0 < \sqrt{x} \leqslant 3$, поэтому можем записать, что $F_X(\sqrt{x}) = \frac{\sqrt{x}+1}{4}$. При $x > 9$ выполнено неравенство $\sqrt{x} > 3$, в этом случае $F_X(\sqrt{x}) = 1$.

Аналогично разбираем значения $F_X(-\sqrt{x})$ в разных диапазонах x: при $0 < x \leqslant 1$ имеет место неравенство $-1 \leqslant -\sqrt{x} < 0$, поэтому $F_X(-\sqrt{x}) = \frac{-\sqrt{x}+1}{4}$, а при $x > 1$ выполнено $-\sqrt{x} < -1$, следовательно, $F_X(-\sqrt{x}) = \mathsf{P}(X < -\sqrt{x} < -1) = 0$. Запишем выражение для $F_Y(x)$ при разных значениях x:

1) если $0 < x \leqslant 1$:

$$F_Y(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x}) = \frac{\sqrt{x}+1}{4} - \frac{-\sqrt{x}+1}{4} = \frac{\sqrt{x}}{2};$$

2) если $1 < x \leqslant 9$:

$$F_Y(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x}) = \frac{\sqrt{x}+1}{4} - 0 = \frac{\sqrt{x}+1}{4};$$

3) если $x > 9$:

$$F_Y(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x}) = 1 - 0 = 1.$$

Таким образом, функция распределения случайной величины Y может быть записана следующим образом:

$$F_Y(x) = \begin{cases} 0, & x \leqslant 0, \\ \frac{\sqrt{x}}{2}, & x \in (0, 1], \\ \frac{\sqrt{x}+1}{4}, & x \in (1, 9], \\ 1, & x > 9. \end{cases}$$ □

Задача 76. Пусть случайная величина X имеет следующую функцию распределения:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{x^2}{4}, & x \in [0, 2], \\ 1, & x > 2. \end{cases}$$

Найдите функции распределения случайных величин:
а) $Y_1 = 2X + 1$, б) $Y_2 = \dfrac{X^2}{2}$, в) $Y_3 = |X - 1|$, г) $Y_4 = \sin\left(\dfrac{X}{4}\right)$.

3.9.2. Суммирование независимых случайных величин

Теперь рассмотрим задачу, в которой необходимо найти закон распределения случайной величины, равной сумме двух случайных величин с известными законами распределения. Для начала отметим, что без дополнительных уточнений эта задача не имеет однозначного решения.

Пример 77. На отрезке AB длиной 10 случайным образом отмечены две точки — C и D. Рассмотрим такие случайные величины: X — длина отрезка AC, Y — длина отрезка BC, Z — длина отрезка AD. Отметим, что случайные величины X, Y и Z равномерно распределены на отрезке $[0, 10]$.

Рассмотрим суммы некоторых из этих величин: $W_1 = X + Y$, $W_2 = X + Z$. Несложно убедиться в том, что величина W_1 всегда принимает значение 10 (это сумма отрезков AC и BC, то есть длина AB), а случайная величина W_2 может принимать значения из отрезка $[0, 20]$.

В обоих случаях в приведенном примере мы складывали две случайные величины, равномерно распределенные на $[0, 10]$, однако получили разные результаты. Поэтому знания законов распределения слагаемых недостаточно для того, чтобы найти закон распределения суммы. В то же время если добавить условие независимости слагаемых, то эта задача становится разрешимой. Так, например, для абсолютно непрерывных независимых случайных величин закон распределения суммы может быть получен в виде достаточно удобного соотношения (формулы свертки).

Теорема 3.23 (формула свёртки)**.** *Пусть независмые непрерывные случайные величины X и Y имеют плотности распределения $f_X(x)$ и $f_Y(x)$ соответственно. Тогда плотность распределения их суммы* *может быть вычислена следующим образом:*

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

Пример 78. Пусть X и Y — независимые случайные величины, распределенные по нормальному закону с параметрами 0 и 1. Доказать, что случайная величина $Z = X + Y$ распределена по нормальному закону с параметрами 0 и 2.

Доказательство. Плотность распределения случайных величин X и Y, имеющих нормальный закон распределения с параметрами 0 и 1, задается следующей формулой:

$$f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Для нахождения плотности распределения их суммы воспользуемся формулой свертки:

$$\begin{aligned}
f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\,dx =& \\
= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-(z-x)^2/2}\,dx =& \\
= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-x^2+xz-z^2/2}\,dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-(x-z/2)^2-z^2/4}\,dx =& \\
= \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{+\infty} e^{-(x-z/2)^2}\,dx = \left| \begin{array}{c} \text{Замена: } t = x - z/2 \\ x = t + z/2,\ dx = dt \end{array} \right| =& \\
= \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{+\infty} e^{-t^2}\,dt = \left| \begin{array}{c} \text{Интеграл Эйлера-Пуассона:} \\ \displaystyle\int_{-\infty}^{+\infty} e^{-t^2}\,dt = \sqrt{\pi} \end{array} \right| =& \\
= \frac{1}{2\pi} e^{-z^2/4} \cdot \sqrt{\pi} = \frac{1}{\sqrt{2}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z}{\sqrt{2}}\right)^2}.&
\end{aligned}$$

