Уральский
федеральный
университет
имени первого Президента
России Б.Н.Ельцина
Институт экономики
и управления

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА: РЕШЕНИЕ ЗАДАЧ

Учебное пособие

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б. Н. ЕЛЬЦИНА

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА: РЕШЕНИЕ ЗАДАЧ

Учебное пособие

Рекомендовано
методическим советом Уральского федерального университета
в качестве учебного пособия для студентов вуза,
обучающихся по направлениям подготовки 38.03.01 «Экономика»,
38.03.02 «Менеджмент», 38.03.05 «Бизнес-информатика»,
по специальностям 38.05.01 «Экономическая безопасность»,
38.05.02 «Таможенное дело»

Екатеринбург
Издательство Уральского университета
2021

УДК 330.4:519.2(075.8)
ББК 65.051+22.17я73
Т33

Авторы:
О. Я. Шевалдина, Е. В. Выходец, О. Л. Кузнецова,
Е. А. Трофимова, Д. В. Гилёв, Н. В. Кисляк

Под общей редакцией
Е. А. Трофимовой

Рецензенты:
отдел аппроксимации и приложений Института математики и механики УрО РАН
(заведующий отделом доктор физико-математических наук А. Г. Бабенко);
Г. Б. Захарова, кандидат технических наук, доцент,
ведущий научный сотрудник НИЧ Уральского государственного
архитектурно-художественного университета

Т33 **Теория вероятностей** и математическая статистика: решение задач : учебное пособие / О. Я. Шевалдина, Е. В. Выходец, О. Л. Кузнецова [и др.] ; под ред. Е. А. Трофимовой ; Министерство науки и высшего образования Российской Федерации, Уральский федеральный университет. — Екатеринбург : Изд-во Урал. ун-та, 2021. — 220 с. : ил. — 100 экз. — ISBN 978-5-7996-3189-5. — Текст : непосредственный.

ISBN 978-5-7996-3189-5

Все главы учебного пособия включают теоретический блок — определения основных понятий, формулировки необходимых теорем и утверждений. Ключевые слова и понятия выделены в тексте. Представлены задачи для решения на практических занятиях и самостоятельно, приведено большое количество примеров и разборов задач. Каждому математическому понятию дается экономическая интерпретация.

Для студентов, изучающих дисциплину «Теория вероятностей и математическая статистика», а также дисциплины в рамках модулей «Математические методы анализа», «Математические методы анализа и основы информационных технологий».

УДК 330.4:519.2(075.8)
ББК 65.051+22.17я73

ISBN 978-5-7996-3189-5

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Данное учебное пособие написано в соответствии с требованиями государственных образовательных стандартов третьего поколения по экономическим специальностям. Включает в себя кратко, но всесторонне изложенный теоретический материал с разобранными на каждую тему практическими заданиями, с объяснением экономического смысла каждого введенного понятия, а также задачи для самостоятельного решения. Пособие может быть использовано в качестве основной литературы для проведения лекций и практических занятий.

Предпосылками написания учебного пособия послужили необходимость систематизировать накопленный материал при многолетнем прочтении лекций и проведении практических занятий у авторов пособия, а также возможность иметь полный комплект наработанных материалов, учитывающий новые разработки и обеспечивающий дисциплину «Теория вероятностей и математическая статистика». При написании пособия учтены современные требования и компетенции, предъявляемые к бакалавру экономики. Материал подобран так, чтобы можно было не только уловить суть предмета, но и понять его назначение в современном мире. Особый уклон сделан на экономические приложения. Содержание пособия целиком соответствует рабочей программе по дисциплине и охватывает объем шире необходимого минимума. Некоторые темы приведены для самостоятельного разбора студентами.

Дисциплина «Теория вероятностей и математическая статистика» является важнейшей частью модуля «Математические методы анализа». Ее прикладная значимость в экономике достаточно велика. На ней зиждется эконометрика, многомерный статистический анализ, нейронные сети, распознавание образов и многие другие научные области. Современный экономист должен уметь использовать аппарат математической статистики на высоком уровне.

ВВЕДЕНИЕ

Теория вероятностей и ее значение для экономической науки

Многие явления в природе, технике, экономике и в других областях носят случайный характер, т. е. невозможно точно предсказать, как явление будет происходить. Оказывается, однако, что течение и таких явлений может быть описано количественно, если только они наблюдались достаточное число раз при неизменных условиях.

Теория вероятностей — математическая наука, изучающая закономерности массовых случайных явлений (событий), способных многократно повторяться при воспроизведении определенного комплекса условий. Так как многие реальные процессы подвержены случайным воздействиям, то основы этой теории важно знать специалистам, занимающимся естественными, техническими, экономическими, а также общественными науками.

Математическая статистика есть также раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов.

За несколько последних десятилетий от теории вероятностей «отпочковались» такие отрасли науки, как теория случайных процессов, теория массового обслуживания, теории надежности, теория информации, эконометрическое моделирование и др.

Одной из важнейших сфер приложения теории вероятностей является экономика. В настоящее время трудно себе представить исследование и прогнозирование экономических явлений без использования эконометрического моделирования, регрессионного анализа и других методов, опирающихся на теорию вероятностей.

Краткая историческая справка

Первые работы, в которых появились основные понятия теории вероятностей, представляли собой попытки создания теории азартных игр (Кардано, Гюйгенс, Паскаль, Ферма и др., XVI–XVII вв.).

Следующий этап развития теории вероятностей связан с именем швейцарского математика Я. Бернулли (1654–1705). Доказанная им теорема (1713), получившая впоследствии название «закона больших чисел», была первым теоретическим обоснованием накопленных ранее фактов. Дальнейшими успехами теория вероятностей обязана А. Муавру (Англия), Лапласу (Франция), Гауссу (Германия), Пуассону (Франция) и др. К этому периоду относится доказательство первых предельных теорем, носящих теперь названия теорем Лапласа (1812) и Пуассона (1837); в это же время А. Лежандром (Франция, 1806) и К. Гауссом (1808) был разработан метод наименьших квадратов.

Новый период развития теории вероятностей связан с именами русских математиков П. Л. Чебышёва (1821–1894), А. А. Маркова (1856–1922), А. М. Ляпунова (1857–1918). В это время теория вероятностей становится стройной математической наукой. Чебышёв чрезвычайно просто доказал (1867) закон больших чисел. Он же впервые сформулировал (1887) центральную предельную теорему для независимых случайных величин. Последующее развитие тории вероятностей обязано в России математикам С. Н. Бернштейну, В. И. Романовскому, А. Н. Колмогорову, А. Я. Хинчину и др., во Франции — Э. Борелю, П. Леви, М. Фреше, в Германии — Р. Мизесу, в США — Н. Винеру, В. Феллеру, Дж. Дубу, в Швеции — Г. Крамеру. Позднее А. Я. Хинчин, А. Н. Колмогоров и Е. Е. Слуцкий заложили основы теории случайных процессов. Далее большая работа проделана по применению методов теории вероятностей к задачам математической статистики.

1. ОСНОВНЫЕ ПОНЯТИЯ И ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.1. Классификация событий

Теория вероятностей базируется на опыте, наблюдении различных процессов, происходящих в окружающем нас мире.

Под *опытом*, *испытанием* G понимается воспроизведение определенного комплекса условий для наблюдения исследуемого явления.

Исходы (результаты) опытов, наблюдений, испытаний называют *событиями*.

Пример 1.1. *Испытание* — бросание монеты. Выпадение герба (или цифры) является *событием*.

Пример 1.2. *Испытание* — студенты сдают экзамен по теории вероятностей. Наудачу выбранный студент получил оценку «отлично» — *событие*.

Обычно считается, что событие в опыте *случайно*, если при неоднократном воспроизведении опыта оно иногда происходит, а иногда нет, причем невозможно заранее предсказать возможный исход этого опыта.

Пример 1.3. Пусть опыт G состоит в подбрасывании игральной кости и наблюдении числа выпавших очков X. Тогда можно ввести следующие случайные события:

$$\{X=1\},\ \{X=2\},\ldots,\ \{X=6\},\ \{X\le 2\},\ \{X \text{ четно}\},\ \{X \text{ нечетно}\} \text{ и т. д.}$$

Возможные исходы ω опыта G называются *элементарными событиями* (*элементарными исходами*), если они являются *взаимно исключающими* и в результате опыта G одно из них обязательно происходит.

Совокупность Ω всех элементарных событий ω в опыте G называется *пространством элементарных событий* (*полной группой событий*).

Пространство элементарных событий — это математическая модель опыта, в которой любому событию ставится в соответствие некоторое подмножество пространства Ω.

В примере 1.1 пространство элементарных событий $\Omega = \{\omega_1,\ \omega_2\}$, где $\omega_1 = \{\text{выпадение герба}\}$, $\omega_2 = \{\text{выпадение решки}\}$.

В примере 1.3 пространство элементарных событий $\Omega = \{\omega_1, \ldots, \omega_6\}$, где элементарное событие ω_i состоит в том, что $\{X = i\}$, $i = 1, \ldots, 6$.

Составные события, или просто *события*, могут быть описаны как подмножества множества элементарных событий Ω. Случайные события будем обозначать заглавными буквами латинского алфавита: A, B, C, …

Так, в примере 1.3 события A = {выпало нечетное число очков} = $\{\omega_1, \omega_3, \omega_5\}$, B = {выпало четное число очков} = $\{\omega_2, \omega_4, \omega_6\}$, C = {выпало число очков, кратное трем} = $\{\omega_3, \omega_6\}$ являются подмножествами пространства Ω.

Событие называется *достоверным* в опыте G, если при повторении опыта оно обязательно происходит. Ему соответствует пространство Ω.

Событие называется *невозможным* в опыте G, если при повторении опыта оно никогда не происходит. Ему соответствует пустое подмножество в Ω, которое обозначают $\varnothing$.

Пример 1.4. В коробке находятся шары красного цвета. Событие A — извлечение наудачу из коробки шара *красного* цвета — *достоверное* событие. Событие B — извлечение наудачу из коробки шара *синего* цвета — *невозможное* событие.

Говорят, что в опыте G событие A влечет появление события B ($A \subset B$), если из осуществления события A следует наступление события B.

Пример 1.5. Пусть опыт G состоит в подбрасывании игральной кости один раз. Событие A = {выпало число очков, кратное трем}, событие B = {выпало не менее трех очков}. Здесь $A = \{\omega_3, \omega_6\}$, $B = \{\omega_3, \omega_4, \omega_5, \omega_6\}$. Поэтому $A \subset B$.

Если одновременно $A \subset B$ и $B \subset A$, то события A и B называются *равносильными* и пишут $A = B$.

События называют *равновозможными*, если по условию испытания нет оснований считать какое-либо из них более возможным, чем другое. Так, появление «орла» или «решки» при бросании монеты — равновозможные события, если считать монету симметричной.

События A и B называются *совместными*, если наступление (появление) одного из них не исключает возможность наступления (появления) в одном опыте и другого.

События A и B называются *несовместными*, если они одновременно не могут произойти в одном опыте.

События A и $\overline{A}$ называются *противоположными*, если тот факт, что одно из них не наступило в результате данного испытания, влечет наступление другого. Очевидно, что противоположные события несовместны.

Пример 1.6. Пусть при подбрасывании игральной кости событие A = {выпало число очков, кратное трем}, событие B = {выпало число очков, кратное двум}, событие C = {выпало число очков, кратное пяти}, событие D = {выпало нечетное число очков}.

Тогда события A и B совместны, так как в случае выпадения шести очков оба события наступают. События B и D несовместны и противоположны. События B и C несовместны, но и не являются противоположными, так как в случае, если B не наступает, может не наступить и событие C (например, если выпадет три очка).

1.2. Классическое определение вероятности

Понятие *вероятности*, очевидно, является одним из ключевых в теории вероятностей. В целом вероятность события можно определить как численную меру объективной возможности его появления в данном испытании. Для вычисления этой меры используются различные подходы. Начнем с классического определения вероятности.

Пусть полная группа элементарных событий $\Omega = \{\omega_1, \ldots, \omega_n\}$ конечна и все ω_i в ней равновозможны. Элементарные исходы, в которых интересующее нас событие A наступает, называются *благоприятствующими* этому событию или *благоприятными*.

Вероятность события A — это отношение числа элементарных исходов, благоприятствующих A, к общему числу элементарных исходов, т. е.

$$\mathrm{P}(A) = \frac{m}{n}, \qquad (1.1)$$

где $\mathrm{P}(A)$ — вероятность события A; m — число элементарных исходов, благоприятствующих событию A; n — общее число элементарных исходов.

Пример 1.7. Подбрасывается игральная кость и наблюдается число выпавших очков X. Найти вероятность следующих событий: A = {выпало число очков, кратное трем}, B = {выпало четное число очков}.

Решение. Полная группа элементарных исходов здесь $\Omega = \{\omega_1, \ldots, \omega_6\}$, где элементарный исход ω_i состоит в том, что $\{X = i\}$, $i = 1, \ldots, 6$. Таким образом, всего исходов шесть и $n = 6$.

Событию A благоприятствует два элементарных исхода ω_3 и ω_6, следовательно, $m = 2$. Тогда

$$\mathrm{P}(A) = \frac{2}{6} = \frac{1}{3}.$$

Событию B благоприятствует три элементарных исхода: ω_2, ω_4 и ω_6, следовательно, $m = 3$. Тогда

$$\mathrm{P}(B) = \frac{3}{6} = \frac{1}{2}.$$

Пример 1.8. Наудачу выбрано натуральное число, не превосходящее 10. Найти вероятность того, что оно простое.

Решение. Испытанием в данном случае является выбор одного из чисел: 1, 2, 3, … 10 и, следовательно, $n = 10$. Пусть событие C = {выбранное число является простым}. Среди возможных вариантов простыми являются числа 2, 3, 5, 7 и, таким образом, $m = 4$. Тогда

$$\mathrm{P}(C) = \frac{4}{10} = 0{,}4.$$

Пример 1.9. Подбрасываются две симметричные монеты. Найти вероятность того, что на обеих выпадет «орел».

Решение. Испытанием в данном случае является подбрасывание двух монет. В этом испытании четыре равновозможных исхода $n = 4$: («орел», «орел»), («решка», «орел»), («орел», «решка»), («решка», «решка»). Событию D благоприятствует только одно из них — («орел», «орел») и, следовательно, $m = 1$. Тогда

$$\mathrm{P}(D) = \frac{1}{4} = 0{,}25.$$

Пример 1.10. В старинной игре в кости для выигрыша необходимо было получить при бросании трех игральных костей сумму очков, превосходящую 10. Найти вероятность выпадения 11 очков.

Решение. Выпишем подходящие (с суммой, равной 11) наборы из трех чисел от 1 до 6:

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

Всевозможные варианты можно получить как перестановки из этих чисел, при этом там, где два одинаковых числа, будет 3 варианта (например, (1, 5, 5); (5, 1, 5); (5, 5, 1)), а где все три разные — 6 вариантов (например, (1, 4, 6); (1, 6, 4); (4, 1, 6); (4, 6, 1); (6, 1, 4); (6, 4, 1)). Итого вариантов: $6 + 3 + 6 + 6 + 3 + 3 = 27$. Общее число исходов при бросании трех игральных костей равно $6 \cdot 6 \cdot 6 = 216$. Поэтому вероятность выпадения 11 очков на трех костях

$$\mathrm{P}(A) = \frac{27}{216} = 0{,}125.$$

Свойства вероятности:

1. Вероятность достоверного события равна 1: $\mathrm{P}(\Omega) = 1$.

Доказательство. В этом случае $m = n$, так как достоверное событие включает в себя все n элементарных событий. Тогда $\mathrm{P}(\Omega) = n / n = 1$.

2. Вероятность невозможного события равна 0: $\mathrm{P}(\varnothing) = 0$.

Доказательство. В этом случае $m = 0$ и $\mathrm{P}(\varnothing) = 0 / n = 0$.

3. Вероятность случайного события есть положительное число, заключенное между нулем и единицей.

Доказательство. Действительно, случайному событию благоприятствуют лишь часть из общего числа элементарных исходов испытания. В этом случае

$0 < m < n$ и $0 < \dfrac{m}{n} < 1,$ и, следовательно, $0 < \mathrm{P}(A) < 1.$

Итак, вероятность любого события удовлетворяет неравенству:

$$0 \le \mathrm{P}(A) \le 1.$$

1.3. Комбинаторика и вероятность

Комбинаторика — раздел математики, который изучает количества различных комбинаций, подчиненных тем или иным условиям, которые можно составить из элементов заданного конечного множества.

Правила и основные формулы комбинаторики используют при вычислении вероятностей. Сформулируем два универсальных правила, применяемых при решении комбинаторных задач.

Правило умножения. Если из множества A элемент a_1 может быть выбран n_1 способами, после каждого такого выбора элемент a_2 может быть выбран n_2 способами и так до элемента a_k, который может быть выбран n_k способами, то выбор всех элементов a_1, a_2, …, a_k в указанном порядке может быть сделан $n = n_1 \cdot n_2 \cdot \ldots \cdot n_k$ способами.

Пример 1.11. В группе 30 студентов. Необходимо выбрать старосту и профорга. Сколькими способами это можно сделать?

Решение. Выберем для начала, например, старосту. Это можно сделать $n_1 = 30$ способами. После этого выбрать профорга из оставшихся 29 студентов можно $n_2 = 29$ способами (на каждый вариант выбора старосты). Таким образом, по правилу умножения количество способов выбора профорга и старосты равно $n = n_1 \cdot n_2 = 30 \cdot 29 = 870$.

Правило сложения. Если из множества A элемент a_1 может быть выбран n_1 способами, другой элемент a_2 может быть выбран n_2 способами и так до отличного от предыдущих элемента a_k, который может быть выбран n_k способами, то выбор одного из элементов: или a_1, или a_2, …, или a_k может быть сделан $n = n_1 + n_2 + \ldots + n_k$ способами.

Пример 1.12. Имеется 20 изделий первого сорта и 30 изделий второго сорта. Необходимо выбрать два изделия одного сорта. Сколько способов выбора существует, если учитывается порядок выбора изделий?

Решение. Если выбираются два изделия первого сорта, то по правилу умножения это можно сделать $n_1 = 20 \cdot 19 = 380$ способами. Для изделий второго сорта существует $n_2 = 30 \cdot 29 = 870$ способов. Необходимо выбрать изделия одного сорта — неважно какого. Таким образом, могут быть выбраны два изделия либо первого, либо второго сорта. По правилу сложения это можно сделать $n = n_1 + n_2 = 380 + 870 = 1\,250$ способами.

Рассмотрим некоторое конечное множество из n различных элементов. Пусть из числа его элементов выбрано подмножество из m различных элементов ($m \leq n$). Если важен порядок, в котором произведена выборка элементов, то говорят об упорядоченном подмножестве, если порядок не важен, то о неупорядоченном.

Размещением из n элементов по m называется любое упорядоченное подмножество из m элементов множества, содержащего n различных элементов.

Пример 1.13. Рассмотрим множество $X = \{A, B, C, D\}$, состоящее их четырех различных элементов (букв). Размещениями из четырех элементов по два в данном случае являются всевозможные «слова», состоящие из двух различных букв, которые можно сложить из элементов данного множества: *AB*, *AC*, *AD*, *BA*, *BC*, *BD*, *CA*, *CD*, *CB*, *DA*, *DB*, *DC*. Всего получилось 12 размещений. Размещений из четырех элементов по три будет, очевидно, больше.

На практике обычно важнее найти количество размещений, а не их вид. Количество размещений из n элементов по m обозначается A_n^m и вычисляется по формуле

$$A_n^m = \frac{n!}{(n-m)!}. \tag{1.2}$$

Пример 1.14. Сколько можно записать четырехзначных чисел, используя без повторения все 10 цифр?

Решение. Чтобы определить, каким количеством способов можно составить упорядоченную четверку цифр, воспользуемся формулой (1.2):

$$n_1 = A_{10}^4 = \frac{10!}{(10-4)!} = \frac{10!}{6!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 7 \cdot 8 \cdot 9 \cdot 10 = 5\,040.$$

Из этих «чисел» необходимо теперь исключить те, которые начинаются с нуля и не являются, собственно, числами. Их количество определяется тройкой цифр, стоящей после нуля (эти цифры уже не могут содержать ноль):

$$n_2 = A_9^3 = \frac{9!}{(9-3)!} = \frac{9!}{6!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6} = 7 \cdot 8 \cdot 9 = 504.$$

Тогда общее количество четырехзначных чисел:

$$n = n_1 - n_2 = 5\,040 - 504 = 4\,536.$$

Рассмотрим теперь размещения, содержащие в себе все элементы исходного множества ($m = n$).

Перестановками из n элементов называются любые упорядоченные подмножества, в которые входят по одному все n различных элементов данного множества.

Таким образом, перестановки отличаются друг от друга только порядком расположения элементов.

Количество перестановок из n элементов обозначается P_n. Для его нахождения необходимо учесть, что по определению факториалом нуля является единица ($0! = 1$):

$$P_n = A_n^n = \frac{n!}{(n-n)!} = n! \tag{1.3}$$

Пример 1.15. Найдем всевозможные сочетания букв (перестановки), которые можно получить из различных элементов множества $X = \{A, B, C\}$: ABC, ACB, BAC, BCA, CAB, CBA. Убедимся, что мы нашли все имеющиеся варианты. По формуле (1.3) $P_3 = 3! = 6$.

Пусть теперь при выборе подмножеств из m элементов нам не важен порядок элементов, а важен только их состав. Такие подмножества называются *сочетаниями.*

Сочетанием из n элементов по m называется любое подмножество из m элементов, принадлежащих множеству, состоящему из n различных элементов.

Пример 1.16. Рассмотрим множество $X = \{A, B, C, D\}$. Сочетаниями по три элемента здесь будут четыре неупорядоченных подмножества: $\{A, B, C\}$, $\{A, B, D\}$, $\{B, C, D\}$ и $\{A, C, D\}$.

Количество сочетаний из n элементов по m обозначается C_n^m и вычисляется по формуле

$$C_n^m = \frac{n!}{m!(n-m)!}. \tag{1.4}$$

Пример 1.17. В урне 10 белых и 5 черных шаров. Сколькими способами можно выбрать из урны 7 шаров, чтобы среди них было 3 черных?

Решение. В данном случае не имеет значения порядок выбора шаров, а только их цвет: 3 черных и 4 белых. Найдем сначала количество способов выбора черных шаров. По формуле (1.4) находим:

$$n_1 = C_5^3 = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{1\cdot2\cdot3\cdot4\cdot5}{1\cdot2\cdot3\cdot2} = 10;$$

для белых шаров

$$n_2 = C_{10}^4 = \frac{10!}{4!(10-4)!} = \frac{10!}{4!6!} = 210.$$

По формуле умножения получаем общее количество способов выбора: $n = n_1 \cdot n_2 = 210 \cdot 10 = 2\,100$.

Замечание. Пусть теперь в размещениях из n элементов по m каждый элемент исходного множества может появляться несколько раз. Такие подмножества называются *размещениями с повторениями*. Их количество обозначается $\overline{A}_n^m$ и определяется по формуле

$$\overline{A}_n^m = n^m. \qquad (1.5)$$

Пример 1.18. Вернемся к множеству $X = \{A, B, C, D\}$. Размещениями из четырех элементов по два с повторениями в данном случае являются всевозможные «слова», состоящие из двух букв, которые можно сложить из элементов данного множества: *AB*, *AC*, *AD*, *BA*, *BC*, *BD*, *CA*, *CD*, *CB*, *DA*, *DB*, *DC*, *AA*, *BB*, *DD*, *CC*. Всего получилось 16 размещений с повторениями. Убедимся, что найдены все возможные варианты. По формуле (1.5) $\overline{A}_4^2 = 4^2 = 16$.

При вычислении вероятностей событий часто используются формулы комбинаторики.

Пример 1.19. На шести одинаковых карточках написаны буквы Т, Е, О, Р, И, Я. Карточки перемешиваются наугад и раскладываются в ряд. Какова вероятность того, что получится слово ТЕОРИЯ?

Решение. Испытанием в данном случае является раскладывание карточек. Поскольку все буквы разные, это можно сделать $n = P_6 = 6!$ способами. При этом только один исход будет благоприятствовать событию $A = \{\text{слово ТЕОРИЯ}\}$.

Таким образом, по формуле (1.1) имеем: $\mathrm{P}(A) = \frac{1}{6!} = \frac{1}{720}$.

Пример 1.20. В ряд из 20 мест рассаживаются произвольным образом 10 девочек и 10 мальчиков. Какова вероятность того, что никакие две девочки не окажутся рядом?

Решение. Общее число элементарных событий равно количеству различных размещений 20 человек на 20 местах: $n = P_{20} = 20!$ Будем обозначать буквой М какого-либо мальчика, а буквой Д — девочку. Рассмотрим сначала различные размещения, при которых два мальчика могут оказаться рядом, но никакие две девочки не окажутся рядом. Таким размещениям соответствуют следующие последовательности из букв М и Д:

Д$\underline{\text{ММ}}$ДМДМД…МД,

ДМД$\underline{\text{ММ}}$ДМД…МД,

………………………

ДМДМ…ДМД$\underline{\text{ММ}}$Д.

Поскольку два мальчика подряд могут находиться или после первой, или после второй и т. д., или после 9-й девочки, то таких последовательностей 9. Кроме указанных последовательностей есть еще две:

ДМДМДМ…ДМДМ и

МДМДМД…МДМД.

Всего $9 + 1 + 1 = 11$ последовательностей, в каждой из которых девочки могут размещаться 10! различными способами и мальчики тоже 10! способами. Поэтому количество благоприятных исходов равно $n = 11 \cdot 10! \cdot 10!$ Искомая вероятность

$$\mathrm{P}(A) = \frac{11 \cdot (10!)^2}{20!} = 5{,}95 \cdot 10^{-5}.$$

Рассмотрим еще один случай использования формул комбинаторики для нахождения вероятности (урновая схема).

Пусть в урне содержится N шаров, из них R красного цвета. Наугад, без возвращения извлекают n шаров. Требуется найти вероятность того, что в выборке содержится ровно r красных шаров.

Будем считать, что совокупность шаров — это множество, состоящее из N различных элементов. Нас не будет интересовать порядок расположения шаров в выборке, поэтому общее число исходов (количество способов выбрать из N шаров n шаров) равно C_N^n.

Благоприятными исходами являются те, когда выбрано ровно r красных шаров и, следовательно, $n - r$ шаров другого цвета. Вначале найдем число способов, какими можно выбрать r шаров из множества R шаров красного цвета, т. е. C_R^r. К этой комбинации шаров красного цвета присоединим комбинацию $n - r$ шаров другого цвета, таких комбинаций C_{N-R}^{n-r}. По правилу произведения общее число благоприятных исходов равно $m = C_R^r C_{N-R}^{n-r}$. Тогда вероятность того, что в выборке содержится ровно r красных шаров, вычисляется по формуле

$$\mathrm{P}(A) = \frac{C_R^r C_{N-R}^{n-r}}{C_N^n}. \qquad (1.6)$$

Пример 1.21. Среди 25 студентов группы, из которых 20 девушек, разыгрывается 5 билетов в кино. Найти вероятность того, что среди обладателей билетов будет ровно один юноша.

Решение. Это задача на формулу (1.6). Здесь из 25 студентов необходимо выбрать 5 и, следовательно, $n = C_{25}^5$. Так как юношей в группе всего пять, а выбрать необходимо одного, то количество вариантов выбора юношей $C_5^1 = 5$.

Каждый юноша может оказаться в компании четырех девушек из 20: C_{20}^4. Таким образом, $m = C_5^1 C_{20}^4 = 5C_{20}^4$. Тогда

$$\mathrm{P}(A) = \frac{5C_{20}^4}{C_{25}^5} = 5 \cdot \frac{20!}{4! \cdot 16!} \cdot \frac{5! \cdot 20!}{25!} = \frac{5 \cdot 17 \cdot 18 \cdot 19 \cdot 20 \cdot 5}{21 \cdot 22 \cdot 23 \cdot 24 \cdot 25} = \frac{1\,615}{3\,542} = 0{,}456.$$

1.4. Относительная частота события. Статистическое определение вероятности

Существуют события, вероятность которых по разным причинам невозможно вычислить с помощью классического определения вероятности. Если, допустим, испытание — это бросание кости, то не всегда можно быть уверенным в том, что кубик симметричный и без дефектов, а следовательно, в том, что элементарные исходы равновозможны. В таком случае, классическую формулу для вычисления вероятности применять нельзя.

Статистическое определение вероятности дает возможность найти вероятность опытным путем.

Отношение числа m наступлений данного случайного события A в данной серии испытаний к общему числу n испытаний этой серии называется *частотой* (или *относительной частотой*) события A и обозначается $W(A)$. Таким образом:

$$W(A) = \frac{m}{n}. \tag{1.7}$$

Чем отличается относительная частота от вероятности? Относительная частота — результат многократных испытаний. С увеличением числа испытаний относительная частота проявляет свойство *устойчивости*: всегда найдется такое число, от которого относительная частота будет отличаться сколь угодно мало при неограниченном увеличении количества испытаний. Это число и называется *статистической вероятностью события*.

Так, например, французский естествоиспытатель Бюффон[1] подбрасывал монету 4 040 раз, при этом относительная частота появления герба оказалась равной 0,50693. У английского статистика Пирсона[2] по результатам 23 тыс. под-

[1] Жорж Луи Леклерк Бюффон (Buffon) (1707–1788) — французский естествоиспытатель, иностранный почетный член Петербургской академии наук (1776). В основном труде «Естественная история» высказал представления о развитии земного шара и его поверхности, о единстве плана строения органического мира.

[2] Карл Пирсон (1857–1936) — английский математик, статистик, биолог и философ; основатель математической статистики. Опубликовал основополагающие труды по математической статистике (более 400 работ по этой теме). Разработал теорию корреляции, критерии согласия, алгоритмы принятия решений и оценки параметров.

брасываний монеты относительная частота появления герба оказалась равной 0,5005. Демографам хорошо известна цифра 0,514 (на 1 000 рождающихся детей приходится в среднем 514 мальчиков).

Видно, что относительные частоты незначительно отклоняются от числа 0,5 (значения вероятности событий), причем тем меньше, чем больше число опытов.

Статистической вероятностью события A называется относительная частота этого события, вычисленная по результатам большого числа испытаний. Будем обозначать ее $\mathrm{P}^*(A)$, т. е.

$$\mathrm{P}^*(A) \approx \underbrace{W(A)}_{n\to\infty} = \frac{m}{\underbrace{n}_{n\to\infty}}. \tag{1.8}$$

В приведенных примерах мы заметили, что статистическая вероятность приблизительно равна классической: $\mathrm{P}^*(A) \approx \mathrm{P}(A)$.

В приведенном примере вероятность выпадения герба определялась приблизительно по результатам опыта — это *апостериорная* (*после опыта*) вероятность. Но определить вероятность выпадения герба можно и теоретически, без подбрасывания монеты, — это *априорная* (*до опыта*) вероятность.

Пример 1.22. Среди 300 деталей, изготовленных на станке, 12 оказались не соответствующими стандарту качества. Найти частоту появления нестандартных деталей.

Решение. По формуле (1.8) находим:

$$W(A) = \frac{12}{300} = 0{,}04.$$

1.5. Геометрическое определение вероятности

Пусть теперь число возможных исходов испытания бесконечно, как, например, при стрельбе по мишени (можно попасть в одну из бесконечного числа точек на мишени). В данном случае использование классического определения вероятности также не представляется возможным, и при вычислении вероятности используют *геометрическое* определение вероятности.

Пусть каждый результат испытаний определяется случайным положением точки в некоторой области G, мера которой G (mes G). Под мерой области G будем понимать длину (для отрезка), площадь (для плоской фигуры), объем (для трехмерного тела). Если G_0 — мера той области, попадание в которую благоприятствует наступлению события A, то

$$\mathrm{P}(A) = \frac{G_0}{G}. \tag{1.9}$$

Пример 1.23. Перед окопами вдоль прямой линии через каждые 10 м установлены противотанковые мины. Перпендикулярно этой линии движется танк, ширина которого 3 м. Какова вероятность того, что танк пересечет линию установки мин невредимым, т. е. не взорвется?

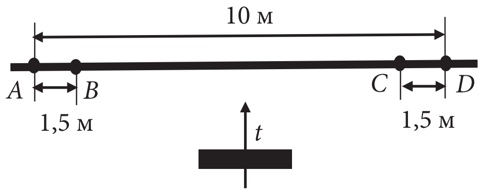


Рис. 1.1. Геометрическая иллюстрация к примеру 1.23

Решение. Танк может пересечь линию установки мин в бесконечном числе точек, поэтому данная задача решается с помощью геометрического определения вероятностей. Пусть AD — отрезок между двумя соседними минами (рис. 1.1). Ось симметрии танка t пересекает один из таких отрезков в любой его точке (т. е. отрезок соответствует множеству всех исходов испытания). Длина AD равна 10 м. $AB = CD = 1{,}5$ м. Поскольку длина танка 3 м, то при пересечении танком линии минирования его ось симметрии (середина) должна находиться на расстоянии от мин, большем 1,5 м, чтобы танк не взорвался. Таким образом, областью, благоприятствующей наступлению события A = {танк не взорвался при пересечении линии}, будет отрезок BC, длина которого 7 м. Тогда по формуле (1.9) находим:

$$\mathrm{P}(A) = \frac{BC}{AD} = \frac{7}{10} = 0{,}7.$$

Пример 1.24. Стержень длиной a наудачу разломан на три части. Определить вероятность того, что из полученных частей можно составить треугольник (событие A).

Решение. Стержень одинаково возможно поломать в любых точках, т. е. множество равновероятных исходов бесконечно. Применимо геометрическое определение вероятности. Пусть x, y — длины двух частей, тогда $a - x - y$ — длина третьей части. Все возможные исходы описываются системой неравенств:

$$\begin{cases} x \ge 0, \\ y \ge 0, \\ a - x - y \ge 0. \end{cases}$$

Записанной системе неравенств соответствует на плоскости треугольник ABO (рис. 1.2), его площадь $S_{\Delta ABO} = \frac{1}{2}a^2$. В любом треугольнике сумма длин двух сторон больше длины третьей стороны, поэтому благоприятные исходы задаются системой неравенств:

$$\begin{cases} x+y \geq a-x-y, \\ x+(a-x-y) \geq y, \\ y+(a-x-y) \geq x \end{cases} \quad \text{или} \quad \begin{cases} x \leq a/2, \\ y \leq a/2, \\ x+y \geq a/2. \end{cases}$$

На плоскости этой системе соответствует треугольник *DKN*. Его площадь: $S_{\Delta DKN} = \frac{1}{2}\left(\frac{a}{2}\right)^2$. Тогда $\mathrm{P}(A) = \frac{S_{\Delta DKN}}{S_{\Delta ABO}} = \frac{1}{4}$.

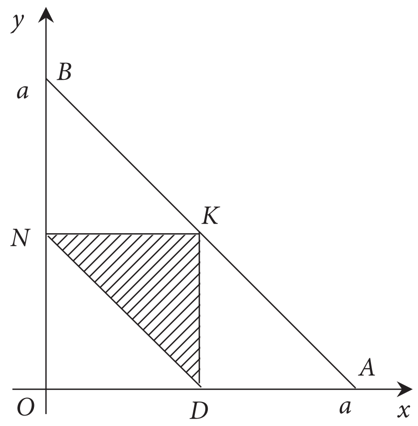


Рис. 1.2. Геометрическая иллюстрация к примеру 1.24

1.6. Действия над событиями

Пусть события A, $B \subset \Omega$ (Ω — пространство элементарных событий).

Суммой двух событий A и B называется событие $A + B$, состоящее в том, что в результате испытания произойдет хотя бы одно из этих событий.

Таким образом, если события A и B совместны, то их сумма $A + B$ означает, что наступило или событие A, или событие B, или оба события вместе. Если события несовместны, то событие $A + B$ заключается в том, что наступит или событие A, или событие B, так как совместное наступление событий невозможно.

Пример 1.25. В урне находятся красные, белые и синие шары. Вынимается один шар. Пусть события: A = {вынули белый шар}, B = {вынули красный шар}. Тогда событие $A + B$ = {вынули не синий шар}.

Пример 1.26. Подбрасывается игральный кубик. Пусть события: A = {выпадение числа очков, кратного трем}, B = {выпадение числа очков, большего трех}. Тогда событие $A + B$ = {выпадение числа очков, большего двух}.

Для наглядного представления событий, операций над событиями и отношений между ними используются диаграммы Эйлера[3] — Венна[4] (рис. 1.3–1.10). На этих диаграммах достоверное событие Ω изображается в виде некоторой области на плоскости, а случайному событию A будет соответствовать некоторая геометрическая фигура внутри области Ω (рис. 1.3).

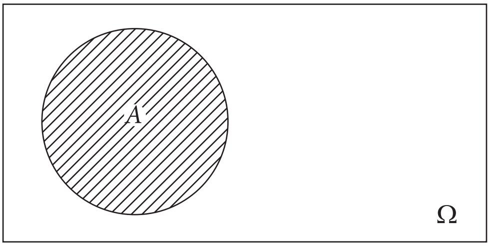


Рис. 1.3. Диаграмма Эйлера — Венна:
A — случайное событие, Ω — достоверное событие

Событию $A + B$ соответствует множество, элементы которого принадлежат хотя бы одному из множеств A или B, т. е. объединение множеств A и B (рис. 1.4 и 1.5).

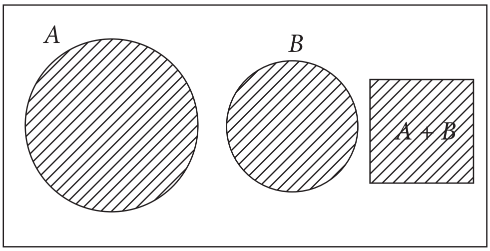


Рис. 1.4. Сумма несовместных событий A и B

[3] Леонард Эйлер (1707–1783) — швейцарский, немецкий и российский математик и механик, внесший фундаментальный вклад в развитие этих наук (а также физики, астрономии и ряда прикладных наук). В 1768 г. описал так называемые круги Эйлера для наглядного представления отношений между подмножествами.

[4] Джон Венн (1834–1923) — английский логик и философ. Известен тем, что усовершенствовал и развил идеи Эйлера, создал особый графический аппарат (диаграммы Венна), которые используются во многих областях, таких как теория множеств, теория вероятностей, логика, статистика и информатика.

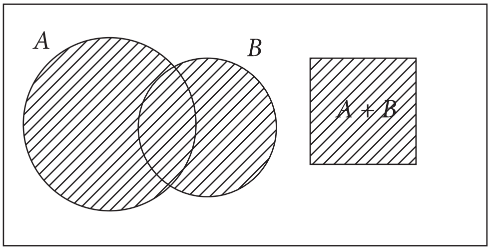


Рис. 1.5. Сумма совместных событий A и B

Для нескольких событий сумма определяется аналогичным образом. Например, событие $A + B + C$ заключается в наступлении одного из событий A, B, C, или в одновременном наступлении A и B, B и C, C и A, или в совместном наступлении всех трех событий.

Для суммы событий выполняются соотношения:

$$A + B = B + A;$$
$$(A + B) + C = A + (B + C);$$
$$A + A = A;$$
$$A + \Omega = \Omega;$$
$$A + \varnothing = A;$$
$$\overline{A} + A = \Omega.$$

Произведением событий A и B называется событие AB, состоящее в одновременном наступлении событий A и B.

Событию AB соответствует множество с элементами, принадлежащими одновременно множествам A и B, т. е. пересечению множеств A и B (рис. 1.6). Если события A и B несовместны, то их произведение является невозможным событием (рис. 1.7).

Пример 1.27. Пусть имеются события: A = {из колоды карт вынули «даму»} и B = {из колоды карт вынули карту пиковой масти}. Тогда произведение событий AB = {из колоды карт вынули «даму пик»}.

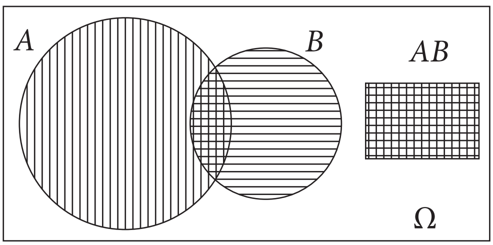


Рис. 1.6. Произведение совместных событий A и B

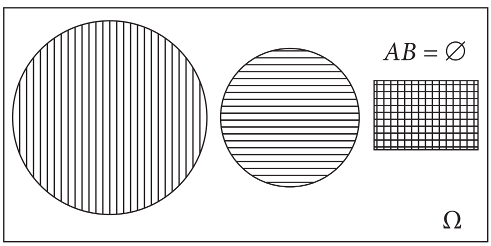


Рис. 1.7. Произведение несовместных событий A и B: $AB = \varnothing$

Произведением нескольких событий называется событие, состоящее в совместном наступлении всех этих событий в результате испытания.

Если события в группе не являются совместными в совокупности, то их произведение будет событием невозможным. Из попарной совместности событий из группы не следует совместность в совокупности. В схеме на рис. 1.8 событие $A_1A_2A_3$ невозможное ($A_1 \cap A_2 \cap A_3 = \varnothing$), поскольку события не являются совместными в совокупности, хотя часть из них попарно несовместны, так как $A_1A_2 \neq \varnothing, A_2A_3 \neq \varnothing$.

События $A_1, A_2, \ldots, A_n$ называются *совместными в совокупности*, если каждое из них и произведение остальных являются совместными событиями (рис. 1.9).

Переформулируем теперь понятие «полной группы событий».

События $A_1, A_2, \ldots, A_n$ в испытании G образуют *полную группу событий,* если они попарно несовместны ($A_iA_j = \varnothing$ при $i \neq j$) и их сумма эквивалентна достоверному событию: $A_1 + A_2 + \ldots + A_n = \Omega$, т. е. в испытании одно из событий обязательно произойдет.

Любые два противоположных события образуют полную группу событий: $A\overline{A} = \varnothing$ и $A + \overline{A} = \Omega$.

Для произведения событий справедливы соотношения:

$$AB = BA;$$

$$(AB)C = A(BC);$$

$$AA = A;$$

$$A\Omega = A;$$

$$A\overline{A} = \varnothing;$$

$$A\varnothing = \varnothing.$$

Операции сложения и умножения удовлетворяют *свойству дистрибутивности*:

$$(A + B)C = AC + BC.$$

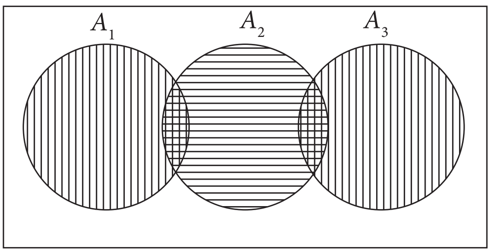


Рис. 1.8. Геометрическая иллюстрация несовместных в совокупности событий

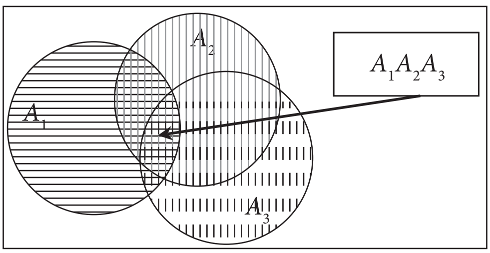


Рис. 1.9. Геометрическая иллюстрация совместных в совокупности событий

Операции над событиями удовлетворяют *формулам Моргана*:

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

Разностью событий A и B называется событие A/B, состоящее в том, что происходит событие A и при этом не происходит событие B (рис. 1.10).

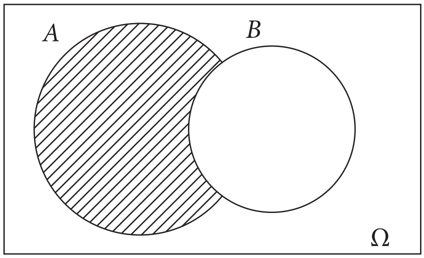


Рис. 1.10. Разность событий A и B

Событию A/B соответствует множество, состоящее из элементов множества A и при этом не принадлежащих B.

1.7. Теоремы сложения вероятностей

С помощью теории вероятностей можно определять вероятность события по известным вероятностям других событий, если они связаны с первым. Для этого используются теоремы сложения и умножения вероятностей.

Теорема 1.1. Если два события A и B являются несовместными, то вероятность появления их суммы $A+B$ равна сумме вероятностей этих событий, т. е.

$$\mathrm{P}(A+B)=\mathrm{P}(A)+\mathrm{P}(B). \tag{1.10}$$

Пример 1.28. В урне 30 шаров: 5 синих, 10 белых и 15 красных. Найти вероятность того, что из урны будет извлечен цветной шар.

Решение. Обозначим события: $A=\{$извлечен синий шар$\}$, $B=\{$извлечен белый шар$\}$, $C=\{$извлечен красный шар$\}$. Извлечение цветного шара означает извлечение либо синего, либо красного шара, т. е. сумму событий $A+C$. Найдем вероятность этих событий с помощью классического определения вероятностей. Имеем: $\mathrm{P}(A)=5/30$, $\mathrm{P}(C)=15/30$. Так как события A и C несовместны, то по формуле (1.10) получаем:

$$\mathrm{P}(A+C)=\mathrm{P}(A)+\mathrm{P}(C)=\frac{5}{30}+\frac{15}{30}=\frac{20}{30}=\frac{2}{3}.$$

Обобщим теорему на случай нескольких событий.

Следствие 1. Вероятность одного из нескольких попарно несовместных событий равна сумме вероятностей этих событий, т. е.

$$\mathrm{P}\left(A_1+A_2+\ldots+A_n\right)=\mathrm{P}\left(A_1\right)+\mathrm{P}\left(A_2\right)+\ldots+\mathrm{P}\left(A_n\right). \tag{1.11}$$

Следствие 2. Сумма вероятностей событий A_1, A_2, ..., A_n, образующих полную группу, равна единице:

$$P(A_1)+P(A_2)+\ldots+P(A_n)=1. \tag{1.12}$$

Следствие 3. Сумма вероятностей противоположных событий равна 1:

$$P(A)+P(\overline{A})=1$$

или

$$P(A)=1-P(\overline{A}). \tag{1.13}$$

Пример 1.29. Вероятность попадания в мишень при одном выстреле равна 0,7. Какова вероятность того, что стрелок промахнется?

Решение. Пусть событие A = {стрелок попал в мишень}, $P(A) = 0{,}7$. Тогда $\overline{A}$ = {стрелок промахнулся} и по формуле (1.13) имеем:

$$P(\overline{A})=1-P(A)=1-0{,}7=0{,}3.$$

Пример 1.30. Среди 11 изделий находятся три бракованных. Произвольно извлекают три изделия. Найти вероятность того, что среди них хотя бы одно бракованное.

Решение. Пусть событие A = {среди трех извлеченных изделий хотя бы одно бракованное}. Тогда $\overline{A}$ = {среди трех извлеченных изделий нет ни одного бракованного}. Найдем вероятность $\overline{A}$ по классической формуле. Число всех исходов испытания равно C_{11}^3, а число исходов, благоприятствующих $\overline{A}$, равно C_8^3 (изделия выбираются из 8 стандартных). Тогда

$$P(\overline{A})=\frac{C_8^3}{C_{11}^3}=\frac{8!}{3!\cdot 5!}\cdot\frac{3!\cdot 8!}{11!}=\frac{56}{165},$$

откуда по формуле (1.13) получаем:

$$P(A)=1-P(\overline{A})=1-\frac{56}{165}=\frac{109}{165}.$$

Теорема 1.2. Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления:

$$P(A+B)=P(A)+P(B)-P(AB). \tag{1.14}$$

Формулу (1.14) легко проиллюстрировать с помощью диаграммы (рис. 1.11). Вероятность события пропорциональна площади фигуры, которая соответствует данному событию. Событию $A + B$ соответствует вся заштрихованная площадь. Событию AB — площадь «в клеточку». Из рисунка видно, что площади этих фигур связаны соотношением $S_{A+B} = S_A + S_B - S_{AB}$ (S_{AB} необходимо вычесть, так как в сумме она присутствует в обоих слагаемых). Данное соотношение соответствует формуле (1.14).

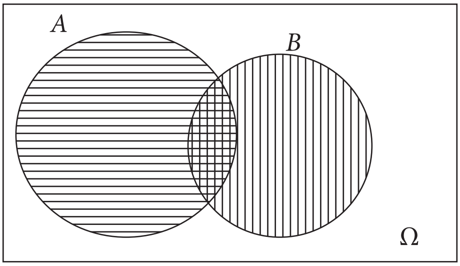


Рис. 1.11. Геометрическая иллюстрация к теореме 1.2

Аналогично формулируется теорема сложения для трех совместных событий.

Теорема 1.3. Вероятность суммы трех совместных событий равна:

$$\begin{aligned} \mathrm{P}(A+B+C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \\ - \mathrm{P}(AB) - \mathrm{P}(BC) - \mathrm{P}(AC) + \mathrm{P}(ABC). \end{aligned} \tag{1.15}$$

Пример 1.31. Отдел технического контроля проверяет изделия на стандартность по двум параметрам. Установлено, что 10 из 100 деталей не проходят контроль только по первому параметру, 15 из 100 — только по второму, а 6 из 100 — и по первому, и по второму. Найти вероятность того, что наудачу выбранная деталь нестандартна.