Заметим, что полученная в результате функция является плотностью распределения случайной величины, распределенной по нормальному закону с параметрами 0 и 2. Отсюда непосредственно следует справедливость доказываемого утверждения. □

По аналогии с только что разобранным примером может быть решена следующая

Задача 79. Пусть X и Y — независимые случайные величины. Случайная величина X имеет нормальный закон распределения с параметрами a_1 и σ_1^2, а случайная величина Y имеет нормальный закон распределения с параметрами a_2 и σ_2^2. Докажите, что случайная величина $Z = X + Y$ распределена по нормальному закону с параметрами $a_1 + a_2$ и $\sigma_1^2 + \sigma_2^2$.

Задача 80. Пусть X и Y — независимые случайные величины. Случайная величина X имеет равномерный закон распределения на отрезке $[0, 2]$, а случайная величина Y распределена по показательному закону с параметром 3. Найдите закон распределения случайной величины $Z = X + Y$.

3.9.3. Генерирование случайных величин

Ранее была рассмотрена задача нахождения закона распределения зависимой случайной величины. А именно, известен закон распределения случайной величины X, а также функциональная зависимость $Y = g(X)$; требуется найти закон распределения величины Y.

Попробуем несколько видоизменить данную задачу. Пусть известен закон распределения исходной случайной величины X, а также закон распределения функционально зависящей от неё случайной величины Y. Требуется найти вид функциональной зависимости между этими величинами. Иначе говоря, требуется выяснить, какой функцией g необходимо подействовать на случайную величину X, чтобы получившаяся случайная величина $Y = g(X)$ имела требуемый закон распределения.

Итак, рассмотрим задачу *генерирования случайной величины* с нужным законом распределения по исходной случайной величине с известным законом распределения.

Теорема 3.24. *Пусть случайная величина X имеет строго монотонную функцию распределения $F_X(x)$. Тогда случайная величина $Y = F_X(X)$ распределена равномерно на отрезке* $[0, 1]$.

Доказательство. По определению функции распределения случайной величины Y: $F_Y(x) = \mathsf{P}(Y < x) = \mathsf{P}(F_X(X) < x)$.

Так как $F_X(X) \in [0,1]$ по свойству функции распределения, то неравенство $F_X(X) < x$ не выполняется при $x \leqslant 0$ (значение функции распределения не может быть меньше 0) и гарантированно выполняется при $x > 1$ (так как $F_X(X) \leqslant 1$, то $F_X(X) < x$ при всех $x > 1$). Для $x \in (0,1]$ преобразуем функцию распределения:

$$F_Y(x) = \mathsf{P}\left(F_X(X) < x\right) = \mathsf{P}\left(X < F_X^{-1}(x)\right) = F_X\left(F_X^{-1}(x)\right) = x.$$

Таким образом, функция распределения случайной величины Y записывается следующим образом:

$$F_Y(x) = \begin{cases} 0, & x \leqslant 0, \\ x, & x \in (0,1], \\ 1, & x > 1. \end{cases}$$

Но это функция распределения случайной величины, равномерно распределенной на отрезке $[0,1]$, что и требовалось доказать. $\square$

Доказанную теорему можно проинтерпретировать следующим образом: при некоторых ограничениях непрерывную случайную величину с каким-либо законом распределения можно преобразовать в случайную величину, распределенную равномерно на отрезке $[0,1]$ (причем в качестве преобразующей функции используется функция распределения исходной случайной величины).

Замечание. Для выполнения утверждения предыдущей теоремы достаточно, чтобы функция F_X была строго возрастающей не на всей числовой прямой, а на множестве значений, которые может принимать случайная величина X. Например, если исходная случайная величина имеет показательное распределение, то она принимает только положительные значения, а ее функция распределения

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\alpha x}, & x \geqslant 0, \end{cases}$$

строго возрастает на множестве положительных чисел (т. е. на множестве значений случайной величины X). Поэтому утверждение теоремы остается верным.

Теперь рассмотрим, как из случайной величины, равномерно распределенной на отрезке $[0,1]$, получить случайную величину с требуемым законом распределения.

Теорема 3.25. *Пусть случайная величина X распределена по равномерному закону на отрезке $[0,1]$, а строго монотонная функция $F(x)$ обладает всеми свойствами функции распределения. Тогда случайная величина $Y = F^{-1}(X)$ имеет функцию распределения $F_Y(x) \equiv F(x)$.*

Доказательство. Фактически это обратное утверждение к предыдущей теореме: если случайная величина с функцией распределения F под действием этой же функции F преобразуется к случайной величине, распределенной по равномерному закону на отрезке $[0,1]$, то применением обратной функции F^{-1} к случайной величине, равномерно распределенной на отрезке $[0,1]$, можно получить случайную величину с функцией распределения F. □

Замечание. Аналогично замечанию к предыдущей теореме: достаточно, чтобы функция F была строго возрастающей не на всей числовой прямой, а на множестве значений, которые должна принимать случайная величина Y.