Решение. Рассмотрим события: A = {деталь не удовлетворяет стандарту по первому параметру}, B = {деталь не удовлетворяет стандарту по второму параметру}, C = {деталь нестандартна}. Тогда событие AB = {деталь не удовлетворяет стандарту по обоим параметрам}. Имеем:

$$\mathrm{P}(A) = 10/100 = 0{,}1;$$

$$\mathrm{P}(B) = 15/100 = 0{,}15;$$

$$\mathrm{P}(AB) = 6/100 = 0{,}06.$$

Событие C происходит, когда происходит хотя бы одно из событий A и B, т.е. $C = A + B$. Тогда по формуле (1.14) имеем:

$$\mathrm{P}(C) = \mathrm{P}(A+B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(AB) = 0{,}1 + 0{,}15 - 0{,}06 = 0{,}19.$$

1.8. Теоремы умножения вероятностей. Условная вероятность

Вероятность некоторого случайного события B, как правило, изменяется, если известно, что произошло некоторое другое случайное событие A.

Условной вероятностью называется вероятность события B при условии, что событие A наступило с вероятностью $P(A) > 0$. Условная вероятность обозначается $P(B/A)$ или $P_A(B)$.

Пример 1.32. В урне 10 шаров с номерами от 1 до 10. Наудачу один за другим берут два шара, не возвращая их обратно. Найти вероятность того, что второй шар будет иметь нечетный номер, если первый вынутый шар имеет номер 5.

Решение. Пусть событие $A = \{$первый вынутый шар имеет номер 5$\}$, $B = \{$второй вынутый шар имеет нечетный номер$\}$. В задаче требуется найти условную вероятность $P(B/A)$. Воспользуемся классическим определением вероятности. Если вынули шар с номером 5, то в урне осталось 9 шаров и $n = 9$. Из них шаров с нечетными номерами 4 штуки (1, 3, 7, 9), т. е. $m = 4$. Получается, $P(B/A) = 4/9$.

Теорема 1.4. Вероятность произведения двух случайных событий равна произведению вероятности одного из них на условную вероятность другого при условии, что первое событие произошло, т. е.

$$P(AB) = P(A)P(B/A) = P(B)P(A/B). \tag{1.16}$$

Таким образом,

$$P(A/B) = \frac{P(AB)}{P(B)}. \tag{1.17}$$

Аналогично

$$P(B/A) = \frac{P(AB)}{P(A)}. \tag{1.18}$$

Пример 1.32 (*продолжение*). В урне 10 шаров с номерами от 1 до 10. Наудачу один за другим берут два шара, не возвращая их обратно. Найти вероятность того, что первый вынутый шар имеет номер 5, а второй шар — нечетный номер.

Решение. Пусть событие $A = \{$первый вынутый шар имеет номер 5$\}$, $B = \{$второй вынутый шар имеет нечетный номер$\}$. В задаче требуется найти $P(AB)$. По формуле (1.16) имеем:

$$P(AB) = P(A)P(B/A) = \frac{1}{10} \cdot \frac{4}{9} = \frac{4}{90} = \frac{2}{45}.$$

Два события A и B называются *независимыми*, если наступление одного из них не меняет вероятность другого, т. е.

$$\mathrm{P}\left(A/B\right)=\mathrm{P}\left(A\right)$$

или

$$\mathrm{P}\left(B/A\right)=\mathrm{P}\left(B\right).$$

Сформулируем теперь теорему о вероятности произведения для двух независимых событий.

Теорема 1.5. Вероятность произведения двух независимых событий A и B равна произведению вероятностей этих событий:

$$\mathrm{P}\left(AB\right)=\mathrm{P}\left(A\right)\mathrm{P}\left(B\right). \tag{1.19}$$

Пример 1.33. Бросаются два игральных кубика. Найти вероятность того, что на первом кубике выпадет четное число очков, а на втором — число очков, большее 2.

Решение. Рассмотрим события A = {на первом кубике выпало четное число очков}, B = {на втором кубике выпало число очков, большее 2}. В задаче требуется найти P(AB). Очевидно, что события A и B независимы, ведь количество очков, выпавших на первом кубике, никак не влияет на результат бросания второго. Поскольку P(A) = 3/6 = 1/2 (три из шести возможных результатов бросания кубика четные), а P(B) = 4/6 = 2/3 (4 из 6 возможных результатов больше двух), то

$$\mathrm{P}\left(AB\right)=\mathrm{P}\left(A\right)\mathrm{P}\left(B\right)=\frac{1}{2}\cdot\frac{2}{3}=\frac{1}{3}.$$

Пример 1.34. Два стрелка независимо друг от друга стреляют по мишени. Вероятность попадания в мишень первого стрелка равна 0,7, а второго — 0,9. Найти вероятность того, что мишень будет поражена.

Решение. Пусть события A = {попадание в мишень первым стрелком}, B = {попадание в мишень вторым стрелком}. P(A) = 0,7, P(B) = 0,9. По условию, события независимы. Мишень будет поражена, если в нее попадет либо первый стрелок, либо второй, либо оба вместе, т. е. требуется найти вероятность события $A + B$. По формулам (1.14) и (1.19) имеем:

$$\mathrm{P}\left(A+B\right)=\mathrm{P}\left(A\right)+\mathrm{P}\left(B\right)-\mathrm{P}\left(AB\right)=0{,}7+0{,}9-0{,}7\cdot 0{,}9=0{,}97.$$

Пример 1.35. В первой урне три белых шара, четыре синих и пять красных, а во второй, соответственно, четыре, пять и шесть. Из каждой урны наугад выбирают два шара. Какова вероятность того, что будут выбраны шары одного цвета?

Решение. Пусть события A = {выбрали шары одного цвета}, B_i = {выбрали из i-й урны два белых шара}, C_i = {выбрали из i-й урны два синих шара}, K_i = {выбрали из i-й урны два красных шара}. Событие A произойдет, если из первой урны будут выбраны два белых шара (событие B_1) и из второй урны будут выбраны тоже два белых шара (событие B_2), или из первой урны извлекут два синих шара (событие C_1) и из второй урны будут выбраны тоже два синих шара (событие C_2), или из первой урны будут выбраны два красных шара (событие K_1) и из второй урны будут выбраны тоже два красных шара (событие K_2). Поэтому $A = B_1B_2 + C_1C_2 + K_1K_2$. События независимы и слагаемые несовместны. В итоге получаем, что

$$\mathrm{P}(A) = \mathrm{P}(B_1B_2 + C_1C_2 + K_1K_2) = \mathrm{P}(B_1)\mathrm{P}(B_2) + \mathrm{P}(C_1)\mathrm{P}(C_2) + \mathrm{P}(K_1)\mathrm{P}(K_2) =$$
$$= \left(\frac{3}{12}\cdot\frac{2}{11}\right)\cdot\left(\frac{4}{15}\cdot\frac{3}{14}\right) + \left(\frac{4}{12}\cdot\frac{3}{11}\right)\cdot\left(\frac{5}{15}\cdot\frac{4}{14}\right) + \left(\frac{5}{12}\cdot\frac{4}{11}\right)\cdot\left(\frac{6}{15}\cdot\frac{5}{14}\right) = \frac{38}{1155}.$$

Рассмотрим теперь случай произведения произвольного числа событий.

Теорема 1.6. Для любого конечного числа событий вероятность произведения событий равна произведению вероятностей этих событий, причем вероятность каждого следующего события вычисляется при условии, что предыдущие события произошли, т. е.

$$\mathrm{P}\left(A_1\ A_2 \ldots A_n\right) = \mathrm{P}\left(A_1\right)\mathrm{P}\left(A_2 / A_1\right)\mathrm{P}\left(A_3 / A_1A_2\right)\ldots\mathrm{P}\left(A_n / A_1A_2 \ldots A_{n-1}\right). \quad (1.20)$$

Пример 1.36. Студент знает ответ на 20 вопросов из 25. Найти вероятность того, что он ответит на три вопроса, предложенные преподавателем.

Решение. Пусть событие A_i — ответ студента на i-й вопрос. Имеем: $\mathrm{P}\left(A_1\right) = \frac{20}{25}$, $\mathrm{P}\left(A_2 / A_1\right) = \frac{19}{24}$, $\mathrm{P}\left(A_3 / A_1A_2\right) = \frac{18}{23}$. Тогда по теореме умножения вероятностей получаем:

$$\mathrm{P}\left(A_1A_2A_3\right) = \mathrm{P}\left(A_1\right)\mathrm{P}\left(A_2 / A_1\right)\mathrm{P}\left(A_3 / A_1A_2\right) = \frac{20}{25}\cdot\frac{19}{24}\cdot\frac{18}{23} = \frac{57}{115} \approx 0{,}496.$$

Эту же вероятность можно найти, используя классическое определение вероятности:

$$\mathrm{P}\left(A\right) = \frac{C_{20}^3}{C_{25}^3} = \frac{20!}{3!17!}\cdot\frac{3!22!}{25!} = \frac{20}{25}\cdot\frac{19}{24}\cdot\frac{18}{23} = \frac{57}{115} \approx 0{,}496.$$

Сформулируем теперь теорему о вероятности произведения для нескольких независимых событий.

Случайные события $A_1, A_2, \ldots, A_n$ называются *попарно независимыми*, если для любых $i, j, i \neq j$ события A_i и A_j независимы.

Случайные события $A_1, A_2, \ldots, A_n$ называются независимыми в совокупности, если наряду с их попарной независимостью независимы любое из них и произведение любого числа из остальных.

Замечание. Из независимости в совокупности следует попарная независимость, но из попарной независимости не следует независимость событий в совокупности.

Пример 1.37. Пусть на карточках написаны числа от 1 до 12 и карточки тщательно перемешаны. Наудачу вынимается карточка. Пусть события $A = \{$на карточке четное число$\}$, $B = \{$на карточке число, большее 6$\}$, $C = \{$на карточке число, большее 2 и меньшее 9$\}$. Легко показать, что

$$\mathrm{P}(A) = \mathrm{P}(B) = \mathrm{P}(C) = 6/12 = 1/2.$$

Пусть теперь произошло событие A. Тогда количество возможных вариантов сокращается до шести: 2, 4, 6, 8, 10, 12. Из них три (8, 10, 12) благоприятствуют событию B и три (4, 6, 8) — событию C. Следовательно:

$$\mathrm{P}(B/A) = \mathrm{P}(C/A) = 3/6 = 1/2.$$

Аналогично можно показать, что

$$\mathrm{P}(A/B) = \mathrm{P}(C/B) = 1/2 \quad \text{и} \quad \mathrm{P}(A/C) = \mathrm{P}(B/C) = 1/2.$$

Таким образом, простые и условные вероятности совпадают и события A, B и C являются попарно независимыми. Пусть теперь произошло событие $AB = \{$на карточке четное число, большее 6$\}$. Тогда количество возможных вариантов равно трем (8, 10, 12) и только одно из них (8) благоприятствует событию C. Следовательно: $\mathrm{P}(C/AB) = 1/3 \neq \mathrm{P}(C)$, т. е. события не являются независимыми в совокупности.

Теорема 1.7. Вероятность произведения нескольких событий, независимых в совокупности, равна произведению вероятностей этих событий, т. е.

$$\mathrm{P}(A_1 A_2 \ldots A_n) = \mathrm{P}(A_1)\mathrm{P}(A_2)\ldots\mathrm{P}(A_n). \tag{1.21}$$

Пример 1.38. Вероятность того, что студент сдаст первый экзамен, равна 0,9, второй — 0,8, третий — 0,7. Считая, что результат одного экзамена не влияет на результаты других, найти вероятность того, что студентом будут сданы все три экзамена.

Решение. Пусть события $A_1 = \{$студент сдал первый экзамен$\}$, $A_2 = \{$студент сдал второй экзамен$\}$ и $A_3 = \{$студент сдал третий экзамен$\}$. Тогда A_1, A_2 и A_3 независимы в совокупности и $\mathrm{P}(A_1) = 0{,}9$, $\mathrm{P}(A_2) = 0{,}8$, $\mathrm{P}(A_3) = 0{,}7$. Пусть событие

$B =$ {студент сдал все три экзамена}. Очевидно, что $B = A_1A_2A_3$. Тогда по формуле (1.21) получим:

$$\mathrm{P}(B) = \mathrm{P}\left(A_1A_2A_3\right) = \mathrm{P}\left(A_1\right)\mathrm{P}\left(A_2\right)\mathrm{P}\left(A_3\right) = 0,9 \cdot 0,8 \cdot 0,7 = 0,504.$$

Замечание. Если события A и B *независимы*, то *независимы* также события A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.

Пример 1.38 (*продолжение*). Вероятность того, что студент сдаст первый экзамен, равна 0,9, второй — 0,8, третий — 0,7. Считая, что результат одного экзамена не влияет на результаты других, найти вероятность того, что студентом будут сданы:

а) только второй экзамен (событие C_1);

б) только один экзамен (событие C_2);

в) по крайней мере, два экзамена (событие C_3);

г) хотя бы один экзамен (событие C_4).

Решение. а) Очевидно, что $C_1 = \overline{A}_1A_2\overline{A}_3$ (студент сдаст второй экзамен и не сдаст первый и третий экзамены). Учитывая, что события A_1, A_2, A_3 независимы, то независимы и события $\overline{A}_1$, A_2, $\overline{A}_3$. Тогда

$$\mathrm{P}(C_1) = \mathrm{P}\left(\overline{A}_1A_2\overline{A}_3\right) = \mathrm{P}\left(\overline{A}_1\right)\mathrm{P}\left(A_2\right)\mathrm{P}\left(\overline{A}_3\right) = \left(1-0,9\right) \cdot 0,8 \cdot \left(1-0,7\right) = 0,024.$$

б) Событие C_2 означает, что студент сдаст или только первый экзамен из трех, или только второй, или только третий, т. е.

$$\begin{aligned}\mathrm{P}(C_2) &= \mathrm{P}\left(A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3\right) = \\ &= \mathrm{P}\left(A_1\right)\mathrm{P}\left(\overline{A}_2\right)\mathrm{P}\left(\overline{A}_3\right) + \mathrm{P}\left(\overline{A}_1\right)\mathrm{P}\left(A_2\right)\mathrm{P}\left(\overline{A}_3\right) + \mathrm{P}\left(\overline{A}_1\right)\mathrm{P}\left(\overline{A}_2\right)\mathrm{P}\left(A_3\right) = \\ &= 0,9 \cdot 0,2 \cdot 0,3 + 0,1 \cdot 0,8 \cdot 0,3 + 0,1 \cdot 0,2 \cdot 0,7 = 0,092.\end{aligned}$$

в) Событие C_3 означает, что студент сдаст либо два экзамена, либо все три экзамена:

$$\begin{aligned}\mathrm{P}(C_3) &= \mathrm{P}\left(A_1A_2A_3 + A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3\right) = \\ &= \mathrm{P}\left(A_1A_2A_3\right) + \mathrm{P}\left(A_1A_2\overline{A}_3\right) + \mathrm{P}\left(A_1\overline{A}_2A\right) + \mathrm{P}\left(\overline{A}_1A_2A_3\right) = 0,902.\end{aligned}$$

г) Событие C_4 означает, что студент сдаст либо один экзамен, либо два экзамена, либо все три экзамена:

$$\begin{aligned}\mathrm{P}(C_4) &= \mathrm{P}\left(A_1A_2A_3 + A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3 + A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3\right) = \\ &= 0,9 \cdot 0,8 \cdot 0,7 + 0,9 \cdot 0,8 \cdot 0,3 + 0,9 \cdot 0,2 \cdot 0,7 + 0,1 \cdot 0,8 \cdot 0,7 + \\ &+ 0,9 \cdot 0,2 \cdot 0,3 + 0,1 \cdot 0,8 \cdot 0,3 + 0,1 \cdot 0,2 \cdot 0,7 = 0,994.\end{aligned}$$

Вероятность события C_4 можно найти другим способом. Рассмотрим противоположное событие $\overline{C}_4$ = {студент не сдал ни одного экзамена}. Очевидно, что $\overline{C}_4 = \overline{A}_1\overline{A}_2\overline{A}_3.$ Тогда

$$\mathrm{P}\left(C_4\right) = 1 - \mathrm{P}\left(\overline{C}_4\right) = 1 - \mathrm{P}\left(\overline{A}_1\overline{A}_2\overline{A}_3\right) = 1 - 0{,}1 \cdot 0{,}2 \cdot 0{,}3 = 0{,}994.$$

Последний пример иллюстрирует один из эффективных приемов для вычисления вероятности суммы нескольких ($n \geq 2$) совместных событий.

Теорема 1.8. Вероятность появления хотя бы одного из нескольких событий $A_1, A_2, \ldots, A_n$, независимых в совокупности, равна разности между единицей и произведением вероятностей противоположных событий:

$$\mathrm{P}\left(A_1 + A_2 + \ldots + A_n\right) = 1 - \mathrm{P}\left(\overline{A}_1\right)\mathrm{P}\left(\overline{A}_2\right)\mathrm{P}\left(\overline{A}_3\right)\ldots\,\mathrm{P}\left(\overline{A}_n\right). \qquad (1.22)$$

1.9. Формула полной вероятности

Рассмотрим события $A, H_1, H_2, \ldots, H_n$, связанные с некоторым опытом. Пусть $H_1, H_2, \ldots, H_n$ образуют полную группу событий. Известно, что событие A может произойти только с одним из событий $H_1, H_2, \ldots, H_n$ полной группы (рис. 1.12). Иначе, событие A произойдет, если произойдет событие H_1 и при этом появится событие A, или произойдет событие H_2 и при этом появится событие A и т. д. Символически представим событие A в виде

$$A = AH_1 + AH_2 + \ldots + AH_n.$$

События AH_i и AH_j несовместны при $i \neq j$, поэтому по теореме сложения вероятностей несовместных событий имеем:

$$\mathrm{P}\left(A\right) = \mathrm{P}\left(AH_1\right) + \mathrm{P}\left(AH_2\right) + \ldots + \mathrm{P}\left(AH_n\right).$$

Далее, используя теорему умножения вероятностей, получим:

$$\mathrm{P}\left(A\right) = \mathrm{P}\left(H_1\right)\mathrm{P}\left(A/H_1\right) + \mathrm{P}\left(H_2\right)\mathrm{P}\left(A/H_2\right) + \ldots + \mathrm{P}\left(H_n\right)\mathrm{P}\left(A/H_n\right). \qquad (1.23)$$

Формула (1.23) носит название *формулы полной вероятности*.

Замечание. Так как среди событий $H_1, H_2, \ldots, H_n$, образующих полную группу, в результате опыта должно наступить одно и только одно, то эти события H_i ($i = 1, 2, \ldots, n$) называют *гипотезами*.

Пример 1.39. В торговую фирму поступили телевизоры от трех поставщиков в отношении 1 : 4 : 5. Известно, что телевизоры, поставленные от первого поставщика, не требуют ремонта в течение гарантийного срока в 98 % случаев, от второго поставщика — в 88 % случаев, от третьего поставщика — в 92 % случаев. Найти вероятность того, что поступивший в фирму телевизор не потребует ремонта в течение гарантийного срока.

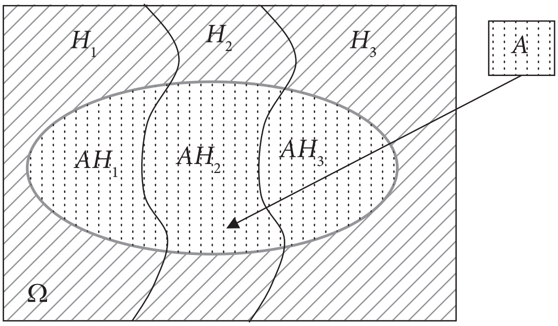


Рис. 1.12. Событие A может произойти одновременно с одним из событий (гипотез) $H_1, H_2, \ldots, H_n$ полной группы

Решение. Путь событие A = {поступивший в фирму телевизор не потребует ремонта в течение гарантийного срока}. Здесь возможны три гипотезы:

H_1 = {телевизор поступил от первого поставщика} и $P(H_1) = 1/10 = 0{,}1$;

H_2 = {телевизор поступил от второго поставщика} и $P(H_2) = 4/10 = 0{,}4$;

H_3 = {телевизор поступил от третьего поставщика} и $P(H_3) = 5/10 = 0{,}5$.

События H_1, H_2, H_3 образуют полную группу попарно несовместных событий, так как они не могут произойти одновременно и сумма их вероятностей равна единице: $P(H_1) + P(H_2) + P(H_3) = 1$. Событие A происходит одновременно с одним и только с одним из событий H_1, H_2, H_3.

Условные вероятности события A при этих гипотезах соответственно равны:

$$P\left(A/H_1\right) = 0{,}98, \quad P\left(A/H_2\right) = 0{,}88 \quad \text{и} \quad P\left(A/H_3\right) = 0{,}92.$$

По формуле полной вероятности получаем:

$$P\left(A\right) = P\left(H_1\right)P\left(A/H_1\right) + P\left(H_2\right)P\left(A/H_2\right) + P\left(H_3\right)P\left(A/H_3\right) =$$
$$= 0{,}1 \cdot 0{,}98 + 0{,}4 \cdot 0{,}88 + 0{,}5 \cdot 0{,}92 = 0{,}91.$$

Пример 1.40. Имеются две урны с шарами: в первой урне 6 черных и 4 белых шара, во второй — 7 черных и 3 белых шара. Наудачу из первой урны извлекается шар и перекладывается во вторую. После этого из второй урны наудачу извлекают два шара. Найти вероятность того, что оба этих шара черные.

Решение. Пусть событие A = {из второй урны извлекли два черных шара после того, как один шар переложили из первой урны во вторую}. Здесь возможны две гипотезы:

H_1 = {из первой урны во вторую переложили черный шар}, $P(H_1) = 6/10 = 3/5$,

$P(A/H_1) = \dfrac{8}{11} \cdot \dfrac{7}{10} = \dfrac{28}{55}$ (поскольку после перекладывания черного шара во второй урне стало 8 черных и 3 белых шара);

$H_2 =$ {из первой урны во вторую переложили белый шар}, $P(H_2) = 4/10 = 2/5$,

$P(A/H_2) = \frac{7}{11} \cdot \frac{6}{10} = \frac{21}{55}$ (поскольку после перекладывания белого шара во второй урне стало 7 черных и 4 белых шара).

Тогда по формуле полной вероятности имеем:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) = \frac{3}{5} \cdot \frac{28}{55} + \frac{2}{5} \cdot \frac{21}{55} = \frac{126}{275}.$$

1.10. Формулы Байеса

Пусть событие A, которое может появиться только с одним из событий H_1, H_2, …, H_n полной группы, произошло в результате испытания. Выясним, изменились ли после этого вероятности гипотез $P(H_i/A)$. По теореме умножения имеем:

$$P(AH_i) = P(A)P(H_i / A) = P(H_i)P(A / H_i), \quad i = 1, 2, 3 \ldots n$$

или

$$P(A)P(H_i/A) = P(H_i)P(A/H_i).$$

Откуда

$$P(H_i/A) = \frac{P(H_i)P(A/H_i)}{P(A)}.$$

Применяя в знаменателе формулу полной вероятности, получим формулу Байеса[5]:

$$P(H_i/A) = \frac{P(H_i)P(A/H_i)}{\sum_{i=1}^{n} P(H_i)P(A/H_i)}. \qquad (1.24)$$

Формулы Байеса позволяют переоценить вероятности гипотез с учетом информации о появлении события A.

[5] Томас Байес (1702–1761) — английский математик, пресвитерианский священник, член Лондонского королевского общества. Математические интересы Байеса относились к теории вероятностей. Он сформулировал и решил одну из основных задач этого раздела математики (теорема Байеса). Работа, посвященная этой задаче, была опубликована в 1763 г., посмертно. Формула Байеса, дающая возможность оценить вероятность событий эмпирическим путем, играет важную роль в современной математической статистике и теории вероятностей. Другая крупная его работа — «Очерки к решению проблемы доктрины шансов».

Пример 1.41. На предприятии изготовляются изделия определенного вида на трех поточных линиях. На первой линии производится 40 % изделий от всего объема их производства, на второй — 30 %, на третьей — 30 %. Каждая из линий характеризуется соответственно следующими процентами брака изделий: 5 %, 3 %, 2 %. Требуется определить вероятность того, что наугад взятое бракованное изделие, выпущенное предприятием, сделано на первой линии.

Решение. Пусть событие A = {наудачу взятое изделие оказалось бракованным}. Гипотезы: H_1, H_2, H_3 — наугад взятое изделие произведено соответственно на 1, 2, 3-й поточных линиях. По условию, $\mathrm{P}(H_1)=\frac{4}{10}$, $\mathrm{P}(H_2)=\frac{3}{10}$, $\mathrm{P}(H_3)=\frac{3}{10}$. Условные вероятности события A:

$$\mathrm{P}(A/H_1)=0{,}05,\ \mathrm{P}(A/H_2)=0{,}03,\ \mathrm{P}(A/H_3)=0{,}02.$$

Используя формулу полной вероятности, находим:

$$\mathrm{P}(A)=0{,}4\cdot 0{,}05+0{,}3\cdot 0{,}03+0{,}3\cdot 0{,}02=0{,}035.$$

Определим теперь апостериорную (посчитанную после того, как брак был выявлен) вероятность того, что это изделие изготовлено на первой линии. По формуле Байеса имеем:

$$\mathrm{P}(H_1/A)=\frac{\mathrm{P}(H_1)\mathrm{P}(A/H_1)}{\mathrm{P}(A)}=\frac{0{,}05\cdot 0{,}4}{0{,}035}=\frac{20}{35}=\frac{4}{7}.$$

Вероятность того, что изделие изготовлено на первой линии, увеличилась (4/7 > 4/10) после появления информации о том, что изделие браковано.

Пример 1.42. Пусть в условиях примера 1.39 торговой фирмой был продан телевизор. Пусть этот телевизор потребовал ремонта в течение гарантийного срока. Выяснить, от какого поставщика вероятнее всего поступил данный телевизор.

Решение. Пусть событие B = {поступивший в фирму телевизор потребует ремонта в течение гарантийного срока} произошло. $B=\overline{A}$ и по условию:

$$\mathrm{P}(B/H_1)=1-\mathrm{P}(A/H_1)=1-0{,}98=0{,}02;$$

$$\mathrm{P}(B/H_2)=1-\mathrm{P}(A/H_2)=1-0{,}88=0{,}12;$$

$$\mathrm{P}(B/H_3)=1-\mathrm{P}(A/H_3)=1-0{,}92=0{,}08.$$

По формуле Байеса имеем:

$$\mathrm{P}(H_1/B)=\frac{\mathrm{P}(H_1)\mathrm{P}(B/H_1)}{\mathrm{P}(H_1)\mathrm{P}(B/H_1)+\mathrm{P}(H_2)\mathrm{P}(B/H_2)+\mathrm{P}(H_3)\mathrm{P}(B/H_3)}=\frac{0{,}1\cdot 0{,}02}{0{,}09}=0{,}022$$

— вероятность того, что проданный телевизор от первого поставщика;

$$P(H_2/B)=\frac{P(H_2)P(B/H_2)}{P(H_1)P(B/H_1)+P(H_2)P(B/H_2)+P(H_3)P(B/H_3)}=\frac{0,4\cdot 0,12}{0,09}=0,533$$

— вероятность того, что проданный телевизор от второго поставщика;

$$P(H_3/B)=\frac{P(H_3)P(B/H_3)}{P(H_1)P(B/H_1)+P(H_2)P(B/H_2)+P(H_3)P(B/H_3)}=\frac{0,5\cdot 0,08}{0,09}=0,444$$

— вероятность того, что проданный телевизор от третьего поставщика.

Таким образом, учитывая, что телевизор сломался в течение гарантийного срока, он, вероятнее всего, от второго поставщика. При этом первоначально наибольшая вероятность была у третьей гипотезы.

Задачи для самостоятельного решения

1.1. Все натуральные числа от 1 до 40 записаны на одинаковых карточках и помещены в урну. После тщательного перемешивания карточек из урны извлекается одна карточка. Найти вероятность того, что число на взятой карточке окажется кратным 7.

1.2. Подбрасываются два игральных кубика, подсчитывается сумма очков на верхних гранях. Найти вероятность того, что в сумме получится 10 очков.

1.3. В книге 300 страниц. Найти вероятность того, что наугад открытая страница будет иметь номер, содержащий цифру ноль.

1.4. Подбрасываются три симметричные монеты. Найти вероятность того, что все монеты упали одинаково.

1.5. На четырех одинаковых карточках написаны буквы У, Р, А, Л. Карточки перемешиваются и наугад раскладываются в ряд. Найти вероятность того, что получится слово УРАЛ.

1.6. На четырех одинаковых карточках написаны буквы У, У, Ф, Р. Карточки перемешиваются и наугад раскладываются в ряд. Найти вероятность того, что получится слово УРФУ.

1.7. На шести одинаковых карточках написаны все буквы слова ТОСТЕР. Карточки перемешиваются, наугад извлекаются 4 карточки и раскладываются в ряд. Найти вероятность того, что получится слово ТОСТ.

1.8. В урне 9 шаров, из которых 5 красных. Наудачу извлекают 4 шара. Найти вероятность того, что все извлеченные шары красные.

1.9. В урне 9 красных, 8 синих и 5 зеленых шаров. Наудачу извлекают 9 шаров. Найти вероятность того, что извлекли 3 красных, 3 синих и 3 зеленых шара.

1.10. Относительная частота нормального всхода семян равна 0,98. Из высеянных семян взошло 1 470. Сколько семян было высеяно?

1.11. На плоскости начерчены две концентрические окружности с радиусами 8 см и 10 см. Найти вероятность того, что точка, брошенная наудачу в большой круг, попадет в кольцо, образованное двумя окружностями.

1.12. В квадрат с вершинами A (0, 0), B (1, 0), C (1, 1), D (0, 1) наудачу брошена точка с координатами (x, y). Найти вероятность того, что координаты этой точки удовлетворяют уравнению $y < 2x$.

1.13. Вероятность попадания в мишень для первого спортсмена равна 0,75, для второго — 0,8, для третьего — 0,9. Спортсмены независимо друг от друга делают по одному выстрелу. Какова вероятность того, что попадет хотя бы один спортсмен? Ровно один спортсмен? Только первый спортсмен?

1.14. Вероятность попадания в мишень для первого спортсмена равна 0,8, для второго — 0,9. Спортсмены независимо друг от друга два раза стреляют по мишени. Найти вероятность того, что количество попаданий у них будет одинаково.

1.15. В урне 5 желтых, 8 красных и 7 зеленых шаров. Из урны наудачу поочередно извлекают по одному шару и выкладывают их на столе, причем второй шар кладут под первым, а третий — под вторым. Найти вероятность того, что на столе получится «светофор».

1.16. Рабочий обслуживает три однотипных станка. Вероятность того, что любой станок в течение часа потребует внимания рабочего, равна 0,4. Предполагая, что станки работают независимо, найти вероятность того, что в течение часа потребуют внимания по крайней мере два станка.

1.17. Найти вероятность того, что наудачу взятое двузначное число кратно или 5, или 8, или тому и другому числу одновременно.

1.18. В пирамиде пять винтовок, две из которых снабжены оптическим прицелом. Вероятность того, что стрелок поразит мишень при выстреле из винтовки с оптическим прицелом, равна 0,98; для винтовки без оптического прицела эта вероятность равна 0,6. Найти вероятность того, что мишень будет поражена, если стрелок производит один выстрел из наудачу взятой винтовки.

1.19. Имеются две партии изделий по 12 и 15 штук, причем в каждой партии одно изделие бракованное. Изделие, взятое наудачу из первой партии, переложено во вторую, после чего выбирается наудачу изделие из второй партии. Определить вероятность того, что из второй партии извлечено бракованное изделие.

1.20. Исследованиями психологов установлено, что мужчины и женщины по-разному реагируют на некоторые жизненные обстоятельства. Результаты исследований показали, что 70 % женщин позитивно реагируют на эти ситуации, в то время как 40 % мужчин реагируют на них негативно. 20 женщин и 15 мужчин заполнили анкету, в которой отразили свое отношение к данной ситуации. Случайно извлеченная анкета содержит негативную реакцию. Чему равна вероятность того, что ее заполнил мужчина?

1.21. Среди поступающих на сборку деталей с первого автомата 0,1 % брака, со второго — 0,2 %, с третьего — 0,25 %. Производительности их относятся как 5 : 3 : 3. Найти вероятность того, что взятая наудачу деталь окажется бракованной.

Ответы к задачам для самостоятельного решения

1.1. P = 1/8.
1.2. P = 1/12.
1.3. P = 48/300.
1.4. P = 1/4.
1.5. P = 1/24.
1.6. P = 1/12.
1.7. P = 1/15.
1.8. P = 5/126.
1.9. P = 336/3 553.
1.10. 1 500 семян.
1.11. P = 0,36.
1.12. P = 3/4.
1.13. P_1 = 0,995, P_2 = 0,08, P_3 = 0,015.
1.14. P = 0,5764.
1.15. P = 7/171.
1.16. P = 0,352.
1.17. P = 29/90.
1.18. P = 0,752.
1.19. P = 13/192.
1.20. P = 1/2.
1.21. P ≈ 0,00168.

2. ПОСЛЕДОВАТЕЛЬНОСТИ ИСПЫТАНИЙ

С понятием «независимых событий» тесно связано понятие «независимых испытаний (опытов)».

Если проводится несколько испытаний, т. е. опыт выполняется при данном комплексе условий многократно (проводится «последовательность испытаний»), причем вероятность наступления некоторого события A в каждом испытании не зависит от исходов других испытаний, то такие испытания называются *независимыми*.

Примерами независимых испытаний могут служить: несколько подбрасываний монеты; несколько извлечений из урны занумерованных шаров, если шары каждый раз (после просмотра) возвращаются опять в урну; выстрелы по мишени и т. д.

2.1. Формула Бернулли

Пусть проводится конечное число n последовательных *независимых испытаний* (опытов), в результате каждого из которых может наступить (такую ситуацию назовем *успехом*) или не наступить (такую ситуацию назовем *неудачей*) некоторое случайное событие A, причем в каждом отдельно взятом испытании вероятность успеха $\mathrm{P}(A) = p$, $p \in (0; 1)$ и, следовательно, $\mathrm{P}(\overline{A}) = 1 - p = q$ — вероятность неудачи.

Такая последовательность испытаний называется *схемой Бернулли*[6] или *биномиальной схемой*, сами испытания — *испытаниями Бернулли*.

---

[6] Якоб Бернулли (1654–1705) — швейцарский математик, профессор Базельского университета, иностранный член Парижской (1699) и Берлинской (1702) академий наук, родоначальник знаменитой семьи ученых. Один из основателей теории вероятностей и математического анализа. Доказал частный случай закона больших чисел — теорему Бернулли.

Теорема 2.1. Пусть опыт G производится по схеме Бернулли. Тогда вероятность $P_n(m)$ события, состоящего в том, что при n повторениях опыта G событие A произойдет ровно m раз, вычисляется по формуле Бернулли:

$$P_n(m) = C_n^m p^m q^{n-m} \quad \left(m = 0, 1, 2, \ldots, n\right), \tag{2.1}$$

где $C_n^m = \dfrac{n!}{m!\left(n-m\right)!}$.

Замечание. $C_n^m p^m q^{n-m}$ является $m+1$ членом бинома Ньютона $(p+q)^n$, поэтому вероятности $P_n(m)$ называют *биномиальными вероятностями*. Сумма биномиальных вероятностей равна 1:

$$\sum_{m=0}^{n} P_n\left(m\right) = \sum_{m=0}^{n} C_n^m p^m q^{n-m} = \left(p+q\right)^n = 1^n = 1.$$

Пример 2.1. Монету бросают 7 раз. Найти вероятность того, что 4 раза выпадет герб.

Решение. Воспользуемся формулой Бернулли. Под опытом в задаче понимается бросание монеты. В каждом опыте событие A = {выпадение герба} происходит с вероятностью $P(A) = p = 1/2$. При этом $P(\overline{A}) = q = 1 - p = 1/2$. Опыт проводится $n = 7$ раз при неизменных условиях. Требуется найти вероятность того, что событие A произойдет в $m = 4$ случаях. По формуле (2.1) имеем:

$$P_7(4) = C_7^4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^3 = \frac{7!}{4! \cdot 3!} \cdot \frac{1}{16} \cdot \frac{1}{8} = \frac{35}{128}.$$

Пример 2.2. Применяемый метод лечения приводит к выздоровлению в 90 % случаев. Найти вероятность того, что из пяти больных поправятся не менее четырех.

Решение. В данной задаче под опытом понимается лечение больного определенным методом. В каждом опыте событие A = {больной выздоровел} происходит с вероятностью $P(A) = p = 0{,}9$. При этом $P(\overline{A}) = q = 1 - p = 0{,}1$. Опыт проводится $n = 5$ раз при неизменных условиях. Требуется найти вероятность события B = {событие A произойдет не менее чем в четырех случаях}. Событие B можно представить в виде суммы двух несовместных событий: B_1 = {ровно $m = 4$ больных поправятся} и B_2 = {ровно $m = 5$ больных поправятся}. По формуле (2.1) имеем:

$$P(B) = P(B_1) + P(B_2) = P_5(4) + P_5(5) = C_5^4 0{,}9^4 0{,}1^1 + C_5^5 0{,}9^5 0{,}1^0 =$$
$$= 0{,}328 + 0{,}590 = 0{,}918.$$

Пример 2.3. Средний процент невозвращения в срок кредита, выдаваемого банком, составляет 5 %. Найти вероятность того, что при выдаче банком 10 кредитов проблемы с возвратом денег возникнут не менее чем в двух случаях.

Решение. Под опытом здесь понимается получение кредита, выданного банком. В каждом опыте событие $A = \{$кредит не возвращен в срок$\}$ происходит с вероятностью $\mathrm{P}(A) = p = 0{,}05$. При этом $\mathrm{P}(\overline{A}) = q = 1 - p = 0{,}95.$ Этот опыт проводится $n = 10$ раз при неизменных условиях. Требуется найти вероятность события $B = \{$событие A произойдет хотя бы в двух случаях$\}$. Перейдем к вероятности противоположного события: кредит не будет возвращен в срок менее чем в двух случаях. Это событие можно представить в виде суммы двух несовместных событий: $B_1 = \{$ровно в 1 случае из 10 кредит не будет возвращен в срок$\}$ и $B_2 = \{$во всех 10 случаях кредит будет возвращен в срок$\}$. По теореме сложения вероятностей несовместных событий получим:

$$\mathrm{P}\left(\overline{B}\right) = \mathrm{P}\left(B_1\right) + \mathrm{P}\left(B_2\right) = \mathrm{P}_{10}\left(1\right) + \mathrm{P}_{10}\left(0\right) = C_{10}^{1}\left(0{,}05\right)^1\left(0{,}95\right)^9 + C_{10}^{0}(0{,}05)^0(0{,}95)^{10} =$$
$$= 10 \cdot 0{,}05 \cdot 0{,}95^9 + 0{,}95^{10} \approx 0{,}9137.$$

Тогда вероятность искомого события равна:

$$\mathrm{P}\left(B\right) = 1 - \mathrm{P}\left(\overline{B}\right) = 1 - 0{,}9137 \approx 0{,}0863.$$

2.2. Наивероятнейшее число событий

Число m_0 наступлений события A в n независимых испытаниях называется *наивероятнейшим,* если вероятность осуществления этого события $\mathrm{P}_n(m_0)$ по крайней мере не меньше вероятностей других событий $\mathrm{P}_n(m)$ при любом m. Число m_0 удовлетворяет неравенствам:

$$n \cdot p - q \le m_0 \le n \cdot p + p. \tag{2.2}$$

Заметим, что разность $np + p - (np - q) = p + q = 1.$ Поэтому всегда существует целое число m_0, удовлетворяющее неравенству (2.2). При этом если $np + p$ — целое число, то наивероятнейших чисел два: $m_0 = np + p$ и $m_0^* = np - q$.

Пример 2.4. В летнюю сессию студент Вася Зубрилкин должен был сдать четыре экзамена. Однако по некоторым причинам он подготовил только 3/4 билетов по каждому предмету. Найти наивероятнейшее число экзаменов, которые сдаст Вася.

Решение. Будем считать, что каждый экзамен есть испытание в схеме Бернулли с вероятностью успеха (сдачи) $p = 3/4$. Наивероятнейшее число экзаменов подчиняется неравенствам (2.2), в которых $n = 4$, $p = 3/4$, $q = 1/4$, т. е.

$$4 \cdot 3/4 - 1/4 \le m_0 \le 4 \cdot 3/4 + 3/4$$

или

$$2{,}75 \le m_0 \le 3{,}75,$$

откуда

$$m_0 = 3.$$

Таким образом, наивероятнейшее число экзаменов, которые Вася сдаст, равно 3.

2.3. Асимптотические формулы в схеме Бернулли

При больших n ($n > 30$) подсчет вероятностей по формуле Бернулли сопряжен с громоздкими вычислениями. В связи с этим возникла необходимость в построении асимптотических (приближенных) формул, позволяющих с достаточной степенью точности определить вероятность $P_n(m)$, не прибегая к сложным вычислениям. Рассмотрим некоторые из них.

2.3.1. Локальная теорема Муавра — Лапласа

Теорема 2.2. Пусть вероятность $p \in (0; 1)$ наступления события A в каждом из n независимых испытаний постоянна и не меняется от опыта к опыту, а число испытаний достаточно велико. Тогда вероятность $P_n(m)$ того, что в n испытаниях событие A наступит m раз, приближенно равна:

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{m - np}{\sqrt{npq}}\right), \tag{2.3}$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Формулу (2.3) называют *локальной формулой Муавра — Лапласа*. На практике формулой (2.3) пользуются при $npq \ge 10$. График функции $\varphi(x)$ называют *кривой Гаусса*[7] (рис. 2.1).

Для упрощенных расчетов, связанных с применением формулы (2.3), составлена таблица значений функции $\varphi(x)$ (см. Приложение, табл. 1).

Отметим основные свойства функции $\varphi(x)$:

1) $\varphi(-x) = \varphi(x)$, т. е. функция Гаусса четная;

---

[7] Иоганн Карл Фридрих Гаусс (1777–1855) — немецкий математик, механик, физик, астроном и геодезист. Считается одним из величайших математиков всех времен, «королем математиков». Лауреат медали Копли, иностранный член Шведской и Российской академий наук, английского Королевского общества. В теории вероятностей исследования связаны с нормальным законом распределения, график которого часто называют кривой Гаусса.

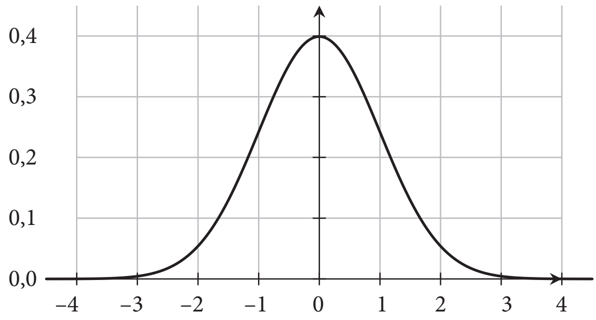


Рис. 2.1. Кривая Гаусса

2) $\varphi_{\max} = \varphi(0) = \dfrac{1}{\sqrt{2\pi}} \approx 0,3989;$

3) функция $\varphi(x)$ монотонно убывает при $x > 0$;

4) $\lim\limits_{x\to\infty} \varphi(x) = 0.$ При решении задач можно полагать $\varphi(x) \approx 0$ при $x > 5$.

Пример 2.5. Вероятность того, что книга, выпущенная тиражом 400 экз., сброшюрована неправильно, составляет 0,1. Найти вероятность того, что тираж содержит 30 бракованных книг.

Решение. Здесь $n = 400$, $p = 0,1$, $q = 1 - p = 0,9$, $npq = 400 \cdot 0,1 \cdot 0,9 = 36$ ($npq \geq 10$). Воспользуемся локальной теоремой Муавра — Лапласа. Требуется найти вероятность того, что тираж содержит 30 бракованных книг, т. е. $m = 30$:

$$\mathrm{P}_{400}(30) \approx \frac{1}{\sqrt{400 \cdot 0,1 \cdot 0,9}} \varphi\left(\frac{30 - 40}{\sqrt{400 \cdot 0,1 \cdot 0,9}}\right) = \frac{1}{6}\varphi\left(\frac{-10}{6}\right) = \frac{1}{6}\varphi(1,67) \approx \frac{0,0989}{6} \approx 0,0165.$$

2.3.2. Теорема Пуассона

Теорема 2.3. Пусть вероятность p наступления события A в каждом из n независимых испытаний постоянна и близка к нулю, а число испытаний достаточно велико. Тогда вероятность $\mathrm{P}_n(m)$ того, что в n испытаниях событие A наступит m раз, приближенно равна:

$$\mathrm{P}_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}, \tag{2.4}$$

где $\lambda = np$.

Теорема Пуассона[8] применяется, как правило, если n велико и при этом p очень мало, $np < 10$. Вероятность $\mathrm{P}_n(m)$ в этом случае называют вероятностью *массовых* (n велико) и *редких* (p мало) событий.

---

[8] Симеон Дени Пуассон (1781–1840) — французский ученый, выдающийся математик и физик; почетный член Петербургской академии наук (1826); член Парижской академии наук (1812); профессор Парижского университета (с 1809). В истории науки Пуассон стоит в одном ряду с его выдающимися современниками — Лапласом, Лагранжем, Фурье, Коши, Ампером,

Пример 2.6. На факультете 1 825 студентов. Найти вероятность того, что 1 сентября является днем рождения одновременно четырех студентов факультета.

Решение. Для каждого студента факультета событие $A = \{$день рождения 1 сентября$\}$ происходит с вероятностью $\mathrm{P}(A) = p = 1/365$. При этом $n = 1\,825$, $\lambda = np = (1/365)\cdot 1\,825 = 5 < 10$. Тогда по формуле (2.4) имеем:

$$\mathrm{P}_{1825}(4) \approx \frac{5^4}{4!}\mathrm{e}^{-5} = \frac{625}{24\cdot \mathrm{e}^5} \approx 0{,}1755.$$

Попробуем решить задачу (2.6) при помощи локальной теоремы Муавра — Лапласа:

$$\mathrm{P}_{1825}(4) \approx \frac{1}{\sqrt{1825\cdot(1/365)\cdot(364/365)}}\varphi\left(\frac{4-5}{\sqrt{1825\cdot(1/365)\cdot(364/365)}}\right) \approx 0{,}1614.$$

Точное значение вероятности найдем по формуле Бернулли:

$$\mathrm{P}_{1825}(4) = C_{1825}^{4}\left(\frac{1}{365}\right)^4\left(\frac{364}{365}\right)^{1821} = 0{,}1756.$$

Таким образом, при малых p формула Пуассона является более предпочтительной, чем формула Муавра — Лапласа.

2.3.3. Интегральная формула Муавра — Лапласа

Пусть теперь в схеме Бернулли требуется вычислить вероятность того, что число m наступлений события A в n испытаниях окажется в промежутке от m_1 до m_2:

$$\mathrm{P}_n(m_1 \le m \le m_2) = \mathrm{P}_n(m_1) + \mathrm{P}_n(m_1+1) + \ldots + \mathrm{P}_n(m_2).$$

Очевидно, что при больших n и m даже применение теоремы Пуассона или локальной теоремы Муавра — Лапласа не избавляет от громоздких вычислений.

Для решения подобных задач на практике используется *интегральная теорема Муавра — Лапласа.*

Теорема 2.4. Пусть вероятность $p\in(0;\,1)$ наступления события A в каждом из n независимых испытаний постоянна и не меняется от опыта к опыту, а число испытаний достаточно велико. Тогда вероятность $\mathrm{P}_n(m_1 \le m \le m_2)$ того, что в n испытаниях число наступлений события A заключено в пределах от m_1 до m_2, приближенно равна:

$$\mathrm{P}_n(m_1 \le m \le m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right), \qquad (2.5)$$

Гей-Люссаком, Френелем. Его именем названы: интеграл Пуассона в интегральном исчислении, теорема Пуассона и распределение Пуассона в теории вероятностей.

где $\Phi\left(x\right)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt$ (*функция Лапласа*). Формулу (2.5) называют интегральной формулой Муавра — Лапласа (рис. 2.2).

Отметим о с н о в н ы е с в о й с т в а функции Лапласа:

1) $\Phi(x)$ — нечетная функция, т. е. $\Phi(-x) = -\Phi(x)$;

2) $\Phi(x)$ монотонно возрастающая функция, так как $\Phi'(x) > 0, \forall x \in \mathbf{R}$;

3) $\lim\limits_{x\to+\infty}\Phi(x)=0,5$ (при $x \geq 5$ $\Phi(x) \approx 0,5$).

Для значений функции Лапласа имеются специальные таблицы (см. Приложение, табл. 2).

Пример 2.7. При обследовании уставных фондов банков установлено, что пятая часть банков имеют уставный фонд свыше 100 млн руб. Найти вероятность того, что среди 625 банков имеют уставный фонд свыше 100 млн руб.: а) не менее 150; б) от 100 до 150 включительно; в) не более 125.

Решение. По условию $p = 1/5 = 0,2$ — вероятность того, что уставный фонд отдельно взятого банка свыше 100 млн руб. Имеем: $n = 625$, $q = 1 - p = 0,8$, $npq = 625 \cdot 0,2 \cdot 0,8 = 100 \geq 10$. Воспользуемся интегральной теоремой Муавра — Лапласа:

1. $m_1 = 150$, $m_2 = 625$. Тогда искомая вероятность

$$P_{625}\left(150\leq m\leq 625\right)\approx\Phi\left(\frac{625-625\cdot 0,2}{10}\right)-\Phi\left(\frac{150-625\cdot 0,2}{10}\right)=$$

$$=\Phi\left(50\right)-\Phi\left(2,5\right)=0,5-0,4938=0,0062.$$

2. $m_1 = 100$, $m_2 = 150$. Тогда искомая вероятность

$$P_{625}\left(100\leq m\leq 150\right)\approx\Phi\left(\frac{150-625\cdot 0,2}{10}\right)-\Phi\left(\frac{100-625\cdot 0,2}{10}\right)=$$

$$=\Phi\left(2,5\right)-\Phi\left(-2,5\right)=\Phi\left(2,5\right)+\Phi\left(2,5\right)=$$

$$=0,4938+0,4938=0,9876.$$

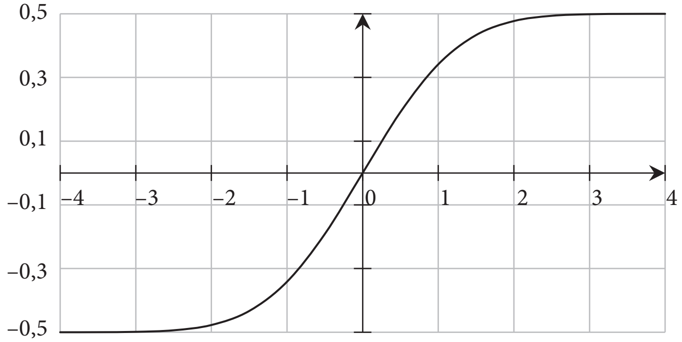


Рис. 2.2. График функции Лапласа

3. $m_1 = 0$, $m_2 = 125$. Тогда искомая вероятность

$$P_{625}\left(0 \le m \le 125\right) \approx \Phi\left(\frac{125 - 625 \cdot 0,2}{10}\right) - \Phi\left(\frac{0 - 625 \cdot 0,2}{10}\right) =$$
$$= \Phi\left(0\right) - \Phi\left(-12,5\right) = \Phi\left(0\right) + \Phi\left(12,5\right) = 0 + 0,5 = 0,5.$$

Следствие. Если вероятность p, $p \in (0; 1)$, наступления события A в каждом испытании постоянна и не меняется от опыта к опыту, а число n независимых испытаний достаточно велико, то вероятность того, что в n независимых испытаниях модуль отклонения относительной частоты события A от его вероятности p не превышает заданного $\varepsilon > 0$, приближенно равна:

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \tag{2.6}$$

Пример 2.8. При обработке линз в среднем 2 из 100 имеют брак. Сколько линз следует обработать, чтобы с вероятностью 0,95 можно было ожидать, что отклонение доли брака от его вероятности не превысит 0,01 (по абсолютной величине)?

Решение. Проводятся независимые испытания, состоящие в проверке брака обработки линз. Для каждой линзы вероятность события A = {линза обработана с браком} по условию постоянна и равна: $P(A) = p = 0,02$; $P(\overline{A}) = q = 0,98$. В задаче вероятность (надежность) $P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = \beta = 0,95$, отклонение $\varepsilon = 0,01$. Требуется определить число линз n, которые необходимо обработать.

Имеем:

$$P\left(\left|\frac{m}{n} - p\right| \le 0,01\right) = 0,95,$$

с другой стороны,

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Отсюда:

$$2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) \approx 0,95; \;\; \Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) \approx 0,475.$$

По таблице значений функции Лапласа находим:

$$\varepsilon\sqrt{\frac{n}{pq}} = \Phi^{-1}(0,475) = 1,96,$$

откуда

$$n = \frac{1,96^2}{\varepsilon^2} \cdot pq = \frac{1,96^2 \cdot 0,02 \cdot 0,98}{0.01^2} = 752,95.$$

Следовательно, нужно обработать 753 линзы, чтобы с вероятностью 0,95 можно было утверждать, что отклонение доли появления бракованной линзы от вероятности брака (0,02) не превысит 0,01 (по абсолютной величине).

Пример 2.9. Французский естествоиспытатель Бюффон подбросил монету 4 040 раз, причем герб появился 2 048 раз. Найти вероятность того, что при повторении опыта Бюффона относительная частота появления герба отклонится по модулю от вероятности его появления не более чем в опыте Бюффона.

Решение. В опыте Бюффона относительная частота появления герба (событие A) отклоняется от его вероятности по модулю на число, равное $\left|\frac{2\,048}{4\,040} - \frac{1}{2}\right| = \frac{7}{1\,010}$. Задача сводится к нахождению вероятности $\mathrm{P}\left(\left|\frac{m}{n} - \frac{1}{2}\right| \le \frac{7}{1\,010}\right)$. Имеем: $\mathrm{P}(A) = p = 0,5$; $\mathrm{P}(\overline{A}) = q = 0,5$, отклонение $\varepsilon = \frac{7}{1\,010}$. По формуле (2.6) вычислим:

$$\mathrm{P}\left(\left|\frac{m}{n} - \frac{1}{2}\right| \le \frac{7}{1\,010}\right) \approx 2\Phi\left(\frac{7}{1\,010}\sqrt{\frac{4\,040}{\frac{1}{2}\cdot\frac{1}{2}}}\right) = 2\Phi(0,88) = 2\cdot 0,31057 = 0,62114.$$

Задачи для самостоятельного решения

2.1. Вероятность попадания в мишень при одном выстреле для данного стрелка равна 0,8 и не зависит от номера выстрела. Найти вероятность того, что при 5 выстрелах произойдет ровно 3 попадания в мишень.

2.2. Игральная кость бросается шесть раз. Найти вероятность того, что в половине случаев выпадет число, большее четырех.

2.3. Всхожесть семян данного растения имеет вероятность 0,9. Какова вероятность, что из 5 посеянных семян взойдет не менее 4?

2.4. Подбрасывается монета. Что вероятнее: герб выпадет при одном бросании из двух или при двух бросаниях из четырех?

2.5. Вероятность банкротства каждой из четырех фирм к концу года равна 0,2. Какова вероятность того, что к концу года обанкротится не более одной фирмы?

2.6. Для нормальной работы автобазы на линии должно быть не менее восьми машин, а имеется их десять. Вероятность невыхода каждой машины на линию равна 0,1. Найти вероятность нормальной работы автобазы на ближайший день.

2.7. Доля изделий высшего сорта на предприятии составляет 30 %. Чему равно наивероятнейшее число изделий высшего сорта в случайно отобранной партии из 95 изделий?

2.8. Сколько раз надо подбросить игральный кубик, чтобы наивероятнейшее число выпадений шестерки было равно 20?

2.9. Вероятность найти белый гриб среди прочих равна 0,25. Какова вероятность того, что из 80 грибов белых будет 20?

2.10. Известно, что 80 % подержанных автомобилей попадали в ДТП. Найти вероятность того, что среди 900 автомобилей в ДТП попадали: а) 750 автомобилей; б) 710 автомобилей.

2.11. Сколько раз с вероятностью 0,0484 можно ожидать появления события A в 100 независимых испытаниях, если вероятность его появления в отдельном испытании равна 0,5?

2.12. В среднем вступительные экзамены в университет выдерживают 25 % абитуриентов. Пусть в приемную комиссию было подано 1 800 заявлений. Чему равна вероятность того, что хотя бы 450 поступающих наберут проходной балл?

2.13. Вероятность, что деталь пройдет проверку ОТК, равна 0,9. Найти вероятность того, что среди 900 деталей, отправленных на проверку, число успешно ее прошедших будет заключено между 790 и 830.

2.14. Известно, что из 100 семей только одна не имеет холодильника. Какова вероятность, что среди 1 100 исследуемых семей без холодильника окажутся не более 17?

2.15. Вероятность появления положительного результата в каждом из n опытов равна 0,9. Сколько нужно произвести опытов, чтобы с вероятностью 0,98 можно было ожидать, что не менее 150 опытов дадут положительный результат?

2.16. С конвейера выходит в среднем 85 % изделий первого сорта. Сколько изделий необходимо взять, чтобы с вероятностью 0,997 отклонение частоты изделий первого сорта от их вероятности по модулю не превосходило 0,01?

2.17. Производится 500 бросаний симметричной монеты. В каких пределах будет находиться отклонение частоты выпадения герба от 0,5 с вероятностью 0,99?

2.18. Завод отправил на базу 5 тыс. доброкачественных изделий. Вероятность того, что в пути изделие повредится, равна 0,0002. Какова вероятность того, что на базу прибудут 3 поврежденных изделия?

2.19. При введении вакцины иммунитет создается в 99,99 % случаев. Какова вероятность того, что из 1 000 вакцинированных детей иммунитет не выработался: а) у 1 ребенка; б) у 2 детей; в) у 3 детей?

2.20. Магазин получил 1 000 бутылок минеральной воды. Вероятность того, что бутылка не заполнена до конца, равна 0,003. Найти вероятность того, что магазин получит: а) хотя бы одну неполную бутылку; б) менее двух неполных бутылок; в) более двух неполных бутылок.

Ответы к задачам для самостоятельного решения

2.1. $P = 0{,}2048$.

2.2. $P = 160/729$.

2.3. $P = 0{,}91854$.

2.4. При одном бросании из двух.

2.5. $P = 0{,}8192$.

2.6. $P = 0{,}9298$.

2.7. $m_0 = 28$.

2.8. $119 \le n \le 125$.

2.9. $P = 0{,}103$.

2.10. а) $P \approx 0{,}00146$; б) $P \approx 0{,}0236$.

2.11. $m = 55$.

2.12. $P = 0{,}5$.

2.13. $P = 0{,}9736$.

2.14. $P = 0{,}9651$.

2.15. $n = 169$.

2.16. $n = 11\,171$.

2.17. $\left|\dfrac{m}{n} - p\right| \le 0{,}057$.

2.18. $P \approx 0{,}0613$.

2.19. а) $P = 0{,}3679$; б) $P = 0{,}1839$; в) $P = 0{,}0613$.

2.20. а) $P = 0{,}95$; б) $P = 0{,}1992$; в) $P = 0{,}577$.

3. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

3.1. Определение случайной величины и способы ее задания

Случайной величиной $X = X(\omega)$ называется функция $X\colon \Omega \to \mathbf{R}$ элементарного события ω с областью определения Ω $(\omega \in \Omega)$ и областью значений $\mathbf{R}$.

Иначе, случайная величина — это числовая функция $X(\omega)$, описывающая результаты эксперимента и определенная на пространстве элементарных событий.

Значения x функции $X(\omega)$ называются *реализациями случайной величины* $X(\omega)$.

Случайные величины обозначают как греческими буквами $\xi, \eta, \zeta, \ldots$, так и прописными латинскими буквами X, Y, Z, …, а их значения (реализации) — соответствующими строчными буквами x, y, z, …

Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Закон распределения может быть задан с помощью *таблицы*, *графически* или *аналитически*.

Различают *дискретные* и *непрерывные* случайные величины.

Дискретной случайной величиной называется случайная величина $X(\omega)$, которая в зависимости от элементарных исходов ω принимает конечное или бесконечное счетное число изолированных значений (т. е. их можно перенумеровать натуральными числами). Например:

1) если опыт состоит в подбрасывании двух монет, а элементарным событием ω является положение упавших монет, то число выпавших «гербов» есть случайная величина $X(\omega)$ с конечным числом возможных значений $\{0, 1, 2\}$;

2) если случайная величина $X(\omega)$ — число родившихся мальчиков в ближайший день в городе, то ее возможные значения: 0, 1, 2, 3, …

Дискретная случайная величина характеризуется значениями $x_1, x_2, \ldots, x_n, \ldots$, которые она может принимать, а также вероятностями $p_i = \mathrm{P}(x_i)$ $(i = 1, 2, \ldots)$, с которыми она принимает эти значения.

Для *дискретной случайной величины* простейшей формой закона распределения является *ряд распределения*, $p_i = \mathrm{P}(X = x_i)$ $(i = 1, 2, \ldots)$, который задается таблицей (табл. 3.1).

Таблица 3.1

X	x_1	x_2	…	x_n	…
P	p_1	p_2	…	p_n	…

Здесь в верхней строке расположены по возрастанию все возможные различные значения $x_1, x_2, \ldots, x_n, \ldots$ дискретной случайной величины $X(\omega)$, а в нижней — соответствующие им вероятности $p_1, p_2, \ldots, p_n, \ldots$ Так как события $\{X = x_i\}$ $(i = 1, 2, \ldots)$ попарно несовместны и образуют полную группу, то $p_1 + p_2 + \ldots + p_n + \ldots = 1$.

В случае графического задания дискретной случайной величины по оси абсцисс откладывают значения случайной величины, а по оси ординат — соответствующие им вероятности. Ломаная линия, соединяющая точки (x_i, p_i), называется *многоугольником распределения* (рис. 3.1).

Пример 3.1. В урне 5 белых и 20 черных шаров. Вынули один шар, отметили его цвет и вернули в урну. Затем еще раз вынули один шар. Случайная величина $X(\omega)$ — число вынутых белых шаров. Построить ряд и многоугольник распределения.

Решение. Вероятность p появления белого шара при каждом извлечении $p = \dfrac{1}{5}$, следовательно, вероятность появления черного шара при каждом извлечении равна: $q = 1 - \dfrac{1}{5} = \dfrac{4}{5}$.

Случайная величина X — число белых из двух вынутых шаров — может принимать значения $x_1 = 0$, $x_2 = 1$, $x_3 = 2$. Событие $X = x_1 = 0$ состоит в том, что среди вынутых двух шаров нет белых, событие $X = x_2 = 1$ — среди вынутых двух шаров один белый, наконец, событие $X = x_3 = 2$ — среди вынутых двух шаров два белых. Испытания удовлетворяют схеме Бернулли (извлечение первого и извлечение второ-

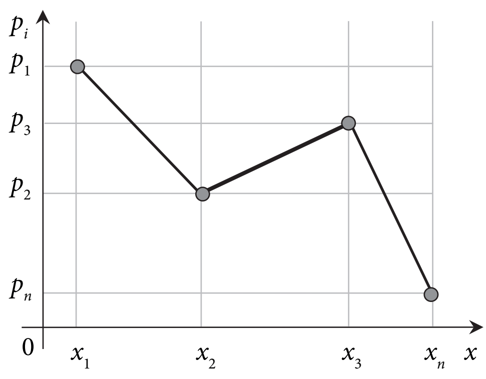


Рис. 3.1. Многоугольник распределения

го шара — события независимые). Поэтому вероятность того, что среди двух вынутых шаров нет белых, определим по формуле Бернулли:

$$P(X=0)=P_2(0)=C_2^0p^0q^2=1\cdot 1\cdot\left(\frac{4}{5}\right)^2=\frac{16}{25};$$

$$P(X=1)=P_2(1)=C_2^1p^1q^1=2\cdot\frac{1}{5}\cdot\frac{4}{5}=\frac{8}{25};$$

$$P(X=2)=P_2(2)=C_2^2p^2q^0=1\cdot\left(\frac{1}{5}\right)^2\cdot 1=\frac{1}{25}.$$

Закон распределения числа вынутых белых шаров имеет вид:

X	0	1	2	Σ
P	16/25	8/25	1/25	1

Построим многоугольник распределения (рис. 3.2).

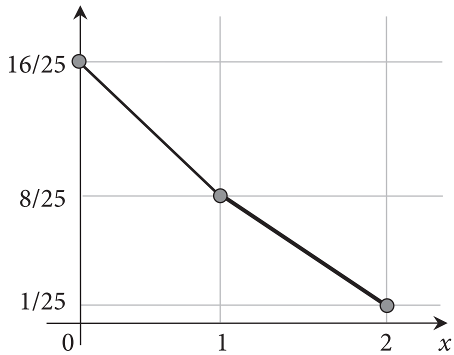


Рис. 3.2. Многоугольник распределения к примеру 3.2

Пример 3.2. В магазине имеется 15 автомобилей определенной марки. Среди них 7 черного цвета, 6 серого и 2 белого. Представители фирмы обратились в магазин с предложением о продаже им 3 автомобилей этой марки, безразлично какого цвета. Составить ряд распределения числа проданных автомобилей черного цвета при условии, что автомобили отбирались случайно.

Решение. Пусть X — число проданных автомобилей черного цвета. Случайная величина X может принимать значения 0, 1, 2 и 3. Найдем соответствующие вероятности по классическому определению вероятности. Всего способов выбрать 3 любых автомобиля из 15 равно:

$$n=C_{15}^3=\frac{15!}{3!\cdot 12!}=\frac{13\cdot 14\cdot 15}{1\cdot 2\cdot 3}=455.$$

Событие $X = x_1 = 0$ состоит в том, что все проданные автомобили не черные, таких 8 штук, поэтому

$$\mathrm{P}\left(X = 0\right) = \frac{C_7^0 \cdot C_8^3}{C_{15}^3} = \frac{56}{455} = \frac{8}{65}.$$

Событие $X = x_2 = 1$ — один автомобиль черный (из 7) и еще два — не черные (из 8 остальных), тогда

$$\mathrm{P}\left(X = 1\right) = \frac{C_7^1 \cdot C_8^2}{C_{15}^3} = \frac{7 \cdot 28}{455} = \frac{28}{65}.$$

Событие $X = x_3 = 2$ — два автомобиля черные (из 7) и еще один — не черный (из 8 остальных), поэтому

$$\mathrm{P}\left(X = 2\right) = \frac{C_7^2 \cdot C_8^1}{C_{15}^3} = \frac{21 \cdot 8}{455} = \frac{24}{65}.$$

Наконец, событие $X = x_4 = 3$ — все проданные автомобили черного цвета, вероятность этого равна:

$$\mathrm{P}\left(X = 3\right) = \frac{C_7^3 \cdot C_8^0}{C_{15}^3} = \frac{35 \cdot 1}{455} = \frac{1}{13}.$$

Таким образом, закон распределения числа проданных автомобилей черного цвета имеет вид:

X	0	1	2	3	Σ
P	8/65	28/65	24/65	1/13	1

При этом $\sum\limits_{i=1}^{n} p_i = 8/65 + 28/65 + 24/65 + 5/65 = 1.$

Пример 3.3. Идет стрельба по мишени до первого промаха с вероятностью попадания 0,7. Составить закон распределения случайной величины X — числа выстрелов.

X	1	2	3	…	n	…	Σ
P	0,3	$0{,}7 \cdot 0{,}3$	$0{,}7^2 \cdot 0{,}3$	…	$0{,}7^{n-1} \cdot 0{,}3$	…	1

$$\sum_{i=1}^{\infty} p_i = 0{,}3 + 0{,}7 \cdot 0{,}3 + 0{,}7^2 \cdot 0{,}3 + \ldots + 0{,}7^{n-1} \cdot 0{,}3 + \ldots =$$

$$= 0{,}3 \cdot \left(1 + 0{,}7 + 0{,}7^2 + 0{,}7^3 + \ldots + 0{,}7^n + \ldots\right) = \frac{0{,}3 \cdot 1}{1 - 0{,}7} = 1.$$

3.2. Функция распределения

Функция $F_X(x) = \mathrm{P}\{\omega\colon X(\omega) < x\}$, определенная для всех $x \in \mathbf{R}$, называется *функцией распределения вероятностей случайной величины* $X(\omega)$.

Далее, если ясно, о какой случайной величине идет речь, будем писать кратко:

$$F(x) = \mathrm{P}\{X < x\}.$$

Под $\{X < x\}$ (или $(X < x)$) понимается событие $\{\omega\colon X(\omega) < x\}$, состоящее в том, что случайная величина примет значение меньшее, чем заданное число x.

Функция распределения является одной из форм закона распределения для случайных величин всех типов и однозначно определяет случайную величину.

Как числовая функция от числового аргумента x, функция распределения $F(x)$ обладает следующими свойствами:

1) $F(x)$ определена для всех $x \in \mathbf{R}$;

2) функция распределения может принимать любые значения от 0 до 1, так как по определению является вероятностью: $0 \le F(x) \le 1$;

3) $F(x)$ неубывающая функция, т. е. если $x_1 < x_2$, то $F(x_1) \le F(x_2)$;

4) $\mathrm{P}\{a \le X < b\} = F(b) - F(a)$, если $a < b$.

Если X — дискретная случайная величина, то

$$\mathrm{P}\{a \le X \le b\} = F(b) - F(a) + \mathrm{P}\{X = b\};$$

5) $\mathrm{P}\{X \ge x\} = 1 - \mathrm{P}\{X < x\} = 1 - F(x)$;

6) функция распределения непрерывна слева, т. е. $\lim\limits_{\varepsilon \to +0} F(x - \varepsilon) = F(x)$.

Если все возможные значения случайной величины $X = X(\omega)$ принадлежат промежутку $[a; b]$, то при $x \le a$ $F(x) = 0$, при $x > b$ $F(x) = 1$;

7) если все возможные значения случайной величины $X = X(\omega)$ расположены на всей числовой прямой ($x \in \mathbf{R}$), то

$$\mathrm{F}(-\infty) = \lim_{x \to -\infty} \mathrm{F}(x) = \lim_{x \to -\infty} \mathrm{P}\{X < x\} = \mathrm{P}\{X < -\infty\} = \mathrm{P}(\varnothing) = 0;$$

$$\mathrm{F}(+\infty) = \lim_{x \to +\infty} \mathrm{F}(x) = \lim_{x \to +\infty} \mathrm{P}\{X < x\} = \mathrm{P}\{X < +\infty\} = \mathrm{P}(\Omega) = 1.$$

Рассмотрим функцию распределения $F(x)$ для *дискретной случайной величины* $X = X(\omega)$.

Если $x \le x_1$, то $F(x) = \mathrm{P}\{X < x\} = \mathrm{P}\{X < x_1\} = \mathrm{P}(\varnothing) = 0$ (значений меньших, чем x_1, случайная величина X не принимает, поэтому событие $\{X < x_1\}$ в этом случае невозможно).

Если $x_1 < X \le x_2$, то событие $\{X < x\}$ наступит тогда и только тогда, когда наступит событие $\{X = x_1\}$. Поэтому $F(x) = \mathrm{P}\{X < x\} = \mathrm{P}\{X < x_1\} = p_1$.

Если $x_2 < X \le x_3$, то событие $\{X < x\} = \{X = x_1\} + \{X = x_2\}$ и $F(x) = \mathrm{P}\{X < x\} = \mathrm{P}\{X = x_1\} + \mathrm{P}\{X = x_2\} = p_1 + p_2$.

Аналогично, если $x_i < X \le x_{i+1}$, то $F(x) = p_1 + p_2 + \ldots + p_i$. Таким образом, функция распределения дискретной случайной величины равна:

$$F(x) = \sum_{i:\, x_i < x} p_i,$$

где $p_i = \mathrm{P}\{X = x_i\}$ и суммирование проводится по тем индексам i, для которых $x_i < x$. Функция распределения дискретной случайной величины является разрывной ступенчатой функцией, причем в точках разрыва $F(x)$ величины скачков равны вероятностям p_i соответствующих реализаций x_i случайной величины X. Сумма всех скачков равна 1 (рис. 3.3).

Пример 3.4. Найти функцию распределения $F(x)$ для случайной величины X из примера 3.1.

Возможные значения случайной величины X: 0, 1, 2. Ясно, что если $x \le 0$, то $F(x) = 0$, так как значений меньших, чем 0, случайная величина X не принимает. Если $0 < x \le 1$, то $F(x) = \mathrm{P}\{X < x\} = \mathrm{P}\{X = 0\} = \frac{16}{25}$. Если $1 < x \le 2$, то $F(x) = \mathrm{P}\{X < x\} = \mathrm{P}\{X = 0\} + \mathrm{P}\{X = 1\} = \frac{16}{25} + \frac{8}{25} = \frac{24}{25}$. Если $x > 2$, то в этом случае любые возможные значения случайной величины 0, 1, 2 менее x и $F(x) = 1$.

Итак:

$$F(X) = \begin{cases} 0,\ x \le 0, \\ 16/25,\ 0 < x \le 1, \\ 24/25,\ 1 < x \le 2, \\ 1,\ x > 2. \end{cases}$$

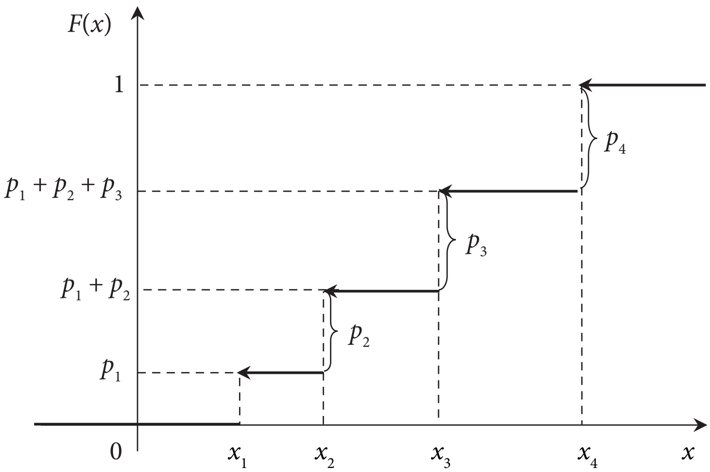


Рис. 3.3. Функция распределения дискретной случайной величины

График функции распределения изображен на рис. 3.4.

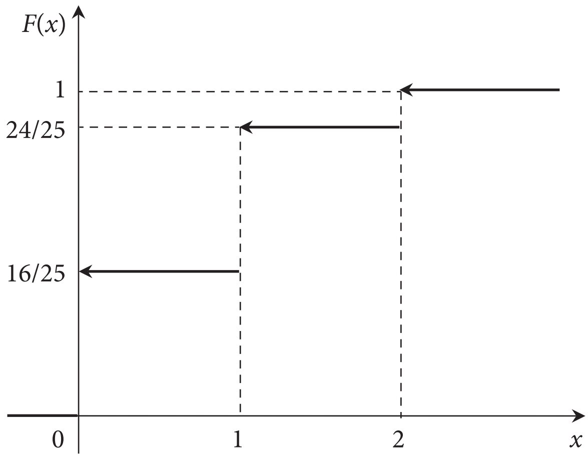


Рис. 3.4. График функции распределения для случайной величины из примера 3.1

3.3. Непрерывные случайные величины

Случайная величина называется *непрерывной*, если ее значения непрерывно заполняют конечный или бесконечный промежуток числовой оси с *плотностью распределения* $f(x)$, связанной с функцией распределения вероятностей равенством

$$F(x) = \int_{-\infty}^{x} f(t)dt. \tag{3.1}$$

Ясно, что функция распределения непрерывной случайной величины есть непрерывная, кусочно-дифференцируемая функция с кусочно-непрерывной неотрицательной производной:

$$F'(x) = f(x). \tag{3.2}$$

Плотность распределения $f(x)$ называют еще *дифференциальной функцией* (см. (3.2)), а функцию распределения, являющуюся первообразной для плотности (см. (3.1)), называют *интегральной функцией распределения*.

График плотности $f(x)$ непрерывной случайной величины X называют *кривой распределения*.

Теорема. Вероятность любого отдельно взятого значения непрерывной случайной величины равна нулю.

Поэтому для *непрерывной* случайной величины

$$\mathrm{P}\{a \le X < b\} = \mathrm{P}\{a \le X \le b\} = \mathrm{P}\{a < X < b\} = \mathrm{P}\{a < X \le b\} = F(b) - F(a). \quad (3.3)$$

С в о й с т в а плотности распределения вероятностей:

1) $f(x) \ge 0$, так как $f(x) = F'(x)$ и $F(x)$ — неубывающая функция;

2) $F'(x) = \lim\limits_{\Delta x \to 0} \dfrac{F(x + \Delta x) - F(x)}{\Delta x} = \lim\limits_{\Delta x \to 0} \dfrac{\mathrm{P}(x \le X < x + \Delta x)}{\Delta x}$, т. е. производная функции распределения $F(x)$ является пределом средней вероятности, приходящейся на единицу длины рассматриваемого участка, и характеризует плотность, с которой распределяются значения непрерывной случайной величины в данной точке. Поэтому функция $f(x) = F'(x)$ и была названа функцией плотности распределения вероятностей;

3) вероятность попадания случайной величины X на отрезок $[a; b]$ вычисляется по формуле

$$\mathrm{P}\{a \le X \le b\} = \int_a^b f(x)dx. \quad (3.4)$$

Геометрически это означает, что вероятность того, что непрерывная случайная величина примет значение, принадлежащее $[a; b]$, равна площади криволинейной трапеции, ограниченной осью Ox, графиком функции $f(x)$ и прямыми $x = a$, $x = b$ (рис. 3.5);

4) функция распределения непрерывной случайной величины $F(x)$ выражается через плотность распределения вероятности $f(x)$ формулой $F(x) = \int_{-\infty}^{x} f(t)dt$ (рис. 3.6);

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

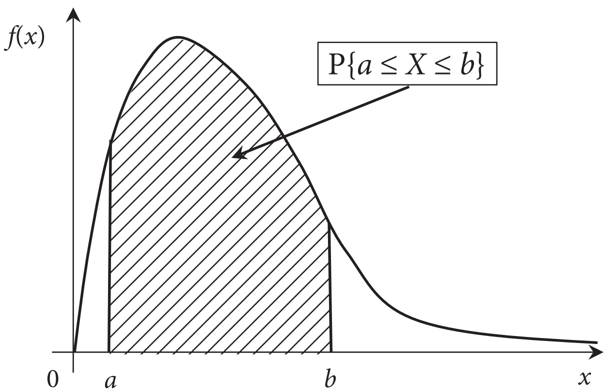


Рис. 3.5. Вероятность попадания непрерывной случайной величины X в интервал

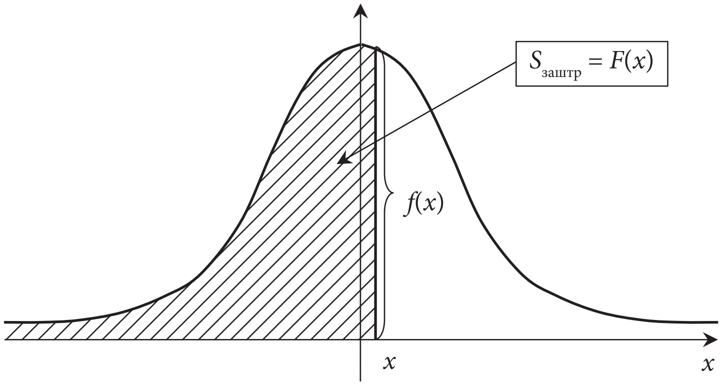


Рис. 3.6. Вычисление функции распределения непрерывной случайной величины через ее функцию плотности вероятности

Геометрическая интерпретация этого равенства: площадь области, ограниченной кривой распределения и осью абсцисс, равна 1 (рис. 3.7).

Следствие. Если все возможные значения случайной величины принадлежат $[a; b]$, то $\int\limits_a^b f(x)dx = 1.$

Пример 3.5. Непрерывная случайная величина $X = X(\omega)$ задана плотностью

$$f(x) = \begin{cases} 0, & x < a, \\ C, & a \le x \le b, \\ 0, & x > b. \end{cases}$$

Найти постоянную C.

Решение. Учитывая свойство плотности вероятности $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$, находим:

$$\int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_{-\infty}^{a} 0 \cdot dx + \int\limits_{a}^{b} C \cdot dx + \int\limits_{b}^{+\infty} 0 \cdot dx = \int\limits_{a}^{b} C \cdot dx = C\big(b - a\big) = 1.$$

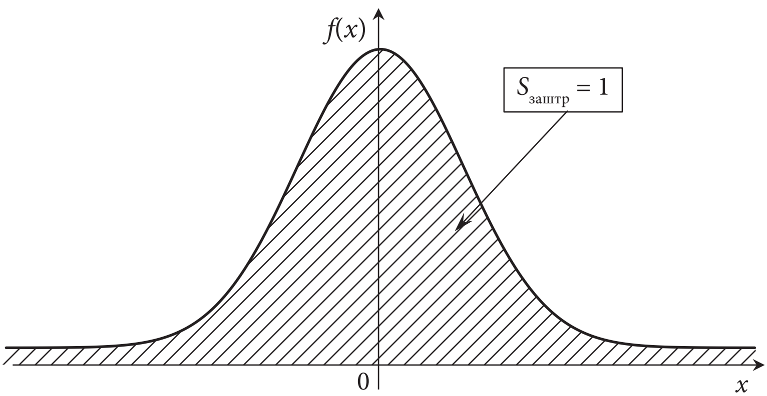


Рис. 3.7. Геометрическая интерпретация свойства 5

Отсюда:

$$C = \frac{1}{b-a} \quad \text{и} \quad f(x) = \begin{cases} 0, & x < a, \\ \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & x > b. \end{cases}$$

Найдем *функцию распределения* $F(x)$ для случайной величины, распределенной по *равномерному закону*. Для этого воспользуемся формулой, связывающей плотность $f(x)$ с функцией распределения $F(x)$:

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

Если $x < a$, то $F(x) = \int_{-\infty}^{x} 0 \cdot dt = 0.$

Для значений $a \le x \le b$ $\quad F(x) = \int_{-\infty}^{a} 0 \cdot dt + \int_{a}^{x} \frac{1}{b-a} dt = \frac{1}{b-a} t \Big|_a^x = \frac{x-a}{b-a}.$

При $x > b$ $\quad F(x) = \int_{-\infty}^{a} 0 \cdot dt + \int_{a}^{b} \frac{1}{b-a} dt + \int_{b}^{x} 0 \cdot dt = \frac{1}{b-a} t \Big|_a^b = \frac{b-a}{b-a} = 1.$

Таким образом, $F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \le x \le b, \\ 1, & x > b. \end{cases}$

Пример 3.6. Случайная величина задана функцией распределения вероятностей (рис. 3.8):

$$F(x) = \begin{cases} 0, & x < -1, \\ a + b \cdot \arcsin x, & -1 \le x \le 1, \\ 1, & x > 1. \end{cases}$$

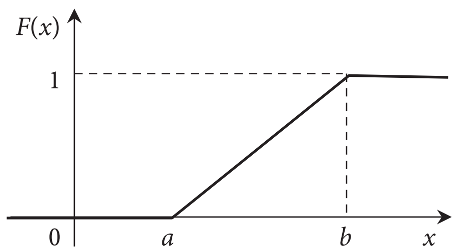


Рис. 3.8. Функция распределения вероятностей из примера 3.4

Найти:

а) неизвестные параметры a, b;

б) плотность распределения вероятностей $f(x)$;

в) вероятность попадания случайной величины в интервал $\left(-\frac{1}{2};\frac{1}{2}\right)$.

Решение. Так как функция $F(x)$ непрерывна, то $F(-1-0)=F(-1+0)=F(-1)$:

$\lim\limits_{x\to-1+0}\left(a+b\arcsin x\right)=a-b\cdot\frac{\pi}{2}=0.$ Аналогично, $F(1-0)=F(1+0)=F(1)$:

$\lim\limits_{x\to1-0}\left(a+b\arcsin x\right)=a+b\cdot\frac{\pi}{2}=1.$ Решая систему уравнений $\begin{cases} a-b\cdot\frac{\pi}{2}=0, \\ a+b\cdot\frac{\pi}{2}=1, \end{cases}$ находим:

$a=\frac{1}{2},\ b=\frac{1}{\pi}.$ Таким образом:

$$F(x)=\begin{cases} 0, & x<-1, \\ \frac{1}{2}+\frac{1}{\pi}\arcsin x, & -1\le x\le 1, \\ 1, & x>1. \end{cases}$$

Так как $f(x)=F'(x),$ то

$$f(x)=\begin{cases} 0, & x<-1 \\ \frac{1}{\pi\sqrt{1-x^2}}, & -1<x<1, \\ 0, & x>1. \end{cases}$$

$$\mathrm{P}\left(-\frac{1}{2}<X<\frac{1}{2}\right)=F\left(\frac{1}{2}\right)-F\left(-\frac{1}{2}\right)=$$
$$=\left(\frac{1}{2}+\frac{1}{\pi}\arcsin\frac{1}{2}\right)-\left(\frac{1}{2}+\frac{1}{\pi}\arcsin\left(-\frac{1}{2}\right)\right)=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}.$$

3.4. Функции от случайных величин

Пусть $X=X(\omega)$ — случайная величина, определенная на множестве элементарных событий Ω; Σ — множество возможных значений случайной величины X; f — некоторая функция, определенная на множестве Σ. Тогда $Y=f(X)=f\left(X(\omega)\right)=Y(\omega)$ — случайная величина: каждому $\omega\in\Omega$ ставит в соответствие число $Y(\omega)=f\big(X(\omega)\big)$. Функция f называется функцией случайного аргумента.

Аналогично вводится функция двух или более случайных аргументов.

Пусть в общем случае $X_1=X_1(\omega),\ldots,X_n=X_n(\omega)$ — случайные величины, определенные на множестве элементарных событий Ω; $f(x)=f\big(x_1,\ldots,x_n\big)$ — некоторая функция, определенная на Σ. Тогда $Y=f(\overline{X})=f\big(X_1,\ldots,X_n\big)$ — случайная величина: каждому $\omega\in\Omega$ ставит в соответствие число $Y(\omega)=f\big(\overline{X}(\omega)\big)$.

Например, если $f(X)=X_1+X_2$, то новую случайную величину X_1+X_2 называют суммой случайных величин X_1 и X_2.

Если $X=X(\omega)$ — дискретная случайная величина, то и $Y=\varphi(X)$ — дискретная случайная величина с законом распределения (табл. 3.2).

Таблица 3.2

Y	$\varphi(X_1)$	$\varphi(X_2)$	…	$\varphi(X_n)$
P	p_1	p_2	…	p_n

Где все $\varphi(X_1)$, $i=1,2,\ldots,n$ различны. Если какие-то два (или более) значения $\varphi(X_i)$ и $\varphi(X_j)$ $(i\neq j)$ равны, то эти столбцы объединяются и соответствующие вероятности складываются: p_i+p_j.

Пример 3.7. Даны законы распределения независимых случайных величин X и Y:

X	–3	0	1
P	0,1	0,3	0,6

,

Y	0	3
P	0,2	0,8

Составить законы распределения случайных величин:

а) Y^3;

б) $X+Y$.

Решение. а) $P(Y^3=0)=P(Y=0)=0{,}2$, $P(Y^3=27)=P(Y=3)=0{,}8$. Поэтому закон распределения случайной величины Y^3:

Y^3	0	27
P	0,2	0,8

б) Случайная величина $X+Y$ принимает значения: –3, 0, 1, 3, 4. Найдем соответствующие им вероятности:

$$P\big(X+Y=-3\big)=P\big(X=-3,\ Y=0\big)=P\big(X=-3\big)P\big(Y=0\big)=0{,}1\cdot 0{,}2=0{,}02;$$

$$P\big(X+Y=0\big)=P\big(X=-3,\ Y=3\big)+P\big(X=0,\ Y=0\big)=0{,}1\cdot 0{,}8+0{,}3\cdot 0{,}2=0{,}14;$$

$$P\big(X+Y=1\big)=P\big(X=1,\ Y=0\big)=0{,}6\cdot 0{,}2=0{,}12;$$

$$P\big(X+Y=3\big)=P\big(X=0,\ Y=3\big)=0{,}3\cdot 0{,}8=0{,}24;$$

$$P(X+Y=4)=P(X=1,\ Y=3)=0,6\cdot 0,8=0,48.$$

Таким образом, случайная величина $X+Y$ имеет ряд распределения, указанный в табл. 3.3.

Таблица 3.3

X	–3	0	1	3	4	$\sum$
P	0,02	0,14	0,12	0,24	0,48	1

Заметим, что событие $X+Y=\{x_i\}+\{y_j\}$ равносильно событию $X+Y=\{X=x_i, Y=Y_i\}$, а $P(X+Y)=P(X=x_i)P(Y=y_j/X=x_i)$.

Рассмотрим непрерывную случайную величину

$$Y=\varphi(X)$$

(называемую *функцией от непрерывной случайной величины* X), где $y=\varphi(x)$ — строго монотонная дифференцируемая функция, а X — непрерывная случайная величина с плотностью распределения $f(x)=f_X(x)$. Тогда *плотность распределения вероятностей* $g(y)=f_Y(y)$ непрерывной случайной величины Y имеет вид:

$$g(y)=f\left(\psi(y)\right)\left|\psi'(y)\right|,$$

где $\psi(y)=\varphi^{-1}(y)$ — обратная по отношению к $\varphi(x)$ функция.

Дискретные случайные величины X, Y, Z называют *независимыми*, если события $\{X=x_i\}$, $\{Y=y_j\}$, $\{Z=z_k\}$ независимы для $\forall i,\ j, k$, т. е.

$$P\left(X=x_i, Y=y_j, \ldots, Z=z_k\right)=P\left(X=x_i\right)P\left(Y=y_j\right)\ldots P\left(Z=z_k\right).$$

Если случайные величины $X_1, X_2, \ldots, X_n$ *независимы*, то *независимы* также случайные величины $f(X_1), f(X_2), \ldots, f(X_n)$.

3.5. Числовые характеристики случайных величин

Важнейшей числовой характеристикой центра группирования значений случайной величины является математическое ожидание. Математическим ожиданием дискретной случайной величины X называется число

$$E(X)=\sum_{i=1}^{n} p_i x_i,$$

где $p_i=P(X=x_i)$. Предполагается, что ряд $\sum\limits_{i=1}^{\infty} p_i\left|x_i\right|$ сходится.

Для математического ожидания используют обозначения:

$$\mathrm{M}(X), \mathrm{M}X, \mathrm{E}(X), \mathrm{E}X, m_X.$$

Пусть плотность $f(x)$ непрерывной случайной величины X такова, что сходится интеграл $\int\limits_{-\infty}^{+\infty}|x|f(x)dx,$ тогда число

$$\mathrm{E}(X)=\int\limits_{-\infty}^{+\infty}xf(x)dx \tag{3.5}$$

называется *математическим ожиданием непрерывной случайной величины* X.

Для случайной величины $Y=\varphi(X)$, где X — непрерывная случайная величина ($y=\varphi(x)$ — строго монотонная дифференцируемая функция), математическое ожидание случайной величины Y вычисляется следующим образом:

$$\mathrm{E}\big(\varphi(x)\big)=\int\limits_{-\infty}^{+\infty}\varphi(x)f(x)dx,$$

если указанный интеграл абсолютно сходится.

Если X — дискретная случайная величина, то $\mathrm{E}\big(\varphi(X)\big)=\sum\limits_{i=1}^{\infty}\varphi(x_i)p_i.$

С в о й с т в а математического ожидания:

1) размерность случайной величины X совпадает с размерностью $\mathrm{E}(X)$. Математическое ожидание называют *центром распределения* или *средним значением случайной величины* X;

2) для *дискретной случайной величины* $\mathrm{E}(X)=\dfrac{x_1p_1+\ldots+x_np_n}{p_1+\ldots+p_n}$ — центр тяжести системы материальных точек с координатами $x_1, \ldots, x_n$, расположенными на оси Ox, в которых сосредоточены массы $p_1, \ldots, p_n$ соответственно (*механико-вероятностный смысл* математического ожидания);

3) для *непрерывной случайной величины* $\mathrm{E}(X)$ — центр тяжести стержня, масса которого равна 1, а плотность определяется функцией $f(x)$;

4) $\mathrm{E}(C)=C$;

5) $\mathrm{E}(CX)=C\mathrm{E}(\mathrm{X})$, $C=\mathrm{const}$;

6) $\mathrm{E}(X_1+\ldots+X_n)=\mathrm{E}(X_1)+\ldots+\mathrm{E}(X_n) \quad \forall\, X_1,\ldots,X_n$;

7) если $X_1, \ldots, X_n$ независимы, то $\mathrm{E}(X_1\ldots X_n)=\mathrm{E}(X_1)\ldots\mathrm{E}(X_n)$.

Следствие. $\mathrm{E}(C_1X_1+\ldots+C_nX_n)=C_1\mathrm{E}(X_1)+\ldots+C_n\mathrm{E}(X_n) \quad \forall X_1,\ldots,X_n.$

Пример 3.8. Найти математическое ожидание случайной величины X, равной числу выпадений «герба» при подбрасывании двух монет.

Случайная величина X принимает три значения: 0, 1, 2. Найдем соответствующие им вероятности:

$$\mathrm{P}(X=0)=\mathrm{P}_2(0)=C_2^0p^0q^2=1\cdot 1\cdot\frac{1}{4}=\frac{1}{4};$$

$$\mathrm{P}(X=1)=\mathrm{P}_2(1)=C_2^1p^1q^1=2\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2};$$

$$\mathrm{P}(X=2)=\mathrm{P}_2(2)=C_2^2p^2q^0=1\cdot\frac{1}{4}\cdot 1=\frac{1}{4}.$$

Ряд распределения случайной величины X имеет вид:

X	0	1	2
P	1/4	1/2	1/4

Математическое ожидание $\mathrm{E}(X)=0\cdot\frac{1}{4}+1\cdot\frac{1}{2}+2\cdot\frac{1}{4}=1.$

Пример 3.9. Найти среднее значение случайной величины X, имеющей равномерное распределение на отрезке $[a, b]$:

$$f(x)=\begin{cases} 0, & x<a,\\ \dfrac{1}{b-a}, & a\le x\le b,\\ 0, & x>b.\end{cases}$$

Решение. Воспользуемся формулой (3.5):

$$\mathrm{E}(X)=\int_{-\infty}^{a}x\cdot 0\cdot dx+\int_{a}^{b}x\cdot\frac{1}{b-a}dx+\int_{b}^{+\infty}x\cdot 0\cdot dx=\frac{1}{b-a}\cdot\left.\frac{x^2}{2}\right|_a^b=\frac{1}{2}\cdot\frac{b^2-a^2}{b-a}=\frac{1}{2}(a+b).$$

Итак, среднее значение случайной величины X, имеющей равномерное распределение, совпадает с серединой отрезка $[a, b]$ (рис. 3.9).

Дисперсия случайной величины

Во многих случаях даже для самого краткого описания случайной величины недостаточно указать ее центр распределения.

Например, найдем математические ожидания случайных величин X и Y, заданных соответственно законами распределения:

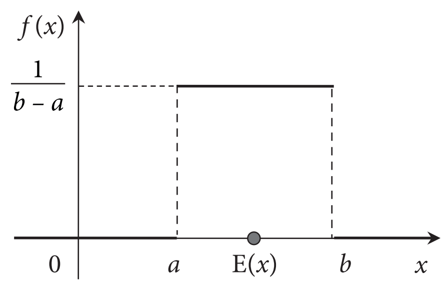


Рис. 3.9. Математическое ожидание равномерного распределения

X	–0,1	0,2
P	2/3	1/3

,

Y	–100	100
P	1/2	1/2

$$\mathrm{E}(X) = -0{,}1 \cdot \frac{2}{3} + 0{,}2 \cdot \frac{1}{3} = 0 \quad \text{и} \quad \mathrm{E}(Y) = -100 \cdot \frac{1}{2} + 100 \cdot \frac{1}{2} = 0.$$

Мы видим, что $\mathrm{E}(X) = \mathrm{E}(Y) = 0$. Но значения случайной величины X мало отклоняются от своего центра, а значения случайной величины Y имеют довольно большие отклонения.

Представление о *степени рассеивания* реализаций (значений) случайной величины относительно ее центра (математического ожидания) дает дисперсия случайной величины X.

Дисперсией случайной величины X называют число $\mathrm{D}(X)$, равное математическому ожиданию квадрата отклонения случайной величины X от своего математического ожидания:

$$\mathrm{D}(X) = \mathrm{E}\left(X - \mathrm{E}(X)\right)^2. \tag{3.6}$$

Для дисперсии используют обозначения:

$$\mathrm{D}(X),\, \mathrm{D}X,\, \mathrm{Var}\,(X),\, \mathrm{Var}\,X,\, d_X.$$

Дисперсия вычисляется по формуле

$$\mathrm{D}(X) = \begin{cases} \sum\limits_{i=1}^{\infty} \left[x_i - \mathrm{E}(X)\right]^2 \cdot p_i & \text{для дискретной случайной величины;} \\ \int\limits_{-\infty}^{+\infty} \left[x - \mathrm{E}(X)\right]^2 \cdot f(x)dx & \text{для непрерывной случайной величины.} \end{cases} \tag{3.7}$$

Дисперсию удобно находить по формуле

$$\mathrm{D}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X). \tag{3.8}$$

С учетом (3.8) формулы (3.7) приводятся к виду:

$$\mathrm{D}(X) = \begin{cases} \sum\limits_{i=1}^{\infty} x_i^{\,2} p_i - \mathrm{E}^2(X) & \text{для дискретной случайной величины;} \\ \int\limits_{-\infty}^{+\infty} x^2 f(x)dx - \mathrm{E}^2(X) & \text{для непрерывной случайной величины.} \end{cases} \tag{3.9}$$

Для случайной величины $Y = \varphi(X)$, где X — непрерывная случайная величина ($y = \varphi(x)$ — строго монотонная дифференцируемая функция), дисперсия случайной величины Y вычисляется следующим образом:

$$\mathrm{D}\big(\varphi(X)\big) = \int\limits_{-\infty}^{+\infty} \big(\varphi(x) - \mathrm{E}(Y)\big)^2 f(x)dx.$$

С в о й с т в а дисперсии:

1) $\mathrm{D}(X) \geq 0$;

2) $\mathrm{D}(C) = 0$;

3) $\mathrm{D}(CX) = C^2\mathrm{D}(X)$;

4) $\mathrm{D}\big(X \pm Y\big) = \mathrm{D}(X) + \mathrm{D}(Y) \pm 2\mathrm{E}\big(X - \mathrm{E}(X)\big) \cdot \mathrm{E}\big(Y - \mathrm{E}(Y)\big)$;

5) если X и Y — независимые случайные величины, то

$$\mathrm{D}\big(X \pm Y\big) = \mathrm{D}(X) + \mathrm{D}(Y).$$

Следствие 1. Если $X_1, \ldots, X_n$ — независимые случайные величины, то

$$\mathrm{D}\big(X_1 + \ldots + X_n\big) = \mathrm{D}\big(X_1\big) + \ldots + \mathrm{D}\big(X_n\big).$$

Следствие 2. $\mathrm{D}\big(X - Y\big) = \mathrm{D}\big(X + (-Y)\big) = \mathrm{D}(X) + (-1)^2\mathrm{D}(Y) = \mathrm{D}(X) + \mathrm{D}(Y)$.