Применение преобразований, которые были описаны в утверждениях двух последних теорем, позволяет решить задачу нахождения такой функции g, чтобы случайная величина $Y = g(X)$ имела требуемую функцию распределения F:

1. Случайная величина $Z = F_X(X)$ является равномерно распределенной на отрезке $[0,1]$;

2. Случайная величина $Y = F^{-1}(Z)$ имеет функцию распределения F;

3. Таким образом, в качестве искомой функции g можно взять следующую: $g(x) = F^{-1}(F_X(x))$.

Описанный метод имеет, например, следующее приложение: если при моделировании каких-либо процессов требуется генерировать (разыгрывать) псевдослучайные числа, распределение которых подчинено некоторому закону (например, требуется датчик чисел, распределенных по показательному закону с каким-нибудь параметром), то для этого не нужно создавать специализированный датчик псевдослучайных чисел — достаточно обойтись стандартным датчиком чисел, равномерно распределенных на отрезке $[0,1]$, но применять к полученным числам соответствующую функцию, преобразующую равномерное распределение на $[0,1]$ в случайную величину с требуемым распределением.

Пример 81. Пусть случайная величина X равномерно распределена на отрезке $[0,1]$. Найдите такую функцию g, чтобы случайная величина $Y = g(X)$ имела показательный закон распределения с параметром α.

Решение. Требуемая функция распределения случайной величины Y имеет следующий вид:

$$F_Y(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\alpha x}, & x \geqslant 0. \end{cases}$$

Достаточно рассмотреть только ту часть функции F_Y, которая определена на множестве значений случайной величины Y, то есть при $x \geqslant 0$. Это функция $F(x) = 1 - e^{-\alpha x}$, $x \geqslant 0$. Обратной к ней является функция $F^{-1}(x) = -\frac{1}{\alpha}\ln(1-x)$. Согласно утверждению второй доказанной теоремы, именно эту функцию можно взять в качестве искомой:

$$g(x) = -\frac{1}{\alpha}\ln(1-x). \qquad \square$$

Замечание. В предыдущем примере доказано следующее: если X равномерно распределена на отрезке $[0,1]$, то случайная величина $Y = -\frac{1}{\alpha}\ln(1-X)$ распределена по показательному закону с параметром α. При этом несложно заметить, что преобразование $1-X$ переводит случайную величину с равномерным распределением на $[0,1]$ в случайную величину с таким же (равномерным на $[0,1]$) законом распределения. Значит, вполне можно обойтись и без этого «лишнего» преобразования и рассмотреть величину $Y = -\frac{1}{\alpha}\ln X$ — она тоже будет распределена по показательному закону с параметром α. Поэтому в качестве искомой функции g можно взять следующую:

$$g(x) = -\frac{1}{\alpha}\ln x.$$

Пример 82. Пусть случайная величина X имеет показательный закон распределения с параметром 3. Найдите такую функцию g, чтобы случайная величина $Y = g(X)$ имела функцию распределения

$$F(x) = \begin{cases} 0, & x < 0, \\ \sin\left(\frac{\pi x^2}{8}\right), & x \in [0, 2], \\ 1, & x > 2. \end{cases}$$

Решение. Сначала найдем такую функцию h, чтобы случайная величина $Z = h(X)$ имела равномерное распределение на отрезке $[0, 1]$. Выше было показано, что в качестве h можно выбрать функцию распределения случайной величины X (причем на той числовой области, на которой принимает свои значения величина X, то есть на множестве положительных чисел). Так как функция распределения случайной величины X равна

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-3x}, & x \geqslant 0, \end{cases}$$

то $Z = 1 - e^{-3X}$ равномерно распределена на отрезке $[0, 1]$.

Для того чтобы получить случайную величину с функцией распределения F, подействуем функцией, обратной к F, на случайную величину Z, равномерно распределенную на отрезке $[0, 1]$. Опять же можно взять обратную функцию только на том множестве, на котором принимает значения предполагаемая случайная величина. По виду функции F можно сделать вывод, что случайная величина с такой функцией распределения будет принимать значения только на отрезке $[0, 2]$ (достаточно убедиться, что на этот отрезок соответствующая случайная величина попадает с вероятностью $F(2) - F(0) = 1$). На отрезке $[0, 2]$ функция, обратная к F, имеет вид:

$$F^{-1}(x) = \sqrt{\frac{8\arcsin x}{\pi}}.$$

Следовательно, случайная величина

$$Y = F^{-1}(Z) = \sqrt{\frac{8\arcsin Z}{\pi}} = \sqrt{\frac{8\arcsin(1 - e^{-3X})}{\pi}}$$

имеет функцию распределения $F(x)$.