Следствие 3. $\mathrm{D}(X + C) = \mathrm{D}(X) + \mathrm{D}(C) = \mathrm{D}(X) + 0 = \mathrm{D}(X)$.

Дисперсия случайной величины X имеет размерность квадрата случайной величины, что не всегда удобно. Поэтому в качестве меры рассеивания случайной величины X наряду с $\mathrm{D}(X)$ используют среднее квадратичное отклонение

$$\sigma_X = \sigma(X) = \sqrt{\mathrm{D}(X)}.$$

Размерность среднего квадратичного отклонения совпадает с размерностью случайной величины X.

Приведем интерпретацию математического ожидания и дисперсии в финансовом анализе. Если X — доходность некоторого актива (например, акции), то математическое ожидание отражает *среднюю* (*прогнозную*) доходность актива, а дисперсия или среднее квадратичное отклонение — меру отклонения, колебания доходности от ожидаемого среднего значения, т. е. *риск данного актива*.

Определение. Случайная величина $\overset{\circ}{X} = X - m_X$, где $m_X = \mathrm{E}(X)$, называется *центрированной*, а случайная величина $\overset{*}{X} = \dfrac{\overset{\circ}{X}}{\sigma_X} = \dfrac{X - m_X}{\sigma_X}$ — *нормированной*.

Для нормированной случайной величины: $\mathrm{E}\overset{*}{X} = 0$, $\mathrm{D}\overset{*}{X} = 1$.

Пример 3.10. Найти дисперсию случайной величины, определенной в примере 3.8.

Ряд распределения случайной величины X имеет вид:

X	0	1	2
P	1/4	1/2	1/4

Математическое ожидание $\mathrm{E}(X) = 1$.

$$\mathrm{D}(X) = \mathrm{E}(X^2) - \mathrm{E}^2(X) = 0^2 \cdot \frac{1}{4} + 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} - 1^2 = \frac{1}{2}.$$

Пример 3.11. Найти дисперсию случайной величины, имеющей равномерное распределение.

Решение. Воспользуемся формулой

$$\mathrm{D}(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mathrm{E}^2(X) =$$

$$= \int_{-\infty}^{a} x^2 \cdot 0 \cdot dx + \int_{a}^{b} x^2 \cdot \frac{1}{b-a} dx + \int_{b}^{+\infty} x^2 \cdot 0 \cdot dx - \frac{(a+b)^2}{4} = \frac{1}{b-a} \frac{x^3}{3}\bigg|_a^b - \frac{1}{4}\left(a^2 + 2ab + b^2\right) =$$

$$= \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} - \frac{1}{4}\left(a^2 + 2ab + b^2\right) = \frac{(b-a)^2}{12}.$$

Пример 3.12. Функция распределения случайной величины X равна:

$$F(x) = \begin{cases} 0, & x \le 0, \\ ax^4, & 0 < x \le 1, \\ 1, & x > 1. \end{cases}$$

Найти: а) параметр a; б) плотность вероятности $f(x)$; в) $\mathrm{E}(X)$, $\mathrm{D}(X)$ и $\mathrm{E}(X^2 - 1)$;

г) $\mathrm{P}\left(0 < X < \frac{1}{2}\right)$; $\mathrm{P}\left(\left|X - \frac{4}{5}\right| \ge \frac{1}{5}\right)$.

Решение. а) Функция $F(x)$ непрерывна при любом a во всех точках, кроме $x = 1$. Найдем a из условия: $F(1 + 0) = F(1)$, $\lim\limits_{x \to 1+0} F(x) = 1$, $F(1) = a \Rightarrow a = 1$.

б) $F(x) = \begin{cases} 0, & x \le 0, \\ x^4, & 0 < x \le 1, \\ 1, & x > 1. \end{cases}$ Так как $f(x) = F'(x)$, то $f(x) = \begin{cases} 0, & x \le 0, \\ 4x^3, & 0 < x \le 1, \\ 0, & x > 1. \end{cases}$

в) $\mathrm{E}(X) = \int\limits_{-\infty}^{+\infty} x f(x)dx = \int\limits_0^1 4x^4 dx = 4 \cdot \frac{x^5}{5}\bigg|_0^1 = \frac{4}{5}$;

$$\mathrm{D}(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \mathrm{E}^2(X) = \int_{-\infty}^{0} x^2 \cdot 0 \cdot dx + \int_0^1 4x^5 dx + \int_1^{+\infty} x^2 \cdot 0 \cdot dx - \left(\frac{4}{5}\right)^2 =$$

$$= 4\frac{x^6}{6}\bigg|_0^1 - \frac{16}{25} = \frac{2}{3} - \frac{16}{25} = \frac{2}{75};$$

$$\mathrm{E}\left(X^2-1\right)=\mathrm{E}\left(X^2\right)-1=\int_{-\infty}^{+\infty}x^2f(x)dx-1=\int_0^1 4x^5dx-1=4\cdot\left.\frac{x^6}{6}\right|_0^1-1=\frac{2}{3}-1=-\frac{1}{3};$$

$$\mathrm{P}\left(0<X<\frac{1}{2}\right)=F\left(\frac{1}{2}\right)-F(0)=\left(\frac{1}{2}\right)^4-0=\frac{1}{16};$$

$$\mathrm{P}\left(\left|X-\frac{4}{5}\right|\geq\frac{1}{5}\right)=1-\mathrm{P}\left(\left|X-\frac{4}{5}\right|<\frac{1}{5}\right)=1-\mathrm{P}\left(\frac{3}{5}<X<1\right)=$$

$$=1-\left(F(1)-F\left(\frac{3}{5}\right)\right)=1-\left(1^4-\left(\frac{3}{5}\right)^4\right)=1-1+\frac{81}{625}=\frac{81}{625}.$$

Мода дискретной случайной величины X есть ее наиболее вероятное значение $\mathrm{M}_0(X)$. *Мода непрерывной случайной величины* X с плотностью $f(x)$ есть то ее значение, при котором $f(x)$ достигает максимума.

Медиана случайной величины X (обозначается $\mathrm{M}_e(X)$) есть также ее значение x_p, для которого одинаково вероятно, окажется ли случайная величина X меньше x_p или больше x_p, т. е.

$$\mathrm{P}\left(X<x_p\right)=\mathrm{P}\left(X>x_p\right)=\frac{1}{2}.$$

Квантилем уровня p называется число x_p, удовлетворяющее уравнению

$$\mathrm{P}\left(X<x_p\right)=p.$$

Таким образом, x_p является решением уравнения $F(x_p)=p$. В частности, квантиль уровня $p=\frac{1}{2}$ является медианой: $x_{\frac{1}{2}}=\mathrm{M}_e(X)$.

Пример 3.13. Непрерывная случайная величина $X=X(\omega)$ распределена по закону Коши[9]. Ее плотность имеет вид: $f(x)=\frac{1}{\pi}\cdot\frac{a}{x^2+a^2}$, $a>0$. Найти моду, медиану, квантили $x_{0{,}25}$, $x_{0{,}5}$, $x_{0{,}75}$.

Решение. Найдем точку максимума функции $f(x)$:

$$f'(x)=\frac{a}{\pi}\cdot\left(\frac{1}{x^2+a^2}\right)'=-\frac{2a}{\pi}\cdot\frac{x}{\left(x^2+a^2\right)^2}=0 \quad \text{при } x=0.$$

[9] Огюстен Луи Коши (1789–1857) — французский математик и механик, член Парижской академии наук, Лондонского королевского общества, Петербургской академии наук и других академий. Разработал фундамент математического анализа, внес огромный вклад в анализ, алгебру, математическую физику и многие другие области математики; один из основоположников механики сплошных сред. Его имя внесено в список величайших ученых Франции, помещенный на первом этаже Эйфелевой башни.

Причем $x = 0$ является точкой максимума, так как слева от нее функция возрастает, а справа убывает. При $x < 0$ $f'(x) > 0$, а при $x > 0$ $f'(x) < 0$.

Таким образом, $\mathrm{M}_0(X) = 0$.

Найдем медиану $\mathrm{M}_e(X) = x_1$. Прямая $x = x_1$ делит площадь фигуры, ограниченной графиком функции $f(x)$, на две равные части. А поскольку функция $f(x)$ четная $\left(f(-x) = \frac{1}{\pi} \cdot \frac{a}{(-x)^2 + a^2} = \frac{1}{\pi} \cdot \frac{a}{x^2 + a^2} = f(x) \right)$, график функции симметричен относительно оси Oy, следовательно, $\mathrm{M}_e(X) = 0$.

Убедимся, что мы найдем то же значение медианы, если будем искать его по определению. $\mathrm{P}(X < x_1) = \mathrm{P}(X > x_1) = \frac{1}{2}$. Воспользуемся первым равенством:

$$\mathrm{P}(X < x_1) = \int_{-\infty}^{x_1} f(x)dx = \int_{-\infty}^{x_1} \frac{1}{\pi} \cdot \frac{a}{x^2 + a^2} dx = \frac{1}{\pi} \cdot \operatorname{arctg} \frac{x}{a} \bigg|_{-\infty}^{x_1} =$$
$$= \frac{1}{\pi} \cdot \operatorname{arctg} \frac{x_1}{a} - \frac{1}{\pi} \cdot \operatorname{arctg}(-\infty) = \frac{1}{\pi} \cdot \operatorname{arctg} \frac{x_1}{a} + \frac{1}{2} = \frac{1}{2}.$$

Так как $\frac{1}{\pi} \cdot \operatorname{arctg} \frac{x_1}{a} = 0$, то $x_1 = 0$. Таким образом, $\mathrm{M}_e(X) = 0$.

Найдем квантили $x_{0,25}$, $x_{0,5}$, $x_{0,75}$. Поскольку $\mathrm{M}_e(X) = 0$, то $x_{0,5} = 0$.

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{1}{\pi} \cdot \frac{a}{t^2 + a^2} dt = \frac{1}{\pi} \cdot \operatorname{arctg} \frac{t}{a} \bigg|_{-\infty}^{x} =$$
$$= \frac{1}{\pi} \cdot \operatorname{arctg} \frac{x}{a} - \frac{1}{\pi} \cdot \operatorname{arctg}(-\infty) = \frac{1}{\pi} \cdot \operatorname{arctg} \frac{x}{a} + 0,5.$$

Квантиль $x_{0,25}$ является решением уравнения $F(x_{0,25}) = 0,25$.

$$F(x_{0,25}) = \frac{1}{\pi} \cdot \operatorname{arctg} \frac{x_{0,25}}{a} + 0,5 = 0,25 \quad \Rightarrow \quad \operatorname{arctg} \frac{x_{0,25}}{a} = -\frac{\pi}{4} \quad \Rightarrow \quad x_{0,25} = -a.$$

Аналогично

$$F(x_{0,75}) = \frac{1}{\pi} \cdot \operatorname{arctg} \frac{x_{0,75}}{a} + 0,5 = 0,75 \quad \Rightarrow \quad \operatorname{arctg} \frac{x_{0,75}}{a} = \frac{\pi}{4} \quad \Rightarrow \quad x_{0,75} = a.$$

В практических задачах встречаются так называемые *усеченные распределения*, у которых из общего множества значений случайной величины устранены значения, большие или меньшие некоторого порогового уровня C_0. В частности, такое распределение будет в случае, когда рассматривается величина заработной платы работников при условии, что ее значение не может быть менее некоторой заданной величины.

Распределением Парето[10] называется такое распределение, для которого функция и плотность распределения вероятностей имеют вид:

$$F(x)=\begin{cases}0, & x\le C_0,\\ 1-\left(\dfrac{C_0}{x}\right)^{\alpha}, & x\le C_0,\ \alpha>0;\end{cases}$$

$$f(x)=\begin{cases}0, & x\le C_0,\\ \dfrac{\alpha}{C_0}\left(\dfrac{C_0}{x}\right)^{\alpha+1}, & x\le C_0.\end{cases}$$

Числовые характеристики:

$$\mathrm{E}(X)=\frac{\alpha}{\alpha-1}C_0,\quad \mathrm{D}\left(X\right)=\frac{\alpha}{\left(\alpha-1\right)^2\left(\alpha-2\right)}C_0\ ,\quad \alpha>2.$$

Пример 3.14. Годовой доход случайно выбранного налогоплательщика описывается случайной величиной *X* с плотностью распределения

$$f(x)=\begin{cases}0, & x<1,\\ \dfrac{c}{x^4}\ , & x\ge 1.\end{cases}$$

Найти: а) значение параметра *c*, функцию распределения годового дохода; б) средний годовой доход и среднее квадратичное отклонение годового дохода; в) размер годового дохода $x_{\min}$, не ниже которого с вероятностью 0,6 окажется годовой доход случайно выбранного налогоплательщика.

Решение. а) Найдем *c* из свойства 5 (п. 3.3) плотности вероятности:

$$\int_{-\infty}^{+\infty} f(x)dx=\int_{-\infty}^{1}0\cdot dx+\int_{1}^{+\infty}\frac{c}{x^4}dx=c\cdot\int_{1}^{+\infty}x^{-4}dx=-\frac{c}{3x^3}\bigg|_1^{+\infty}=\frac{c}{3}=1.$$

Отсюда $c=3$ и $f(x)=\begin{cases}0, & x<1,\\ \dfrac{3}{x^4}\ , & x\ge 1.\end{cases}$

Для построения функции распределения годового дохода воспользуемся формулой $F(x)=\int_{-\infty}^{x}f(t)dt.$

[10] Вильфредо Федерико Дамасо Парето (1848–1923) — итальянский инженер, экономист и социолог. Один из основоположников теории элит. Разработал теории, названные впоследствии его именем: статистическое парето-распределение и парето-оптимум, широко используемые в экономической теории и иных научных дисциплинах.

Если $x < 1$, то $F(x) = \int\limits_{-\infty}^{x} 0 \cdot dt = 0.$

Если $x \geq 1$, то $F(x) = \int\limits_{-\infty}^{1} 0 \cdot dt + 3 \cdot \int\limits_{1}^{x} t^{-4} dt = -\frac{1}{t^3}\Big|_1^x = -\frac{1}{x^3} + 1.$

Таким образом, $F(x) = \begin{cases} 0, & x < 1, \\ 1 - \dfrac{1}{x^3}, & x \geq 1. \end{cases}$

б) Найдем средний годовой доход и среднее квадратичное отклонение годового дохода:

$$\mathrm{E}(X) = \int\limits_{-\infty}^{+\infty} x f(x) dx = \int\limits_{1}^{+\infty} x \cdot \frac{3}{x^4} dx = \int\limits_{1}^{+\infty} \frac{3}{x^3} dx = 3 \cdot \frac{x^{-2}}{-2}\Big|_1^{+\infty} = \frac{3}{2} = 1{,}5;$$

$$\mathrm{D(X)} = \int\limits_{-\infty}^{+\infty} x^2 f(x) dx - \mathrm{E}^2(X) = \int\limits_{1}^{+\infty} x^2 \cdot \frac{3}{x^4} dx - \left(\frac{3}{2}\right)^2 = \int\limits_{1}^{+\infty} \frac{3}{x^2} dx - \frac{9}{4} = -\frac{3}{x}\Big|_1^{+\infty} - \frac{9}{4} = 0{,}75;$$

$$\sigma_x = \sqrt{\mathrm{D}(X)} = \sqrt{0{,}75} \approx 0{,}87.$$

в) Размер годового дохода $x_{\min}$, не ниже которого с вероятностью 0,6 окажется годовой доход случайно выбранного налогоплательщика, равен:

$$\mathrm{P}\left(X \geq x_{\min}\right) = 1 - F\left(x_{\min}\right) = 0{,}6 \quad \Rightarrow \quad F\left(x_{\min}\right) = 1 - 0{,}6 = 0{,}4 \quad \Rightarrow$$

$$\Rightarrow x_{\min}^3 = 1/0{,}6 = \frac{10}{6} = \frac{5}{3} \approx 1{,}(6) \quad \Rightarrow \quad x_{\min} = \sqrt[3]{\frac{5}{3}} \approx 1{,}185.$$

3.6. Понятие о моментах распределения

Начальным моментом порядка k случайной величины X называется математическое ожидание случайной величины X^k, т. е.

$$\nu_k = \mathrm{E}X^k.$$

В частности: $\nu_1 = \mathrm{E}X$, $\nu_2 = \mathrm{E}X^2$, $\mathrm{D}X = \nu_2 - \left(\nu_1\right)^2$.

Пример 3.15. Рассмотрим случайную величину X, заданную законом распределения:

X	1	2	5	100
P	0,6	0,2	0,19	0,01

Найдем $\mathrm{E}(X) = 2{,}9$; $\nu_2 = \mathrm{E}X^2 = 106{,}15$.

Переход от X к X^2 позволяет учитывать влияние на математическое ожидание значений случайной величины, которые велики, но имеют малую вероятность. Поэтому оказалось целесообразным рассматривать математическое ожидание целой положительной степени случайной величины.

Центральным моментом порядка k случайной величины X называется математическое ожидание величины $\left[(X-\mathrm{E}X)\right]^k$:

$$\mu_k = \mathrm{E}\left(X-\mathrm{E}X\right)^k.$$

В частности: $\mu_1 = \mathrm{E}\left(X-\mathrm{E}X\right)=0,\ \mu_2 = \mathrm{E}\left(X-\mathrm{E}X\right)^2 = \mathrm{D}(X)$.

Упражнение. Докажите формулы: $\mu_2 = \nu_2 - \nu_1^2;\ \mu_3 = \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3$.

Асимметрия — нормированный центральный момент третьего порядка:

$$As = \frac{\mu_3}{\sigma^3} = \frac{\mathrm{E}\left(X-\mathrm{E}X\right)^3}{\sigma^3}.$$

Для симметричных распределений асимметрия равна нулю, для несимметричных отлична от нуля.

Эксцесс распределения — нормированный центральный момент четвертого порядка:

$$K = \frac{\mu_4}{\sigma^4} = \frac{\mathrm{E}\left(X-\mathrm{E}X\right)^4}{\sigma^4}.$$

Эксцесс служит для сравнения какого-либо распределения с нормальным распределением. Для нормального распределения эксцесс равен трем. Он больше трех, если пик распределения около математического ожидания острый, и меньше трех, если пик очень гладкий.

Задачи для самостоятельного решения

3.1. Пять юношей и три девушки случайным образом делятся на две команды для игры в волейбол. Случайная величина X — число девушек в первой команде. Составить ряд распределения случайной величины X, найти функцию распределения $F(x)$ и построить ее график.

3.2. Среди 12 студентов, из которых 6 изучают английский язык, 4 немецкий, а остальные китайский, случайным образом набирают четырех волонтеров для работы на международной конференции. Составить закон распределения случайной величины X — числа волонтеров, не знающих китайский язык, среди выбранных.

3.3. По статистике сезона биатлонист попадает в мишень в 90 % случаев. Найти ряд распределения для случайной величины X — число промахов по пяти мишеням, среднее значение числа промахов и дисперсию случайной величины X.

3.4. Студент, готовясь к тестированию по математике, считает, что он способен правильно ответить на 8 вопросов из 10 по алгебре, на половину вопросов по математическому анализу и на 3 вопроса из 10 по теории вероятностей. Составить закон распределения случайной величины X — числа правильных ответов студента в тесте, который содержит по одному вопросу из каждого раздела. Найти математическое ожидание и дисперсию случайной величины X.

3.5. Заемщик с плохой кредитной историей хочет взять очередной кредит. Для этого он обращается в банк, если ему отказывают, то обращается в следующий городской банк. Так он делает до тех пор, пока ему либо дадут кредит, либо он обойдет все 5 банков своего города. Вероятность, что ему одобрят кредит, равна 0,3. Составить закон распределения случайной величины X — число банков, в которые обращался заемщик. Найти математическое ожидание случайной величины X.

3.6. Случайная величина задана функцией распределения вероятностей:

$$F(x)=\begin{cases}0, & x<-0,5,\\ x^2+b\cdot x+c, & -0,5\le x\le 0,5,\\ 1, & x>1.\end{cases}$$

Найти:

а) неизвестные параметры b, c;

б) плотность распределения вероятностей $f(x)$;

в) вероятность попадания случайной величины в интервал $\left(-\frac{1}{3};\frac{1}{3}\right)$.

3.7. Случайная величина задана плотностью вероятностей:

$$f(x)=\begin{cases}0, & x<0,\\ a\cdot x^2, & 0\le x\le 3,\\ 0, & x>3.\end{cases}$$

Найти: а) параметр a; б) функцию распределения $F(x)$; в) $\mathrm{E}(X)$.

3.8. Известно, что случайные величины X и Y независимы: $\mathrm{E}(X)=2$, $\mathrm{D}(X)=5$, $\mathrm{E}(Y)=4$, $\mathrm{D}(Y)=3$. Случайная величина $Z=5X-2Y+1$.

Найти математическое ожидание и дисперсию случайной Z.

3.9. Доказать формулы: $\mu_2=\nu_2-\nu_1^2$; $\mu_3=\nu_3-3\nu_1\nu_2+2\nu_1^3$.

3.10. В условиях примера 3.7 имеем:

X	–3	0	1
P	0,1	0,3	0,6

,

Y	0	3
P	0,2	0,8

Составить закон распределения случайной величины $W=X\cdot Y$.

3.11. Доходность акций первого предприятия при позитивной для этого предприятия ситуации оценивается как 15 %, при негативной — как 10 %, для второго предприятия соответственно — 20 % и 1 %. Вероятность позитивной ситуации для первого предприятия составляет 80 %, для второго — 60 %. Инвестор вкладывает свой капитал в равных долях в акции этих предприятий. Составить закон распределения случайной величины *X* — доходность данного инвестиционного портфеля. Найти среднюю доходность и дисперсию случайной величины *X*. Как изменится средняя доходность, если вложить в акции первого предприятия 60 % капитала, а в акции второго — 40 %?

3.12. Квантиль уровня 0,15 нормально распределенной случайной величины *X* равен 12, а квантиль уровня 0,6 равен 16. Найти математическое ожидание и среднее квадратичное отклонение случайной величины.

Ответы к задачам для самостоятельного решения

3.1.

X	0	1	2	3
P	1/14	3/7	3/7	1/14

$$F(x)=\begin{cases} 0,\ x\le 0,\\ 1/14,\ 0<x\le 1,\\ 1/2,\ 1<x\le 2,\\ 13/14,\ 2<x\le 3,\\ 1,\ x>3.\end{cases}$$

3.2.

X	2	3	4
P	1/11	16/33	14/33

3.3.

X	0	1	2	3	4	5
P	0,59049	0,32805	0,0729	0,0081	0,00045	0,00001

$\mathrm{E}X = 0{,}5$; $\mathrm{D}X = 0{,}45$.

3.4.

X	0	1	2	3
P	0,07	0,38	0,43	0,12

3.5.

X	1	2	3	4	5
P	0,3	0,21	0,147	0,1029	0,2401

3.6. а) $b=1,\ c=\dfrac{1}{4}$; б) $f(x)=\begin{cases}0, & x<-0,5,\\ 2x+1, & -0,5\le x\le 0,5,\\ 0, & x>1;\end{cases}$ в) $\dfrac{2}{3}$.

3.7. а) $a=\dfrac{1}{9}$; б) $F(x)=\begin{cases}0, & x<0,\\ \dfrac{x^3}{27}, & 0\le x\le 3,\\ 1, & x>3;\end{cases}$ в) 2,25.

3.8. EZ = 3; DZ = 137.

3.10.

W	–9	0	3
P	0,08	0,44	0,48

3.11.

X	5,5	8	15	17,5
P	0,08	0,32	0,12	0,48

EX = 13,2; DX = 22,66. Средняя доходность увеличится на 0,16 %.

4. ОСНОВНЫЕ ДИСКРЕТНЫЕ И НЕПРЕРЫВНЫЕ РАСПРЕДЕЛЕНИЯ

Рассмотрим некоторые законы распределения, наиболее часто используемые для построения теоретико-вероятностных моделей реальных социально-экономических явлений. Начнем с *дискретных распределений*.

4.1. Распределение Бернулли

Случайная величина X имеет *распределение Бернулли* ($X \sim \mathrm{Bi}\,(1; p)$) с параметром p $(0 < p < 1)$, если $\mathrm{P}\,(X = 1) = p$, $\mathrm{P}(X = 0) = 1 - p = q$, т. е. закон распределения случайной величины X имеет вид, представленный в табл. 4.1.

Таблица 4.1

X	0	1
P	q	p

Функция распределения Бернулли имеет вид: $F(x) = \begin{cases} 0 \text{ при } x \le 0, \\ q \text{ при } 0 \le x \le 1, \\ 1 \text{ при } x > 1. \end{cases}$

График функции распределения изображен на рис. 4.1. Числовые характеристики распределения Бернулли:

$$\mathrm{E}(X) = p, \mathrm{D}(X) = pq. \tag{4.1}$$

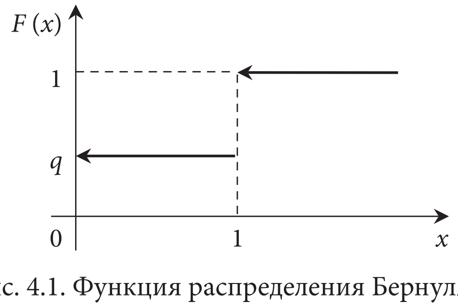


Рис. 4.1. Функция распределения Бернулли

Распределение Бернулли играет фундаментальную роль в теории вероятностей и математической статистике, являясь математической моделью опыта с двумя исходами.

4.2. Биномиальное распределение

Случайная величина X имеет *биномиальное распределение* ($X \sim \mathrm{Bi}\ (n;\ p)$) с параметрами n ($n \geq 1$) и p ($0 < p < 1$), если

$$\mathrm{P}\big(X = k\big) = C_n^k p^k q^{n-k}, \qquad (4.2)$$

где $k = 0, 1, \ldots, n$.

Заметим, что если $n = 1$ и $0 < p < 1$, то имеем распределение Бернулли: $X \sim \mathrm{Bi}\ (1;\ p)$.

Закон распределения случайной величины $X \sim \mathrm{Bi}\ (n;\ p)$ представлен в табл. 4.2.

Таблица 4.2

X	0	1	…	k	…	n	Σ
P	q^n	npq^{n-1}	…	$C_n^k p^k q^{n-k}$	…	p^n	1

Заметим, что

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

Функция распределения биномиального распределения равна:

$$F(x) = \begin{cases} 0 \ \text{ при } \ x \leq 0, \\ \sum\limits_{k=0}^{l} C_n^k p^k q^{n-k} \quad \text{при } \ l < x \leq l + 1, \\ 1 \ \text{ при } \ \ x > n. \end{cases}$$

Математическое ожидание, дисперсия и среднее квадратичное отклонение случайной величины X, имеющей биномиальное распределение, равны:

$$\mathrm{E}(m) = \mathrm{E}X = np, \quad \mathrm{D}(m) = \mathrm{D}X = npq, \quad \sigma_X = \sigma(m) = \sqrt{npq}. \qquad (4.3)$$

Пример 4.1. Финансовая операция *форвард* состоит в заключении сделки на продажу (или покупку) в будущем некоторого товара по цене, определяемой сторонами в настоящий момент времени. Фермер предполагает, что через месяц, когда он соберет урожай, цена пшеницы в каждом из десяти регионов, куда он обычно ее продает, может с вероятностью 0,8 понизиться и с вероятностью 0,2 повыситься. Поэтому он заключает с десятью мельниками в этих регионах десять форвардов на поставку им пшеницы через месяц по сегодняшней цене. Цены в регионах изменяются независимо. Найти математическое ожидание числа форвардов, которые окажутся выгодными для фермера, и вероятность того, что:

а) все десять проданных форвардов окажутся для него выгодными (форвард окажется выгодным, если в данном регионе за месяц цена понизится);

б) по меньшей мере пять форвардов окажутся для него выгодными.

Построить ряд и многоугольник распределения.

Решение. Рассматривается схема 10 независимых испытаний (заключение 10 форвардов) с вероятностью выгоды каждого форварда, равной 0,8: $X \sim \mathrm{Bi}\,(10; 0{,}8)$. Математическое ожидание числа форвардов, которые окажутся выгодными для фермера, равно $\mathrm{E}(X) = 10 \cdot 0{,}8 = 8$.

а) Вероятность того, что все десять проданных форвардов окажутся для фермера выгодными, равна:

$$\mathrm{P}\left(X = 10\right) = \mathrm{P}_{10}(10) = C_{10}^{10} \cdot 0{,}8^{10} \cdot 0{,}2^{0} = 0{,}8^{10} \approx 0{,}1074.$$

Тот же результат получим, используя функцию Microsoft Excel БИНОМ.РАСП (значение интегральной величины — 0):

$$\mathrm{P}_{10}(10) = \text{БИНОМ.РАСП}\ (10;\ 10;\ 0{,}8;\ 0) = 0{,}1074.$$

б) Вероятность того, что по меньшей мере пять форвардов окажутся для фермера выгодными, равна:

$$\mathrm{P}_{10}(5) + \mathrm{P}_{10}(6) + \mathrm{P}_{10}(7) + \mathrm{P}_{10}(8) + \mathrm{P}_{10}(9) + \mathrm{P}_{10}(10) = C_{10}^{5} \cdot 0{,}8^{5} \cdot 0{,}2^{5} + C_{10}^{6} \cdot 0{,}8^{6} \cdot 0{,}2^{4} +$$
$$+ C_{10}^{7} \cdot 0{,}8^{7} \cdot 0{,}2^{3} + C_{10}^{8} \cdot 0{,}8^{8} \cdot 0{,}2^{2} + C_{10}^{9} \cdot 0{,}8^{9} \cdot 0{,}2^{1} + C_{10}^{10} \cdot 0{,}8^{10} \cdot 0{,}2^{0}.$$

Для вычисления последней суммы воспользуемся в Microsoft Excel функцией БИНОМ.РАСП (значение интегральной величины — 1). Для этого перепишем последнюю сумму в виде

$$\mathrm{P}_{10}(5) + \mathrm{P}_{10}(6) + \mathrm{P}_{10}(7) + \mathrm{P}_{10}(8) + \mathrm{P}_{10}(9) + \mathrm{P}_{10}(10) =$$
$$= 1 - \left(\mathrm{P}_{10}(0) + \mathrm{P}_{10}(1) + \mathrm{P}_{10}(2) + \mathrm{P}_{10}(3) + \mathrm{P}_{10}(4)\right) = 1 - \mathrm{P}\left(k \le 4\right).$$

Для вычисления вероятности $\mathrm{P}\ (k \le 4)$ используем функцию Microsoft Excel БИНОМ.РАСП (значение интегральной величины — 1):

$$\mathrm{P}\ (k \le 4) = \text{БИНОМ.РАСП}(4;\ 10;\ 0{,}8;\ 1) = 0{,}0064.$$

Искомая вероятность равна: $1 - \mathrm{P}(k \le 4) \approx 0{,}9936$.

Построим *ряд распределения* случайной величины X — числа успехов в серии n испытаний. Массив A3:A13 содержит значения случайной величины k (число успехов). В ячейку B3 занесите формулу биномиального распределения: =БИНОМ.РАСП (A3;10; 0,8; 0). В результате вычислений в ячейке B3 появится значение вероятности $p = 0{,}0000001$. Далее следует размножить результат вычислений вероятности p в ячейки B4:B13. Полученный таким образом ряд распределения показан на рис. 4.2 (ячейки B3:B13). Для построения *многоугольника распределения* выделяем два столбца исходных данных (A3:A13, B3:B13). В главном меню выберите закладку Вставка → Рекомендуемые диаграммы → Точечная с прямыми отрезками и маркерами → ОК.

Пример 4.2. Построить функцию распределения числа успехов в серии $n = 10$ независимых испытаний с вероятностью успеха $p = 0{,}5$. Для построения воспользуемся в Microsoft Excel функцией БИНОМ.РАСП (значение интегральной величины — 1). Для этого следует заполнить поля ввода панели Аргументы функции БИНОМ.РАСП, как показано на рис. 4.3, и нажать клавишу ОК. В ячейке C18 и C19 значения $F(x)$ полагаем равными нулю. В ячейке C20 появится значение функции $F(x) = 0{,}00098$. Далее результат вычисления значения функции распределения размножаем в ячейки C21:C29. Функция распределения и ее график приведены на рис. 4.4.

К сожалению, Excel не располагает процедурой построения функции распределения, поэтому график строится в виде точечной диаграммы, а затем вручную достраивается с помощью выбора Работы с диаграммами → Формат → Вставка фигур…

Пример 4.3. Вычислить математическое ожидание, дисперсию случайной величины, заданной в примере 4.1 с помощью Microsoft Excel.

Для вычисления математического ожидания необходимо воспользоваться формулой СУММПРОИЗВ. Выбираем ячейку H2, в которой будет вычислено

	A	B
2	k	P_{10}^{k} (p=0,8)
3	0	0,0000001
4	1	0,0000041
5	2	0,0000737
6	3	0,0007864
7	4	0,0055050
8	5	0,0264241
9	6	0,0880804
10	7	0,2013266
11	8	0,3019899
12	9	0,2684355
13	10	0,1073742

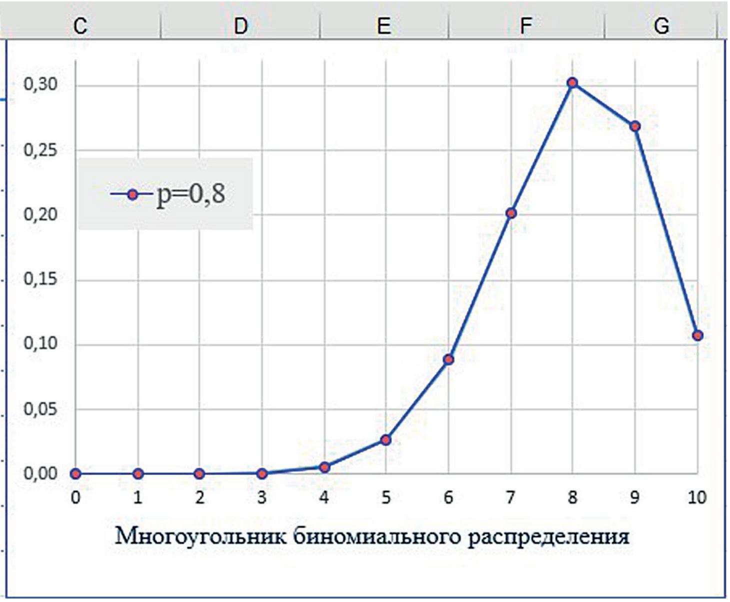


Рис. 4.2. Ряд и многоугольник биномиального распределения

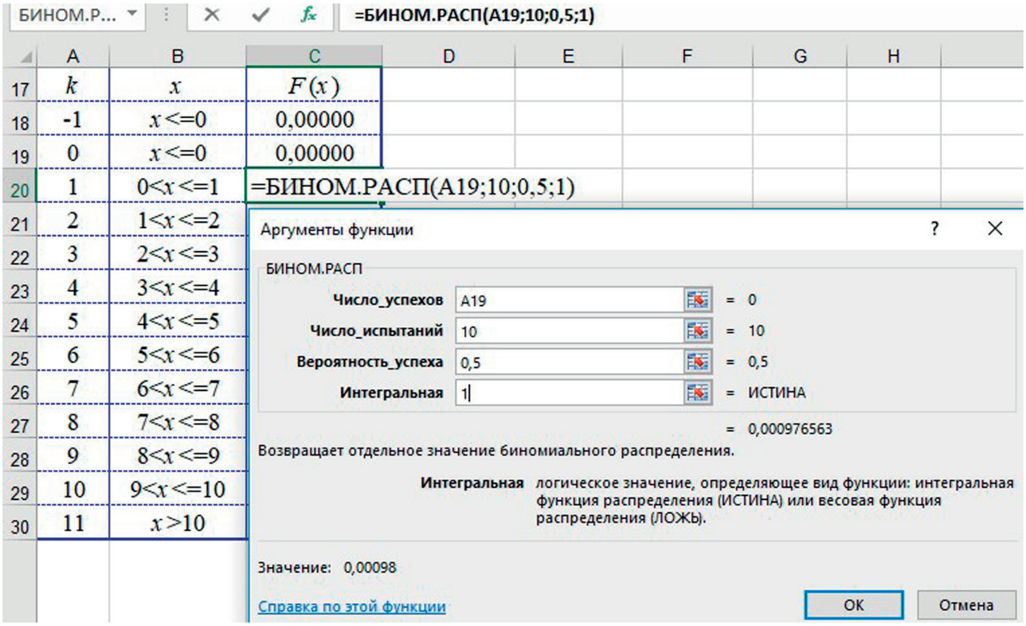
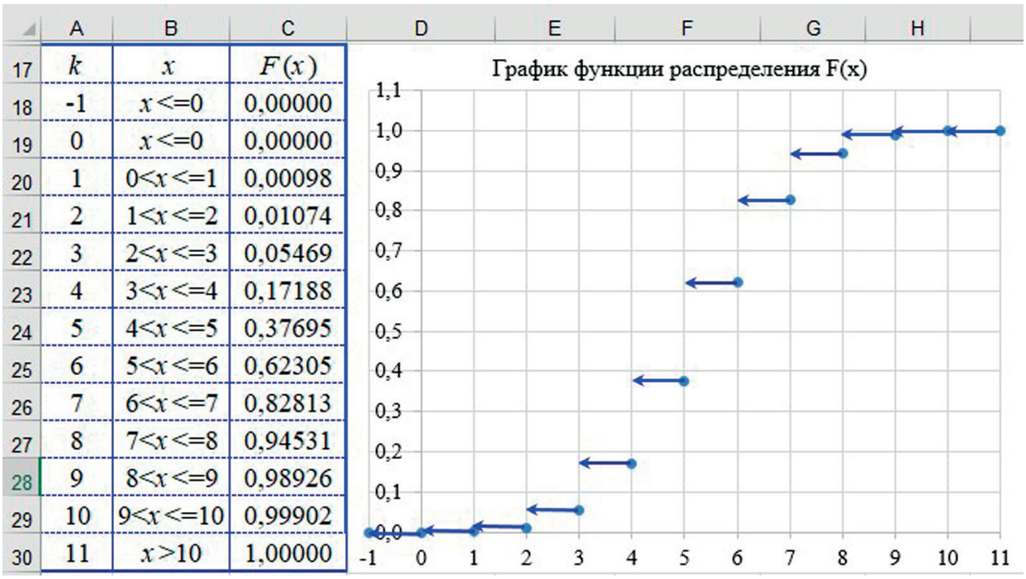


Рис. 4.3. Заполнение полей ввода функции БИНОМ.РАСП для построения функции распределения $F(x)$ в ячейках C19:C29



	A	B	C
17	k	x	F(x)
18	-1	x<=0	0,00000
19	0	x<=0	0,00000
20	1	0<x<=1	0,00098
21	2	1<x<=2	0,01074
22	3	2<x<=3	0,05469
23	4	3<x<=4	0,17188
24	5	4<x<=5	0,37695
25	6	5<x<=6	0,62305
26	7	6<x<=7	0,82813
27	8	7<x<=8	0,94531
28	9	8<x<=9	0,98926
29	10	9<x<=10	0,99902
30	11	x>10	1,00000

Рис. 4.4. График функции распределения $F(x)$

математическое ожидание, заполняем поля диалогового окна, как показано на рис. 4.5.

В ячейке H2 получаем результат вычисления математического ожидания EX = 8.

Для вычисления дисперсии в ячейку J2 поместим результат вычисления дисперсии. Для этого вновь воспользуемся функцией СУММПРОИЗВ. В ячейку J2 поместим формулу = СУММПРОИЗВ(СТЕПЕНЬ(A3:A13;2); B3:B13) — СТЕПЕНЬ(I2;2), как показано на рис. 4.6. В результате получим DX = 1,6.

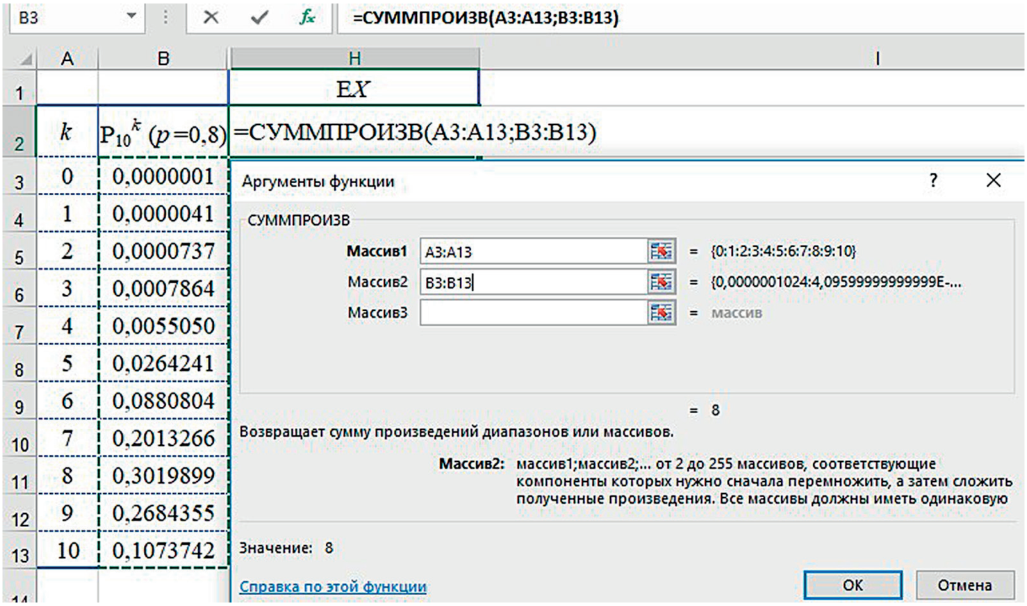


Рис. 4.5. Диалоговое окно функции СУММПРОИЗВ с заполненными полями ввода

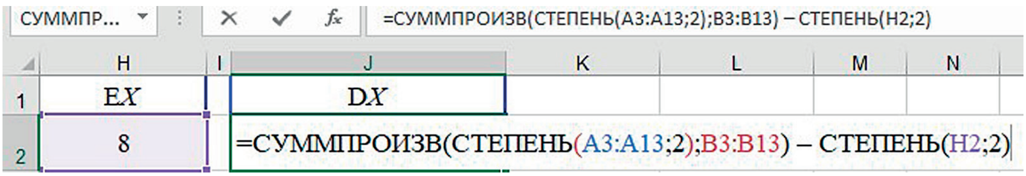


Рис. 4.6. Вычисление дисперсии DX

4.3. Распределение Пуассона

Случайная величина X имеет *распределение Пуассона* ($X \sim \Pi(\lambda)$) с параметром λ ($\lambda > 0$), если

$$\mathrm{P}\left(X = k\right) = \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}, \tag{4.4}$$

где $k = 0, 1, 2, \ldots$

Закон распределения Пуассона определяется интенсивностью наступления события A: $\lambda \approx np$. Ряд распределения случайной величины имеет вид (табл. 4.3).

Таблица 4.3

X	0	1	…	k	…	Σ
P	$\mathrm{e}^{-\lambda}$	$\lambda \mathrm{e}^{-\lambda}$	…	$\frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}$	…	1

Заметим, что $\sum\limits_{k=0}^{\infty} \mathrm{P}\left(X = k\right) = \sum\limits_{k=0}^{\infty} \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} = 1.$

Числовые характеристики случайной величины, имеющей распределение Пуассона

Функция распределения Пуассона равна:

$$F(x) = \begin{cases} 0 \ \text{ при } \ x \le 0, \\ \sum\limits_{k=0}^{l} \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \ \text{ при } \ l < x \le l + 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{cases}$$

Замечание. Закон Пуассона является *предельным* для биномиального распределения: $np \to \lambda$, $n \to \infty$. Для биномиального распределения имеем:

$$\mathrm{E}(X) = np \to \lambda, \ n \to \infty;$$

$$\mathrm{D}(X) = npq \to \lambda, \ n \to \infty \ \left(p \to 0, \ np \to \lambda, \ q \to 1\right).$$

Так как вероятность наступления события A в каждом испытании мала ($p \to 0$), то закон распределения Пуассона называют *законом редких событий*.

Распределение Пуассона имеет место, когда встречается однородный поток независимых событий, регистрируемых в заданном промежутке времени или пространства при заданной интенсивности λ, например, поток отказов абонентов от телефонной сети.

Пример 4.4. В банк поступило 4 тыс. пакетов денежных знаков. Вероятность того, что пакет содержит недостаточное или избыточное количество денежных знаков, равна 0,001.

Составить закон распределения числа ошибочно укомплектованных пакетов. Найти:

а) вероятность того, что при проверке будет обнаружен хотя бы один ошибочно укомплектованный пакет;

б) вероятность того, что при проверке будет обнаружено не более трех ошибочно укомплектованных пакетов;

в) математическое ожидание и дисперсию числа ошибочно укомплектованных пакетов.

Решение. Найдем пуассоновский параметр распределения $\lambda = np = 4000 \cdot 0{,}001 = 4$. Ряд распределения случайной величины имеет вид:

X	0	1	…	k	…	Σ
P	e^{-4}	$4e^{-4}$	…	$\dfrac{4^k e^{-4}}{k!}$	…	1

а) Перейдем к вероятности противоположного события: $1 - P_{4000}(0) = 1 - e^{-4} \approx 0{,}9817$.

б) Искомая вероятность равна сумме вероятностей:

$$P_{4000}(0) + P_{4000}(1) + P_{4000}(2) + P_{4000}(3) = \frac{e^{-4} \cdot 4^0}{0!} + \frac{e^{-4} \cdot 4^1}{1!} + \frac{e^{-4} \cdot 4^2}{2!} + \frac{e^{-4} \cdot 4^3}{3!} =$$

$$= e^{-4}\left(1 + 4 + 8 + \frac{32}{3}\right) \approx 0{,}4335.$$

в) Воспользуемся формулами $E(X) = D(X) = \lambda$. То есть математическое ожидание и дисперсия числа ошибочно укомплектованных пакетов равны: $E(X) = D(X) = 4$.

Пример 4.5. Менеджер телекоммуникационной компании решил рассчитать вероятность того, что в некотором небольшом городе в течение пяти минут поступят 0, 1, 2, … вызовов. Выбраны случайные интервалы в пять минут, подсчитано число вызовов в каждый из интервалов и рассчитано среднее число вызовов: $\lambda = 4{,}8$. Вычислить вероятность того, что в течение пяти минут поступят: а) шесть вызовов; б) не более шести вызовов.

Решение. а) По формуле Пуассона получаем:

$$P(X = 6) = \frac{4{,}8^6 e^{-4{,}8}}{6!} \approx 0{,}1398.$$

Тот же результат получим, используя в Microsoft Excel функцию ПУАССОН.РАСП (значение интегральной величины — 0):

$P(X=6) = $ ПУАССОН.РАСП (6; 4,8; 0) = 0,1398.

Кроме того, удобно пользоваться значениями из Приложения (табл. 3).

б) Искомая вероятность равна сумме вероятностей:

$$\mathrm{P}(0)+\mathrm{P}(1)+\mathrm{P}(2)+\mathrm{P}(3)+\mathrm{P}(4)+\mathrm{P}(5)+\mathrm{P}(6)=$$

$$=\frac{\mathrm{e}^{-4,8}\cdot 4,8^0}{0!}+\frac{\mathrm{e}^{-4,8}\cdot 4,8^1}{1!}+\frac{\mathrm{e}^{-4,8}\cdot 4,8^2}{2!}+\frac{\mathrm{e}^{-4,8}\cdot 4,8^3}{3!}+\frac{\mathrm{e}^{-4,8}\cdot 4,8^4}{4!}+\frac{\mathrm{e}^{-4,8}\cdot 4,8^5}{5!}+$$

$$+\frac{\mathrm{e}^{-4,8}\cdot 4,8^6}{6!}=\mathrm{e}^{-4,8}\left(1+4,8+\frac{4,8^2}{2}+\frac{4,8^3}{6}+\frac{4,8^4}{24}+\frac{4,8^5}{120}+\frac{4,8^6}{720}\right)\approx 0,7908.$$

Тот же результат получим, используя функцию Microsoft Excel (значение интегральной величины — 1):

$P(X \le 6) = $ ПУАССОН.РАСП(6; 4,8; 1) = 0,7908.

4.4. Геометрическое распределение

Случайная величина X имеет *геометрическое распределение* ($X \sim \mathrm{G}(p)$), если она принимает счетное число значений: 1, 2, …, m, … с соответствующими вероятностями: $p_m = \mathrm{P}(X = m) = q^{m-1}p$, где $m = 1, 2, \ldots, 0 < p < 1, q = 1 - p$.

Случайную величину, распределенную по геометрическому закону, можно интерпретировать как число m опытов, проведенных по схеме Бернулли до первого успеха.

Закон распределения случайной величины X представлен в табл. 4.4.

Таблица 4.4

X	1	2	3	…	m	…	Σ
P	p	qp	q^2p	…	$q^{m-1}p$	…	1

Заметим, что

$$\sum_{m=1}^{\infty} q^{m-1}p = p\left(1+q+q^2+\ldots\right) = p\cdot\frac{1}{1-q} = \frac{p}{p} = 1.$$

Функция распределения геометрического распределения равна:

$$F(x) = \begin{cases} 0 \ \text{ при } \ x \le 1, \\ \sum_{m=1}^{l} q^{m-1}p \quad \text{при } \ l < x \le l+1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{cases}$$

Числовые характеристики случайной величины, имеющей геометрическое распределение:

$$\mathrm{E}\left(X\right)=\frac{1}{p},\ \mathrm{D}\left(X\right)=\frac{q}{p^2}. \tag{4.5}$$

Пример 4.6. Среди выпускаемых заводом автомобилей 75 % некомплектны. Определить, сколько автомобилей должен в среднем осмотреть покупатель, чтобы выбрать комплектный автомобиль.

Решение. Число автомобилей — случайная величина, имеющая геометрическое распределение: $X \sim \mathrm{G}(0{,}25)$. Ее среднее значение оценивается математическим ожиданием $\mathrm{E}(X)=\frac{1}{p}=\frac{1}{0{,}25}=4.$ Итак, покупатель в среднем должен осмотреть 4 автомобиля.

4.5. Гипергеометрическое распределение

Случайная величина X имеет *гипергеометрическое распределение* с параметрами n, M, N, если она принимает значения $m = 0, 1, 2, \ldots, m, \ldots, \min(n, M)$ с вероятностями:

$$\mathrm{P}\left(X=m\right)=\frac{C_M^m\,C_{N-M}^{n-m}}{C_N^n}, \tag{4.6}$$

где $m \le n$, $n \le N$.

Гипергеометрическое распределение имеет случайная величина $X = m$ — число объектов, обладающих заданным свойством, среди n объектов, случайно извлеченных (без возврата) из совокупности N объектов, M из которых обладают этим свойством.

Математическое ожидание и дисперсия случайной величины, распределенной по гипергеометрическому закону, определяются формулами:

$$\mathrm{E}X=n\frac{M}{N};\quad \mathrm{D}X=n\frac{M}{N-1}\left(1-\frac{M}{N}\right)\left(1-\frac{n}{N}\right). \tag{4.7}$$

Если n мало по сравнению с N, то гипергеометрическое распределение приближается к биномиальному распределению с параметрами n и $p=\frac{M}{N}$, т. е.

$$\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}\approx C_n^k p^k q^{n-k}.$$

Пример 4.7. В комитете каждый из 50 американских штатов представлен двумя сенаторами. Найти вероятность того, что в комиссии из 50 случайно выбранных сенаторов представлен конкретный штат.

Решение. Обозначим X — число сенаторов, представляющих данный штат. Из условия задачи следует, что $N = 100$ (всего сенаторов), $M = 2$ (число сенаторов, представляющих данный штат), $m = 50$ (число сенаторов в комиссии). Данный штат будет представлен в комиссии, если $X = 1$ или $X = 2$, т. е. искомая вероятность равна:

$$\mathrm{P}\left(X=1\right)+\mathrm{P}\left(X=2\right)=\frac{C_2^1 C_{98}^{49}}{C_{100}^{50}}+\frac{C_2^2 C_{98}^{48}}{C_{100}^{50}}=\frac{50}{49}+\frac{49\cdot 50}{99\cdot 100}=0{,}72525.$$

Пример 4.8. Из 10 новых телевизоров на выставке представлены пять телевизоров типа QLED. Наудачу для магазина выбрано четыре. Составить закон распределения числа телевизоров типа QLED среди 4 отобранных. Найти математическое ожидание и дисперсию числа выбранных телевизоров.

Решение. Пусть X — число телевизоров типа QLED среди 4 отобранных. Случайная величина X может принимать значения 0, 1, 2, 3 и 4. Найдем соответствующие вероятности, используя формулу (4.6). Из условия задачи следует, что $N = 10$ (всего телевизоров), $M = 5$ (количество телевизоров типа QLED), $n = 4$ (количество отобранных телевизоров). Тогда:

$$\mathrm{P}\left(X=0\right)=\frac{C_5^0\cdot C_5^4}{C_{10}^4}=\frac{5}{210}=\frac{1}{42},\quad \mathrm{P}\left(X=1\right)=\frac{C_5^1\cdot C_5^3}{C_{10}^4}=\frac{5\cdot 10}{210}=\frac{5}{21},$$

$$\mathrm{P}\left(X=2\right)=\frac{C_5^2\cdot C_5^2}{C_{10}^4}=\frac{10\cdot 10}{210}=\frac{10}{21},\quad \mathrm{P}\left(X=3\right)=\frac{C_5^3\cdot C_5^1}{C_{10}^4}=\frac{10\cdot 5}{210}=\frac{5}{21},$$

$$\mathrm{P}\left(X=4\right)=\frac{C_5^4\cdot C_5^0}{C_{10}^4}=\frac{5}{210}=\frac{1}{42}.$$

Таким образом, закон распределения количества телевизоров типа QLED среди 4 отобранных имеет вид:

X	0	1	2	3	4	$\sum$
P	1/42	5/21	10/21	5/21	1/42	1

Найдем сначала математическое ожидание и дисперсию по ряду распределения:

$$\mathrm{E}X=0\cdot\left(1/42\right)+1\cdot\left(5/21\right)+2\cdot\left(10/21\right)+3\cdot\left(5/21\right)+4\cdot\left(1/42\right)=2;$$

$$\mathrm{D}X=0^2\cdot\left(1/42\right)+1^2\cdot\left(5/21\right)+2^2\cdot\left(10/21\right)+3^2\cdot\left(5/21\right)+4^2\cdot\left(1/42\right)-2^2=2/3.$$

По формулам (4.7) получаем:

$$\mathrm{E}X = n\frac{M}{N} = 4\cdot\frac{5}{10} = 2,\quad \mathrm{D}X = 4\cdot\frac{5}{9}\left(1-\frac{5}{10}\right)\left(1-\frac{4}{10}\right) = \frac{2}{3} \approx 0,6667.$$

4.6. Производящая функция

Выше были рассмотрены способы определения вероятности события для случаев, когда вероятность события во всех независимых испытаниях одна и та же. На практике вероятность события от опыта к опыту может меняться.

Справедливо следующее **утверждение**: если производится n независимых испытаний в различных условиях, причем вероятность события A в i-м испытании равна p_i, $i = 1,2, \ldots, n$, то вероятность того, что событие A появится в этих событиях ровно m раз, равна коэффициенту при z^m в разложении по степеням z *производящей функции*:

$$\varphi_n(z) = \left(q_1 + p_1 z\right)\left(q_2 + p_2 z\right)\ldots\left(q_n + p_n z\right),\quad \text{где } q_i = 1 - p_i.$$

Пример 4.9. Предприятие производит изделия, каждое из которых должно пройти четыре этапа проверки. Первый этап проверки изделие проходит благополучно с вероятностью $p_1 = 0{,}9$, второй — $p_2 = 0{,}8$, третий — $p_3 = 0{,}75$ и четвертый — $p_4 = 0{,}7$. Составить закон распределения случайной величины X — числа испытаний изделий, закончившихся благополучно. Найти математическое ожидание и дисперсию случайной величины X.

Решение. Найдем вероятности $\mathrm{P}(X = m)$ того, что изделие пройдет благополучно m этапов проверки. Для этого составим производящую функцию

$$\varphi_4(z) = \left(0,1 + 0,9z\right)\left(0,2 + 0,8z\right)\left(0,25 + 0,75z\right)\left(0,3 + 0,7z\right) =$$
$$= 0,0015 + 0,0275z + 0,1685z^2 + 0,4245z^3 + 0,378z^4.$$

Каждый из пяти полученных коэффициентов при z^m в функции $\varphi_4(z)$ выражает соответственно вероятности $\mathrm{P}(X = m)$, т.е. закон (ряд) распределения случайной величины X имеет вид:

X	0	1	2	3	4
P	0,0015	0,0275	0,1685	0,4245	0,378

Найдем математическое ожидание и дисперсию случайной величины X:

$$\mathrm{E}(X) = 0\cdot 0,0015 + 1\cdot 0,0275 + 2\cdot 0,1685 + 3\cdot 0,4245 + 4\cdot 0,378 = 3,15;$$

$$\mathrm{D}(X) = 0^2\cdot 0,0015 + 1^2\cdot 0,0275 + 2^2\cdot 0,1685 + 3^2\cdot 0,4245 + 4^2\cdot 0,378 - \left(3,15\right)^2 = 0,6475.$$

Заметим, что математическое ожидание случайной величины X можно вычислить непосредственно (не находя предварительно закона распределения). Представим X как сумму случайных величин X_m, выражающих число благополучных проверок для m-го этапа: $X = X_1 + X_2 + X_3 + X_4$, где

X_1	0	1
P	0,1	0,9

,

X_2	0	1
P	0,2	0,8

,

X_3	0	1
P	0,25	0,75

,

X_4	0	1
P	0,3	0,7

$$\mathrm{E}(X) = \mathrm{E}(X_1) + \mathrm{E}(X_2) + \mathrm{E}(X_3) + \mathrm{E}(X_4);$$

$$\mathrm{E}(X_1) = 0 \cdot 0{,}1 + 1 \cdot 0{,}9 = 0{,}9; \ \mathrm{E}(X_2) = 0 \cdot 0{,}2 + 1 \cdot 0{,}8 = 0{,}8;$$

$$\mathrm{E}(X_3) = 0 \cdot 0{,}25 + 1 \cdot 0{,}75 = 0{,}75; \ \mathrm{E}(X_4) = 0 \cdot 0{,}3 + 1 \cdot 0{,}7 = 0{,}7.$$

Тогда

$$\mathrm{E}(X) = 0{,}9 + 0{,}8 + 0{,}75 + 0{,}7 = 3{,}15.$$

Таким образом, в общем случае:

$$\boxed{\mathrm{E}(X) = p_1 + p_2 + \ldots + p_n}.$$

Аналогично для дисперсии:

$$\mathrm{D}(X) = \mathrm{D}(X_1) + \mathrm{D}(X_2) + \mathrm{D}(X_3) + \mathrm{D}(X_4) =$$
$$= 0{,}9 - 0{,}9^2 + 0{,}8 - 0{,}8^2 + 0{,}75 - 0{,}75^2 + 0{,}7 - 0{,}7^2 = 0{,}6475.$$

В общем случае:

$$\boxed{\mathrm{D}(X) = p_1q_1 + p_2q_2 + \ldots + p_nq_n}.$$

Далее следуют важнейшие непрерывные распределения.

4.7. Равномерное распределение

Непрерывная случайная величина $X = X(\omega)$ имеет *равномерное распределение* (*закон постоянной плотности*) на отрезке $[a; b]$, если ее плотность вероятности $f(x)$ постоянна на этом отрезке, а вне его равна нулю:

$$f(x) = \begin{cases} 0, & x < a, \\ \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & x > b. \end{cases}$$

К случайным величинам, имеющим равномерное распределение, относятся те случайные величины, о которых известно, что все их значения расположены внутри некоторого промежутка $[a;\ b]$ и при этом одинаково возможны. Например, время ожидания транспорта, ошибка, получающаяся от определения результата измерения до ближайшего целого числа, и т. д.

Функция распределения равномерно распределенной случайной величины имеет вид (рис. 4.7):

$$F(x)=\int_{-\infty}^{x} f(t)dt=\begin{cases} 0, & x<a, \\ \dfrac{x-a}{b-a}, & a\le x\le b, \\ 1, & x>b. \end{cases}$$

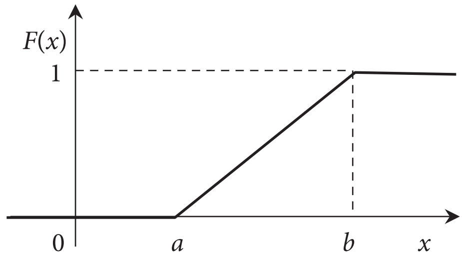


Рис. 4.7. Функция распределения вероятностей равномерного закона

Математическое ожидание и дисперсия случайной величины X, имеющей равномерное распределение, равны:

$$\mathrm{E}(X)=\frac{1}{2}(a+b),\quad \mathrm{D}(X)=\frac{(b-a)^2}{12}.$$

Заметим, что график плотности вероятности равномерного закона симметричен относительно прямой $x=\frac{1}{2}(a+b)$ (см. рис. 3.9).

Линейное преобразование $Y=\dfrac{X-a}{b-a}$ переводит случайную величину X, равномерно распределенную на отрезке $[a;\ b]$, в случайную величину Y, равномерно распределенную на отрезке $[0;\ 1]$.

Пример 4.10. Цена деления шкалы измерительного прибора равна 0,2. Показания прибора округляются до ближайшего целого деления. Найти вероятность того, что при отсчете будет сделана абсолютная ошибка: а) меньшая 0,04; б) большая 0,05.

Решение. Пусть X — показания прибора. Поскольку показания прибора округляются до ближайшего целого деления, то следует считать, что X — случайная величина, равномерно распределенная в промежутке [0; 0,2]. Ее центр находится в точке $x = \frac{1}{2}(0 + 0,2) = 0,1.$ Функция распределения случайной величины имеет вид:

$$F(x) = \begin{cases} 0, & x < 0, \\ \dfrac{x - 0}{0,2 - 0}, & 0 \le x \le 0,2, \\ 1, & x > 0,2. \end{cases}$$

Вероятность того, что при отсчете будет сделана ошибка, меньшая по абсолютной величине 0,04, равна:

$$\mathrm{P}\left(\left|X - 0,1\right| < 0,04\right) = \mathrm{P}\left(0,06 < x < 0,14\right) = F\left(0,14\right) - F\left(0,06\right) = \frac{0,14 - 0}{0,2} - \frac{0,06 - 0}{0,2} = 0,4.$$

Вероятность того, что при отсчете будет сделана абсолютная ошибка, большая 0,05, равна:

$$\mathrm{P}\left(\left|X - 0,1\right| > 0,05\right) = 1 - \mathrm{P}\left(\left|X - 0,1\right| < 0,05\right) = 1 - \mathrm{P}\left(0,05 < x < 0,15\right) =$$
$$= 1 - F(0,15) + F(0,05) = 1 - \frac{0,15 - 0}{0,2} + \frac{0,05 - 0}{0,2} = 0,5.$$

4.8. Показательное (экспоненциальное) распределение

Непрерывная случайная величина X имеет показательное (экспоненциальное) распределение с параметром λ ($\lambda > 0$), если плотность распределения вероятностей имеет вид (рис. 4.8):

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, x \ge 0, \\ 0, x < 0. \end{cases}$$

Так как $\lambda > 0$, то $f(x) \ge 0$. Заметим, что $\int\limits_{-\infty}^{\infty} f(x)dx = 1.$

Функция распределения экспоненциального закона:

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \begin{cases} 0, \ x \le 0, \\ 1 - \mathrm{e}^{-\lambda x}, \ x > 0. \end{cases}$$

Функция $F(x)$ возрастает для $x > 0$. График функции $F(x)$ является выпуклым вверх для $x > 0$. Прямая $y = 1$ является горизонтальной асимптотой графика функции случайной величины, имеющей показательное распределение (рис. 4.9).

Важнейшие числовые характеристики определяются равенствами:

$$\mathrm{E}(X)=\frac{1}{\lambda};\quad \mathrm{D}(X)=\frac{1}{\lambda^2};\quad \sigma(X)=\frac{1}{\lambda}.$$

Показательное распределение используется в теории массового обслуживания и теории надежности.

Пусть X — время обслуживания одного требования каналом обслуживания, $\mathrm{E}(X)=\frac{1}{\lambda}$ — среднее время обслуживания одной заявки (требования). Среднее число требований (заявок), обслуживаемых в единицу времени, равно $\lambda=\frac{1}{\mathrm{E}(X)}$.

Часто важно знать вероятность того, что обслуживание будет продолжаться более длительное время, чем x:

$$\mathrm{P}\left(X>x\right)=\mathrm{P}\left(X\geq x\right)=1-\mathrm{P}\left(X<x\right)=1-F(x)=\mathrm{e}^{-\lambda x}. \qquad (4.8)$$

Пример 4.11. Длительность междугородних телефонных разговоров распределена примерно по показательному закону, разговор продолжается в среднем 3 мин. Найти вероятность того, что очередной разговор будет продолжаться более 3 мин. Определить долю разговоров, которые длятся менее 1 мин. Найти

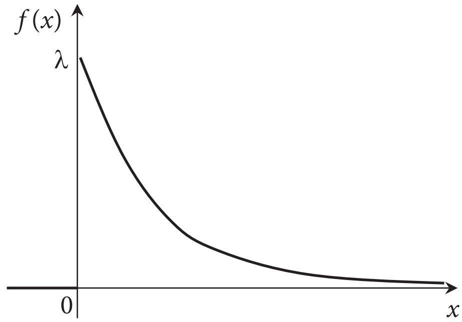


Рис. 4.8. Плотность вероятности показательного закона

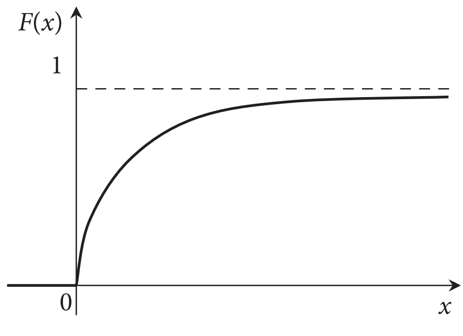


Рис. 4.9. Функция распределения вероятностей показательного закона

вероятность того, что разговор, который длится уже 10 мин, закончится в течение ближайшей минуты, а также математическое ожидание и дисперсию длительности разговора.

Решение. По условию средняя длительность переговоров составляет $E(X)=3$ мин. Но для показательного распределения $E(X)=\dfrac{1}{\lambda}$, откуда $\lambda=\dfrac{1}{E(X)}=\dfrac{1}{3}$. Дисперсия длительности разговора равна $D(X)=\dfrac{1}{\lambda^2}=9$.

Вероятность того, что очередной разговор будет продолжаться более 3 мин, согласно (4.8) равна:

$$P\left(X>3\right)=1-\left(1-e^{-\frac{1}{3}\cdot 3}\right)=e^{-\frac{1}{3}\cdot 3}=e^{-1}\approx 0,368.$$

Вероятность того, что разговор будет длиться менее 1 мин, равна:

$$P\left(X<1\right)=F(1)=1-e^{-\frac{1}{3}\cdot 1}=1-e^{-\frac{1}{3}}\approx 0,283.$$

Вероятность того, что разговор, который длится уже 10 мин, закончится в течение ближайшей минуты, равна:

$$P\left(X<11/X\geq 10\right)=\frac{P\left(\left(X\geq 10\right)\cap\left(X<11\right)\right)}{P\left(X\geq 0\right)}=\frac{P\left(10\leq X<11\right)}{P\left(X>10\right)}=\frac{P\left(10<X<11\right)}{P\left(X>10\right)}=$$

$$=\frac{F(11)-F(10)}{1-F(10)}=\frac{1-e^{-11/3}-1+e^{-10/3}}{1-\left(1-e^{-10/3}\right)}=1-e^{-1/3}\approx 0,283.$$

4.9. Нормальный закон распределения

Особенность *нормального распределения* заключается еще и в том, что это распределение является предельным, к которому с ростом числа наблюдений стремятся другие распределения. Например, закон Пуассона и биномиальный закон при больших n стремятся к нормальному распределению.

Случайная величина X имеет *нормальное* (*гауссовское*) *распределение* с параметрами a и σ ($\sigma>0$) ($X\sim N\,(a;\,\sigma)$), если плотность распределения вероятностей имеет вид:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{\left(x-a\right)^2}{2\sigma^2}},\quad x\in\mathbf{R}.$$

График плотности нормального распределения называется *кривой Гаусса*.

С в о й с т в а функции плотности вероятности нормального закона распределения:

1) так как $\sigma > 0$, то $f(x) > 0$, $\forall x \in \mathbf{R}$;

2) $\lim\limits_{x\to\infty} f(x) = 0$, поэтому ось Ox является горизонтальной асимптотой графика функции (рис. 4.10);

3) максимальное значение функция $f(x)$ принимает в точке $x_0 = a$, при этом $f(a) = \dfrac{1}{\sigma\sqrt{2\pi}}$;

4) кривая плотности нормального закона распределения симметрична относительно прямой $x = a$ и имеет две точки перегиба с координатами: $\left(a \pm \sigma;\ \dfrac{1}{\sigma\sqrt{2\pi e}}\right)$;

5) $\int\limits_{-\infty}^{\infty} f(x)dx = 1$;

6) параметр a характеризует положение кривой (рис. 4.11), а параметр σ — форму кривой плотности распределения вероятности (рис. 4.12). Если σ возрастает, то кривая сжимается к оси Ox, если σ убывает, пик становится более острым.

Функция распределения случайной величины, распределенной по нормальному закону

Функция распределения случайной величины, распределенной по нормальному закону, имеет вид:

$$F(x) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right). \tag{4.9}$$

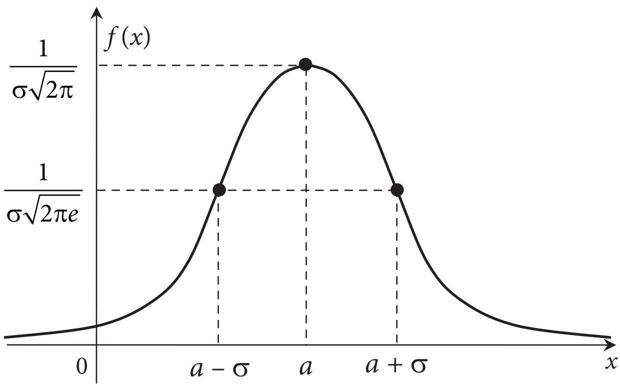


Рис. 4.10. Функция плотности вероятности нормального распределения

Здесь

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

— функция Лапласа. Таблицу значений функции Лапласа можно найти в Приложении (табл. 2).

Если $a = 0$, $\sigma = 1$, т. е. случайная величина $X \sim N(0; 1)$, то соответствующее нормальное распределение называется *стандартным* (рис. 4.13).

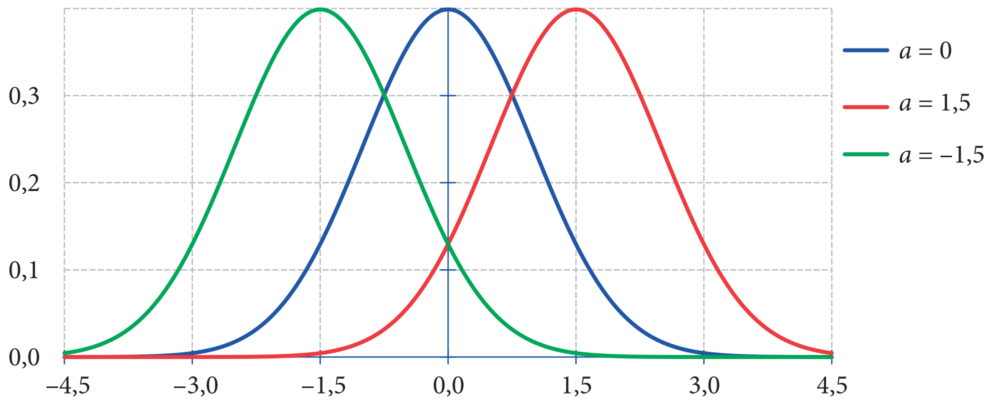


Рис. 4.11. Сравнение плотности нормального распределения с разными математическими ожиданиями

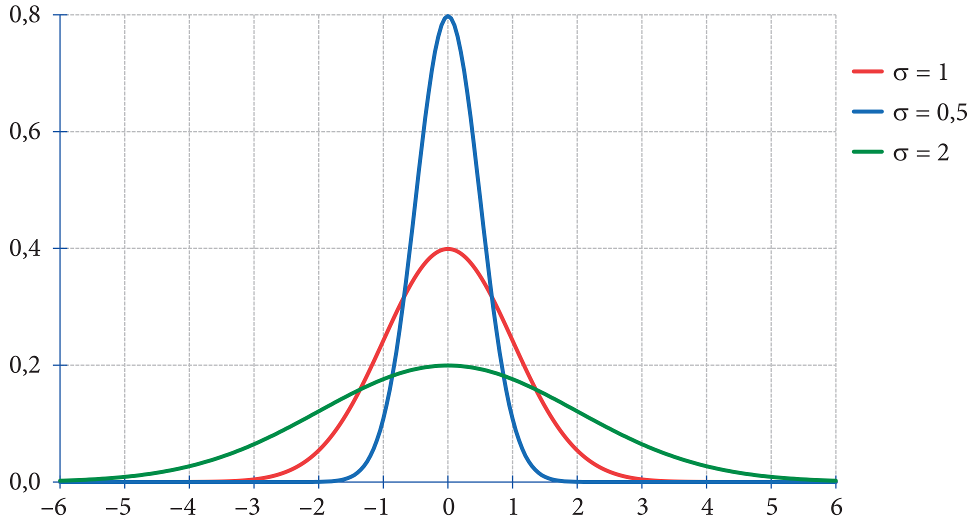


Рис. 4.12. Сравнение плотности вероятности нормального распределения с разными дисперсиями

Функция распределения такой случайной величины имеет вид:

$$F(x) = \frac{1}{2} + \Phi(x).$$

Плотность вероятностей случайной величины $X \sim \mathrm{N}\ (0; 1)$ равна:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}.$$

Таблицы для этой функции имеются в Приложении (табл. 1).

Легко видеть, что для случайной величины $X \sim \mathrm{N}\ (a; \sigma)$

$$f(x) = \frac{1}{\sigma}\varphi\left(\frac{x-a}{\sigma}\right), \quad F(x) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right).$$

Числовые характеристики случайной величины X ~ N (a; σ)

Параметры a и σ представляют собой соответственно математическое ожидание и среднеквадратичное отклонение случайной величины X:

$$\mathrm{E}(X) = a, \quad \sigma(X) = \sigma.$$

Отсюда $\mathrm{D}(X) = \sigma^2$.

Если $X \sim \mathrm{N}\ (a; \sigma)$, то для нахождения вероятности попадания этой случайной величины в заданный интервал (α, β) используется формула

$$\mathrm{P}(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right). \tag{4.10}$$

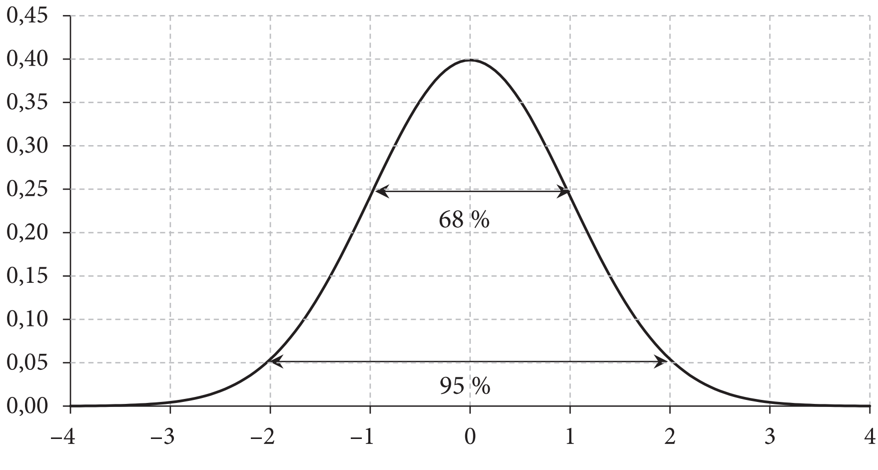


Рис. 4.13. Стандартное нормальное распределение с областями, содержащими 68 % и 95 % всех наблюдений

В частности, вероятность попадания случайной величины $X \sim N\,(a;\,\sigma)$ в интервал $(a-\varepsilon,\,a+\varepsilon)$, симметричный относительно центра распределения a, находится по формуле

$$P\left(a-\varepsilon < X < a+\varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)$$

или

$$P\left(\left|X-a\right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right). \tag{4.11}$$

Эта формула носит название *формулы Лапласа.*

В частности, если $\varepsilon = 3\sigma$, то

$$P\left(\left|X-a\right| < 3\sigma\right) = 2\Phi\left(\frac{3\sigma}{\sigma}\right) = 2\Phi(3) \approx 0{,}9973 \approx 1,$$

т. е. практически достоверно, что случайная величина $X \sim N\,(a;\,\sigma)$ принимает значения в промежутке $(a-3\sigma,\,a+3\sigma)$. Это утверждение называют «правилом трех сигм».

Для односторонних интервалов с использованием формулы (4.9) находим:

$$P\left(X < \beta\right) = F(\beta) = \frac{1}{2} + \Phi\left(\frac{\beta-a}{\sigma}\right); \tag{4.12}$$

$$P\left(X > \alpha\right) = \frac{1}{2} - \Phi\left(\frac{\alpha-a}{\sigma}\right). \tag{4.13}$$

Пример 4.12. Стоимость некоторой ценной бумаги нормально распределена. В течение последнего года 20 % рабочих дней она была ниже 88 ден. ед., а 75 % — выше 90 ден. ед. Найти:

а) математическое ожидание и среднее квадратичное отклонение стоимости ценной бумаги;

б) вероятность того, что в день покупки стоимость будет заключена в пределах от 85 до 96 ден. ед.;

в) с надежностью 0,95 определить максимальное отклонение стоимости ценной бумаги от среднего (прогнозного) значения (по абсолютной величине).

Решение. а) По условию

$$0{,}2 = P(X < 88) = 0{,}5 + \Phi\left(\frac{88-a}{\sigma}\right) \Rightarrow \Phi\left(\frac{88-a}{\sigma}\right) = -0{,}3$$

и

$$0{,}75 = P(X > 90) = 0{,}5 - \Phi\left(\frac{90-a}{\sigma}\right) \Rightarrow \Phi\left(\frac{90-a}{\sigma}\right) = -0{,}25.$$

По таблицам значений функции Лапласа (Приложение, табл. 2) находим

$$\frac{a-88}{\sigma}=\Phi^{-1}(0,3)=0,84 \quad \text{и} \quad \frac{a-90}{\sigma}=\Phi^{-1}(0,25)=0,67.$$

Из системы уравнений $\begin{cases} a-90=0,67\sigma, \\ a-88=0,84\sigma \end{cases}$ получаем математическое ожидание

и среднее квадратичное отклонение стоимости ценной бумаги: $a = 97,8824$; $\sigma = 11,7647$.

б) $P\left(85<X<96\right)=\Phi\left(\frac{96-97,8824}{11,7647}\right)-\Phi\left(\frac{85-97,8824}{11,7647}\right)=$

$=\Phi(-0,16)-\Phi(-1,09)=-0,06356+0,36214=0,29858.$

в) По условию $P(|X-a|<\varepsilon) = 0,95$. С другой стороны, по формуле (4.11)

$$P\left(\left|X-a\right|<\varepsilon\right)=2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Отсюда

$$2\Phi\left(\frac{\varepsilon}{\sigma}\right)=0,95 \Rightarrow \Phi\left(\frac{\varepsilon}{\sigma}\right)=0,475 \quad \text{и} \quad \frac{\varepsilon}{\sigma}=\Phi^{-1}(0,475)=1,96.$$

Следовательно, максимальное отклонение стоимости ценной бумаги от прогнозного значения (по абсолютной величине) составит: $\varepsilon = 1,96$, $\sigma = 23,06$ ден. ед.

Пример 4.13. Значения теста *IQ* (коэффициента интеллекта) Стэнфорда — Бине распределены приблизительно по нормальному закону с математическим ожиданием $a = 100$ и средним квадратичным отклонением $\sigma = 16$. Записать выражения для функции распределения коэффициента интеллекта и плотности его распределения. Найти долю людей, у которых коэффициент интеллекта окажется: а) меньше 60; б) меньше 100; в) в пределах от 80 до 120; г) коэффициент интеллекта отклонится от 100 менее чем на 48; д) найти вероятность того, что из пяти независимо отобранных человек у двоих коэффициент интеллекта будет выше 92.

Решение. $F(x)=\frac{1}{16\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{(t-100)^2}{512}}dt=\frac{1}{2}+\Phi\left(\frac{x-100}{16}\right),$

$$f(x)=\frac{1}{16\sqrt{2\pi}}e^{-\frac{(x-100)^2}{512}}=\frac{1}{16}\varphi\left(\frac{x-100}{16}\right), \quad x\in\mathbf{R}.$$

а) Воспользуемся определением функции распределения $F(x) = P(X < x)$ и формулой (4.9):

$$P\left(X<60\right)=F(60)=\frac{1}{2}+\Phi\left(\frac{60-100}{16}\right)=\frac{1}{2}+\Phi(-2,5)=0,5-0,4938=0,0062.$$

б) Аналогично $P\left(X < 100\right) = F(100) = \frac{1}{2} + \Phi\left(\frac{100-100}{16}\right) = \frac{1}{2} + \Phi(0) = 0,5.$

в) Воспользуемся формулой (4.10):

$$P\left(80 < X < 120\right) = \Phi\left(\frac{120-100}{16}\right) - \Phi\left(\frac{80-100}{16}\right) = 2\Phi(1,25) = 0,7887.$$

г) Так как отклонение составляет $3\sigma = 48$, воспользуемся «правилом трех сигм»:

$$P\left(\left|X-100\right| < 3\cdot 16\right) = 2\Phi\left(\frac{48}{16}\right) = 2\Phi(3) \approx 0,9973.$$

д) Найдем сначала вероятность того, что у наудачу отобранного человека коэффициент интеллекта выше 92. Для этого воспользуемся формулой (4.13):

$$P\left(X > 92\right) = 1 - P\left(X \le 92\right) = 1 - P\left(X < 92\right) = 1 - F(92) =$$
$$= 1 - \left(\frac{1}{2} + \Phi\left(\frac{92-100}{16}\right)\right) = \frac{1}{2} + \Phi(0,5) = 0,5 + 0,19146 = 0,69146.$$

Тогда вероятность того, что из пяти независимо отобранных человек у двоих коэффициент интеллекта будет выше 92, по формуле Бернулли равна:

$$P_5(2) = C_5^2 \cdot 0,69146^2 \cdot 0,30854^3 = 0,14043.$$

4.10. Основные распределения в статистике

4.10.1. Распределение *хи*-квадрат

Пусть $X_1, \ldots, X_n$ — независимые стандартные нормально распределенные случайные величины, т.е. $X_k \sim N(0, 1)$. Тогда случайная величина

$$X_n = \sum_{k=1}^{n} X_k^2$$

имеет распределение *хи-квадрат* (χ^2-*распределение*)[11] с *n степенями свободы,* что обозначается как $X_n \sim \chi^2(n)$.

Плотность вероятности случайной величины, имеющей χ^2-распределение, зависит только от числа слагаемых n.

[11] Критерий χ^2 был предложен К. Пирсоном в 1900 г. Его работа рассматривается как фундамент современной математической статистики.

С в о й с т в а распределения *хи*-квадрат $\chi^2(n)$:

1) случайная величина $X_n = \sum_{k=1}^{n} X_k^2$ имеет плотность вероятности:

$$f(x,n)=\begin{cases}0, x \le 0,\\ \dfrac{(1/2)^{\frac{n}{2}}}{\Gamma(n/2)}\cdot x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, x>0,\end{cases}$$

где $\Gamma(m)=\int\limits_0^{+\infty} y^{m-1}\mathrm{e}^{-y}dy$ — гамма-функция Эйлера. В частности, $\Gamma(n+1)=n!$

Впервые χ^2-распределение было рассмотрено Р. Хельмертом (1876) и К. Пирсоном (1900);

2) если независимые случайные величины $X_k \sim \chi^2(n_k)$, $k=\overline{1,m}$, то их сумма $\sum_{k=1}^{n} X_k$ имеет распределение $\chi^2(n)$ с $n=\sum_{k=1}^{m} n_k$ степенями свободы:

$$\sum_{k=1}^{n} X_k \sim \chi^2\left(\sum_{k=1}^{m} n_k\right);$$

3) если $X_n \sim \chi^2(n)$, то $\mathrm{E}X_n = n$, $\mathrm{D}X_n = 2n$;

4) при $n = 1$ $\chi^2(1) = X^2$, где $X \sim \mathrm{N}(0, 1)$, плотность вероятности

$$f(x,1)=\frac{1}{\sqrt{2\pi x}}\mathrm{e}^{-\frac{x}{2}}, \quad x>0;$$

5) если $n = 1$, то распределение *хи*-квадрат совпадает с экспоненциальным распределением: $\chi^2(2) = \mathrm{Exp}(1/2)$;

6) распределение $\chi^2(n)$ обладает свойством асимптотической нормальности:

$$\frac{X_n - n}{\sqrt{2n}} \xrightarrow{F} U, \quad n\to\infty,$$

где случайная величина $U \sim \mathrm{N}(0, 1)$. Это означает, что при достаточно большом объеме n выборки можно приближенно считать $X_n \sim \mathrm{N}(n, 2n)$. Фактически эта аппроксимация имеет место уже при $n \ge 30$.

Графики плотности распределения $\chi^2(n)$ при различных n изображены на рис. 4.14.

На практике часто используют не плотность вероятности, а квантили распределения $\chi^2(n)$. *Квантилем распределения* $\chi^2(n)$, соответствующим уровню значимости α, называется такое значение $\chi^2_\alpha(n)$, при котором вероятность того, что значение случайной величины превысит данное значение, равна α:

$$P\left(\chi^2(n) > \chi^2_\alpha(n)\right) = \int_{\chi^2_\alpha(n)}^{\infty} f(x,n)\,dx = \alpha.$$

Значения квантилей приводятся в Приложении (табл. 4). В Microsoft Excel имеются функции ХИ2ОБР или ХИ2.ОБР. Для стандартного нормального распределения квантили уровня α (обозначаем u_α) являются решением уравнения

$$\Phi(u_\alpha) = \frac{1-\alpha}{2}.$$

С геометрической точки зрения нахождение квантиля $\chi^2_\alpha(n)$ заключается в нахождении такого значения $\chi^2(n) = \chi^2_\alpha(n)$, что площадь заштрихованной на рис. 4.15 фигуры была равна α.

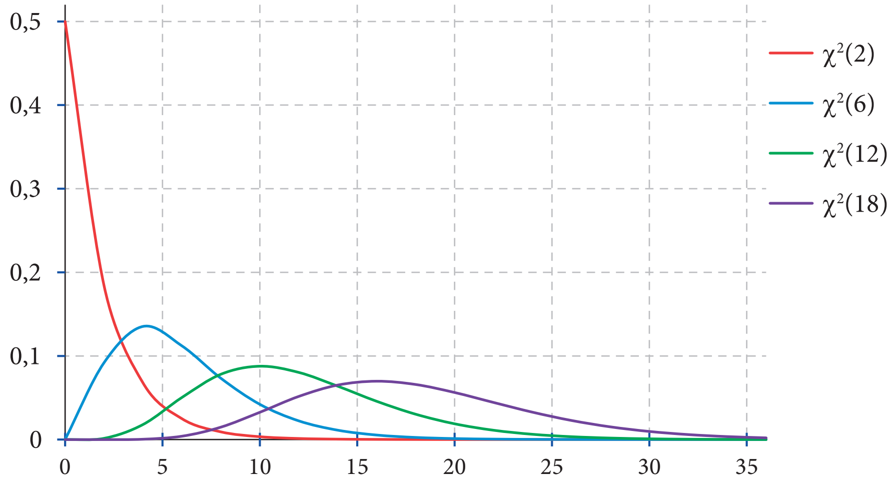


Рис. 4.14. Плотность распределения *хи*-квадрат

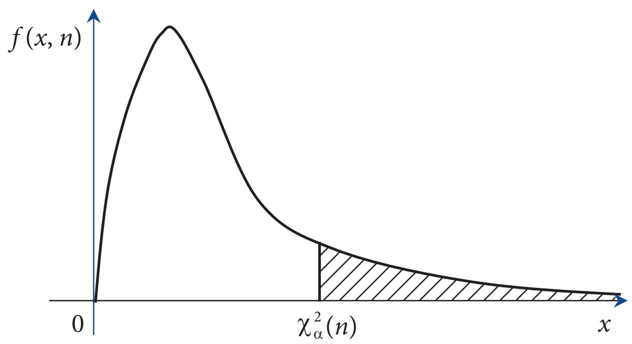


Рис. 4.15. Геометрическая иллюстрация нахождения квантиля *хи*-квадрат

4.10.2. Распределение Стьюдента

Рассмотрим две независимые случайные величины: Z, имеющую нормальное распределение $Z \sim \mathrm{N}(0; 1)$, и $X_n \sim \chi^2(n)$, распределенную по закону «*хи*-квадрат» с n степенями свободы. Тогда случайная величина $T_n = \dfrac{Z}{\sqrt{X_n/n}}$ имеет *распределение Стьюдента*[12] с *n степенями свободы*, что обозначают как $T_n \sim \mathrm{T}(n)$.

Свойства распределения Стьюдента $\mathrm{T}(n)$:

1. Случайная величина $T_n = \dfrac{Z}{\sqrt{X_n/n}}$ имеет плотность распределения:

$$f(t,n) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)} \cdot \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}},$$

где $\Gamma(m) = \int\limits_0^{+\infty} y^{m-1} \mathrm{e}^{-y} dy$ — гамма-функция Эйлера.

Графики плотностей $f(t, n)$, называемые кривыми Стьюдента, симметричны при всех $n = 1, 2, \ldots$ относительно оси ординат.

Частные случаи:

- распределение Стьюдента с одной степенью свободы ($n = 1$) — это распределение Коши, плотность которого равна:

$$f(t,1) = \frac{1}{\pi\left(1+t^2\right)};$$

- распределение Стьюдента с двумя степенями свободы ($n = 2$) — это распределение с плотностью вероятности

$$f(t,2) = \frac{1}{\left(2+t^2\right)^{\frac{3}{2}}};$$

[12] Это распределение было введено в 1908 г. английским статистиком В. Госсетом, работавшем на фабрике, выпускающей пиво. Вероятностно-статистические методы использовались для принятия экономических и технических решений на этой фабрике, поэтому ее руководство запрещало В. Госсету публиковать научные статьи под своим именем. Таким способом охранялась коммерческая тайна, «ноу-хау» в виде вероятностно-статистических методов, разработанных В. Госсетом. Однако он имел возможность публиковаться под псевдонимом «Стьюдент». История Госсета — Стьюдента показывает, что еще сто лет назад менеджерам Великобритании была очевидна большая экономическая эффективность вероятностно-статистических методов принятия решений.

• распределение Стьюдента с бесконечным числом степеней свободы асимптотически нормально, т. е.

$$T_n \xrightarrow{F} U, \quad n \to \infty,$$

где случайная величина $U \sim \mathrm{N}\,(0; 1)$. При $n \geq 30$ распределение Стьюдента $\mathrm{T}(n)$ практически не отличается от $\mathrm{N}\,(0; 1)$ (рис. 4.16).

2. Случайная величина T_n имеет математическое ожидание $\mathrm{E}T_n = 0$ для всех $n \geq 2$ и дисперсию $\mathrm{D}T_n = \dfrac{n}{n-2}$ при $n > 2$.

Приведем **теорему Фишера**. Пусть $X_1, \ldots, X_n$ — независимые нормально распределенные случайные величины, т. е. $X_i \sim \mathrm{N}\,(a, \sigma)$, $\overline{i = 1, n}$. Тогда:

• случайная величина $\dfrac{\left(\overline{X} - a\right)\sqrt{n}}{\sigma}$ имеет распределение $\mathrm{N}\,(0, 1)$;

• случайная величина $Y = \sum\limits_{k=1}^{n} \left(\dfrac{X_i - \overline{X}}{\sigma} \right)^2$ имеет распределение $\chi^2(n-1)$;

• случайная величина $\left(\overline{X} - a\right)\dfrac{\sqrt{n}}{s} = \dfrac{\left(\overline{X} - a\right)\sqrt{n}}{\sqrt{\dfrac{1}{n-1}\sum\limits_{k=1}^{n}\left(X_i - \overline{X}\right)^2}}$ имеет распределение $\mathrm{T}(n-1)$.

На практике используют квантили t-распределения. *Квантилем t-распределения*, соответствующим уровню значимости α, называется такое значение

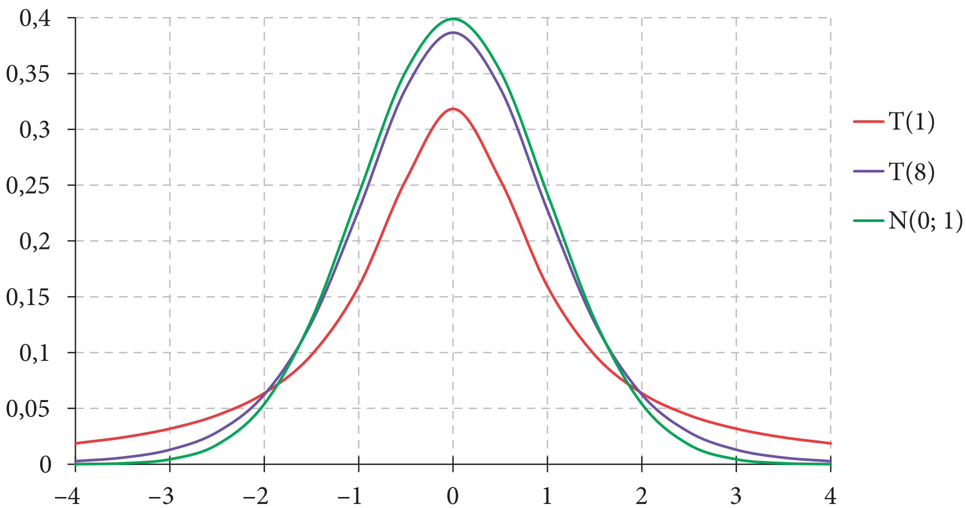


Рис. 4.16. Плотность вероятности распределения Стьюдента $\mathrm{T}(n)$

$t = t_{\alpha/2,n}$, при котором вероятность того, что модуль значения случайной величины превысит $t_{\alpha/2,n}$, равна α:

$$P\left(|t| > t_{\alpha/2,n}\right) = 2\int\limits_{t_{\alpha/2,n}}^{\infty} f(t)\,dt = \alpha.$$

Значения квантилей приводятся в Приложении (табл. 6). В Microsoft Excel имеется функция СТЬЮДРАСПОБР.

С геометрической точки зрения нахождение квантиля $t_{\alpha/2,n}$ заключается в нахождении таких значений $t = \pm t_{\alpha/2,n}$, чтобы площадь заштрихованной на рис. 4.17 фигуры была равна α.

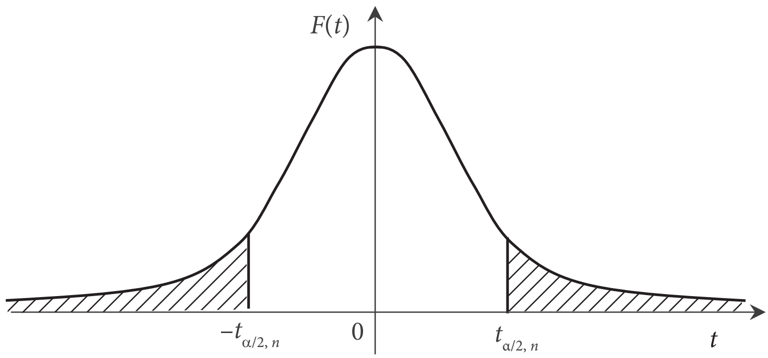


Рис. 4.17. Геометрическая иллюстрация нахождения квантиля распределения Стьюдента

4.10.3. Распределение Фишера — Снедекора

Распределением Фишера — Снедекора (или *F-распределением*) с m и n степенями свободы называется распределение случайной величины

$$F(m,n) = \frac{\frac{1}{m}\chi^2(m)}{\frac{1}{n}\chi^2(n)},$$

где $\chi^2(m)$ и $\chi^2(n)$ — независимые случайные величины, имеющие χ^2-распределение соответственно с m и n степенями свободы.

Плотность распределения Фишера — Снедекора:

$$f_F(x) = \begin{cases} 0, x \le 0; \\ C_0 \dfrac{x^{\frac{m-1}{2}}}{\left(n+mx\right)^{\frac{m+n}{2}}}, x > 0, \end{cases} \quad \text{где } C_0 = \frac{\Gamma\left(\dfrac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}.$$

График плотности распределения Фишера — Снедекора изображен на рис. 4.18. При $n \to \infty$ F-распределение стремится к нормальному закону.

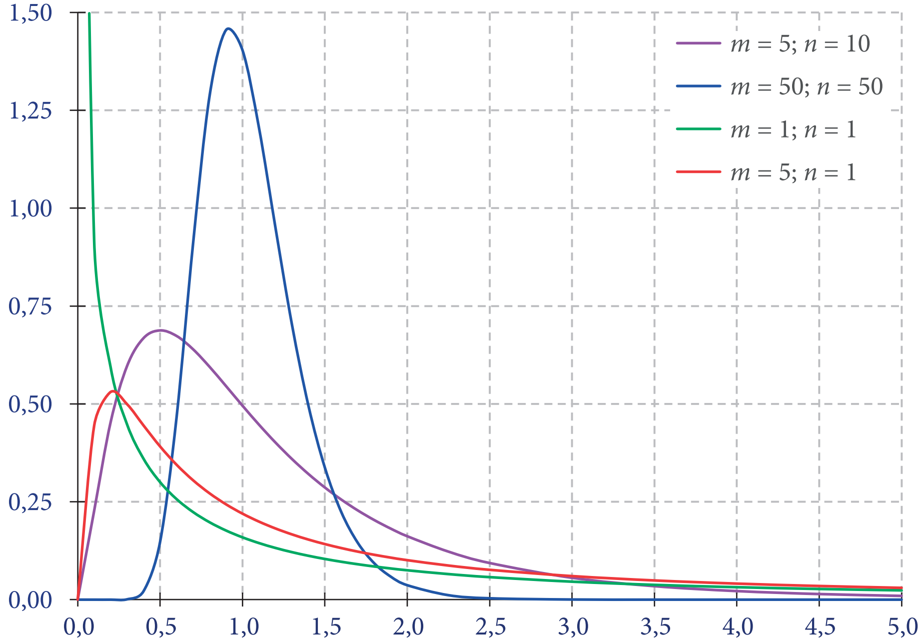


Рис. 4.18. Плотность вероятности распределения Фишера — Снедекора $F(m, n)$

Случайная величина $F(m, n)$ имеет математическое ожидание $\mathrm{E}F = \dfrac{n}{n-2}$ для всех $n > 2$ и дисперсию $\mathrm{D}F = \dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ при $n > 4$.