Таким образом, искомая функция g может быть задана так:

$$g(x) = \sqrt{\frac{8 \arcsin(1 - e^{-3x})}{\pi}}. \qquad \square$$

3.10. Закон больших чисел

Ранее рассмотренные приближенные формулы Пуассона и Муавра–Лапласа для оценки вероятности количества успехов в серии испытаний Бернулли являются примером того, как в серии массовых случайных явлений возникают новые закономерности. Фактически было показано, что при выполнении некоторых условий биномиальный закон распределения можно достаточно точно приблизить законом Пуассона, а при выполнении других условий — получить в качестве *предельного распределения* закон нормального распределения вероятности. Таким образом, мы уже сталкивались с примерами так называемых *предельных теорем*, в которых рассматривались закономерности, возникающие при проведении большого количества испытаний (и даже при числе испытаний, стремящемся к бесконечности).

В этом разделе мы рассмотрим универсальные (не привязанные к конкретному виду закона распределения случайной величины) закономерности, возникающие при проведении серии из большого количества испытаний, — различные формы *закона больших чисел* и следствия из него.

Для начала сформулируем и докажем два вспомогательных неравенства, которые носят имя П. Л. Чебышёва[13].

3.10.1. Неравенства Чебышёва

Первое из упомянутых неравенств рассмотрим в том виде, в котором оно было сформулировано и доказано А. А. Марковым[14] (поэтому достаточно часто именуется неравенством Маркова).

[13] Пафнутий Львович Чебышёв (1821–1894) — русский математик и механик.
[14] Андрей Андреевич Ма́рков (1856–1922) — русский математик.

Теорема 3.26 (неравенство Маркова (первое неравенство Чебышёва)). *Пусть* $\mathsf{M}\,|X| < \infty$. *Тогда для любого* $t > 0$ *верно неравенство:*

$$\mathsf{P}(|X| \geqslant t) \leqslant \frac{\mathsf{M}\,|X|}{t}.$$

Доказательство. Выберем произвольно $t > 0$. Для выбранного значения t рассмотрим вспомогательную случайную величину Y, определенную следующим образом:

$$Y = \begin{cases} t, & \text{при } |X| \geqslant t, \\ 0, & \text{иначе.} \end{cases}$$

Очевидно, $Y \leqslant |X|$ всегда. Следовательно,

$$\mathsf{M}(Y) \leqslant \mathsf{M}\,|X|.$$

С другой стороны, математическое ожидание $\mathsf{M}(Y)$ равно

$$\mathsf{M}(Y) = t \cdot \mathsf{P}(|X| \geqslant t) + 0 \cdot \mathsf{P}(|X| < t) = t \cdot \mathsf{P}(|X| \geqslant t).$$

Таким образом,

$$t \cdot \mathsf{P}(|X| \geqslant t) = \mathsf{M}(Y) \leqslant \mathsf{M}\,|X|.$$

Остается разделить обе части неравенства на $t > 0$. □

Заметим, что в условиях доказанной теоремы правая часть неравенства стремится к 0 при $t \to \infty$ (так как $\mathsf{M}\,|X|$ — константа). Отсюда получаем

Следствие 3.27. *Пусть* $\mathsf{M}\,|X| < \infty$. *Тогда для любого* $t > 0$:

$$\lim_{t \to \infty} \mathsf{P}(|X| \geqslant t) = 0.$$

Это можно интерпретировать так: если случайная величина имеет конечное математическое ожидание, то для любого наперед заданного положительного числа ε (как бы мало оно ни было) всегда можно заключить случайную величину в такой (достаточно большой) интервал, чтобы она принимала значения вне этого интервала с вероятностью меньше ε (то есть можно сделать так, чтобы эта вероятность была бы сколь угодно близка к нулю).

Следующее неравенство было доказано независимо П. Л. Чебышёвым и И. Бьенэме[15].

Теорема 3.28 (неравенство Чебышёва–Бьенэме (второе неравенство Чебышёва))**.** *Пусть* $\mathsf{M}\,X^2 < \infty$*. Тогда для любого* $t > 0$ *верно неравенство:*

$$\mathsf{P}(|X - \mathsf{M}\,X| \geqslant t) \leqslant \frac{\mathsf{D}\,X}{t^2}.$$

Доказательство. Предварительно заметим, что при $t > 0$ неравенство $|X - \mathsf{M}\,X| \geqslant t$ эквивалентно $(X - \mathsf{M}\,X)^2 \geqslant t^2$. Тогда доказательство рассматриваемого в теореме утверждения несложно получается с помощью неравенства Маркова:

$$\mathsf{P}(|X - \mathsf{M}\,X| \geqslant t) = \mathsf{P}((X - \mathsf{M}\,X)^2 \geqslant t^2) \leqslant \frac{\mathsf{M}(X - \mathsf{M}\,X)^2}{t^2} = \frac{\mathsf{D}\,X}{t^2}. \qquad \square$$

Частным случаем неравенства Чебышёва является соотношение, известное под названием «правило трёх сигм»:

Следствие 3.29 (правило трёх сигм)**.** *Пусть* $\mathsf{D}\,X = \sigma^2$*. Тогда*

$$\mathsf{P}(|X - \mathsf{M}\,X| \geqslant 3\,\sigma) \leqslant \frac{1}{9}.$$

Доказательство. Из неравенства Чебышёва:

$$\mathsf{P}(|X - \mathsf{M}\,X| \geqslant 3\,\sigma) \leqslant \frac{\mathsf{D}\,X}{(3\,\sigma)^2} = \frac{\sigma^2}{9\,\sigma^2} = \frac{1}{9}. \qquad \square$$

Иногда «правило трёх сигм» применяется в такой нестрогой формулировке: вероятность того, что случайная величина отклонится от своего математического ожидания больше, чем на три среднеквадратических отклонения («три сигмы»), очень мала. Правда, доказанное следствие показывает, что вероятность такого отклонения не превосходит $\frac{1}{9}$ (можно ли считать это «очень маленькой величиной» — ответ на данный вопрос весьма неоднозначен). Однако заметим, что в формулировке и доказательстве не было указаний на

[15] Иренэ-Жюль Бьенэме (1796–1878) — французский статистик и администратор.

конкретный вид закона распределения случайной величины X, достаточно было только существования конечной дисперсии $\mathsf{D}\, X$ (или конечного $\mathsf{M}\, X^2$). Поэтому значение $\frac{1}{9}$ будет универсальной верхней границей для вероятности упомянутого события.

Побочным эффектом такой универсальности является то, что оценка является грубой — для отдельных случайных величин (с учетом специфики закона их распределения) можно получить значительно меньшую оценку данной вероятности.

Пример 83 (правило трёх сигм для нормального распределения)**.** Пусть случайная величина X распределена по нормальному закону с параметрами a и σ^2. Тогда

$$\mathsf{P}(|X-a| \geqslant 3\,\sigma) = 2\,\Phi(3) \approx 0{,}0027.$$

Задача 84. Пусть случайная величина X распределена равномерно на $[0,1]$. Вычислите значение $\sigma = \sqrt{\mathsf{D}\, X}$ и найдите $\mathsf{P}(|X - \mathsf{M}\, X| \geqslant 3\,\sigma)$.

Задача 85. Пусть случайная величина X распределена равномерно на $[a,b]$. Вычислите значение $\sigma = \sqrt{\mathsf{D}\, X}$ и найдите $\mathsf{P}(|X - \mathsf{M}\, X| \geqslant 3\,\sigma)$.

Задача 86. Пусть случайная величина X распределена по показательному закону с параметром α. Вычислите значение $\sigma = \sqrt{\mathsf{D}\, X}$ и найдите $\mathsf{P}(|X - \mathsf{M}\, X| \geqslant 3\,\sigma)$.

Пример 87. Автобус в среднем опаздывает на 2 минуты. Более того, будем предполагать, что автобус никогда не приходит раньше времени, указанного в расписании. С помощью неравенств Чебышёва оцените вероятность того, что он опоздает больше, чем на 10 минут.

Решение. Обозначим через X время опоздания автобуса (за единицу времени выберем одну минуту). Из условия задачи $\mathsf{M}\, X = 2$ и $X \geqslant 0$. Требуется найти вероятность события $\{X \geqslant 10\}$.

Воспользуемся первым неравенством Чебышёва для заданной случайной величины X и $t = 10$. Итак,

$$\mathsf{P}(X \geqslant 10) \leqslant \frac{\mathsf{M}\, X}{10} = \frac{2}{10} = 0{,}2. \qquad \square$$

Обращаем внимание на то, что вторым неравенством Чебышёва при решении предыдущей задачи мы воспользоваться не смогли бы, так как дисперсия или среднеквадратическое отклонение нам не были известны (и вывести их значения, зная только величину математического ожидания, нельзя). В следующей задаче такая дополнительная информация будет предоставлена.

Пример 88. В течение дня в автобусах продается в среднем 20000 билетов. Среднеквадратическое отклонение (разброс) количества проданных билетов составляет 1000. С помощью неравенств Чебышёва оцените вероятность того, что количество проданных за день билетов лежит в интервале:

а) от 18000 до 22000,
б) от 17000 до 23000,
в) от 15000 до 25000.

Решение. Обозначим через X случайную величину — количество билетов, продаваемых в течение дня. По условию, $\mathsf{M}\,X = 20000$, $\sigma(X) = 1000$, откуда $\mathsf{D}\,X = (1000)^2 = 10^6$.

Для оценки вероятностей указанных событий воспользуемся вторым неравенством Чебышёва.

а) Преобразуем неравенство, описывающее заданное событие:

$$\mathsf{P}(18000 < X < 22000) = \mathsf{P}(-2000 < X - 20000 < 2000) = \\ = \mathsf{P}(|X - 20000| < 2000).$$

Перейдем к противоположному событию и воспользуемся вторым неравенством Чебышёва:

$$\mathsf{P}(|X - 20000| < 2000) = 1 - \mathsf{P}(|X - 20000| \geqslant 2000) = \\ = 1 - \mathsf{P}(|X - \mathsf{M}\,X| \geqslant 2000) \geqslant 1 - \frac{\mathsf{D}\,X}{2000^2} = \\ = 1 - \frac{10^6}{2000^2} = 1 - \frac{1}{4} = \frac{3}{4} = 0{,}75.$$

Итак, $\mathsf{P}(18000 < X < 22000) \geqslant 0{,}75$.