На практике используют квантили распределения. *Квантилем F-распределения*, соответствующим уровню значимости α, называется такое значение $F = F_\alpha(m,n)$, при котором вероятность того, что значение случайной величины превысит данное значение $F_\alpha(m,n)$, равна α:

$$\mathrm{P}\left(F > F_\alpha(m,n)\right) = \int_{F_\alpha(m,n)}^{\infty} f_F(x)dx = \alpha.$$

Значения квантилей приводятся в Приложении (табл. 6).

С геометрической точки зрения нахождение квантиля заключается в нахождении такого значения $F = F_\alpha(m,n)$, чтобы площадь заштрихованной на рис. 4.19 фигуры была равна α.

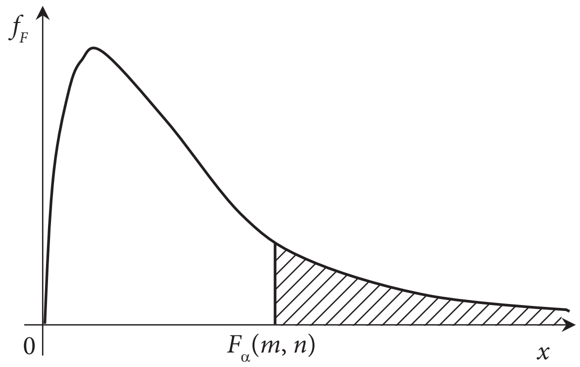


Рис. 4.19. Геометрическая иллюстрация нахождения квантиля распределения Фишера — Снедекора $F(m, n)$

Задачи для самостоятельного решения

4.1. $X \sim \text{Bi}\ (4; 0{,}3)$. Найти:

а) математическое ожидание и дисперсию случайной величины X;

б) ряд распределения случайной величины X;

в) функцию распределения $F(x)$ и построить ее график.

4.2. $X \sim \text{Bi}\ (8; 0{,}55)$. С помощью Microsoft Excel построить ряд распределения, многоугольник и функцию распределения случайной величины X числа успехов. Найти:

а) математическое ожидание, дисперсию, среднее квадратичное отклонение;

б) вероятности: $\text{P}(X = 5)$, $\text{P}(X \le 6)$, $\text{P}(1 \le X \le 5)$, $\text{P}(X > 5)$.

4.3. Количество обанкротившихся компаний по отношению к общему числу российских компаний в 2018 г. составило 0,3 %. Считая случайную величину X распределенной по закону Пуассона, найти:

а) вероятность того, что из 1 000 предприятий хотя бы одно обанкротится;

б) вероятность того, что из 1 000 предприятий обанкротится не более двух;

в) математическое ожидание и дисперсию случайной величины X.

4.4. В экзаменационном тесте 25 вопросов. В 20-ти из них 4 варианта ответов, а в остальных — 5. За каждый правильный вариант начисляется 4 балла. Сколько баллов в среднем получит студент, который на все вопросы отвечал наугад?

4.5. На предприятии для учета рабочего времени сотрудников была внедрена система биометрического контроля. Все 40 тыс. человек, работающих на данном предприятии, заходили на работу и уходили с нее по отпечаткам пальцев. Данная система контроля может допускать ошибочную идентификацию сотрудника с вероятностью 0,0001. Найти:

а) вероятность того, что за день система неправильно идентифицирует хотя бы одного сотрудника;

б) вероятность того, что за день будет допущено не более трех ошибок системы;

в) математическое ожидание и дисперсию числа ошибок системы за один рабочий день.

4.6. Выпускница экономического факультета ищет в интернете платье. Для этого она рассматривает платье и, если оно ей нравится и подходит по всем параметрам, покупает его, а если нет, то продолжает поиски. В среднем, прежде чем решиться на покупку, девушка оценивает 10 платьев. Рассматривая количество перебранных платьев как случайную величину, имеющую геометрическое распределение, найти ее дисперсию.

4.7. Полный цикл светофора, который регулирует перекресток, составляет 3 мин. Из них для автомобилей 1,5 мин горит зеленый свет, 30 сек — желтый и 1 мин — красный. Случайная величина X — время ожидания разрешающего сигнала светофора в случае остановки автомобиля, подъехавшего к данному перекрестку. Найти:

а) среднее время ожидания автомобиля и дисперсию случайной величины X;

б) функцию распределения и плотность распределения случайной величины X;

в) вероятность того, что автомобиль простоит на перекрестке меньше 10 сек.

4.8. Время обслуживания покупателя на кассе супермаркета является случайной величиной, распределенной по показательному закону с параметром $\lambda = 0{,}5$. Найти:

а) среднее время обслуживания покупателя;

б) вероятность того, что покупатель задержится у кассы более чем на пять минут;

в) вероятность того, что покупателя обслужат менее чем за минуту.

4.9. В центр звонков клиентов поступает поток звонков с интенсивностью 0,8 звонков в минуту. Найти вероятность того, что за 2 мин:

а) не придет ни одного звонка;

б) придет ровно один звонок;

в) придет хотя бы один звонок.

4.10. При продаже автомобиля марки N автосалон предоставляет гарантийное обслуживание сроком на один год. Известно, что время до первой поломки автомобиля данной марки является нормально распределенной случайной величиной с математическим ожиданием $a = 2$ и средним квадратичным отклонением $\sigma = 1$. Найти:

а) вероятность того, что покупатель воспользуется правом на гарантийное обслуживание;

б) вероятность того, что автомобиль потребует ремонта в течение года после окончания гарантийного обслуживания;

в) вероятность того, что автомобиль не потребует ремонта в течение трех лет эксплуатации.

4.11. Профессор математики обычно возвращался из университета на автобусе. Автобусы ходили с интервалом 10 мин. Найти:

а) вероятность того, что автобуса надо будет ждать более 8 мин;

б) вероятность того, что автобус придет в течение минуты;

в) среднее время ожидания автобуса.

4.12. Профессор математики узнал расписание автобуса и теперь не ждет его, а приходит по расписанию. Оказалось, что, хотя автобусы в дневные часы ходят точно с интервалом 10 мин, в «час пик» они выбиваются из графика и приходят позже. Профессор заметил, что в это время точно по расписанию автобус приходит только в 16 % случаев, и в среднем автобус опаздывает на 5 мин. Предположив, что время опоздания автобуса является нормально распределенной случайной величиной, найти:

а) вероятность того, что автобус опоздает более чем на 8 мин;

б) вероятность того, что автобус опоздает не более чем на одну минуту;

в) среднее квадратичное отклонение времени опоздания автобуса.

4.13. Случайная величина $Y \sim \mathrm{N}\,(2; 1)$. Пусть $X = 2Y + 5$. Найти вероятности $\mathrm{P}(X > 10)$, $\mathrm{P}(2 < X < 5)$, $\mathrm{P}(X < 2)$, $\mathrm{P}(X = 3)$. Записать функции плотности и функции распределения вероятностей для X и построить их графики. Сформулировать для случайной величины X «правило трех сигм».

4.14. По правилам рыбной ловли во внутренних водоемах разрешается вылавливать только взрослых язей, длина которых достигла 35 см. В выбранном водоеме доля молодых особей составляет 22,7 %. Считая длину язя нормально распределенной случайной величиной и зная, что среднее квадратичное отклонение длины равно 12 см, найти:

а) среднюю длину язя;

б) процент рыб размером от 35 до 53 см;

в) вероятность выловить язя длиной более 90 см.

Ответы к задачам для самостоятельного решения

4.1. а) $\mathrm{E}X = 1{,}2$, $\mathrm{D}X = 0{,}84$;

б)

X	0	1	2	3	4
P	0,2401	0,4116	0,2646	0,0756	0,0081

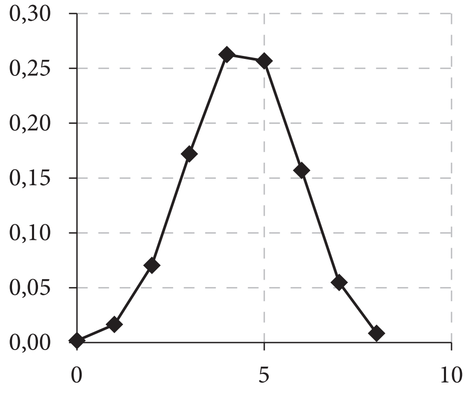
в) $F(x)=\begin{cases}0, x\le 0,\\ 0{,}2401, 0<x\le 1,\\ 0{,}6517, 1<x\le 2,\\ 0{,}9163, 2<x\le 3,\\ 0{,}9919, 3<x\le 4,\\ 1, x>4.\end{cases}$

4.2.

k	P(k)	$F(k)$
0	0,001682	0,001682
1	0,016441	0,018123
2	0,070333	0,088456
3	0,171925	0,260381
4	0,262663	0,523044
5	0,256826	0,779870
6	0,156949	0,936819
7	0,054808	0,991627
8	0,008373	1,000000

а) $\mathrm{E}(X)=4{,}4$; $\mathrm{D}(X)=1{,}98$; $\sigma(X)\approx 1{,}4$;

б) $\mathrm{P}(X=5)=0{,}256826$, $\mathrm{P}(X\le 6)=0{,}936819$, $\mathrm{P}(1\le X\le 5)=0{,}778188$, $\mathrm{P}(X>5)=0{,}22013$.



Многоугольник биномиального распределения

4.3. а) 0,9502; б) $\mathrm{P}(X\le 2)\approx 0{,}4232$; в) $\mathrm{E}(X)=\mathrm{D}(X)=\dfrac{1}{3}$.

4.4. 24 балла.

4.5. а) 0,9817; б) 0,4335; в) 4.

4.6. $\mathrm{D}X=90$.

4.7. а) $\mathrm{E}(X)=\dfrac{3}{4}$, $\mathrm{D}(X)=\dfrac{3}{16}$;

б) $f(x)\begin{cases}0, \; x\notin[0;1,5]\\ \dfrac{2}{3}, \; x\in[0;1,5]\end{cases}$, $F(x)=\begin{cases}0, \; x\notin[0;1,5]\\ \dfrac{2x}{3}, \; x\in[0;1,5]\end{cases}$; в) $\dfrac{1}{9}$.

4.8. а) $\mathrm{E}(X)=2$; б) $\mathrm{P}(X>5)\approx 0{,}0821$; в) $\mathrm{P}(X<1)\approx 0{,}3935$.

4.9. Случайная величина X — число звонков за 2 мин с параметром $\lambda=0{,}8\cdot 2=1{,}6$ распределена по закону Пуассона.

а) $\mathrm{P}(0)\approx 0{,}2019$; б) $\mathrm{P}(1)\approx 0{,}3230$; в) $\mathrm{P}(X\geq 1)\approx 0{,}7981$.

4.10. а) $\mathrm{P}(X\leq 1)\approx 0{,}1587$; б) $\mathrm{P}(1<X\leq 2)\approx 0{,}3413$; в) $\mathrm{P}(X>3)\approx 0{,}1587$.

4.11. а) 0,2; б) 0,1; в) 5.

4.12. $X\sim \mathrm{N}\,(5;5)$; а) $\mathrm{P}(X>8)\approx 0{,}2742$; б) $\mathrm{P}(X\leq 1)\approx 0{,}2119$; в) $\sigma\approx 5$.

4.13. $X\sim \mathrm{N}\,(9;2)$; $\mathrm{P}(X>10)\approx 0{,}3085$; $\mathrm{P}(2<X<5)\approx 0{,}0225$; $\mathrm{P}(X<2)\approx 0{,}0002$; $\mathrm{P}(X=3)=0$.

«Правило трех сигм»: практически достоверно, что случайная величина $X\sim \mathrm{N}\,(9;2)$ принимает значения в промежутке (3, 15).

4.14. а) $\mathrm{E}(X)\approx 44$; б) $\mathrm{P}(35<X<53)\approx 0{,}547$; в) $\mathrm{P}(X>90)\approx 0{,}0001$.

5. МНОГОМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

5.1. Законы распределения системы случайных величин

Пусть $X_1 = X_1(\omega)$, …, $X_n = X_n(\omega)$ — случайные величины, определенные на одном и том же пространстве элементарных событий Ω. Под *n-мерной случайной величиной* (*случайным вектором)* понимается упорядоченный набор n случайных величин $X = (X_1, \ldots, X_n)$.

Закон распределения вероятностей случайного вектора $X = (X_1, \ldots, X_n)$ задается в виде функции совместного распределения

$$F_X\left(x_1, x_2, \ldots, x_n\right) = \mathrm{P}\left(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n\right),$$

где $x_1, \ldots, x_n$ — любые действительные числа, или в виде плотности распределения

$$f_X\left(x_1, \ldots, x_n\right) = \frac{\partial^n\left(F_X\left(x_1, x_2, \ldots, x_n\right)\right)}{\partial x_1 \partial x_2 \ldots \partial x_n}.$$

Многомерные случайные величины, так же как и одномерные, могут быть *дискретными* (когда наборы возможных значений образуют конечное или счетное множество) или *непрерывными* (когда множество наборов возможных значений несчетно).

Ниже рассматриваются *двумерные* случайные величины.

Функция распределения вероятностей двумерной случайной величины представляет собой вероятность совместного выполнения двух неравенств $X < x$ и $Y < y$:

$$F\left(x, y\right) = F_{XY}\left(x, y\right) = \mathrm{P}\left(X < x, Y < y\right),$$

где x и y — любые действительные числа.

Вероятность попадания случайной величины (X, Y) в прямоугольник со сторонами, параллельными координатным осям, находится по формуле

$$\mathrm{P}\left(x_1 \le X \le x_2,\ \ y_1 \le Y \le y_2\right) = F\left(x_2, y_2\right) - F\left(x_2, y_1\right) - F\left(x_1, y_2\right) + F\left(x_1, y_1\right).$$

Свойства функции распределения $F(x, y)$:

1) $0 \le F(x, y) \le 1$;

2) функция $F(x, y)$ — неубывающая функция по каждому из аргументов;

3) $F(x, y)$ непрерывна слева по каждому из своих аргументов;

4) $F(+\infty, +\infty) = 1$ или $\lim\limits_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1$;

5) $F(x, +\infty) = F_X(x)$, $F(+\infty, y) = F_Y(y)$ — функции распределения случайных величин X и Y соответственно.

Закон распределения двумерной *дискретной* случайной величины (X, Y) можно задать формулой

$$p_{ij} = \mathrm{P}\left(X = x_i, Y = y_j\right), \; i, j = 1, 2, \ldots$$

В случае конечного числа значений распределение задается множеством

$$p_{ij} = \mathrm{P}\left(X = x_i, Y = y_j\right), \; i = 1, \ldots, n, \; j = 1, \ldots, m$$

или с помощью табл. 5.1, причем $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{m} p_{ij} = 1$.

Таблица 5.1

X	Y					
	y_1	y_2	…	y_j	…	y_m
x_1	p_{11}	p_{12}	…	p_{1j}	…	p_{1m}
x_2	p_{21}	p_{22}	…	p_{2j}	…	p_{2m}
…	…	…	…	…	…	…
x_i	p_{i1}	p_{i2}	…	p_{ij}	…	p_{im}
…	…	…	…	…	…	…
x_n	p_{n1}	p_{n2}	…	p_{nj}	…	p_{nm}

Вероятность совместного появления пары случайных величин (x_i, y_j) можно записать в виде

$$p_{ij} = \mathrm{P}\left(X = x_i, Y = y_j\right) = p\left(x_i, y_j\right) = p\left(x_i\right) p\left(y_j \mid x_i\right) = p\left(y_j\right) p\left(x_i \mid y_j\right),$$

где $p(x_i \mid y_j)$, $p(y_j \mid x_i)$ — условные вероятности.

Значение $F(x, y)$ функции распределения двумерной *дискретной* случайной величины находится суммированием вероятностей p_{ij} с индексами i, j, для которых $x_i < x$, $y_j < y$, т. е.

$$F(x, y) = \sum_{i:\, x_i < x} \sum_{j:\, y_j < y} p_{ij}.$$

Законы распределения каждой из компонент двумерной случайной величины (так называемые *маргинальные* законы распределения) восстанавливаются по таблице распределения 5.1 при помощи формул:

$$p(x_i) = p(X = x_i) = p_{i1} + p_{i2} + \ldots + p_{im} = \sum_{j=1}^{m} p_{ij} \quad \left(i = \overline{1,n}\right); \tag{5.1}$$

$$p(y_j) = p(Y = y_j) = p_{1j} + p_{2j} + \ldots + p_{nj} = \sum_{i=1}^{m} p_{ij} \quad \left(j = \overline{1,m}\right). \tag{5.2}$$

Здесь $\sum_{i=1}^{n} p(x_i) = 1, \ \sum_{j=1}^{m} p(y_j) = 1.$

Интегральные функции каждой из одномерных случайных величин X и Y определяются соответственно суммированием вероятностей:

$$F_X(x) = F(x, +\infty) = \sum_{i:\, x_i < x} p(x_i) = \sum_{i:\, x_i < x} \sum_{j=1}^{m} p_{ij};$$

$$F_Y(y) = F(+\infty, y) = \sum_{j:\, y_j < y} p(y_j) = \sum_{j:\, y_j < y} \sum_{i=1}^{n} p_{ij}.$$

Двумерная случайная величина (X, Y) называется *непрерывной*, если существует неотрицательная кусочно-непрерывная функция $f(x, y) = f_{XY}(x, y)$ такая, что для любых действительных чисел x и y функцию распределения $F(x, y)$ можно представить в виде

$$F(x, y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f(t, \tau) d\tau \right) dt. \tag{5.3}$$

При этом функцию $f(x, y)$ называют *плотностью распределения* (*плотностью вероятности*) случайной величины (X, Y).

В точках непрерывности $f(x, y)$ плотность распределения выражается через функцию распределения формулой

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Если $G \subset \mathrm{R}^2$, то вероятность попадания случайной величины (X, Y) в область G равна двойному интегралу от плотности по этой области:

$$\mathrm{P}\big((X, Y) \in G\big) = \iint_G f(x, y)\, dx\, dy. \tag{5.4}$$

Плотность распределения *непрерывной* двумерной случайной величины обладает следующими свойствами:

1) $f(x, y) \geq 0, \ \forall x, y;$

2) $\int\limits_{-\infty}^{+\infty}\left(\int\limits_{-\infty}^{\infty} f(x,y)\,dy\right)dx = 1.$

На основании (5.3) и свойства 5 функции $F(x, y)$ получаем:

$$F_X(x) = \int\limits_{-\infty}^{x}\left(\int\limits_{-\infty}^{+\infty} f(t,\, y)\,dy\right)dt,\;\; F_Y(y) = \int\limits_{-\infty}^{y}\left(\int\limits_{-\infty}^{+\infty} f(x,\, \tau)\,dx\right)d\tau,$$

где $F_X(x)$, $F_Y(y)$ — функции распределения случайных величин X и Y.

По плотности вероятности двумерной случайной величины (X, Y) находим плотности вероятности одномерных случайных величин X и Y:

$$f_X(x) = F_X'(x) = \int\limits_{-\infty}^{\infty} f(x,y)\,dy;$$

$$f_Y(y) = F_Y'(y) = \int\limits_{-\infty}^{\infty} f(x,y)\,dx.$$

Случайные величины X и Y называются *независимыми*, если для любых действительных чисел x и y

$$\mathrm{P}\Big((X<x)\cap(Y<y)\Big) = \mathrm{P}(X<x)\mathrm{P}(Y<y)$$

или

$$F_{XY}(x,y) = F_X(x)F_Y(y). \tag{5.5}$$

Для *дискретных* случайных величин X и Y необходимым и достаточным условием *независимости* (выполнения (5.5)) является

$$\mathrm{P}\left(X = x_i, Y = y_j\right) = \mathrm{P}\left(X = x_i\right)\mathrm{P}\left(Y = y_j\right) = p(x_i)\,p(y_j)\;\; \forall i = 1, \ldots, n, j = 1, \ldots, m,$$

для *непрерывных*

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

во всех точках непрерывности $f_{XY}(x, y)$.

Пример 5.1. Два стрелка независимо один от другого производят по одному выстрелу, каждый по своей мишени. Вероятность попадания в мишень для первого стрелка p_1, для второго — p_2. Случайная величина X — число попаданий первого стрелка; Y — второго стрелка. Найти:

а) таблицу распределения случайной величины (X, Y);

б) функцию распределения $F(x, y)$ двумерной случайной величины (X, Y);

в) функции распределения $F_X(x)$ и $F_Y(y)$ случайных величин X и Y соответственно.

Решение. Приведем законы распределения случайных величин X и Y соответственно:

X	0	1
$p(x_i)$	$1-p_1$	p_1

,

Y	0	1
$p(y_j)$	$1-p_2$	p_2

Таблица распределения случайной величины (X, Y) имеет вид:

X	Y	
	0	1
0	$(1-p_1)(1-p_2)$	$(1-p_1)p_2$
1	$p_1(1-p_2)$	p_1p_2

При составлении таблицы мы воспользовались независимостью случайных величин X и Y: $p(x_i, y_j) = p(x_i)p(y_j)$.

Замечание. Зная законы распределения отдельных случайных величин X и Y, входящих в систему, найти закон распределения всей системы в общем случае нельзя. Но это можно сделать всегда, когда случайные величины X и Y, образующие систему, независимы.

Функцию распределения $F(x, y)$ находим, используя геометрическую интерпретацию определения функции распределения (рис. 5.1). Значение $F(x, y)$ — это вероятность попадания случайной точки (X, Y) в бесконечный квадрант с вершиной (x, y). Для вершины этого квадранта, согласно условию задачи, есть девять областей, образованных двумя вертикальными прямыми $x = 0$, $x = 1$ и двумя горизонтальными — $y = 0$, $y = 1$. На рис. 5.1 показан случай, когда вершина (x, y) находится внутри прямоугольника $0 < y \le 1$, $x > 1$. При этом внутри квадранта находятся две точки с координатами (0; 0) и (1; 0), соответственно, с вероятностями $(1-p_1)(1-p_2)$ и $p_1(1-p_2)$. Поэтому значение $F(x, y)$ при $0 < y \le 1$, $x > 1$ равно $(1-p_1)(1-p_2) + p_1(1-p_2) = 1 - p_2$. Аналогично рассматриваем остальные случаи. Функцию распределения $F(x, y)$ удобно представить в виде таблицы (табл. 5.2).

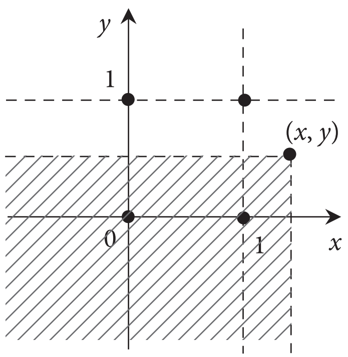


Рис. 5.1. Геометрическая интерпретация к определению $F(x, y)$

Таблица 5.2

X	Y		
	$y \le 0$	$0 < y \le 1$	$y > 1$
$x \le 0$	0	0	0
$0 < x \le 1$	0	$(1-p_1)(1-p_2)$	$1-p_1$
$x > 1$	0	$1-p_2$	1

Частные интегральные функции распределения $F_X(x)$ и $F_Y(y)$ находим из условий $F_X(x) = F(x, +\infty)$ и $F_Y(y) = F(+\infty, y)$. В итоге (сравнивая, соответственно, с последним столбцом и последней строкой таблицы) получим:

$$F_X(x) = \begin{cases} 0, x \le 0, \\ 1 - p_1, 0 < x \le 1, \\ 1, x > 1, \end{cases} \quad F_Y(y) = \begin{cases} 0, y \le 0, \\ 1 - p_2, 0 < y \le 1, \\ 1, y > 1. \end{cases}$$

Случайные величины X и Y независимы, поэтому $F_{XY}(x, y) = F_X(x)F_Y(y)$.

Пример 5.2. Закон распределения двумерного дискретного случайного вектора (X, Y) задан таблицей:

X	Y			
	0	1	2	3
1	0,1	0,1	0,1	0,1
2	0,05	0,05	0,05	0,05
3	0,1	0,1	0,1	0,1

Определить, зависимы или независимы компоненты X и Y.

Решение. Составим законы распределения компонент X и Y:

X	Y				
	0	1	2	3	$p(x_i)$
1	0,1	0,1	0,1	0,1	0,4
2	0,05	0,05	0,05	0,05	0,2
3	0,1	0,1	0,1	0,1	0,4
$p(y_j)$	0,25	0,25	0,25	0,25	1

Проверим теперь выполнение условия $p_{ij} = p(x_i)p(y_j)$ для всех пар индексов $i = 1, 2, 3$, $j = 1, 2, 3, 4$. Очевидно, что это условие выполнено для любых индексов i и j. Значит, составляющие X и Y независимы.

Пример 5.3. Совместная плотность распределения случайных величин X и Y имеет вид:

$$f(x, y) = \begin{cases} C(x + y), 0 \le x \le 2, 0 \le y \le 2, \\ 0 \quad \text{(в остальных случаях)}. \end{cases}$$

Найти:

а) параметр C;

б) плотности распределения случайных величин X и Y;

в) вероятности: $\mathrm{P}(X > 1)$, $\mathrm{P}(X + Y > 1)$;

г) функцию распределения $F(x, y)$;

д) функции распределения отдельных компонент X и Y.

Исследовать случайные величины X и Y на независимость.

Решение. а) Для нахождения константы C воспользуемся свойством плотности

$$\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{\infty} f(x,y)\,dy\right)dx = 1.$$

Вычисляем интеграл:

$$\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{\infty} f(x,y)\,dy\right)dx = \int_{0}^{2}\left(\int_{0}^{2} C(x+y)\,dy\right)dx = C\int_{0}^{2}\left(xy+\frac{y^2}{2}\right)\Bigg|_{0}^{2}dx =$$

$$= C\int_{0}^{2}(2x+2)\,dx = C\left(x^2+2x\right)\Big|_{0}^{2} = 8C.$$

Следовательно, $C = \dfrac{1}{8}$.

б) Найдем плотность распределения случайной величины X:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{0}^{2}\frac{1}{8}(x+y)\,dy = \frac{1}{8}\left(xy+\frac{y^2}{2}\right)\Bigg|_{0}^{2} = \frac{1}{4}(x+1), \quad \text{если } x \in [0;2];$$

$f_X(x) = 0$, если $x \notin [0;2]$,

т. е. $f_X(x) = \begin{cases} \dfrac{1}{4}(x+1), & x \in [0,2], \\ 0, & x \notin [0,2]. \end{cases}$

В силу симметрии заданной области и плотности $f(x, y)$ относительно переменных x и y следует, что

$$f_Y(y) = \begin{cases} \dfrac{1}{4}(y+1), & y \in [0;2], \\ 0, & y \notin [0;2]. \end{cases}$$

в) Для нахождения вероятности $\mathrm{P}(X > 1)$ воспользуемся формулой (5.4) и рис. 5.2:

$$\mathrm{P}(X>1) = \int_{1}^{2}\left(\int_{0}^{2}\frac{1}{8}(x+y)\,dy\right)dx = \frac{1}{8}\int_{1}^{2}\left(xy+\frac{y^2}{2}\right)\Bigg|_{0}^{2}dx = \frac{1}{8}\left(x^2+2x\right)\Big|_{1}^{2} = \frac{5}{8}.$$

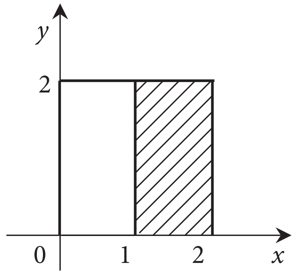


Рис. 5.2. Геометрическая иллюстрация вероятности P($X > 1$)

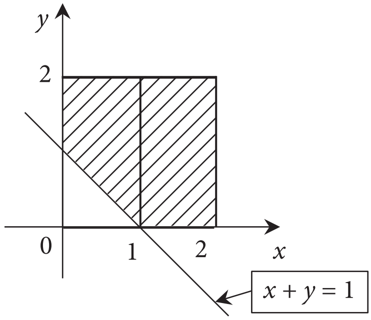


Рис. 5.3. Геометрическая иллюстрация вероятности P($X + Y > 1$)

Аналогично (рис. 5.3)

$$\mathrm{P}\left(X+Y>1\right)=\int_0^1\left(\int_{1-x}^2\frac{1}{8}\left(x+y\right)dy\right)dx+\int_1^2\left(\int_0^2\frac{1}{8}\left(x+y\right)dy\right)dx=\frac{1}{8}\int_0^1\left(xy+\frac{y^2}{2}\right)\Bigg|_{1-x}^{2}dx+\frac{5}{8}=$$

$$=\frac{1}{8}\int_0^1\left(2x+2-x\left(1-x\right)-\frac{\left(1-x\right)^2}{2}\right)dx+\frac{5}{8}=\frac{23}{24}.$$

Замечание. Вероятность P($X + Y > 1$) можно найти и по формуле

$$\mathrm{P}\left(X+Y>1\right)=1-\int_0^1\left(\int_0^{1-x}\frac{1}{8}\left(x+y\right)dy\right)dx=\frac{23}{24}.$$

г) По определению $F\left(x,y\right)=\int\limits_{-\infty}^{x}\left(\int\limits_{-\infty}^{y}f(t,\tau)d\tau\right)dt.$

Если $x \le 0$ или $y \le 0$, то $F(x, y) = 0$. При $0 \le x \le 2$ и $0 \le y \le 2$ получаем:

$$F\left(x,y\right)=\int_0^x\left(\int_0^y\frac{1}{8}\left(t+\tau\right)d\tau\right)dt=\frac{1}{8}\int_0^x\left(ty+\frac{y^2}{2}\right)dt=\frac{1}{16}\left(x^2y+xy^2\right).$$

Если $0 \le x \le 2$ и $y > 2$, то

$$F(x,y)=\int_0^x\left(\int_0^2 \frac{1}{8}(t+\tau)d\tau\right)dt+\int_0^x\left(\int_2^y 0\cdot d\tau\right)dt=\frac{1}{8}\int_0^x\left(t\tau+\frac{\tau^2}{2}\right)\Bigg|_0^2 dt=\frac{1}{8}\left(x^2+2x\right).$$

Если $x > 2$ и $0 \le y \le 2$, то аналогично

$$F(x,y)=\int_0^2\left(\int_0^y \frac{1}{8}(t+\tau)d\tau\right)dt=\frac{1}{8}\left(y^2+2y\right).$$

Если $x > 2$ и $y > 2$, то

$$F(x,y)=\int_0^2\left(\int_0^2 \frac{1}{8}(t+\tau)d\tau\right)dt=1.$$

Итак,

$$F(x,y)=\begin{cases} 0, & x\le 0 \text{ или } y\le 0,\\ xy(x+y)/16, & 0<x\le 2, 0<y\le 2,\\ x(x+2)/8, & 0<x\le 2,\ y>2,\\ y(y+2)/8, & x>2, 0<y\le 2,\\ 1, & x>2,\ y>2. \end{cases}$$

По найденным ранее плотностям $f_X(x)$ и $f_Y(y)$ находим функции распределения этих случайных величин:

$$F_X(x)=\begin{cases} 0, & x\le 0,\\ x(x+2)/8, & 0<x\le 2,\\ 1, & x>2, \end{cases}\quad F_Y(y)=\begin{cases} 0, & y\le 0,\\ y(y+2)/8, & 0<y\le 2,\\ 1, & y>2. \end{cases}$$

Наконец, заметим, что $f_{XY}(x, y) \ne f_X(x)f_Y(y)$ (так же как и $F_{XY}(x, y) \ne F_X(x)F_Y(y)$) для любых чисел x и y. Поэтому случайные величины X и Y зависимы.

Пример 5.4. Пусть $f(x, y)$ — совместная плотность случайных величин X и Y. Найти плотность распределения $f(z)$ случайной величины $Z = X + Y$.

Решение. Сначала запишем функцию распределения $F(z)$ случайной величины Z:

$$F(z)=\mathrm{P}(Z<z)=\mathrm{P}(X+Y<z).$$

Пользуясь формулой функции распределения $F(z)$ непрерывной двумерной случайной величины, получим:

$$F(z)=\mathrm{P}(X+Y<z)=\iint\limits_{x+y<z} f(x,y)dxdy=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{z-x} f(x,y)dy=\int_{-\infty}^{+\infty}dy\int_{-\infty}^{z-y} f(x,y)dx.$$

Так как $f(z) = F'(z)$, то

$$f(z) = f_Z(z) = \int\limits_{-\infty}^{+\infty} f\big(x, z-x\big)dx = \int\limits_{-\infty}^{+\infty} f\big(z-y, y\big)dy. \tag{5.6}$$

Из решения этого примера следует, что если X и Y — независимые случайные величины, то

$$f(z) = \int\limits_{-\infty}^{+\infty} f_X(x) f_Y\big(z-x\big)dx = \int\limits_{-\infty}^{+\infty} f_X\big(z-y\big) f_Y(y)\,dy,$$

где $f_X(x)$ и $f_Y(y)$ — плотности случайных величин соответственно X и Y. В этом случае последнее равенство называется *сверткой* функций $f_X(x)$ и $f_Y(y)$.

Пример 5.5. Совместная плотность распределения случайных величин X и Y имеет вид:

$$f\big(x, y\big) = \begin{cases} \dfrac{1}{\pi}\sin\big(x+y\big), 0 \le x \le \pi, 0 \le y \le \pi - x, \\ 0 \qquad \text{(в остальных случаях).} \end{cases} \tag{5.7}$$

Найти плотность распределения $f_{X+Y}(z)$ случайной величины $Z = X + Y$ (рис. 5.4).

Решение. По формуле (5.6), с учетом условия задачи, находим:

$$f_Z(z) = \int\limits_{-\infty}^{+\infty} f\big(x, z-x\big)dx = \int\limits_{0}^{\pi} f\big(x, z-x\big)dx.$$

В силу (5.7) $f(x, z-x) > 0$ в случае, когда $x \ge 0$ и $0 \le z - x \le \pi - x$, т.е. когда $0 \le x \le z \le \pi$. Следовательно, если $z \notin [0, \pi]$, то $f_Z(z) = 0$. Если же $z \in [0, \pi]$, то

$$f_Z(z) = \int\limits_0^z f\big(x, z-x\big)dx + \int\limits_0^{\pi} f\big(x, z-x\big)dx = \int\limits_0^z f\big(x, z-x\big)dx =$$

$$= \frac{1}{\pi}\int\limits_0^z \sin\big(x + z - x\big)dx = \frac{1}{\pi}\int\limits_0^z \sin z\,dx = \frac{1}{\pi} z \sin z.$$

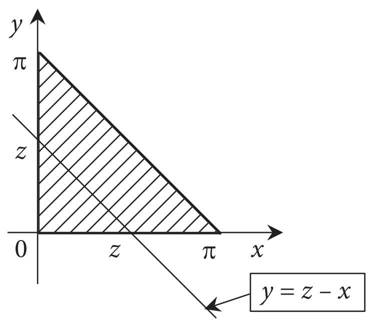


Рис. 5.4. Геометрическая иллюстрация плотности распределения $f_{X+Y}(z)$

Итак,

$$f_Z(z)=\begin{cases}\dfrac{1}{\pi}z\sin z, & 0\le z\le\pi,\\ 0 & \text{(в остальных случаях).}\end{cases}$$

5.2. Числовые характеристики двумерных случайных величин

Математическим ожиданием двумерной случайной величины (X, Y) является точка двумерного пространства $(\mathrm{E}X, \mathrm{E}Y)$, где величины $\mathrm{E}X$ и $\mathrm{E}Y$ — математические ожидания одномерных случайных величин X и Y соответственно.

Математическое ожидание и дисперсия каждой из одномерных *дискретных* случайных величин X и Y вычисляются по формулам:

$$\mathrm{E}X=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{ij}=\sum_{i=1}^{n}x_i p\left(x_i\right),\ \ \mathrm{E}Y=\sum_{j=1}^{m}\sum_{i=1}^{n}y_j p_{ij}=\sum_{j=1}^{m}y_j p\left(y_j\right);$$

$$\mathrm{D}X=\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i-\mathrm{E}X\right)^2 p_{ij}=\sum_{i=1}^{n}x_i^2 p\left(x_i\right)-\mathrm{E}^2X;$$

$$\mathrm{D}Y=\sum_{j=1}^{m}\sum_{i=1}^{n}\left(y_j-\mathrm{E}Y\right)^2 p_{ij}=\sum_{j=1}^{m}y_j^2 p\left(y_j\right)-\mathrm{E}^2Y,$$

где $p(x_i)$ и $p(y_j)$ определяются равенствами (5.1) и (5.2) соответственно.

Математическое ожидание и дисперсия *непрерывных* случайных величин X и Y с совместной плотностью вероятности $f(x, y)$ соответственно равны:

$$\mathrm{E}X=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x\,f(x,y)dxdy=\int_{-\infty}^{+\infty}x\,f_X(x)dx,\ \ \mathrm{E}Y=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y\,f(x,y)dxdy=\int_{-\infty}^{+\infty}y\,f_Y(y)dy;$$

$$\mathrm{D}X=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(x-\mathrm{E}X\right)^2 f(x,y)dxdy=\int_{-\infty}^{+\infty}x^2 f_X(x)dx-\mathrm{E}^2X;$$

$$\mathrm{D}Y=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(y-\mathrm{E}Y\right)^2 f(x,y)dxdy=\int_{-\infty}^{+\infty}y^2 f_Y(y)dy-\mathrm{E}^2Y.$$

Одной из характеристик статистической взаимосвязи двух случайных величин является *ковариация* случайных величин:

$$k_{XY}=\operatorname{cov}\left(X,Y\right)=\mathrm{E}\left[\left(X-\mathrm{E}X\right)\left(Y-\mathrm{E}Y\right)\right].$$

Последняя формула легко преобразуется к виду:

$$\operatorname{cov}\left(X,Y\right)=\mathrm{E}\left(XY\right)-\mathrm{E}\left(X\right)\mathrm{E}\left(Y\right), \tag{5.8}$$

где для двумерной *дискретной* случайной величины (X, Y)

$$\mathrm{E}\left(XY\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_{ij},$$

для двумерной *непрерывной* случайной величины (X, Y)

$$\mathrm{E}\left(XY\right)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xy\,f\left(x,y\right)dxdy.$$

Ковариация случайных величин удовлетворяет следующим свойствам:

$$\operatorname{cov}\left(X,Y\right)=\operatorname{cov}\left(Y,X\right),\ \operatorname{cov}\left(X,X\right)=\mathrm{D}X;$$

$$\operatorname{cov}\left(\alpha X,Y\right)=\alpha\operatorname{cov}\left(X,Y\right),\operatorname{cov}\left(X+Y,Z\right)=\operatorname{cov}\left(X,Z\right)+\operatorname{cov}\left(Y,Z\right);$$

$$\left|\operatorname{cov}\left(X,Y\right)\right|\le\sigma_X\sigma_Y,\quad \text{где}\ \ \sigma_X=\sqrt{\mathrm{D}X},\sigma_Y=\sqrt{\mathrm{D}Y};$$

для *независимых* случайных величин X и Y $\operatorname{cov}(X, Y) = 0$.

Коэффициент корреляции r_{XY} случайных величин X и Y равен ковариации нормированных случайных величин:

$$r_{XY}=\operatorname{corr}\left(X,Y\right)=\left(\operatorname{cov}\left(\frac{X-\mathrm{E}X}{\sigma_X},\frac{Y-\mathrm{E}Y}{\sigma_Y}\right)\right)=\frac{\operatorname{cov}\left(X,Y\right)}{\sigma_X\sigma_Y}. \qquad (5.9)$$

Случайные величины X и Y называют *коррелированными*, если $\operatorname{cov}(X, Y) \neq 0$ (или $r_{XY} \neq 0$), и *некоррелированными*, если $\operatorname{cov}(X, Y) = 0$ (или $r_{XY} = 0$).

Заметим, что коэффициент корреляции r_{XY} — количественная безразмерная характеристика зависимости случайных величин X и Y.

В частности, если $X = Y$, то $r_{XX} = 1$.

Свойства коэффициента корреляции:

1) $|r_{XY}| \le 1$;

2) дисперсия *суммы* произвольных случайных величин равна:

$$\mathrm{D}\left(X\pm Y\right)=\mathrm{D}X+\mathrm{D}Y\pm 2\operatorname{cov}\left(X,Y\right).$$

В частности, если X и Y некоррелированные, а тем более независимы,

$$\mathrm{D}\left(X\pm Y\right)=\mathrm{D}\left(X\right)+\mathrm{D}\left(Y\right);$$

3) если случайные величины X и Y независимы, то они и не коррелированы. Обратное, вообще говоря, неверно;

4) если случайные величины X и Y коррелированы, то они и зависимы. Обратное утверждение, вообще говоря, неверно.

Однако, если обе случайные величины X и Y распределены по *нормальному закону*, то из некоррелированности случайных величин следует их независимость и, наоборот, из независимости случайных величин следует их некоррелированность. К этому вопросу мы вернемся позже (см. п. 5.3);

5) для линейно связанных случайных величин X и $Y = \alpha X + \beta$ (α, $\beta \in \mathrm{R}$, $\alpha \neq 0$), и только для них:

$$|r_{XY}| = 1.$$

Коэффициент корреляции можно рассматривать как меру *линейности* взаимосвязи случайных величин X и Y: при возрастании одной случайной величины другая проявляет тенденцию также возрастать (или убывать). В первом случае $r_{XY} > 0$ и говорят, что случайные величины X и Y связаны *положительной корреляцией*, во втором случае $r_{XY} < 0$ и говорят, что X и Y связаны *отрицательной корреляцией*. Если X и Y связаны линейной функциональной зависимостью $Y = \alpha X + \beta$, то $r_{XY} = 1$, если $\alpha > 0$, и $r_{XY} = -1$, если $\alpha < 0$. Если зависимость близка к линейной, то $|r_{XY}| \cong 1$, если же величины X и Y слабо (линейно) зависимы, то $|r_{XY}| \cong 0$.

Пример 5.6. Имеется оценка совместного вероятностного распределения доходностей от инвестиций в акции компаний А и Б (табл. 5.3).

Таблица 5.3

X	Y		
	–10	10	15
–5	0	0,10	0,30
10	0,10	0,25	0,05
15	0	0,20	0

Случайная величина X описывает доходность инвестирования в акции компании А, случайная величина Y — в акции компании Б. Составить законы распределения случайных величин X и Y, установить наличие зависимости между ними, а также найти ожидаемую доходность и среднее квадратичное отклонение портфеля[13], состоящего на 40 % инвестирования в акции компании А и на 60 % — в акции компании Б. Найти коэффициент корреляции r_{XY}.

Решение. Для составления законов распределения случайных величин X и Y дополним табл. 5.3 строкой и столбцом (табл. 5.4).

Таблица 5.4

X	Y			$p(x_i)$
	–10	10	15	
–5	0	0,10	0,30	0,40
10	0,10	0,25	0,05	0,40
15	0	0,20	0	0,20
$p(y_j)$	0,10	0,55	0,35	$\sum\sum p_{ij} = 1$

[13] *Портфелем* называется пакет ценных бумаг, которым обладает инвестор.

При заполнении правого столбца и нижней строки воспользуемся формулами (5.1)–(5.2). Закон распределения компоненты X представляется рядом распределения:

X	-5	10	15	$\sum p(x_i)$
$p(x_i)$	0,40	0,40	0,20	1

Аналогично закон распределения составляющей Y:

Y	-10	10	15	$\sum p(y_j)$
$p(y_j)$	0,10	0,55	0,35	1

Так как $\mathrm{P}(X = x_i, Y = y_j) \neq \mathrm{P}(X = x_i)\mathrm{P}(Y = y_j)$ (например, $\mathrm{P}(X = 10,\ Y = 15) = 0{,}05 \neq 0{,}40 \cdot 0{,}35 = 0{,}14 = \mathrm{P}(X = 10)\ \mathrm{P}(Y = 15)$), то случайные величины *зависимы*.

По построенным законам распределений случайных величин определим числовые характеристики составляющих X и Y. Сначала определим ожидаемые доходности инвестирования в акции компаний А и Б:

$$\mathrm{E}X = \sum x_i p(x_i) = (-5)\cdot 0{,}40 + 10\cdot 0{,}40 + 15\cdot 0{,}20 = 5;$$

$$\mathrm{E}Y = \sum y_j p(y_j) = (-10)\cdot 0{,}10 + 10\cdot 0{,}55 + 15\cdot 0{,}35 = 9{,}75.$$

Затем определим *дисперсию доходностей* в акции компаний А и Б, которая определяет *риск*:

$$\mathrm{D}X = \sum x_i^2 p(x_i) - \mathrm{E}^2 X = (-5)^2\cdot 0{,}40 + 10^2\cdot 0{,}40 + 15^2\cdot 0{,}20 - 5^2 = 70;$$

$$\mathrm{D}Y = \sum y_j^2 p(y_j) - \mathrm{E}^2 Y = (-10)^2\cdot 0{,}10 + 10^2\cdot 0{,}55 + 15^2\cdot 0{,}35 - (9{,}75)^2 = 48{,}6875.$$

Далее находим среднее квадратичное отклонение, которое также определяет риск, сопоставимый с доходностью:

$$\sigma(X) = \sqrt{\mathrm{D}X} \approx 8{,}367,\ \sigma(Y) = \sqrt{\mathrm{D}Y} \approx 6{,}978.$$

Найдем ожидаемую доходность и среднее квадратичное отклонение портфеля, состоящего на 40 % из первого актива и на 60 % — из второго:

$$\mathrm{E}(0{,}4X + 0{,}6Y) = 0{,}4\mathrm{E}X + 0{,}6\mathrm{E}Y = 0{,}4\cdot 5 + 0{,}6\cdot 9{,}75 = 7{,}85;$$

$$\begin{aligned}\mathrm{D}(0{,}4X + 0{,}6Y) &= 0{,}4^2\,\mathrm{D}X + 0{,}6^2\,\mathrm{D}Y + 2\operatorname{cov}(0{,}4X;\ 0{,}6Y) = \\ &= 0{,}4^2\cdot 70 + 0{,}6^2\cdot 48{,}6875 + 2\cdot 0{,}4\cdot 0{,}6\operatorname{cov}(X,Y) = \\ &= 28{,}7275 + 0{,}48\cdot\operatorname{cov}(X,Y).\end{aligned}$$

Для нахождения cov(X, Y) составим ряд распределения $X \cdot Y$:

$X \cdot Y$	–100	–50	–75	100	150	$\sum\sum p_{ij}$
p_{ij}	0,10	0,10	0,30	0,25	0,25	1

$$\mathrm{E}\left(XY\right) = -100 \cdot 0,1 - 50 \cdot 0,1 + 75 \cdot 0,3 + 100 \cdot 0,25 + 150 \cdot 0,25 = 25.$$

По формуле (5.8) находим $\operatorname{cov}\left(X, Y\right) = 25 - 5 \cdot 9,75 = -23,75.$ Тогда:

$$\mathrm{D}\left(0,4X + 0,6Y\right) = 28,7275 + 0,48 \cdot \left(-23,75\right) = 17,3275;$$

$$\sigma\left(0,4X + 0,6Y\right) = \sqrt{17,3275} \approx 4,1626.$$

Итак, ожидаемая доходность портфеля, состоящего на 40 % инвестирования в акции компании А и на 60 % — компании Б, составляет 7,85 %, что значительно превышает ожидаемую доходность вложения в акции компании А (5 %), но менее доходности вложения в акции компании Б (9,75 %).

Среднее квадратичное отклонение совместного портфеля инвестирования в акции компаний А и Б составляет 4,163 %. Так как $\sigma(0,4X + 0,6Y)$ меньше $\sigma(X)$ ($\approx$8,367) и меньше $\sigma(Y)$ ($\approx$6,978), то вложение в акции двух компаний менее рискованно, чем вложение в акции каждой из компаний порознь.

Наконец, по формуле (5.9) находим:

$$r_{XY} = \operatorname{corr}\left(X, Y\right) = \frac{-23,75}{8,3666 \cdot 6,9776} \approx -0,407.$$

Коэффициент корреляции –0,407 говорит о том, что между случайными величинами X и Y существует *слабая отрицательная линейная зависимость*, т. е. при увеличении одной из них другая имеет тенденцию уменьшаться.

Упражнение. Найти закон распределения суммарного дохода $X + Y$.

Пример 5.7. Совместное распределение случайных величин X_1 и X_2 определяется условиями:

$$\mathrm{P}\left(X_1 X_2 = 0\right) = 1, \quad \mathrm{P}\left(X_i = 1\right) = 1, \quad \mathrm{P}\left(X_i = -1\right) = \frac{1}{4}, \quad i = 1,2.$$

Найти ковариацию cov(X_1, X_2). Являются ли величины X_1 и X_2 независимыми?

Решение. Найдем:

$$\mathrm{E}X_1 = -1 \cdot \frac{1}{4} + 0 \cdot 1 + 1 \cdot \frac{1}{4} = 0, \quad \mathrm{D}X_1 = \left(-1\right)^2 \cdot \frac{1}{4} + 0^2 \cdot 1 + 1^2 \cdot \frac{1}{4} - 0^2 = \frac{1}{2}.$$

Аналогично $EX_2 = 0, DX_2 = \frac{1}{2}.$ Так как $P(X_1X_2 = 0) = 1$, то $P(X_1X_2 \neq 0) = 0$.

Поэтому $E(X_1X_2) = 0 \cdot 1 + P(X_1X_2 \neq 0) \cdot 0 = 0$. Тогда

$$\operatorname{cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = 0.$$

Совместная вероятность $P(X_1 = 1, X_2 = 1) = 0$, так как $P(X_1X_2 \neq 0) = 0$. Но $P(X_1 = 1) \cdot P(X_2 = 1) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}.$ Следовательно:

$$P(X_1 = 1, X_2 = 1) \neq P(X_1 = 1) \cdot P(X_2 = 1),$$

т. е. X_1 и X_2 зависимы, хотя их ковариация равна нулю.

Итак, случайные величины X_1 и X_2 *некоррелированные*, но они *не являются независимыми*.

Упражнение 1. Доказать, что если две случайные величины X и Y принимают только два значения каждая и $\operatorname{cov}(X, Y) = 0$, то X и Y независимы.

Упражнение 2. Пусть случайная величина X распределена по нормальному закону $X \sim N(0; 1)$. Случайная величина $Y = f(X) = X^2 - 1$. Показать, что случайные величины X и Y не коррелированы.

Пример 5.8. Предприятие имеет две поточные линии с совместными технологическими процессами по сборке некоторой продукции. На первой линии выпускается с равной вероятностью либо одно изделие, либо два. На второй линии одно изделие выпускается с вероятностью 2/3, а два изделия — с вероятностью 1/3. Известно, что корреляция поточных линий равна $\frac{\sqrt{2}}{2}.$ Какова вероятность того, что на каждой из поточных линий будет выпускаться одинаковое количество изделий?

Решение. Пусть X — количество изделий, собранных на первой линии, Y — на второй. По условию

X	1	2
P	1/2	1/2

,

Y	1	2
P	2/3	1/3

Найдем математические ожидания и дисперсии случайных величин:

$$EX = \frac{3}{2}, EY = \frac{4}{3} \quad \Rightarrow \quad DX = 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2} - \left(\frac{3}{2}\right)^2 = \frac{1}{4}, \quad DY = 1 \cdot \frac{2}{3} + 4 \cdot \frac{1}{3} - \left(\frac{4}{3}\right)^2 = \frac{2}{9}.$$

Из формулы для корреляции найдем E(XY):

$$\frac{\sqrt{2}}{2}=\frac{\mathrm{E}\left(XY\right)-\frac{3}{2}\cdot\frac{4}{3}}{\sqrt{\frac{1}{4}}\cdot\sqrt{\frac{2}{9}}}\Rightarrow \mathrm{E}\left(XY\right)=\frac{13}{6}.$$

Составим таблицу для совместного распределения случайных величин X и Y:

X	Y		
	1	2	$p(x_i)$
1	p	$1/2-p$	$1/2$
2	$2/3-p$	$p-1/6$	$1/2$
$p(y_j)$	$2/3$	$1/3$	1

Здесь P($X=1$, $Y=1$) $=p$. Остальные ячейки заполняются исходя из законов распределения составляющих X и Y. Распределение $X\cdot Y$ имеет вид:

$X\cdot Y$	1	2	4	Σ
p_{ij}	p	$2/3-p+1/2-p$	$p-1/6$	1

Из полученной таблицы получаем:

$$\mathrm{E}\left(XY\right)=1\cdot p+2\cdot\left(\frac{7}{6}-2p\right)+4\cdot\left(p-\frac{1}{6}\right)=\frac{13}{6}\Rightarrow p=\frac{1}{2}.$$

Тогда искомая вероятность

$$\mathrm{P}\left(X=Y\right)=\mathrm{P}\left(X=1,Y=1\right)+\mathrm{P}\left(X=2,Y=2\right)=p+p-\frac{1}{6}=\frac{5}{6}.$$

5.3. Условные распределения составляющих двумерной случайной величины

Пусть (X, Y) — двумерная *дискретная* случайная величина и P($X=x_i$, $Y=y_j$) $=p_{ij}$, $i=1,\ \ldots,\ n$, $j=1,\ \ldots,\ m$.

Условные распределения $p_X(x_i\mid y_j)$ и $p_Y(y_j\mid x_i)$ определим соответственно равенствами:

$$p\left(x_i\mid y_j\right)=p_X\left(x_i\mid y_j\right)=\mathrm{P}\left(X=x_i\mid Y=y_j\right)=\frac{\mathrm{P}\left(X=x_i,Y=y_j\right)}{\mathrm{P}(Y=y_j)}=\frac{p_{ij}}{p\left(y_j\right)};\quad (5.10)$$

$$p\left(y_j \mid x_i\right) = p_Y\left(y_j \mid x_i\right) = \mathrm{P}\left(Y = y_j \mid X = x_i\right) = \frac{\mathrm{P}\left(Y = y_j, X = x_i\right)}{\mathrm{P}(X = x_i)} = \frac{p_{ij}}{p\left(x_i\right)}. \quad (5.11)$$

При фиксированном значении y_j $\sum_{i=1}^{n} p_X\left(x_i \mid y_j\right) = 1.$ Аналогично $\sum_{j=1}^{m} p_Y\left(y_j \mid x_i\right) = 1.$

Условные законы распределения составляющих X и Y представлены соответственно в табл. 5.5 и 5.6.

Таблица 5.5

X	x_1	x_2	…	x_i	…	x_n
$p_X(x_i \mid y_j)$	$p_X(x_1 \mid y_j)$	$p_X(x_2 \mid y_j)$	…	$p_X(x_i \mid y_j)$	…	$p_X(x_n \mid y_j)$

Здесь $j = \overline{1, m}.$

Таблица 5.6

Y	y_1	y_2	…	y_j	…	y_m
$p_Y(y_j \mid x_i)$	$p_Y(y_1 \mid x_i)$	$p_Y(y_2 \mid x_i)$	…	$p_Y(y_j \mid x_i)$	…	$p_Y(y_m \mid x_i)$

Здесь $i = \overline{1, n}.$

В табл. 5.5 перечислены возможные значения составляющей X и условные вероятности этих значений (таких законов m, так как $j = \overline{1, m}$).

В табл. 5.6 перечислены возможные значения составляющей Y и условные вероятности этих значений (таких законов n, так как $i = \overline{1, n}$).

Если случайные величины X и Y независимы, то условные законы распределения совпадают с безусловными.

Условное математическое ожидание случайной величины Y при условии $X = x_i$ определяется равенством

$$\mathrm{E}\left(Y \mid x_i\right) = \sum_{j=1}^{m} y_i p_Y\left(y_j \mid x_i\right) = \sum_{j=1}^{m} \frac{y_j p_{ij}}{p\left(x_i\right)} = \frac{\sum_{j=1}^{m} y_j p_{ij}}{\sum_{j=1}^{m} p_{ij}}, \quad i = \overline{1, n}. \quad (5.12)$$

В дискретном случае каждому значению x_i ставится в соответствие число: $x_i \to \mathrm{E}(Y \mid x_i).$ Таким образом, на множестве (x_1, …, x_n,) определена функция со значениями $\mathrm{E}(Y \mid x)$. Эту функцию называют регрессией Y на x (или Y от x) и обозначают:

$$\overline{y}_X = \mathrm{E}\left(Y \mid x\right).$$

На плоскости отметим точки (x_i, y_j), а затем те точки, координаты которых удовлетворяют уравнению регрессии $(x_i, \overline{y}_{x_i}).$ График уравнения регрессии —

множество точек $(x_i, \overline{y}_{x_i})$. Соединив их отрезками прямых, получим ломаную, которую называют *линией регрессии*. Она выражает зависимость Y от x в среднем (рис. 5.5).

Аналогично

$$\mathrm{E}\left(X \mid y_j\right) = \sum_{i=1}^{n} x_i p_X\left(x_i \mid y_j\right) = \frac{\sum\limits_{i=1}^{n} x_i p_{ij}}{\sum\limits_{i=1}^{n} p_{ij}}, \quad j = \overline{1, m}. \tag{5.13}$$

Регрессия X от Y: $\overline{x}_Y = \mathrm{E}(X \mid y)$. Эта функция $(y \to \mathrm{E}(X \mid y))$ определена на множестве $(y_1, \ldots, y_m)$. Она выражает зависимость X от y в среднем.

Соотношения (5.12) и (5.13) позволяют трактовать условные математические ожидания как центр «масс» p_{ij}, расположенных на вертикальных прямых $X = x_i$ в точках с ординатами y_j (на горизонтальных прямых $Y = y_j$ в точках с абсциссами $x = x_i$).

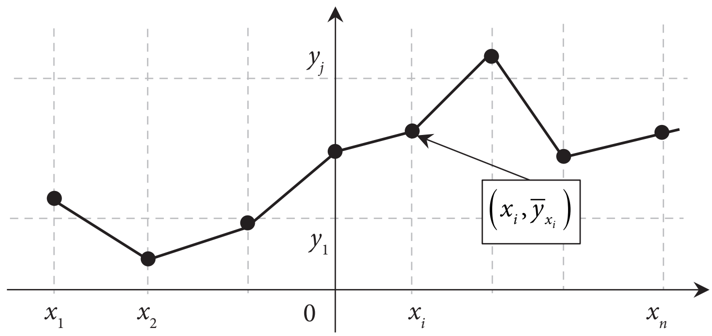


Рис. 5.5. Геометрическая иллюстрация регрессии Y на x

Уравнением *линейной регрессии* Y на (от) x называется уравнение $Y = \alpha X + \beta$, параметры которого минимизируют *остаточную дисперсию* $\mathrm{E}\left(Y - \overline{y}_X\right)^2$:

$$\overline{y}_X - \mathrm{E}Y = r_{XY} \frac{\sigma_Y}{\sigma_X}\left(x - \mathrm{E}X\right).$$

Остаточная дисперсия равна: $S^2_{\text{ост}} = \sigma^2_Y\left(1 - r^2_{XY}\right)$.

Условной дисперсией составляющей Y называется число:

$$\mathrm{D}\left(Y \mid x_i\right) = \sum_{j=1}^{m}\left(y_j - \overline{y}_{x_i}\right)^2 p_Y\left(y_j \mid x_i\right).$$

Аналогично определяется:

$$\mathrm{D}\left(X \mid y\right) = \sum_{i=1}^{n}\left(x_i - \overline{x}_{y_j}\right)^2 p_X(x_i \mid y), \quad \left(y = y_j,\ j = \overline{1, m}\right).$$

Пусть (X, Y) — *непрерывная* случайная величина с известной плотностью $f(x, y)$. *Условная функция распределения* непрерывной случайной величины X при условии, что случайная величина $Y = y$, равна:

$$F_X\left(x / y\right) = \frac{1}{f_Y(y)} \int_{-\infty}^{x} f\left(t, y\right) dt.$$

Аналогично

$$F_Y\left(y / x\right) = \frac{1}{f_X(x)} \int_{-\infty}^{y} f\left(x, \tau\right) d\tau.$$

Условная плотность распределения непрерывной случайной величины X при условии $Y = y$ определяется формулой

$$f(x \mid Y = y) = f_X(x \mid Y = y) = \frac{f_{XY}\left(x, y\right)}{f_Y(y)} = \frac{f\left(x, y\right)}{\int_{-\infty}^{\infty} f\left(x, y\right) dx}, \; f_Y(y) > 0,$$

а *условное математическое ожидание* и *дисперсия* — формулами:

$$\mathrm{E}\left(X \mid Y = y\right) = \int_{-\infty}^{+\infty} x f\left(x \mid Y = y\right) dx = \frac{\int_{-\infty}^{+\infty} x f\left(x, y\right) dx}{\int_{-\infty}^{+\infty} f\left(x, y\right) dx};$$

$$\mathrm{D}\left(X \mid Y = y\right) = \int_{-\infty}^{\infty} \left(x - \mathrm{E}(X \mid Y = y)\right)^2 f\left(x \mid Y = y\right) dx.$$

Аналогично:

$$f\left(y \mid X = x\right) = \frac{f_{XY}\left(x, y\right)}{f_X(x)} = \frac{f\left(x, y\right)}{\int_{-\infty}^{\infty} f\left(x, y\right) dy}, \quad f_X(x) > 0;$$

$$\mathrm{E}\left(Y \mid X = x\right) = \int_{-\infty}^{+\infty} y f\left(y \mid X = x\right) dy = \frac{\int_{-\infty}^{+\infty} y f\left(x, y\right) dy}{\int_{-\infty}^{+\infty} f\left(x, y\right) dy};$$

$$\mathrm{D}\left(Y \mid X = x\right) = \int_{-\infty}^{\infty} \left(y - \mathrm{E}\left(Y \mid X = x\right)\right)^2 f\left(y \mid X = x\right) dy.$$

Функция $\overline{y}_X = \mathrm{E}(Y \mid X = x)$, определенная на множестве возможных значений случайной величины X, является регрессией Y на x.

Заметим, что если непрерывные случайные величины X и Y независимы, то

$$f_Y\left(y \mid X=x\right)=f_Y\left(y\right), \quad f_X\left(x \mid Y=y\right)=f_X\left(x\right).$$

Отметим свойства условного математического ожидания:

$$\mathrm{E}(Y)=\mathrm{E}\left[\mathrm{E}\left(Y \mid x\right)\right].$$

Это равенство называют *формулой полного математического ожидания*. Аналогично

$$\mathrm{E}(X)=\mathrm{E}\left[\mathrm{E}\left(X \mid y\right)\right].$$

В случае *дискретной* случайной величины

$$\mathrm{E}(Y)=\sum_{i=1}^{n} p_i \mathrm{E}\left(Y \mid x_i\right), \quad \text{где } p_i=p\left(x_i\right). \tag{5.14}$$

Аналогично

$$\mathrm{E}(X)=\sum_{j=1}^{m} p_j \mathrm{E}\left(X \mid y_j\right), \quad \text{где } p_j=p\left(y_j\right). \tag{5.15}$$

Приведем два часто встречающихся распределения, задаваемых плотностью совместного распределения.

1. Равномерное распределение на множестве G:

$$f\left(x, y\right)=\begin{cases} \dfrac{1}{|G|}, & \left(x, y\right) \in G, \\ 0, & \left(x, y\right) \notin G. \end{cases}$$

Здесь $|G|$ — площадь множества G.

2. Двумерное нормальное распределение:

$$f_N\left(x, y\right)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-r^2}} \times$$
$$\times \exp\left\{-\frac{1}{2\left(1-r^2\right)}\left[\frac{\left(x-a_X\right)^2}{\sigma_X^2}+\frac{\left(y-a_Y\right)^2}{\sigma_Y^2}-2r\frac{\left(x-a_X\right)\left(y-a_Y\right)}{\sigma_X\sigma_Y}\right]\right\}.$$

Нетрудно показать, что параметры a_X и a_Y выражают математические ожидания случайных величин X и Y соответственно, параметры σ_X^2 и σ_Y^2 — их дисперсии, а $r=r_{XY}$ — коэффициент корреляции X и Y.

Плотности вероятности одномерных случайных величин:

$$f_X(x)=\int_{-\infty}^{+\infty} f_N\left(x, y\right)dy=\frac{1}{\sigma_X\sqrt{2\pi}}\mathrm{e}^{-\frac{\left(x-a_X\right)^2}{2\sigma_X^2}}$$

и аналогично

$$f_Y(y)=\frac{1}{\sigma_Y\sqrt{2\pi}}\mathrm{e}^{-\frac{\left(y-a_Y\right)^2}{2\sigma_Y^2}}.$$

Условные плотности вероятности одномерных случайных величин:

$$f_X\left(x\,|\,Y=y\right)=\frac{f_N\left(x,y\right)}{f_Y(y)}=\frac{1}{\sigma_X\sqrt{2\pi}\sqrt{1-r^2}}\mathrm{e}^{-\frac{1}{2\left(1-r^2\right)}\left(\frac{x-a_X}{\sigma_X}-r\frac{y-a_Y}{\sigma_Y}\right)^2};\tag{5.16}$$

$$f_Y\left(y\,|\,X=x\right)=\frac{f_N\left(x,y\right)}{f_X(x)}=\frac{1}{\sigma_Y\sqrt{2\pi}\sqrt{1-r^2}}\mathrm{e}^{-\frac{1}{2\left(1-r^2\right)}\left(\frac{y-a_Y}{\sigma_Y}-r\frac{x-a_X}{\sigma_X}\right)^2}.\tag{5.17}$$

Каждый из условных законов распределения случайных величин X и Y является нормальным с условным математическим ожиданием и условной дисперсией, определяемым по формулам:

$$\mathrm{E}\left(X\,|\,Y=y\right)=a_X+r\frac{\sigma_X}{\sigma_Y}\left(y-a_Y\right),\ \mathrm{D}\left(X\,|\,Y=y\right)=\sigma_X^2\left(1-r^2\right);\tag{5.18}$$

$$\mathrm{E}\left(Y\,|\,X=x\right)=a_Y+r\frac{\sigma_Y}{\sigma_X}\left(x-a_X\right),\ \mathrm{D}\left(Y\,|\,X=x\right)=\sigma_Y^2\left(1-r^2\right).\tag{5.19}$$

Заметим, что условные дисперсии $\mathrm{D}(X\,|\,Y=y)$ и $\mathrm{D}(Y\,|\,X=x)$ постоянны и не зависят от значений y или x. Это свойство называется *гомоскедастичностью* или *равноизменчивостью* условных нормальных распределений и имеет важное значение в статистическом анализе.

Если две нормально распределенные случайные величины X и Y не коррелированы ($r=r_{XY}=0$), то они и независимы ($f_N(x,y)=f_X(x)f_Y(y)$).

Пример 5.9. Случайные величины X и Y независимы и имеют соответственно нормальные распределения N(–1, 9), N(1, 16). Найти закон распределения случайной величины $Z=X+Y$. Найти вероятности: а) $\mathrm{P}(X-2Y>5)$; б) $\mathrm{P}(X+Y<0)$.

Решение. Используя свойства математического ожидания и дисперсии, находим:

$$a_Z=\mathrm{E}Z=\mathrm{E}\left(X+Y\right)=\mathrm{E}X+\mathrm{E}Y=-1+1=0;$$
$$\sigma_Z^2=\mathrm{D}Z=\mathrm{D}\left(X+Y\right)=\mathrm{D}X+\mathrm{D}Y=9+16=25.$$

Закон распределения случайной величины $Z=X+Y$ можно задать с помощью плотности вероятности нормального распределения:

$$f(z)=\frac{1}{5\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{50}}.$$

Найдем P($X - 2Y > 5$). Числовые характеристики случайной величины $X - 2Y$:

$$\mathrm{E}\left(X-2Y\right)=\mathrm{E}X-2\mathrm{E}Y=-1-2\cdot 1=-3;$$

$$\mathrm{D}\left(X-2Y\right)=\mathrm{D}X+2^2\mathrm{D}Y=9+4\cdot 16=73.$$

Напомним, что здесь случайные величины X и $(-2Y)$ (так же как и случайные величины X и Y) независимы и нормально распределены. Итак, $X - 2Y \sim \mathrm{N}\left(-3; \sqrt{73}\right)$. Воспользовавшись формулой (4.13), находим:

$$\mathrm{P}\left(X-2Y>5\right)=\frac{1}{2}-\Phi\left(\frac{5-\left(-3\right)}{\sqrt{73}}\right)=\frac{1}{2}-\Phi\left(0,9363\right)\approx 0,5-0,3254=0,1746.$$

Аналогично $X+Y \sim \mathrm{N}(0; 5)$ и согласно формуле (4.12):

$$\mathrm{P}\left(X+Y<0\right)=\frac{1}{2}+\Phi\left(\frac{0-0}{5}\right)=\frac{1}{2}.$$

Пример 5.10. Случайный вектор $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ имеет нормальное распределение с математическим ожиданием $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ и ковариационной матрицей $\begin{pmatrix} 9 & -5 \\ -5 & 25 \end{pmatrix}$.

а) Найти P($X_1 + 2X_2 > 20$).

б) Какое распределение имеет X_1 при условии, что $X_2 = 0$?

в) Составить уравнение регрессии Y на x.

г) Найти условную дисперсию компоненты Y.

Решение. а) По условию $\sigma_X^2 = 9, \sigma_Y^2 = 25, \mathrm{cov}(X_1, X_2) = -5.$ Числовые характеристики случайной величины $X_1 + 2X_2$:

$$\mathrm{E}\left(X_1+2X_2\right)=\mathrm{E}X_1+2\mathrm{E}X_2=1+2\cdot 2=5;$$

$$\mathrm{D}\left(X_1+2X_2\right)=\mathrm{D}X_1+4\mathrm{D}X_2+2\mathrm{cov}\left(X_1,2X_2\right)=9+4\cdot 25+2\cdot 2\cdot\left(-5\right)=8.$$

Итак,

$$X_1+2X_2\sim\mathrm{N}\left(5;\sqrt{89}\right);$$

$$\mathrm{P}\left(X_1+2X_2>20\right)=0,5-\Phi\left(\frac{20-5}{\sqrt{89}}\right)=0,5-\Phi\left(1,59\right)\approx 0,0559.$$

б) Используя (5.16), находим:

$$f_{X_1}\left(x\mid X_2=0\right)=\frac{1}{4\sqrt{\pi}}\mathrm{e}^{-\frac{\left(x-1,4\right)^2}{16}}.$$

Следовательно, случайная величина X_1 при условии, что $X_2 = 0$, распределена нормально, причем $(X_1 \mid X_2 = 0) \sim \mathrm{N}(1,4;\sqrt{8})$. Параметры $\mathrm{M}(X_1 \mid X_2) = 0) =$

$$= 1 + \left(\frac{-5}{3 \cdot 5}\right)\frac{3}{5}(0-2) = 1,4; \mathrm{D}(X_1 \mid X_2 = 0) = 9\left(1 - \frac{1}{9}\right) = 8$$ получены с использованием формул (5.18).

в) Условное математическое ожидание на основании (5.19) равно:

$$\mathrm{E}(Y \mid X = x) = 2 + \frac{(-5)}{3 \cdot 5} \cdot \frac{5}{3}(x-1) = 2 - \frac{5}{9}(x-1) = \frac{23}{9} - \frac{5}{9}x.$$

Таким образом, уравнение регрессии Y на x имеет вид: $y = \frac{23}{9} - \left(\frac{5}{9}\right)x$.

г) Условная дисперсия компоненты Y равна:

$$\mathrm{D}(Y \mid X = x) = 25\left(1 - \frac{1}{9}\right) = \frac{200}{9} \approx 22,22.$$

Пример 5.11. Найти условные законы распределения и числовые характеристики составляющих X и Y, заданных в примере 5.2. Построить регрессии Y на x и X на y.

Решение. Условные законы составляющей Y оформим в одной таблице. Воспользуемся формулами (5.11):

Y	-10	10	15	Σ
$p(y_j \mid x_1 = -5)$	$\frac{0}{0,40} = 0$	$\frac{0,10}{0,40} = \frac{1}{4}$	$\frac{0,30}{0,40} = \frac{3}{4}$	1
$p(y_j \mid x_2 = 10)$	$\frac{0,10}{0,40} = \frac{1}{4}$	$\frac{0,25}{0,40} = \frac{5}{8}$	$\frac{0,05}{0,40} = \frac{1}{8}$	1
$p(y_j \mid x_3 = 15)$	$\frac{0}{0,20} = 0$	$\frac{0,20}{0,20} = 0$	$\frac{0}{0,20} = 0$	1

Согласно (5.12):

$$\mathrm{E}(Y \mid x = -5) = -10 \cdot 0 + 10 \cdot \frac{1}{4} + 15 \cdot \frac{3}{4} = \frac{55}{4};$$

$$\mathrm{E}(Y \mid x = 10) = -10 \cdot \frac{1}{4} + 10 \cdot \frac{5}{8} + 15 \cdot \frac{1}{8} = \frac{45}{8};$$

$$\mathrm{E}(Y \mid x = 15) = -10 \cdot 0 + 10 \cdot 1 + 15 \cdot 0 = 10.$$

Ломаная, соединяющая точки $(x_i, \mathrm{E}(Y \mid x_i))$, является линией регрессии Y на x (рис. 5.6).

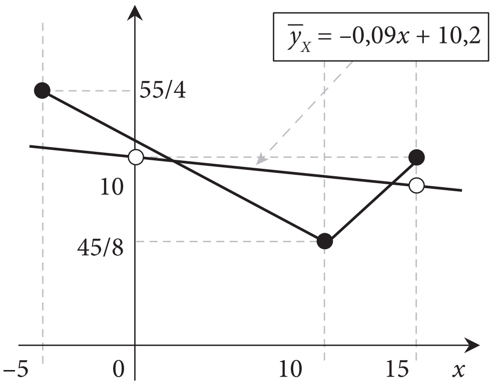


Рис. 5.6. Регрессия Y на x

Для построения уравнения линейной регрессии Y на x воспользуемся формулой

$$\overline{y}_X - \mathrm{E}Y = r_{XY}\frac{\sigma_Y}{\sigma_X}\left(x - \mathrm{E}X\right)$$

и результатами вычислений числовых характеристик примера 5.6:

$$\overline{y}_X - 9{,}75 = -0{,}407\frac{6{,}978}{8{,}367}\left(x-5\right) \Rightarrow \overline{y}_X = -0{,}09x + 10{,}20.$$

При этом остаточная дисперсия равна: $S^2_{\text{ост}} = \sigma^2_Y\left(1 - r^2_{XY}\right) = 40{,}63.$

Аналогично находим условные законы для составляющей X:

X	-5	10	15	Σ
$p\left(x_i \mid y_1 = -10\right)$	$\frac{0}{0{,}1} = 0$	$\frac{0{,}1}{0{,}1} = 1$	$\frac{0}{0{,}1} = 0$	1
$p\left(x_i \mid y_1 = 10\right)$	$\frac{0{,}10}{0{,}55} = \frac{2}{11}$	$\frac{0{,}25}{0{,}55} = \frac{5}{11}$	$\frac{0{,}20}{0{,}55} = \frac{4}{11}$	1
$p\left(x_i \mid y_1 = 15\right)$	$\frac{0{,}30}{0{,}35} = \frac{6}{7}$	$\frac{0{,}05}{0{,}35} = \frac{1}{7}$	$\frac{0}{0{,}35} = 0$	1

Согласно (5.13):

$$\mathrm{E}\left(X \mid y = -10\right) = -5\cdot 0 + 10\cdot 1 + 15\cdot 0 = 10;$$

$$\mathrm{E}\left(X \mid y = 10\right) = -5\cdot\frac{2}{11} + 10\cdot\frac{5}{11} + 15\cdot\frac{4}{11} = \frac{100}{11};$$

$$\mathrm{E}\left(X \mid y = 15\right) = -5\cdot\frac{6}{7} + 10\cdot\frac{1}{7} + 15\cdot 0 = -\frac{20}{7}.$$

Линия регрессии X на y изображена на рис. 5.7.

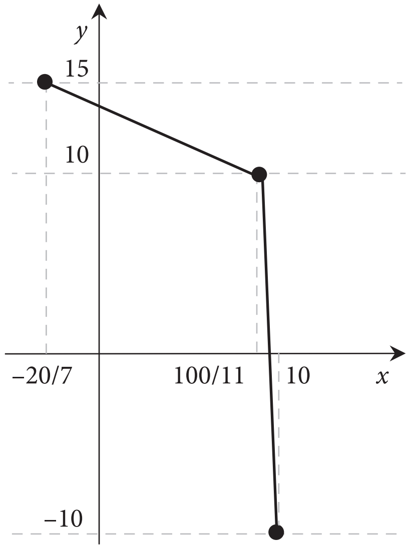


Рис. 5.7. Регрессия X на y

Проверим формулы полного математического ожидания ((5.14), (5,15)) для нашего примера:

$$\mathrm{E}\left[\mathrm{E}\left(X\mid y\right)\right]=0{,}10\cdot\mathrm{E}\left(X\mid y=-10\right)+0{,}55\cdot\mathrm{E}\left(X\mid y=10\right)+0{,}35\cdot\mathrm{E}\left(X\mid y=15\right)=$$

$$=0{,}10\cdot 10+0{,}55\cdot\frac{100}{11}+0{,}35\cdot\left(-\frac{20}{7}\right)=5=\mathrm{E}\left(X\right);$$

$$\mathrm{E}\left[\mathrm{E}\left(Y\mid x\right)\right]=0{,}40\cdot\mathrm{E}(Y\mid x=-5)+0{,}40\cdot\mathrm{E}(Y\mid x=10)+0{,}20\cdot\mathrm{E}(Y\mid x=15)=$$

$$=0{,}40\cdot\left(\frac{55}{4}\right)+0{,}40\cdot\frac{45}{8}+0{,}20\cdot 10=9{,}75=\mathrm{E}\left(Y\right).$$

5.4. Числовые характеристики многомерной случайной величины

Математическим ожиданием случайного вектора $X=(X_1, \ldots, X_n)$ называется вектор $\mathrm{E}X=(\mathrm{E}X_1, \ldots, \mathrm{E}X_n)$.

Ковариационной матрицей случайных величин $X_1, \ldots, X_n$ называется матрица

$$\mathrm{K}=\operatorname{cov}\left(X_i, X_j\right)=\begin{pmatrix}\sigma_1^2 & k_{12} & \ldots & k_{1n}\\ k_{21} & \sigma_2^2 & \ldots & k_{2n}\\ \ldots & \ldots & \ldots & \ldots\\ k_{n1} & k_{n2} & \ldots & \sigma_n^2\end{pmatrix},$$

где

$$k_{ij} = \operatorname{cov}\left(X_i, X_j\right) \equiv \mathrm{M}\left(X_i - \mathrm{M}X_i\right)\left(X_j - \mathrm{M}X_j\right),\ k_{ii} = \sigma_i^2 = \mathrm{D}X_i.$$

Ее определитель называется *обобщенной дисперсией* и может служить мерой рассеяния системы случайных величин $X_1, \ldots, X_n$.

Наряду с ковариационной матрицей рассматривают *корреляционную матрицу* R, составленную из коэффициентов корреляции $r_{ij} = r_{X_i X_j}$:

$$\mathrm{R} = \begin{pmatrix} 1 & r_{12} & \ldots & r_{1n} \\ r_{21} & 1 & \ldots & r_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ r_{n1} & r_{n2} & \ldots & 1 \end{pmatrix}.$$

Матрицы K и R *неотрицательно определены* и, кроме того, *симметричны*:

$$k_{ij} = k_{ji},\ r_{ij} = r_{ji}.$$

Случайные величины $X_1, \ldots, X_n$ называются независимыми, если для любых $x_1, \ldots, x_n \in \mathrm{R}$

$$F_X\left(x_1, \ldots, x_n\right) = F_{X_1}\left(x_1\right) F_{X_2}\left(x_2\right) \ldots F_{X_n}\left(x_n\right).$$

Случайные величины $X_1, \ldots, X_n$ называются попарно некоррелированными, если $k_{ij} = 0,\ \forall\, i \neq j$.

Если случайные величины $X_1, \ldots, X_n$ независимы, то они попарно не коррелированы. Обратное утверждение неверно.

Свойства EX и DX:

1) $\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}X_i$;

2) если случайные величины $X_1, \ldots, X_n$ независимы, то $\mathrm{E}\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} \mathrm{E}X_i$;

3) $\mathrm{D}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{D}X_i + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_{ij}$;

4) если случайные величины $X_1, \ldots, X_n$ не коррелированы, т.е. $k_{ij} = 0,\ \forall\, i \neq j$, то

$$\mathrm{D}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{D}X_i;$$

5) $\left|r_{ij}\right| \leq 1$ при $i \neq j$, $\left|r_{ii}\right| = 1$.

Пример 5.12. Дана матрица $\begin{pmatrix} 1 & x & x \\ x & 1 & x \\ x & x & 1 \end{pmatrix}$. При каких значениях x заданная матрица будет корреляционной матрицей трех случайных величин?

Решение. На основании свойства 5: $|x| \le 1$. Кроме того, корреляционная матрица R неотрицательно определена и симметрична: $r_{ij} = r_{ji}$. Так как условие симметричности выполнено, то остается проверить условие неотрицательности. Для этого воспользуемся критерием Сильвестра: квадратичная форма неотрицательно определена в том и только в том случае, когда ее угловые миноры неотрицательны. Итак, задача сводится к решению системы неравенств:

$$\begin{cases} \begin{vmatrix} 1 & x \\ x & 1 \end{vmatrix} \ge 0, \\ \begin{vmatrix} 1 & x & x \\ x & 1 & x \\ x & x & 1 \end{vmatrix} \ge 0 \end{cases} \Rightarrow \begin{cases} 1 - x^2 \ge 0, \\ 1 + 2x^3 - 3x^2 \ge 0, \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} |x| \le 1, \\ 2x^2(x-1) - (x-1)(x+1) \ge 0, \end{cases} \Rightarrow \begin{cases} |x| \le 1, \\ (x-1)^2(2x+1) \ge 0, \end{cases} \Rightarrow -\frac{1}{2} \le x \le 1.$$

Ответ. $-\frac{1}{2} \le x \le 1.$

Упражнение. При каких значениях x матрица $\begin{pmatrix} 1 & x & -x \\ x & 1 & x \\ -x & x & 1 \end{pmatrix}$ будет корреляционной матрицей трех случайных величин?

Пример 5.13. По заданной ковариационной матрице K найти корреляционную матрицу R.

$$\mathrm{K} = \begin{pmatrix} 9 & 8 & -6 \\ 8 & 16 & 8 \\ -6 & 8 & 25 \end{pmatrix}.$$

Решение. По определению $|r_{ij}| \le 1$ при $i \ne j$, $|r_{ii}| = 1.$

Если

$$r_{ij} = \operatorname{corr}(X_i, X_j) = \frac{\operatorname{cov}(X_i, X_j)}{\sigma_{X_i}\sigma_{X_j}}, \ \operatorname{cov}(X_i, X_i) = \mathrm{D}X_i,$$

то

$$r_{11}=r_{22}=r_{33}=1,\ r_{12}=r_{21}=\frac{8}{\sqrt{9}\sqrt{16}}=\frac{2}{3},$$

$$r_{13}=r_{31}=\frac{-6}{\sqrt{9}\sqrt{25}}=\frac{-2}{5},\ r_{23}=r_{32}=\frac{8}{\sqrt{16}\sqrt{25}}=\frac{2}{5}.$$

Итак,

$$\mathrm{R}=\begin{pmatrix}1 & 2/3 & -2/5\\ 2/3 & 1 & 2/5\\ -2/5 & 2/5 & 1\end{pmatrix}.$$

Заметим, что если $\sigma_{X_1}=\sqrt{\mathrm{D}X_1}, \sigma_{X_2}=\sqrt{\mathrm{D}X_2}, \sigma_{X_3}=\sqrt{\mathrm{D}X_3}$, то ковариационная и корреляционная матрицы случайного вектора $X=(X_1, X_2, X_3)$ связаны соотношением

$$\operatorname{cov}\left(X_i, X_j\right)=\Sigma\mathrm{R}\Sigma, \tag{5.20}$$

где

$$\Sigma=\begin{pmatrix}\sigma_{X_1} & 0 & 0\\ 0 & \sigma_{X_2} & 0\\ 0 & 0 & \sigma_{X_3}\end{pmatrix}.$$

Запишем равенство (5.20) для нашего примера:

$$\begin{pmatrix}9 & 8 & -6\\ 8 & 16 & 8\\ -6 & 8 & 25\end{pmatrix}=\begin{pmatrix}3 & 0 & 0\\ 0 & 4 & 0\\ 0 & 0 & 5\end{pmatrix}\begin{pmatrix}1 & 2/3 & -2/5\\ 2/3 & 1 & 2/5\\ -2/5 & 2/5 & 1\end{pmatrix}\begin{pmatrix}3 & 0 & 0\\ 0 & 4 & 0\\ 0 & 0 & 5\end{pmatrix}.$$

Упражнение. Проверить последнее равенство самостоятельно.

Пример 5.14. По заданной корреляционной матрице R и известным дисперсиям $\mathrm{D}X_1=4$, $\mathrm{D}X_2=9$, $\mathrm{D}X_3=16$ найти матрицу K.

$$\mathrm{R}=\begin{pmatrix}1 & -0{,}4 & 0{,}9\\ -0{,}4 & 1 & -0{,}5\\ 0{,}9 & -0{,}5 & 1\end{pmatrix}.$$

Решение. Так как по определению

$$r_{ij}=\frac{\operatorname{cov}\left(X_i, X_j\right)}{\sigma_{X_i}\sigma_{X_j}},\quad \operatorname{cov}\left(X_i, X_i\right)=\mathrm{D}X_i,$$

то $k_{ij}=\operatorname{cov}\left(X_i, X_j\right)=r_{ij}\sigma_{X_i}\sigma_{X_j}$ и, следовательно,

$$k_{11}=r_{11}\sigma_{X_1}\sigma_{X_1}=1\cdot 4=4,\quad k_{22}=9,\quad k_{33}=16,\quad k_{12}=k_{21}=r_{12}\sigma_{X_1}\sigma_{X_2}=-0,4\cdot 2\cdot 3=-2,4,$$

$$k_{13}=k_{31}=r_{13}\sigma_{X_1}\sigma_{X_3}=0,9\cdot 2\cdot 4=7,2,\quad k_{23}=k_{32}=r_{23}\sigma_{X_2}\sigma_{X_3}=-0,5\cdot 3\cdot 4=-6.$$

Итак,

$$\mathrm{K}=\begin{pmatrix} 4 & -2,4 & 7,2 \\ -2,4 & 9 & -6 \\ 7,2 & -6 & 16 \end{pmatrix}.$$

5.5. Многомерное нормальное распределение

Говорят, что многомерная случайная величина $X=(X_1, \ldots, X_n)$ имеет *многомерное нормальное распределение*, если существует вектор $\boldsymbol{m}=\mathrm{E}X=(m_1, m_2, \ldots, m_n)^{\mathrm{T}}$ $(m_i=\mathrm{E}X_i)$ и невырожденная матрица ковариаций $\mathrm{K}=\mathrm{cov}(X_i, X_j)$ такие, что совместная плотность распределения равна:

$$f_X(x_1,\ldots,x_n)=\frac{1}{\sqrt{2\pi|\mathrm{K}|}}\cdot\exp\left\{-\frac{1}{2}\left(x-\boldsymbol{m}\right)^{\mathrm{T}}\mathrm{K}^{-1}\left(x-\boldsymbol{m}\right)\right\}.$$

Здесь $|\mathrm{K}|$ — определитель ковариационной матрицы.

Свойства многомерного нормального распределения:

1) любая компонента нормального вектора имеет нормальное распределение:

$$X_i \sim \mathrm{N}\left(m_i, \sigma_i\right);$$

2) компоненты нормально распределенной случайной величины $X=(X_1, \ldots, X_n)$ независимы тогда и только тогда, когда они являются попарно некоррелированными случайными величинами, т. е. ковариационная матрица K и обратная к ней K^{-1} являются диагональными:

$$\mathrm{K}=\begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma_n^2 \end{pmatrix},\quad \mathrm{K}^{-1}=\begin{pmatrix} 1/\sigma_1^2 & 0 & \ldots & 0 \\ 0 & 1/\sigma_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1/\sigma_n^2 \end{pmatrix}.$$

Квадратичная форма $(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}}\,\mathrm{K}^{-1}(\boldsymbol{x}-\boldsymbol{m})$ превращается в сумму квадратов:

$$\left(\boldsymbol{x}-\boldsymbol{m}\right)^{\mathrm{T}}\mathrm{K}^{-1}\left(\boldsymbol{x}-\boldsymbol{m}\right)=\sum_i\frac{\left(x_i-m_i\right)^2}{\sigma_i^2}.$$

Следовательно, для нормально распределенной системы некоррелированных случайных величин совместная плотность равна произведению плотностей компонент, входящих в систему:

$$f_X\left(x_1, \ldots, x_n\right) = f_{X_1}\left(x_1\right) f_{X_2}\left(x_2\right) \ldots f_{X_n}\left(x_n\right),$$

где $X_i \sim \mathrm{N}\left(m_i, \sigma_i\right)$;

3) пусть случайная величина $Y = a_1 X_1 + \ldots + a_n X_n$, где $X_i \sim \mathrm{N}\left(m_i, \sigma_i\right)$, а коэффициенты $a_1, \ldots, a_n \in \mathbf{R}$ не все равны нулю. Тогда случайная величина Y распределена нормально;

4) пусть случайная величина $X \sim \mathrm{N}(\boldsymbol{m}, \mathrm{K})$, тогда $Y = AX + \boldsymbol{b}$, где A — матрица размерности $n \times k$ с $\operatorname{rang} A = k$ и $\boldsymbol{b}$ — вектор размерности n, имеет распределение $X \sim \mathrm{N}(\boldsymbol{Am} + \boldsymbol{b}, A\mathrm{K}A^{\mathrm{T}})$;

5) $\left(\boldsymbol{x} - \boldsymbol{m}\right)^{\mathrm{T}} \mathrm{K}^{-1} \left(\boldsymbol{x} - \boldsymbol{m}\right) \sim \chi^2\left(n\right)$.

Пример 5.15. Случайная величина $X = (X_1, X_2, X_3)$ распределена нормально:

$$X \sim \mathrm{N}\left(\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}; \begin{pmatrix} 9 & 8 & -6 \\ 8 & 16 & 8 \\ -6 & 8 & 25 \end{pmatrix}\right).$$

1. Выяснить, являются ли случайные величины $X_1 + X_2$ и $7X_2 - 2X_3$ независимыми?

2. Найти: а) $\mathrm{P}(2X_1 + X_2 > 17)$; б) $\mathrm{P}(X_1 + X_2 + X_3 < 12)$.

Решение. 1. По условию $X_1 \sim \mathrm{N}(1, 3)$; $X_2 \sim \mathrm{N}(2, 4)$; $X_3 \sim \mathrm{N}(3, 5)$. Случайные величины $X_1 + X_2$ и $7X_2 - 2X_3$ нормальные, так как представляют собой линейную комбинацию нормальных векторов. Найдем:

$$\begin{aligned}\operatorname{cov}\left(X_1 + X_2, 7X_2 - 2X_3\right) &= \operatorname{cov}\left(X_1, 7X_2 - 2X_3\right) + \operatorname{cov}\left(X_2, 7X_2 - 2X_3\right) = \\ &= 7\operatorname{cov}\left(X_1, X_2\right) - 2\operatorname{cov}\left(X_1, X_3\right) + 7\operatorname{cov}\left(X_2, X_2\right) - 2\operatorname{cov}\left(X_2, X_3\right) = \\ &= 7 \cdot 8 - 2 \cdot \left(-6\right) + 7 \cdot 16 - 2 \cdot 8 = 164 \neq 0.\end{aligned}$$

Следовательно, $X_1 + X_2$ и $7X_2 - 2X_3$ зависимы.

2. Найдем: а) $\mathrm{P}(2X_1 + X_2 > 6)$.

Числовые характеристики случайной величины $2X_1 + X_2$:

$$\mathrm{E}\left(2X_1 + X_2\right) = 2\mathrm{E}X_1 + \mathrm{E}X_2 = 2 \cdot 1 + 2 = 4;$$

$$\mathrm{D}\left(2X_1 + X_2\right) = 4\mathrm{D}X_1 + \mathrm{D}X_2 + 2 \cdot \operatorname{cov}\left(2X_1, X_2\right) = 36 + 16 + 4 \cdot 8 = 84.$$

Напомним, что здесь случайные величины $2X_1$ и X_2 (так же как и случайные величины X_1 и X_2) независимы и нормально распределены. Итак, $2X_1 + X_2 \sim \mathrm{N}\left(4; 2\sqrt{21}\right)$. Воспользовавшись формулой (4.13), находим:

$$P\left(2X_1+X_2>16\right)=\frac{1}{2}-\Phi\left(\frac{16-4}{2\sqrt{21}}\right)=\frac{1}{2}-\Phi\left(1,3093\right)\approx 0,5-0,4048\approx 0,0052;$$

б) $P(X_1+X_2+X_3<12)$. Числовые характеристики случайной величины $X_1+X_2+X_3$:

$$E\left(X_1+X_2+X_3\right)=1+2+3=6;$$

$$D\left(X_1+X_2+X_3\right)=DX_1+DX_2+DX_3+2\cdot\operatorname{cov}\left(X_1,X_2\right)+$$
$$+2\cdot\operatorname{cov}\left(X_1,X_3\right)+2\cdot\operatorname{cov}\left(X_2,X_3\right)=9+16+25+2\cdot 8+2\cdot\left(-6\right)+2\cdot 8=70.$$

Аналогично

$$P\left(X_1+X_2+X_3<12\right)=\frac{1}{2}+\Phi\left(\frac{12-6}{\sqrt{70}}\right)=\frac{1}{2}+\Phi\left(0,7171\right)\approx 0,5+0,2633\approx 0,7633.$$

Задачи для самостоятельного решения

5.1. По мишени производится один выстрел. Вероятность попадания равна Р. Рассматриваются две случайных величины: X — число попаданий; Y — число промахов. Построить функцию распределения двумерной случайной величины (X, Y). Зависимы ли случайные величины X и Y?

5.2. Два стрелка независимо один от другого производят по два выстрела, каждый по своей мишени. При каждом выстреле вероятность попадания для первого стрелка 0,7, для второго — 0,8. Случайная величина X — число попаданий первого стрелка. Случайная величина Y — суммарное число попаданий двух стрелков. Найти закон распределения случайной величины (X, Y). Найти числовые характеристики (X, Y). Зависимы ли случайные величины X и Y?

5.3. Ожидаемая доходность первого актива равна 6 % со средним квадратичным отклонением 5 %, ожидаемая доходность второго актива равна 8 % со средним квадратичным отклонением 10 %. Коэффициент корреляции между этими активами составляет 0,8. Найти ожидаемую доходность и среднее квадратичное отклонение портфеля, состоящего на 40 % из первого актива и на 60 % — из второго.

5.4. Вероятность дождя в субботу 0,5, вероятность дождя в воскресенье 0,8. Корреляция между наличием дождя в субботу и наличием дождя в воскресенье равна 0,75. Какова вероятность того, что в выходные вообще не будет дождя?

5.5. Совместное распределение случайных величин X и Y задано при помощи таблицы:

X	Y		
	1	2	3
0	0,2	0,1	0,2
1	0,1	0,3	0,1

а) Являются ли случайные величины X и Y независимыми? Ответ обосновать.

б) Построить графики функций распределения $F_X(x)$ и $F_Y(x)$.

в) Построить таблицу распределения случайной величины XY.

г) Найти $\mathrm{E}(X)$, $\mathrm{E}(Y)$, $\mathrm{E}(XY)$, $\mathrm{cov}(X, Y)$ и $\mathrm{cov}(2X+3, 1-3Y)$.

д) Являются ли случайные величины X и Y некоррелированными? Ответ обосновать.

е) Найти все условные распределения случайной величины X и все условные распределения составляющей Y.

ж) Найти $\mathrm{E}(Y \mid X=0)$ и $\mathrm{E}(Y \mid X=1)$. Проверить формулу полного математического ожидания $\mathrm{E}(Y)=\sum_{i=1}^{n} p_i \mathrm{E}\left(Y \mid x_i\right)$ для данной задачи.

з) У инвестора портфель, в котором доля акций X составляет α, а доля акций Y – $(1-\alpha)$. Каковы должны быть доли, чтобы риск портфеля (дисперсия дохода) был минимальным?

5.6. Даны законы распределения двух независимых случайных величин X и Y. Составить закон распределения двумерной случайной величины (X, Y). Найти коэффициент корреляции r_{XY}.

X	1	3	5
P	0,2	0,5	0,3

,

Y	2	3	4	5
P	0,1	0,3	0,4	0,2

5.7. Совместный закон распределения случайных величин X и Y задан таблицей:

X	Y		
	–1	0	1
1	1/6	1/4	1/6
2	1/8	1/6	1/8

Найти: $\mathrm{P}(Y=1 \mid X=1)$, $\mathrm{E}(Y \mid X=1)$, $\mathrm{E}(X \mid Y \le 0)$, $\mathrm{P}(X=1 \mid Y \le 0)$.

5.8. Совместное распределение независимых случайных величин X_1 и X_2 определяется условиями:

$$\mathrm{P}\left(X_1=1\right)=\mathrm{P}\left(X_2=1\right)=p, \quad \mathrm{P}\left(X_1=0\right)=\mathrm{P}\left(X_2=0\right)=1-p.$$

Найти: $\mathrm{E}(X_1 X_2 \mid X_1=1)$, $\mathrm{E}(X_1+X_2 \mid X_1=0)$, $\mathrm{E}(X_1 \mid X_1+X_2=1)$.

5.9. Случайные величины X, Y и Z связаны соотношением

$$3X-2Y+Z=0. \tag{5.21}$$

Известно, что $\mathrm{E}X=\mathrm{E}Y=\mathrm{E}Z=0$, $\mathrm{D}X=2$, $\mathrm{D}Y=3$, $\mathrm{D}Z=6$. Найти корреляционную матрицу этих случайных величин.

5.10. Двухмесячные объемы продаж продукции некоторого предприятия описываются двумерным случайным вектором $\begin{pmatrix} X \\ Y \end{pmatrix}$ с плотностью распределения вероятности

$$f(x,y)=\begin{cases} C, 100\le x\le 150, 50\le y\le 100, \\ 0 \qquad \text{(в остальных случаях)}. \end{cases}$$

Найти:

а) константу C;

б) функцию распределения $F(x, y)$;

в) $\mathrm{P}(100\le X\le 125, 75\le Y\le 100)$.

Исследовать случайные величины X и Y на независимость.

5.11. Плотность совместного распределения непрерывной двумерной случайной величины

$$f(x,y)=\begin{cases} (1/4)\cos x\cos y,\ 0\le x\le \pi/2,\ 0\le y\le \pi/2, \\ \qquad 0 \qquad \text{(в остальных случаях)}. \end{cases}$$

Найти функцию распределения $F(x, y)$. Являются ли составляющие случайного вектора независимыми случайными величинами?

5.12. Совместная плотность распределения случайных величин X и Y имеет вид:

$$f(x,y)=\begin{cases} C(x+y), 0\le x\le 2, 0\le y\le 2, \\ 0 \qquad \text{(в остальных случаях)}. \end{cases}$$

Найти ковариацию и корреляцию случайных величин X и Y.