б) Аналогично предыдущему случаю получаем:

$$\mathsf{P}(17000 < X < 23000) = 1 - \mathsf{P}(|X - \mathsf{M}\,X| \geqslant 3000).$$

Здесь можно заметить, что оценка последней вероятности — это в чистом виде «правило трёх сигм»:

$$\mathsf{P}(|X - \mathsf{M}\,X| \geqslant 3000) = \mathsf{P}(|X - \mathsf{M}\,X| \geqslant 3\,\sigma_X) \leqslant \frac{1}{9}.$$

Отсюда быстро получаем ответ:

$$\mathsf{P}(17000 < X < 23000) \geqslant 1 - \frac{1}{9} = \frac{8}{9}.$$

в) Преобразованиями, аналогичными двум предыдущим случаям, получаем:

$$\mathsf{P}(15000 < X < 25000) = 1 - \mathsf{P}(|X - \mathsf{M}\,X| \geqslant 5000) \geqslant \\ \geqslant 1 - \frac{\mathsf{D}\,X}{5000^2} = 1 - \frac{10^6}{5000^2} = 0{,}96.$$ □

Еще раз отметим, что оценки вероятностей, полученные в предыдущей задаче с помощью второго неравенства Чебышёва, являются достаточно грубыми — они подходят для любых случайных величин с математическим ожиданием 20000 и среднеквадратичным отклонением 1000. Более точные оценки вероятностей пока мы получить не могли, так как закон распределения рассматриваемой случайной величины X не был известен (были известны только некоторые ее числовые характеристики). При наличии дополнительной информации о законе распределения величины X можно построить более точные оценки соответствующих вероятностей.

Например, если в предыдущей задаче было бы дополнительно известно, что случайная величина распределена по нормальному закону (с теми же заданными математическим ожиданием и среднеквадратическим отклонением), то вместо оценок вероятностей 0,75, $\frac{8}{9} \approx 0{,}889$ и 0,96 были бы получены значительно более точные значения — 0,955, 0,9973 и 0,9999994 соответственно.

3.10.2. Закон больших чисел

Как уже было отмечено, оценки вероятностей, полученные непосредственно с помощью неравенств Чебышёва, оказываются не очень точными, но достаточно универсальными (то есть применимыми к довольно широкому классу распределений). Воспользуемся данными результатами для изучения случайных величин, являющихся усреднениями нескольких независимых случайных величин. Соотношения, которые будут выведены в данном разделе, известны как *закон больших чисел.* Для начала сформулируем и докажем закон больших чисел в достаточно общей форме, а в дальнейшем получим из него ряд любопытных следствий.

Теорема 3.30 (закон больших чисел)**.** *Пусть случайные величины $X_1, X_2, \ldots, X_n$ независимы, а их дисперсии ограничены в совокупности (то есть существует такая константа $C > 0$, что $\mathsf{D}\, X_i \leqslant C$ для $1 \leqslant i \leqslant n$). Тогда для любого $\varepsilon > 0$ верно неравенство:*

$$\mathsf{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \frac{\mathsf{M}\, X_1 + \mathsf{M}\, X_2 + \cdots + \mathsf{M}\, X_n}{n}\right| \geqslant \varepsilon\right) \leqslant \frac{C}{n\varepsilon^2}.$$

Доказательство. Введем обозначение:

$$X = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Тогда имеет место следующее соотношение:

$$\mathsf{M}\, X = \mathsf{M}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n}\,\mathsf{M}(X_1 + X_2 + \cdots + X_n) = \\ = \frac{\mathsf{M}\, X_1 + \mathsf{M}\, X_2 + \cdots + \mathsf{M}\, X_n}{n}.$$

Оценим дисперсию величины X:

$$\mathsf{D}\, X = \mathsf{D}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\,\mathsf{D}(X_1 + X_2 + \cdots + X_n).$$

Так как дисперсия суммы независимых случайных величин равна сумме их дисперсий, получаем:

$$\mathsf{D}\, X = \frac{\mathsf{D}\, X_1 + \mathsf{D}\, X_2 + \cdots + \mathsf{D}\, X_n}{n^2} \leqslant \frac{C + C + \cdots + C}{n^2} = \frac{Cn}{n^2} = \frac{C}{n}.$$

Итак, $\mathsf{D}\, X \leqslant \dfrac{C}{n}$. Теперь воспользуемся введенными обозначениями и применим неравенство Чебышёва:

$$\mathsf{P}\left(\left|\frac{X_1+X_2+\dots+X_n}{n}-\frac{\mathsf{M}\,X_1+\mathsf{M}\,X_2+\dots+\mathsf{M}\,X_n}{n}\right|\geqslant\varepsilon\right)=$$

$$=\mathsf{P}(|X-\mathsf{M}\,X|\geqslant\varepsilon)\leqslant\frac{\mathsf{D}\,X}{\varepsilon^2}\leqslant\frac{C}{n\varepsilon^2}. \qquad \square$$

Наиболее часто рассматриваются средние арифметические не произвольных независимых случайных величин, а еще и однотипных, одинаково распределенных (то есть имеющих одинаковые законы распределения и, как следствие, одинаковые числовые характеристики — математические ожидания и дисперсии). Например, с такой ситуацией можно столкнуться при анализе характера распределения среднего арифметического результатов серии однотипных независимых испытаний.

Следствие 3.31 (закон больших чисел для однотипных случайных величин)**.** *Пусть случайные величины X_1, X_2, …, X_n независимы и одинаково распределены, их математические ожидания $\mathsf{M}\,X_i=a$, дисперсии $\mathsf{D}\,X_i=\sigma^2$. Тогда для любого $\varepsilon>0$ верно неравенство:*

$$\mathsf{P}\left(\left|\frac{X_1+X_2+\dots+X_n}{n}-a\right|\geqslant\varepsilon\right)\leqslant\frac{\sigma^2}{n\varepsilon^2}. \tag{3.15}$$

Доказательство. Для указанных случайных величин

$$\frac{\mathsf{M}\,X_1+\mathsf{M}\,X_2+\dots+\mathsf{M}\,X_n}{n}=a.$$

В качестве константы C, ограничивающей дисперсии всех случайных величин X_i, можно выбрать значение σ^2. Для завершения доказательства осталось только подставить данные значения в неравенство из закона больших чисел в общем виде. $\square$

Заметим, что при фиксированном значении $\varepsilon>0$ правая часть неравенства (3.15) стремится к 0 при $n\to\infty$. Это позволяет записать закон больших чисел для одинаково распределенных случайных величин в форме, которая известна как *закон больших чисел Чебышёва*:

Следствие 3.32 (закон больших чисел в форме Чебышёва). *Пусть имеется последовательность независимых и одинаково распределенных случайных величин $X_1, X_2, \ldots, X_n, \ldots$, математические ожидания которых $\mathsf{M}\, X_i = a$ (для всех i), а дисперсии конечны. Тогда для любого $\varepsilon > 0$*

$$\mathsf{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - a\right| > \varepsilon\right) \to 0 \qquad \textit{при } n \to \infty.$$

Менее строго этот результат можно проинтерпретировать следующим образом: в серии независимых однотипных испытаний вероятность даже небольших отклонений среднего арифметического значений случайных величин от математического ожидания этих величин стремится к 0 с ростом количества проведенных испытаний. Иначе говоря, при большом количестве испытаний среднее значение результатов испытаний ведет себя довольно предсказуемо, мало отклоняясь от математического ожидания.

Одним из условий доказанной теоремы и следствий из нее является конечность дисперсий $\mathsf{D}\, X_i$ (или, что то же самое, конечность $\mathsf{M}\, X_i^2$). Оказывается, что закон больших чисел в форме Чебышёва (следствие 3.32) имеет место даже без такого ограничения — полученное утверждение носит название *теоремы Хинчина*[16].

Теорема 3.33 (закон больших чисел в форме Хинчина). *Пусть имеется последовательность независимых и одинаково распределенных случайных величин $X_1, X_2, \ldots, X_n, \ldots$, математические ожидания которых $\mathsf{M}\, X_i = a$ (для всех i). Тогда для любого $\varepsilon > 0$*

$$\mathsf{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - a\right| \geqslant \varepsilon\right) \to 0 \qquad \textit{при } n \to \infty.$$

Доказывать это утверждение мы не будем, но отметим, что доказательство пришлось бы проводить другими средствами (ранее мы пользовались вторым неравенством Чебышёва, а в нем требование конечности дисперсий является ключевым).

Закон больших чисел в форме Чебышёва или в форме Хинчина также обосновывает возможность статистического определения математического ожидания: в качестве оценки математического ожи-

[16] Александр Яковлевич Хинчин (1894–1959) — советский математик.

дания случайной величины можно рассматривать среднее арифметическое значений этой величины в серии независимых однотипных испытаний — вероятность того, что этот приближенный результат будет сильно отклоняться от реального значения математического ожидания, будет достаточно низкой при большом количестве проведенных испытаний.

Ещё одним следствием из закона больших чисел для однотипных испытаний является *закон больших чисел в форме Бернулли.*

Следствие 3.34 (закон больших чисел в форме Бернулли)**.** *Пусть проводятся n испытаний Бернулли с вероятностью успеха p в каждом отдельном испытании. Пусть S_n — количество испытаний, закончившихся успехом. Тогда для любого $\varepsilon > 0$*

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{p(1-p)}{n\varepsilon^2}.$$

Доказательство. Пусть X_i — число успехов в i-м испытании (оно может быть равно 0 или 1), $i = 1, \ldots, n$. Тогда

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Кроме того, нетрудно вычислить $\mathsf{M}\, X_i = p$ и $\mathsf{D}\, X_i = p(1-p)$.

Согласно условию, случайные величины X_1, X_2, …, X_n независимы. Воспользовавшись следствием 3.31, получаем:

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}\, X_i}{n\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}. \qquad \square$$

Следствие 3.35 (закон больших чисел Бернулли в предельной форме)**.** *Пусть проводятся n испытаний Бернулли с вероятностью успеха p в каждом испытании. Пусть S_n — количество испытаний, закончившихся успехом. Тогда для любого $\varepsilon > 0$*

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) \to 0 \qquad \text{\textit{при} } n \to \infty.$$

Рассмотренное в данном утверждении отношение $\dfrac{S_n}{n}$ является *относительной частотой* числа успехов в n испытаниях Бернулли.

Таким образом, закон больших чисел в форме Бернулли фактически является обоснованием статистического определения вероятности: если провести большое количество испытаний, то вероятность больших отклонений относительной частоты числа успехов от действительного значения вероятности успеха стремится к 0 с ростом количества испытаний (и увеличением количества испытаний можно добиться снижения погрешности при оценке вероятности успеха).

Пример 89. Случайная величина X — среднее арифметическое 1000 независимых случайных величин, равномерно распределенных на отрезке [0,1]. Оцените вероятность того, что случайная величина X окажется в диапазоне от 0,45 до 0,55.

Решение. Пусть X_1, X_2, …, X_{1000} — независимые случайные величины, равномерно распределенные на отрезке [0,1]. Несложно установить, что $\mathsf{M}\,X_i = \frac{1}{2}$, $\mathsf{D}\,X_i = \frac{1}{12}$, $i = 1, 2, \ldots, 1000$. Воспользуемся законом больших чисел для однотипных случайных величин:

$$\mathsf{P}(0{,}48 < X < 0{,}52) = 1 - \mathsf{P}\left(\left|\frac{X_1 + \cdots + X_{1000}}{1000} - \frac{1}{2}\right| \geqslant 0{,}05\right) \geqslant$$
$$\geqslant 1 - \frac{\mathsf{D}\,X_i}{1000 \cdot (0{,}05)^2} = 1 - \frac{1/12}{2{,}5} = 1 - \frac{1}{30}.$$

Итак, получили, что указанная случайная величина практически достоверно оказывается в заданной узкой окрестности своего математического ожидания. □

Литература

[1] *Босс, В.* Лекции по математике. Т. 4 : Вероятность, информация, статистика / В. Босс. – М. : КомКнига, 2005. – 216 с.

[2] *Лотов, В. И.* Теория вероятностей и математическая статистика : конспект лекций для студентов факультета информационных технологий / В. И. Лотов. – Новосибирск : НГУ, 2003. – 97 с. – URL : `http://window.edu.ru/resource/248/28248`

[3] *Секей, Г.* Парадоксы в теории вероятностей и математической статистике / Г. Секей; пер. с англ. В. В. Ульянова. – М. : Мир, 1990. – 240 с.

[4] *Феллер, В.* Введение в теорию вероятностей и её приложения : в 2 т. Т. 1 / В. Феллер. – М. : Мир, 1984. – 528 с.

[5] *Феллер, В.* Введение в теорию вероятностей и её приложения : в 2 т. Т. 2 / В. Феллер. – М. : Мир, 1984. – 738 с.

[6] *Чернова, Н. И.* Теория вероятностей : учеб. пособие / Н. И. Чернова. – Новосибирск : НГУ, 2007. – 160 с. – URL : `http://www.nsu.ru/mmf/tvims/chernova/tv/index.html`

Приложение 1

Таблица значений функции $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0170	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Приложение 2

Таблица значений функции $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{\frac{-t^2}{2}}\,dt$

	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2703	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	0,3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3680	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3883	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4230	4251	4265	4279	4292	4305	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	0,4773	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4865	4868	4871	4875	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4980	4980	4981
2,9	4981	4982	4983	4984	4984	4984	4985	4985	4986	4986

$\Phi(3{,}0) = 0{,}49865$; $\Phi(3{,}2) = 0{,}49931$; $\Phi(3{,}4) = 0{,}49966$;
$\Phi(3{,}6) = 0{,}499841$; $\Phi(3{,}8) = 0{,}499928$; $\Phi(4{,}0) = 0{,}499968$;
$\Phi(4{,}5) = 0{,}499997$; $\Phi(5{,}0) = 0{,}4999997$.

Учебное издание

Максименко Александр Николаевич
Богомолов Юрий Викторович

Теория вероятностей

Учебное пособие

Редактор, корректор Л. Н. Селиванова
Компьютерная верстка А. Н. Максименко

Подписано в печать 26.03.2019. Формат 60×84 1/16.
Усл. печ. л. 7,4. Уч.-изд. л. 6,5.
Тираж 22 экз. Заказ .

Ярославский государственный университет им. П. Г. Демидова.
150003, Ярославль, ул. Советская, 14.