5.13. Найти плотность распределения двумерного случайного вектора (X, Y), если известна функция распределения:

$$F(x,y)=\begin{cases} 1+\mathrm{e}^{-2x-5y}-\mathrm{e}^{-2x}-\mathrm{e}^{-5y}, x\ge 0, 0\le y\ge 0, \\ \qquad 0 \qquad \text{(в остальных случаях)}. \end{cases}$$

5.14. Случайная точка (X, Y), распределена по нормальному закону на плоскости:

$$f(x,y)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\left(x^2+y^2\right)/2}.$$

Найти:

1) вероятность попадания точки $(X,\ Y)$ в квадрат: а) $P(|X| \le 1,\ |Y| \le 1)$; б) $P(|X| + |Y| \le 1)$;

2) математическое ожидание $\sqrt{X^2 + Y^2}$.

5.15. Непрерывная случайная величина X распределена равномерно на $[2;\ 4]$, а непрерывная случайная величина Y — нормально с параметрами $a_Y = -1$, $\sigma_Y = 2$. Известно, что коэффициент корреляции $r_{XY} = 0{,}5$. Найти $E(XY)$.

5.16. Случайный вектор $(X,\ Y)$ подчинен закону распределения с плотностью

$$f(x, y) = \begin{cases} 4xy\,\mathrm{e}^{-\left(x^2+y^2\right)}, & x \ge 0,\ y \ge 0, \\ 0 & (\text{иначе}). \end{cases}$$

Найти коэффициент корреляции.

5.17. Доходности акций двух компаний являются случайными величинами X и Y с одинаковым математическим ожиданием и ковариационной матрицей

$$\mathrm{K} = \begin{pmatrix} 4 & -2 \\ -2 & 9 \end{pmatrix}.$$

а) Найти r_{XY}.

б) В какой пропорции нужно приобрести акции этих двух компаний, чтобы дисперсия доходности получившегося портфеля была наименьшей?

в) Можно ли утверждать, что величины $X + Y$ и $7X - 2Y$ независимы?

5.18. Пусть случайная величина $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$, где X — время появления первого покупателя в понедельник, а Y — время появления первого покупателя во вторник, описывается плотностью распределения вероятности

$$f(x, y) = \begin{cases} \mathrm{e}^{-x-y}, & x \ge 0,\ \ y \ge 0, \\ 0 & (\text{иначе}). \end{cases}$$

Найти: $F_{X,Y}(x, y)$; $F_X(x)$; $F_Y(y)$; $f_X(x)$; $f_Y(y)$; $F_X(x|y)$; $F_Y(y|x)$; $f_X(x|y)$; $f_Y(y|x)$. Установить, зависимы ли случайные величины X и Y.

Ответы к задачам для самостоятельного решения

5.1. Значения функции $F(x, y)$ даны в таблице:

X	Y		
	$y \le 0$	$0 < y \le 1$	$y > 1$
$x \le 0$	0	0	0
$0 < x \le 1$	0	0	q
$x > 1$	0	p	1

Случайные величины X и Y зависимы: $X + Y = 1$.

5.2.

X	Y				
	0	1	2	3	4
0	0,0036	0,0288	0,0576	0	0
1	0	0,0168	0,1344	0,2688	0
2	0	0	0,0196	0,1568	0,3136

$\mathrm{E}X = 1{,}4$; $\mathrm{D}X = 0{,}42$; $\mathrm{E}Y = 3$; $\mathrm{D}Y = 0{,}74$; $\mathrm{E}(XY) = 4{,}62$;
$\mathrm{cov}(X, Y) = 0{,}42$; $r_{XY} \approx 0{,}75$. Случайные величины X и Y зависимы.

5.3. Если X — доходность первого актива, Y — доходность второго актива, то $\mathrm{E}(0{,}4X + 0{,}6Y) = 0{,}072$, $\sigma(0{,}4X + 0{,}6Y) \approx 0{,}0769$.

5.4. 0,25.

5.5. а) X и Y зависимы.

б) $F_X(x) = \begin{cases} 0, & x \le 0, \\ 0{,}5, & 0 < x \le 1, \\ 1, & x > 1; \end{cases} \quad F_Y(y) = \begin{cases} 0, & y \le 1, \\ 0{,}3, & 1 < y \le 2, \\ 0{,}7, & 2 < y \le 3, \\ 1, & y > 3. \end{cases}$

в) Распределение XY имеет вид:

XY	0	1	2	3
P_{XY}	0,5	0,1	0,3	0,1

г) $\mathrm{E}X = 0{,}5$; $\mathrm{E}Y = 2$, $\mathrm{E}(XY) = 1$, $\mathrm{cov}(X, Y) = 0$, $\mathrm{cov}(2X + 3, -3Y + 1) = 0$.

д) X и Y являются некоррелированными, при этом они являются зависимыми.

е) Условные распределения Y:

Y	1	2	3	Σ
$\mathrm{P}Y(y_j \mid x = 0)$	0,4	0,2	0,4	1
$\mathrm{P}Y(y_j \mid x = 1)$	0,2	0,6	0,2	1

ж) $\mathrm{E}(Y \mid X = 0) = \mathrm{E}(Y \mid X = 1) = 2$.

з) 12/17; 5/17. Подсказка: если R — доходность портфеля, то $R = \alpha X + (1 - \alpha)Y$.

5.6. $r_{XY} = 0$.

5.7. $P(Y = 1 \mid X = 1) = 2/7$, $E(Y \mid X = 1) = 0$, $E(X \mid Y \le 0) = 24/17$, $P(X = 1 \mid Y \le 0) = 10/17$.

5.8. p, p, 1/2.

5.9. Умножить соотношение (5.21) последовательно на X, Y и Z; далее найти математические ожидания обеих частей получившихся соотношений. Используя данные задачи, найти ковариационную и затем корреляционную матрицы.

$$R = \begin{pmatrix} 1 & 2/\sqrt{6} & -2/\left(3\sqrt{2}\right) \\ 2/\sqrt{6} & 1 & 0 \\ -2/\left(3\sqrt{2}\right) & 0 & 1 \end{pmatrix}.$$

5.10. а) $C = 1/2\,500$;

$$\text{б) } F(x, y) = \begin{cases} 0, & x \le 100 \text{ или } y \le 50, \\ (x-100)(y-50)/2500, & 100 < x \le 150,\ 50 < y \le 100, \\ (x-100)/50, & 100 < x \le 150,\ y > 100, \\ (y-50)/50, & x > 150,\ 50 < y \le 100, \\ 1, & x > 150,\ y > 100; \end{cases}$$

$$\text{в) } P(100 \le X \le 125,\ 75 \le Y \le 100) = \int_{100}^{125} \left(\int_{75}^{100} (1/2500)\,dy \right) dx = \frac{1}{4}.$$

Так как $F_{XY}(x, y) = F_X(x)F_Y(y)$ для всех $x \in R$, $y \in R$, то случайные величины X и Y являются независимыми.

5.11. $F(x, y)$ представлена в виде таблицы:

X	Y: $y \le 0$	$0 < y \le \pi/2$	$y > \pi/2$
$x \le 0$	0	0	0
$0 < x \le \pi/2$	0	$\frac{1}{4}(1+\sin x)(1+\sin y)$	$\frac{1}{2}(1+\sin x)$
$x > \pi/2$	0	$\frac{1}{2}(1+\sin y)$	1

Так как $F(x, y) = F_X(x)F_Y(y)$, то случайные величины X и Y независимы.

5.12. $\operatorname{cov}(X,Y) = -\dfrac{1}{144}, r_{XY} = -\dfrac{1}{11}.$

5.13. $f(x,y) = \begin{cases} 10e^{-2x-5y}, x \geq 0, 0 \leq y \geq 0, \\ 0 \text{ (в остальных случаях).} \end{cases}$

5.14. 1) а) $\mathrm{P}(|X| \leq 1, |Y| \leq 1) = 4\Phi^2(1) \approx 0,4659;$

б) $\mathrm{P}(|X| + |Y| \leq 1) = 4\Phi^2\left(\dfrac{1}{\sqrt{2}}\right) \approx 0,2709;$

2) $\sqrt{\dfrac{\pi}{2}}.$

5.15. $\mathrm{E(XY)} = \sqrt{3} - 1.$

5.16. $r_{XY} = 0.$

5.17. а) $r_{XY} = -\dfrac{1}{3}.$

б) Если R — доходность портфеля, то $\mathrm{R} = \alpha X + (1-\alpha)Y$ и $(\mathrm{D}R)' = 0.$

Акции двух компаний следует приобрести в отношении 11 : 6.

в) Хотя $\operatorname{cov}(X + Y, 7X - 2Y) = 0$, утверждать, что случайные величины $X + Y$ и $7X - 2Y$ независимы, нельзя.

5.18. $F(x,y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}), x \geq 0, y \geq 0, \\ \quad 0 \quad \text{(иначе);} \end{cases}$

$F_X(x) = 1 - e^{-x}, \ x > 0, \ F_Y(y) = 1 - e^{-y}, \ y > 0;;$

$f_X(x) = e^{-x}, \ x > 0, \ f_Y(y) = e^{-y}, \ y > 0;$

$F_X(x \mid y) = 1 - e^{-x}, \ x > 0, \ F_Y(y \mid x) = 1 - e^{-y}, \ y > 0;$

$f_X(x \mid y) = e^{-x}, \ x > 0, \ f_Y(y \mid x) = e^{-y}, \ y > 0.$

X и Y независимы.

6. СЛУЧАЙНЫЕ ПОСЛЕДОВАТЕЛЬНОСТИ

6.1. Понятие о предельных теоремах

Предельными теоремами в теории вероятностей называют группу теорем, устанавливающих связь между теоретическими и экспериментальными характеристиками при большом числе испытаний, а также теоремы о предельных законах распределения. Было замечено, что совокупное действие большого числа случайных факторов приводит к результату, почти не зависящему от случая. Другими словами, при большом числе случайных величин их средний результат перестает быть случайным и может быть предсказан с большой степенью определенности.

Законы больших чисел утверждают, что среднее арифметическое большого числа случайных величин ведет себя как среднее арифметическое их математических ожиданий.

Согласно *центральной предельной теореме* достаточно большая сумма сравнительно малых случайных величин ведет себя приблизительно как нормальная случайная величина.

При доказательстве многих теорем теории вероятностей и математической статистики используется ряд вспомогательных неравенств.

6.2. Вспомогательные неравенства

Неравенство Маркова. Если $\mathrm{E}|X| < \infty$, то для любого $\varepsilon > 0$ справедливо неравенство

$$\mathrm{P}\left(|X| \geq \varepsilon\right) \leq \frac{\mathrm{E}|X|}{\varepsilon}. \tag{6.1}$$

Неравенство Чебышёва. Если $\mathrm{E}X^2 < \infty$, то для любого $\varepsilon > 0$ справедливо неравенство

$$\mathrm{P}\left(|X - \mathrm{M}X| \geq \varepsilon\right) \leq \frac{\mathrm{D}X}{\varepsilon^2}. \tag{6.2}$$

Неравенство Коши — Буняковского[14] **— Шварца**[15]**.** Для любых случайных величин X и Y справедливо неравенство

$$\mathrm{E}\left|XY\right| \le \sqrt{\mathrm{E}X^2} \cdot \sqrt{\mathrm{E}Y^2}.$$

Следствие из неравенства Маркова. Если $X > 0$, то

$$\mathrm{P}\left(X < \varepsilon\right) \ge 1 - \frac{\mathrm{E}X}{\varepsilon}. \tag{6.3}$$

Следствие из неравенства Чебышёва.

$$\mathrm{P}\left(\left|X - \mathrm{M}X\right| < \varepsilon\right) \ge 1 - \frac{\mathrm{D}X}{\varepsilon^2}. \tag{6.4}$$

Применение неравенств Чебышёва и Маркова не требуют знания закона распределения случайной величины.

Выясним точность оценки вероятности попадания в ε-окрестность $\mathrm{E}X$ значений случайной величины X, полученную с помощью неравенства Чебышёва. Положим $\varepsilon = t\sigma$, $t > 0$. Тогда

$$\mathrm{P}\left(\left|X - m_X\right| < t\sigma\right) \ge 1 - \frac{1}{t^2},\ \ \sigma = \sqrt{\mathrm{D}X},\, m_X = \mathrm{E}X.$$

Так, при $t = 3$

$$\mathrm{P}\left(\left|X - m_X\right| < 3\sigma\right) \ge 1 - \frac{1}{3^2} \approx 0{,}8889,$$

т. е. не менее чем 0,889. Причем эта оценка справедлива для любой случайной величины.

Заметим, что если случайная величина $X \sim \mathrm{N}(a; \sigma)$, то

$$\mathrm{P}\left(\left|X - m_X\right| < t\sigma\right) = 2\Phi\left(\frac{t\sigma}{\sigma}\right) = 2\Phi(t).$$

При $t = 3$ получим: $\mathrm{P}\left(\left|X - m_X\right| < 3\sigma\right) = 2\Phi(3) \approx 0{,}9973$.

Если случайная величина распределена по показательному закону, то

$$\mathrm{P}\left(\left|X - m_X\right| < 3\sigma\right) = \mathrm{P}\left(\left|X - \frac{1}{\lambda}\right| < \frac{3}{\lambda}\right) = \mathrm{P}\left(0 < X < \frac{4}{\lambda}\right) =$$

$$= F\left(\frac{4}{\lambda}\right) - F\left(0\right) = \left(1 - \mathrm{e}^{-\lambda x}\right)\Big|_0^{\frac{4}{\lambda}} \approx 0{,}9817.$$

---

[14] Виктор Яковлевич Буняковский (1804–1889) — российский математик, вице-президент Петербургской академии наук в 1864–1889 гг.

[15] Карл Герман Амандус Шварц (1843–1921) — крупный немецкий математик, член Берлинской академии наук, профессор Галльского, Цюрихского, Геттингенского и Берлинского университетов.

Следовательно, «правило трех сигм» (с достаточно большой вероятностью его выполнения) применимо для большинства случайных величин, встречающихся на практике.

Упражнение. Найти $P(|X - m_X| < 3\sigma)$, если случайная величина имеет равномерное распределение на каком-нибудь отрезке.

Пример 6.1. Средние ежемесячные расходы на покупку канцелярских принадлежностей для отделения банка составляют 3 000 руб., а среднее квадратичное отклонение этой случайной величины не превышает 750 руб. Оценить вероятность того, что расходы на канцелярские принадлежности в любой наугад выбранный месяц не превысят 6 000 руб., используя: а) неравенство Маркова; б) неравенство Чебышёва.

Решение. Пусть X — ежемесячные расходы на покупку канцелярских принадлежностей. По условию $\mathrm{E}X = 3\,000$, $\sigma = 750$.

а) Используя неравенство (6.3), получим $\mathrm{P}\left(X < 6\,000\right) \geq 1 - \dfrac{3\,000}{6\,000} = 0{,}5,$ т. е. вероятность того, что расходы на канцелярские принадлежности в любой наугад выбранный месяц не превысят 4 000 руб., будет не менее 0,5.

б) По формуле (6.4) получим:

$$\mathrm{P}\left(0 < X < 6\,000\right) = \mathrm{P}\left(\left|X - 3\,000\right| < 3\,000\right) \geq 1 - \frac{750^2}{3\,000^2} = 0{,}9375.$$

Таким образом, неравенство Чебышёва дает более точную оценку.

Пример 6.2. По статистическим данным в среднем 87 % новорожденных доживают до 50 лет (т. е. вероятность дожития до 50 лет равна 0,87). С помощью неравенства Чебышёва оценить вероятность того, что из 1 000 новорожденных доля (относительная частота) доживших до 50 лет будет отличаться от вероятности не более чем на 0,04 (по модулю).

Решение. Для относительной частоты $\dfrac{m}{n}$: $m_X = p$, $\mathrm{D}\left(\dfrac{m}{n}\right) = \dfrac{npq}{n^2} = \dfrac{pq}{n}$, поэтому неравенство Чебышёва запишем в виде

$$\mathrm{P}\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{pq}{n\varepsilon^2}. \tag{6.5}$$

Полагая $n = 1\,000$, $p = 0{,}87$, $q = 0{,}13$, $\varepsilon = 0{,}04$, по формуле (6.5)

$$\mathrm{P}\left(\left|\frac{m}{n} - p\right| < 0{,}04\right) \geq 1 - \frac{0{,}87 \cdot 0{,}13}{1\,000 \cdot 0{,}04^2} \cong 0{,}9293,$$

т. е. не менее чем 0,9293.

6.3. Закон больших чисел

Последовательность случайных величин $X_1, \ldots, X_n, \ldots$ *сходится по вероятности* к случайной величине X, если для любого $\varepsilon > 0$:

$$\lim_{n\to\infty}\left(\mathrm{P}\left|X_n - X\right| < \varepsilon\right) = 1 \text{ при } n\to\infty \ \left(\text{или} \lim_{n\to\infty}\left(\mathrm{P}\left|X_n - X\right| \geq \varepsilon\right) = 0 \text{ при } n\to\infty\right).$$

Сходимость по вероятности обозначается так:

$$X_n \xrightarrow{\mathrm{P}} X.$$

Значения случайной величины X даже с очень большими номерами могут находиться за пределами ε-окрестности случайной величины X, но вероятность этого события равна нулю.

Говорят, что последовательность случайных величин $X_1, \ldots, X_n, \ldots$ с конечными математическими ожиданиями $\mathrm{E}X_i = a_i$ удовлетворяет *закону больших чисел* (ЗБЧ), если

$$\frac{X_1 + \ldots + X_n}{n} - \frac{a_1 + \ldots + a_n}{n} \xrightarrow{\mathrm{P}} 0 \text{ при } n \to \infty.$$

Теорема 6.1 (*ЗБЧ Чебышёва*, *1867 г.*). Если независимые случайные величины $X_1, \ldots, X_n, \ldots$ имеют математические ожидания $\mathrm{E}X_i = a_i$ и ограниченные в совокупности дисперсии $\mathrm{D}X_i = \sigma_i^2 \leq \sigma^2$, то

$$\mathrm{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - \frac{a_1 + \ldots + a_n}{n}\right| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}. \qquad (6.6)$$

При любом фиксированном $\varepsilon > 0$ с ростом n величина $\frac{\sigma^2}{n\varepsilon^2} \to 0,$ поэтому правая часть (6.6) стремится к 1. Следовательно:

$$\frac{X_1 + \ldots + X_n}{n} - \frac{a_1 + \ldots + a_n}{n} \xrightarrow{\mathrm{P}} 0 \text{ при } n \to \infty.$$

Следствие. Если все независимые случайные величины $X_1, \ldots, X_n, \ldots$ имеют одинаковые математические ожидания $\mathrm{E}X_i = a$ и одинаковые дисперсии $\mathrm{D}X_i = \sigma^2$ (например, одинаково распределены), то среднее арифметическое случайных величин сходится по вероятности к математическому ожиданию a:

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\mathrm{P}} a.$$

При условиях следствия оценка для вероятности отклонения среднего арифметического любого числа попарно независимых и одинаково распределенных величин от $\mathrm{E}X_i = a$ менее чем на заданное ε равна:

$$\mathrm{P}\left(\left|\frac{X_1+\ldots+X_n}{n}-a\right|<\varepsilon\right)\geq 1-\frac{\sigma^2}{n\varepsilon^2}. \quad (6.7)$$

Сходимость по вероятности средних арифметических большого числа независимых случайных величин к их математическому ожиданию означает *устойчивость* этих средних. Отдельные случайные величины могут принимать значения, весьма удаленные от своих математических ожиданий, но их среднее арифметическое с вероятностью, близкой к единице, принимает значение, мало отличающееся от своего математического ожидания.

На практике неизвестные случайные величины заменяют их средними значениями. Это среднее значение вычисляется при повторных измерениях, каждое из которых содержит некоторую ошибку. Теорема Чебышёва подтверждает возможность такой замены.

Статистическая устойчивость *относительной частоты появления успеха* в серии независимых испытаний рассматривалась еще в трудах Я. Бернулли.

Теорема 6.2 (*ЗБЧ Бернулли, 1713 г.*). Пусть m — частота появления события A в серии n независимых испытаний и вероятность успеха постоянна и равна p. Тогда относительная частота $\frac{m}{n}$ при неограниченном увеличении числа независимых испытаний сходится по вероятности к вероятности появления события в каждом испытании, т. е. для любого $\varepsilon > 0$:

$$\lim_{n\to\infty}\mathrm{P}\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1, \quad (6.8)$$

или $\frac{m}{n}\xrightarrow{\mathrm{P}}p$ при $n\to\infty$.

Заметим, что в отличие от ЗБЧ Чебышёва, описывающего предельное поведение среднего арифметического случайных величин с *произвольными* распределениями, ЗБЧ Бернулли имеет дело лишь со схемой Бернулли.

Приведем оценку, соответствующую предельному равенству (6.8):

$$\mathrm{P}\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)\geq 1-\frac{pq}{n\varepsilon^2}. \quad (6.9)$$

Пример 6.3. Для определения среднего дохода налогоплательщиков города налоговой инспекцией была проведена проверка 250 жителей этого города, отобранных случайным образом. Оценить вероятность того, что средний годовой доход жителей города отклонится от среднего арифметического $\overline{X}=\frac{1}{250}\sum_{i=1}^{250}X_i$ годовых доходов выбранных 250 жителей не более чем на 1 000 руб., если известно, что среднее квадратичное отклонение годового дохода не превышает 2 500 руб.

Решение. Согласно неравенству (6.7), поскольку все $DX_i \le (2\,500)^2$, имеем:

$$P\left(\left|\frac{X_1+\ldots+X_n}{n}-\frac{a_1+\ldots+a_n}{n}\right|<1\,000\right)\ge 1-\frac{(2\,500)^2}{250\cdot(1\,000)^2}=0{,}975.$$

Пример 6.4. Монета подбрасывается 10 тыс. раз. Оценить вероятность того, что частота выпадения герба отличается от половины более чем на 0,01.

Решение. Требуется оценить $\lim\limits_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|\ge 0{,}01\right)$, где $n = 10\,000$, $m=\sum\limits_{i=1}^{n} X_i$ — число выпадений герба, а X_i — независимые случайные величины, каждая из которых имеет распределение Бернулли с параметром $p = 1/2$. Поскольку $DX_i = p(1-p) = 1/2\cdot 1/2 = 1/4$, искомая оценка сверху согласно (6.9) равна:

$$P\left(\left|\frac{m}{n}-\frac{1}{2}\right|\ge 0{,}01\right)\le\frac{1}{4\cdot 10^4 0{,}01^2}=\frac{1}{4}.$$

Иначе говоря, неравенство Чебышёва позволяет заключить, что в среднем не более чем в четверти случаев при 10 тыс. подбрасываний монеты частота выпадения герба будет отличаться от 1/2 на одну сотую или больше. В следующем параграфе мы сравним полученный результат с более точным, полученным с помощью центральной предельной теоремы.

6.4. Теорема Ляпунова (центральная предельная теорема)

Законы больших чисел устанавливают факт приближения среднего значения большого числа случайных величин к некоторым постоянным в виде сходимости последовательностей случайных величин по вероятности.

Центральная предельная теорема (ЦПТ) представляет собой группу теорем, утверждающих, что достаточно большая сумма сравнительно малых случайных величин распределена приближенно по нормальному закону.

Справедлива следующая теорема.

Теорема 6.3 (*теорема Ляпунова*[16]). Если случайные величины X_1, X_2, … X_n — независимые случайные величины, у которых существуют математические

[16] Александр Михайлович Ляпунов (1857–1918) — русский математик и механик, академик Петербургской академии наук с 1901 г., член-корреспондент Парижской академии наук,

ожидания $\mathrm{E}X_i = a_i$, дисперсии $\mathrm{D}X_i = \sigma_i^2$, абсолютный центральный момент третьего порядка $\sum_{i=1}^{n} \mathrm{E}\left(\left|X_i - a_i\right|^3\right)$ и $\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathrm{E}\left(\left|X_i - a_i\right|^3\right)}{\left(\sum_{i=1}^{n} \sigma_i^2\right)^{3/2}} = 0$, то закон распределения суммы $Y_n = X_1 + X_2 + \ldots + X_n$ при $n \to \infty$ неограниченно приближается к нормальному с математическим ожиданием $\sum_{i=1}^{n} a_i$ и дисперсией $\frac{1}{n}\sum_{i=1}^{n} \sigma_i^{\ 2}$.

В практических приложениях важно следующее следствие из теоремы Ляпунова.

Следствие (*классическая ЦПТ*). Пусть $X_1, X_2, \ldots, X_n, \ldots$ есть последовательность независимых одинаково распределенных случайных величин с $\mathrm{E}X_i = a$, $\mathrm{D}X_i = \sigma^2$. Тогда функция распределения центрированной и нормированной суммы этих случайных величин $Z_n = \frac{\sum_{i=1}^{n} X_i - na}{\sigma\sqrt{n}}$ стремится при $n \to \infty$ к функции распределения стандартной нормальной случайной величины:

$$F_{Z_n} = \mathrm{P}\left(Z_n < x\right) \xrightarrow[n\to\infty]{} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \mathrm{e}^{-t^2/2} dt,$$

т. е. $Z_n \to \mathrm{N}(0, 1)$ по распределению при $n \to \infty$.

Преобразуем $Z_n = \frac{\sum_{i=1}^{n} X_i - na}{\sigma\sqrt{n}}$ как $\sqrt{n} \cdot \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - a}{\sigma}$. Поэтому мы можем переписать результат ЦПТ в следующем виде:

$$\sqrt{n} \cdot \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - a}{\sigma} \to \mathrm{N}(0,1) \text{ по распределению при } n \to \infty.$$

Это означает, что сумма

$$X_1 + X_2 + \ldots + X_n \to \mathrm{N}\left(na, \sigma\sqrt{n}\right), \ n \to \infty;$$

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \to \mathrm{N}\left(a, \sigma\right), \ n \to \infty.$$

член Национальной академии деи Линчеи (Италия) и ряда других академий наук и научных обществ.

Напомним, что если случайные величины $X_1, \dots, X_n$ независимы и одинаково распределены с $\mathrm{E}(X_i) = a$, $\mathrm{D}(X_i) = \sigma^2$, то

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}\left(X_i\right) = na, \quad \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(X_i\right) = a,$$

$$\mathrm{D}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{D}\left(X_i\right) = n\sigma^2, \quad \mathrm{D}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{D}\left(X_i\right) = \frac{\sigma^2}{n},$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x \mathrm{e}^{-t^2/2} dt \text{ — интегральная функция Лапласа.}$$

Утверждение ЦПТ при больших n позволяет вычислять вероятности различных событий, связанных с суммами случайных величин:

$$\mathrm{P}\left(\alpha \le \frac{\sum_{i=1}^{n} X_i - na}{\sigma\sqrt{n}} \le \beta\right) \approx \Phi\left(\beta\right) - \Phi\left(\alpha\right) = \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta \mathrm{e}^{-t^2/2} dt, \ n \to \infty;$$

$$\mathrm{P}\left(\alpha \le \sum_{i=1}^{n} X_i \le \beta\right) \approx \Phi\left(\frac{\beta - na}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{\alpha - na}{\sigma\sqrt{n}}\right), \ n \to \infty.$$

Часто ЦПТ используют, если $n > 10$.

В ряде задач приходится сталкиваться с ситуацией, когда исследуемая случайная величина является суммой большого числа независимых слагаемых, влияние каждого из которых на сумму очень мало. Такими случайными величинами являются, например, капиталы банков и страховых компаний (доля каждого отдельно взятого вкладчика не зависит от доли других вкладчиков и относительно мала, но в сумме все эти доли весьма весомы), выручка торговых предприятий (покупатели действуют независимо друг от друга и покупают товары на относительно небольшие суммы) и др.

На основании центральной предельной теоремы часто можно до наблюдения того или иного явления сказать, что соответствующая случайная величина должна иметь нормальное распределение или близкое к нему.

Пример 6.5. Случайные величины X_i, $i = 1, \dots, 150$, распределены по равномерному закону на отрезке $[-1; 3]$. Найти закон распределения случайной величины $X = \sum_{i=1}^{150} X_i$, а также вероятность того, что $145 < X < 170$.

Решение. По теореме Ляпунова случайная величина X, равная сумме независимых случайных величин X_i, $i = 1, \dots, 150$, имеет распределение, близкое к нормальному с параметрами $a = \mathrm{E}X = \mathrm{E}\left(\sum_{i=1}^{150} X_i\right) = \sum_{i=1}^{150} \mathrm{E}X_i = \sum_{i=1}^{150} \frac{(-1+3)}{2} = 150,$

$\sigma^2 = \mathrm{D}X = \mathrm{D}\left(\sum_{i=1}^{150} X_i\right) = \sum_{i=1}^{150} \mathrm{D}X_i = \sum_{i=1}^{150} \frac{\left(3-(-1)\right)^2}{12} = 150 \cdot \frac{16}{12} = 200.$ Таким образом, $a = 150,\ \sigma = 10\sqrt{2}.$ Плотность вероятности суммарной случайной величины X:

$$f(x) = \frac{1}{10\sqrt{2}\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-150)^2}{2\cdot 200}} = \frac{1}{20\sqrt{\pi}} \mathrm{e}^{-\frac{(x-150)^2}{400}},\ x \in \mathrm{R}.$$

Используя формулу (4.7), находим:

$$\mathrm{P}(145 < X < 170) \approx \Phi\left(\frac{170-150}{10\sqrt{2}}\right) - \Phi\left(\frac{145-150}{10\sqrt{2}}\right) = \Phi\left(\sqrt{2}\right) + \Phi\left(\frac{\sqrt{2}}{4}\right) \approx$$
$$\approx 0,42135 - 0,13816 \approx 0,2832.$$

Пример 6.6. Задача из примера 6.4. Монета подбрасывается 10 тыс. раз. Оценить вероятность того, что частота выпадения герба отличается от половины на одну сотую или более.

Решение. Требуется найти $\lim\limits_{n\to\infty} \mathrm{P}\left(\left|\frac{m}{n} - \frac{1}{2}\right| \geq 0,01\right)$, где $n = 10\,000,\ m = \sum_{i=1}^{n} X_i$ — число выпадений герба, а X_i — независимые случайные величины, каждая из которых имеет распределение Бернулли с параметром $p = 1/2$. Рассмотрим событие $\alpha \leq \frac{\sum_{i=1}^{n} X_i - na}{\sigma\sqrt{n}} \leq \beta$ и сведем событие $\left|\frac{m}{n} - \frac{1}{2}\right| < 0,01$ к этому виду:

$$\left|\frac{m}{n} - \frac{1}{2}\right| < 0,01 \Leftrightarrow \left|\frac{m - 1/2n}{n}\right| < 0,01 \Leftrightarrow \left|\frac{(m-1/2n)\sqrt{n}}{n\cdot(1/2)}\right| < 0,01 \cdot \frac{\sqrt{n}}{1/2}.$$

Так как $\sqrt{n} = 100,$ то имеем:

$$\left|\frac{(m-1/2n)}{(1/2)\sqrt{n}}\right| < 2 \Leftrightarrow -2 < \frac{\sum_{i=1}^{n} X_i - na}{\sigma\sqrt{n}} < 2$$

и

$$\mathrm{P}\left(\left|\frac{m}{n} - \frac{1}{2}\right| < 0,01\right) = \mathrm{P}\left(-2 < \frac{\sum_{i=1}^{n} X_i - na}{\sigma\sqrt{n}} < 2\right) = \Phi(2) - \Phi(-2) = 2\Phi(2) \approx 0,9545.$$

Искомая вероятность $\mathrm{P}\left(\left|\frac{m}{n}-\frac{1}{2}\right|\geq 0{,}01\right)=1-\mathrm{P}\left(\left|\frac{m}{n}-\frac{1}{2}\right|<0{,}01\right)\approx 0{,}0455.$

Сравнить этот результат с ответом примера 6.4.

Впрочем, решение этой задачи занимает одну строку, если обратиться к формуле (6.11).

Приведем два следствия из центральной предельной теоремы, относящиеся к независимым испытаниям.

Теорема 6.4 (*локальная теорема Муавра — Лапласа*). Пусть вероятность $p\in(0;\,1)$ наступления события A в каждом из n независимых испытаний постоянна и не меняется от опыта к опыту, $z_m=\frac{m-np}{\sqrt{npq}}$ ограничена. Тогда при $n\to\infty$

$$\mathrm{P}_n(m)\approx\frac{1}{\sqrt{npq}}\varphi(z_m),$$

где $\varphi(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$.

На практике, очевидно, вероятность появления любого конкретного числа успехов близка к нулю. Это имеет простое объяснение — ведь всего есть $n+1$ различных событий (может наступить 0, 1, 2, …, n успехов), и сумма вероятностей этих $n+1$ событий должна быть равна единице. Поэтому важно уметь вычислять вероятности $\mathrm{P}_n(m_1\leq m\leq m_2)$ того, что число успехов в серии из n испытаний будет заключено между числами m_1 и m_2. Об этом в теореме 6.5.

Теорема 6.5 (*интегральная теорема Муавра — Лапласа*). Если вероятность $p\in(0;\,1)$ успеха в каждом испытании отлична от нуля и единицы, а число испытаний n достаточно велико, то для расчета вероятности $\mathrm{P}_n(m_1\leq m\leq m_2)$ можно пользоваться приближенной формулой

$$P_n(m_1\leq m\leq m_2)=\sum_{k=m_1}^{m_2}C_n^k p^k q^{n-k}\approx\Phi(z_2)-\Phi(z_1),$$

где $z_1=\frac{m_1-np}{\sqrt{npq}}$, $z_2=\frac{m_2-np}{\sqrt{npq}}$, $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x\mathrm{e}^{-\frac{t^2}{2}}dt$ — интегральная функция Лапласа.

В заключение отметим, что в силу нормальности частоты m, а также относительной частоты $\frac{m}{n}$ к ним применима формула Лапласа:

$$\mathrm{P}\left(|X-a|<\varepsilon\right)=2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Для частоты m: $a = np, \sigma = \sqrt{npq}$, поэтому

$$P\left(\left|m - np\right| \le \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right). \qquad (6.10)$$

Для относительной частоты $\frac{m}{n}$: $a = p, \sigma^2 = \frac{pq}{n} \Rightarrow \sigma = \sqrt{\frac{pq}{n}}$, поэтому

$$P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right). \qquad (6.11)$$

Пример 6.7. В страховой компании 10 тыс. клиентов. Страховой взнос каждого клиента составляет 500 руб. Вероятность наступления страхового случая равна (по оценкам экспертов компании) 0,005, а страховая выплата при наступлении страхового случая составляет 50 тыс. руб.

а) Определить, на какую прибыль может рассчитывать страховая компания с вероятностью 0,99.

б) Определить минимальный размер страховой премии, при котором страховая компания получит прибыль, не меньшую 1 млн руб., с вероятностью 0,999.

Решение. а) Размер прибыли компании составляет разность между суммарным взносом всех клиентов и суммарной страховой суммой, выплаченной n_0 клиентам при наступлении страхового случая, т. е.

$$\Pi = 500 \cdot 10 - 50 \cdot n_0 = 50\left(100 - n_0\right) \text{ тыс. руб.}$$

Для определения n_0 воспользуемся интегральной теоремой Муавра — Лапласа. По условию задачи

$$P_n\left(0 \le m \le n_0\right) \approx \Phi\left(x_2\right) - \Phi\left(x_1\right) = 0{,}99, \qquad (6.12)$$

где n_0 — число клиентов, которым будет выплачена страховая сумма;

$$x_1 = \frac{0 - np}{\sqrt{npq}} = -\sqrt{\frac{np}{q}} = -\sqrt{\frac{10\,000 \cdot 0{,}005}{0{,}995}} = -7{,}089 \approx -7{,}09;$$

$$x_2 = \frac{n_0 - np}{\sqrt{npq}} = \frac{n_0 - 10\,000 \cdot 0{,}005}{\sqrt{10\,000 \cdot 0{,}005 \cdot 0{,}995}} = \frac{n_0 - 50}{\sqrt{49{,}75}}.$$

С учетом (6.12) $\Phi(x_2) = 0{,}99 + \Phi(x_1) = 0{,}99 + \Phi(-7{,}09) = 0{,}99 - 0{,}5 = 0{,}49$. Отсюда $x_2 = \Phi^{-1}(0{,}49) = 2{,}32634$, т. е. $\frac{n_0 - 50}{\sqrt{49{,}75}} = 2{,}32634$, откуда $n_0 = 66{,}4085 \approx 66$

и $\Pi = 50(100 - n_0) = 50 \cdot 34 = 1\,700$ тыс. руб., т.е. с надежностью 0,99 ожидаемая прибыль составит 1,7 млн руб.

б) Страховая премия (страховой взнос, страховой платеж) — плата за страхование, которую страхователь обязан внести страховщику в соответствии с договором страхования. Обозначим ее через S. Тогда по условию

$$\Pi = S \cdot 10\,000 - 50\,000 \cdot n_0 = 10\,000\left(S - 5n_0\right) \geq 1\,000\,000.$$

Найдем n_0:

$$P_n\left(0 \leq m \leq n_0\right) \approx \Phi\left(x_2\right) - \Phi\left(x_1\right) = 0{,}999;$$

$$x_1 = \frac{0 - np}{\sqrt{npq}} = -\sqrt{\frac{10\,000 \cdot 0{,}005}{0{,}995}} = -7{,}089 \approx -7{,}09, \quad x_2 = \frac{n_0 - np}{\sqrt{npq}} = \frac{n_0 - 50}{\sqrt{49{,}75}};$$

$$\Phi\left(x_2\right) = 0{,}999 + \Phi\left(x_1\right) = 0{,}999 + \Phi\left(-7{,}09\right) = 0{,}999 - 0{,}5 = 0{,}499.$$

Отсюда $x_2 = \Phi^{-1}(0{,}499) = 3{,}09023$, откуда $n_0 = 71{,}7965 \approx 72$. Тогда $10\,000(S - 5n_0) = 10\,000(S - 360) \geq 1\,000\,000 \Rightarrow S - 360 \geq 100 \Rightarrow S \geq 460.$ Итак, минимальный размер страховой премии составит 460 руб.

Задачи для самостоятельного решения

6.1. Сумма всех вкладов в некоторую сберегательную кассу составляет 20 тыс. руб., а вероятность того, что случайно взятый вклад не превышает 100 руб., равна 0,8. Что можно сказать о числе вкладчиков данной сберкассы?

6.2. Глубина моря измеряется прибором, не имеющим систематической ошибки. Среднее квадратичное отклонение измерений не превосходит 15 м. Сколько нужно сделать независимых измерений, чтобы с вероятностью, не меньшей 0,9, можно было утверждать, что среднее арифметическое этих измерений отличается от a (глубины моря) по модулю меньше чем на 5 м?

6.3. Вероятность наличия опечатки на одной странице рукописи равна 0,2. Найти вероятность того, что в рукописи, содержащей 400 страниц, частота появления опечатки отличается от соответствующей вероятности по модулю меньше чем на 0,05. Ответ дать с помощью неравенства Чебышёва и интегральной теоремы Муавра — Лапласа. Объяснить различие результатов.

6.4. Среднее квадратичное отклонение каждой из 2 500 независимых случайных величин не превосходит 3. Оценить вероятность того, что абсолютная величина отклонения среднего арифметического этих случайных величин от среднего арифметического их математических ожиданий не превосходит 0,3.

6.5. Оценить с помощью неравенства Чебышёва вероятность того, что при одновременном подбрасывании 12 игральных костей сумма выпавших очков отклонится от математического ожидания меньше чем на 15.

6.6. Оценить вероятность того, что, бросив монету 400 раз, относительная частота появления герба при одном испытании по абсолютной величине отклонится не больше чем на 0,05. Ответ дать с помощью неравенства Чебышёва и интегральной теоремы Муавра — Лапласа. Объяснить различие результатов.

6.7. У страховой компании имеются 10 тыс. клиентов. Каждый из них, страхуясь от несчастного случая, вносит 500 руб. Вероятность несчастного случая 0,0055, а страховая сумма, выплачиваемая пострадавшему, составляет 50 тыс. руб. Какова вероятность того, что: а) страховая компания потерпит убыток; б) на выплату страховых сумм уйдет не более половины всех средств, поступивших от клиентов?

6.8. Среднее изменение курса акций компании в течение одних биржевых торгов составляет 0,3 %. Оценить вероятность того, что на ближайших торгах курс изменится более чем на 3 %.

6.9. Статистические исследования, проведенные учебным отделом университета, показали, что среднее время опоздания студента на лекцию составляет 1 мин. Оценить вероятность того, что:

а) студент опоздает на лекцию не менее чем на 5 мин;

б) если, дополнительно, $\mathrm{D}X = 1$, оценить минимальное значение x_0 при котором вероятность опоздания студента на время, не менее x_0, не превышает заданного значения вероятности $\mathrm{P}_0 = 0{,}1$.

6.10. Некоторый период времени на бирже сохранялся относительно стабильный курс валюты. На основании данных биржевой статистики за этот период была составлена таблица возможных значений изменения курса валют:

Возможное изменение курса, %	–1	–0,5	0	0,5	1
Вероятность изменения	0,1	0,3	0,5	0,05	0,05

Оценить вероятность того, что произойдет изменение курса валюты не более чем на 0,6 %. Сравнить полученную оценку с точным значением вероятности.

6.11. Опыт работы страховой компании показывает, что страховой случай приходится примерно на каждый десятый договор. Оценить с помощью неравенства Чебышёва необходимое количество договоров, которые следует заключить, чтобы с вероятностью не меньшей, чем 0,95, можно было утверждать, что доля страховых случаев отклонится от вероятности не более чем на 0,02 (по абсолютной величине). Уточнить ответ с помощью следствия из интегральной теоремы Муавра — Лапласа.

6.12. Вероятность того, что акции, переданные на депозит, будут востребованы, равна 0,08. Оценить с помощью неравенства Чебышёва вероятность того, что среди 1 000 клиентов от 70 до 90 востребуют свои акции.

Ответы к задачам для самостоятельного решения

6.1. $n \le 1\,000$.

6.2. Не менее 90 раз.

6.3. С помощью теоремы Чебышёва: $\mathrm{P}\left(\left|\frac{m}{n}-0,2\right|<0,05\right)\ge 0,84$; с помощью интегральной теоремы Муавра — Лапласа: $\mathrm{P}\left(\left|\frac{m}{n}-0,2\right|<0,05\right)\approx$ $\approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)\approx 0,9876.$

6.4. $p \ge 0{,}96$.

6.5. $p \ge 0{,}844$.

6.6. 0,75; 0,9545.

6.7. а) 0; б) 0,7054.

6.8. $\mathrm{P}(X>3)\le\frac{0,3}{3}=0,1.$

6.9. $\mathrm{P}(X\ge 5)\le 0,2;$ $\min\left(x_0\colon\left(\mathrm{P}(X\ge x_0)\ge 0,1\right)\right)\approx 4,16,$ т. е. вероятность опоздания студента на время более 4,16 мин не более 0,1.

6.10. Нижняя оценка вероятности с помощью неравенства Чебышёва: $\mathrm{P}(|X|<0{,}6)\ge 0{,}34$, а истинное значение вероятности равно 0,85.

6.11. Оценка, полученная с помощью неравенства Чебышёва: $n \ge 4\,500$; уточненная оценка: $n \ge 864$.

6.12. $p \ge 0{,}264$.

7. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

7.1. Выборочный метод математической статистики

Математическая статистика является частью общей прикладной математической дисциплины «Теория вероятностей и математическая статистика», однако задачи, решаемые ею, носят специфический характер. Если теория вероятностей исследует явления, полностью заданные их моделью, то в математической статистике вероятностная модель определена с точностью до неизвестных параметров. Отсутствие сведений о параметрах компенсируется «пробными» испытаниями, на основе которых и восстанавливается недостающая информация. Иными словами, в теории вероятностей рассматриваются величины с заданными распределениями, но часто возникает вопрос, а откуда у нас имеется эта информация? Как раз математическая статистика позволяет по результатам конечного числа экспериментов делать более-менее точные выводы о распределении рассматриваемых случайных величин в определенных экспериментах.

К основным задачам математической статистики можно отнести: определение способов сбора и группировки статистической информации, разработку методов анализа статистических данных, интерпретацию и формирование выводов.

Изучение закономерностей объектов достаточно большой совокупности методами математической статистики основано на использовании статистических данных для некоторой конечной части рассматриваемых объектов.

Допустим, у нас есть некоторая совокупность однородных объектов и нас интересует некоторый количественный или качественный признак, характеризующий эти объекты, например, количество страховых случаев у застрахованных. Данный признак можно интерпретировать как случайную величину, значение которой меняется от объекта к объекту. Иногда проводят сплошное обследование — изучают каждый объект совокупности относительно признака, которым интересуются, но такой способ слишком трудоемок, более того, не всегда его возможно применить. В большинстве случаев из всей совокупности объектов случайным образом отбирают ограниченное число объектов, которые и подвергают изучению. В связи с этим естественно ввести следующие определения.

Генеральная совокупность — это все мыслимые значения (измерения, наблюдения), описывающие поведение исследуемого объекта или явления.

Выборка — ограниченный набор реально наблюдаемых из генеральной совокупности значений, описывающих исследуемый объект или явление.

Количество этих значений называют объемом выборки.

Иными словами, выборкой $X = (X_1, X_2, \dots X_n)$ объема n из некоторого распределения F называется набор из n независимых и одинаково распределенных случайных величин, имеющих распределение F. Отдельно отметим, что для взаимно независимых одинаково распределенных случайных величин используется обозначение *iid*, что является сокращением от англ. *independent identically distributed*.

Понятие генеральной совокупности аналогично понятию случайной величины.

Выборку можно рассматривать как некоторый эмпирический аналог генеральной совокупности.

Пример 7.1. В страховой компании за 10 лет работы накопились данные о 25 тыс. клиентов. Для исследования количества страховых случаев и корректировки страховой премии взято 3 500. В данном случае, 25 тыс. — объем генеральной совокупности, а 3 500 — объем выборки.

Таким образом, мы рассматриваем некоторую часть генеральной совокупности (т. е. выборку) и по ней делаем выводы о свойствах генеральной совокупности. В этом и заключается сущность выборочного метода.

Достоинства выборочного метода:

- позволяет существенно экономить затраты ресурсов;
- является единственно возможным в случае бесконечной генеральной совокупности;
- при тех же затратах ресурсов дает возможность проведения углубленного исследования за счет расширения программы исследования;
- позволяет снизить ошибки регистрации.

Недостатки выборочного метода:

- ошибки исследования, называемые ошибками репрезентативности.

Однако неизбежные ошибки могут быть и заранее оценены с помощью правильной организации выборки и сведены к практически незначительным величинам.

Чтобы по данным выборки иметь возможность судить о генеральной совокупности, она должна быть отобрана случайно.

Выборка называется *репрезентативной* (представительной), если она достаточно хорошо воспроизводит генеральную совокупность.

Различают несколько видов выборки.

Обозначим: x_i — значение признака (т. е. случайной величины X); N, n — объемы генеральной совокупности и выборки.

Средние арифметические распределения признака в генеральной и выборочной совокупности называются генеральным и выборочным средним соответственно. Дисперсии — генеральной и выборочной дисперсией.

Важнейшей задачей выборочного метода является оценка параметров (характеристик) генеральной совокупности по данным выборки.

Теоретическую основу применимости выборочного метода составляет закон больших чисел, согласно которому при неограниченном увеличении выборки практически достоверно, что случайные выборочные характеристики как угодно близко приближаются по вероятности к ограниченным параметрам генеральной совокупности.

Оценка неизвестных параметров переменной происходит на основании анализа материала наблюдения. Однако прежде чем приступить к оцениванию, производят предварительную обработку материала наблюдения — составляют вариационный ряд и рассчитывают некоторые описательные статистики этого ряда, которые будут анализироваться дальше.

7.2. Применение математической статистики

Зачем нужна математическая статистика? Рассмотрим несколько областей, в которых она применяется:

- *Маркетинг*. Изучение окупаемости рекламы, исследование рынка, исследование и анализ целевых аудиторий и потребительских предпочтений, выстраивание прогнозов спроса и предложения.
- *Бизнес*. При разработке бизнес-плана потребуется все детально рассчитывать: за сколько вы сможете купить или продать тот или иной товар или услугу. Необходимо построить (смоделировать) несколько сценариев: оптимистичный, средний, пессимистичный. Затем в каждом из сценариев есть разветвления. Таким образом можно построить некое подобие дерева решений, где в каждой ситуации возможно просчитать ожидаемую прибыль к тому или иному месяцу.
- *Банковское дело*. Построение разумной стратегии по выдаче кредитов. Возникает случайная величина: будет возвращен кредит или нет. Чтобы определить, кому выдать кредит, а кому — нет, банк анализирует статистическую информацию. Сюда входит и кредитная история самого человека, и процент

вернувших кредит в срок и т. д. Этот анализ проводится методами теории вероятностей и математической статистики.

Пример 7.2. Банк выдает кредиты по 1 млн руб. сроком на 1 год. Известно, что в среднем вероятность невозврата кредита = 1 %. Какую процентную ставку должен установить банк, чтобы в среднем иметь прибыль?

Обозначим ставку $P = 100$ %. Прибыль — величина случайная.

P	–1 млн
0,99	0,01

$\mathrm{E}X = 0{,}99p - 0{,}01 > 0 \Rightarrow p > 1/99 \Rightarrow$ ставка должна быть 100/99, т. е. несколько больше одного процента.

• *Страхование.* Наступление страхового случая — величина случайная. Страховая компания анализирует статистические данные по наступлению страхового случая, в которых они наступили. Таким образом можно оценить вероятность наступления страхового случая и назначить для него страховой взнос.

Пример 7.3. Пусть страховая компания заключает договоры страхования сроком на 1 год на S руб. каждый. Страховой случай происходит с вероятностью p и не происходит с вероятностью $q = 1 - p$. Таким образом, имеем закон распределения случайной величины X — количество страховых случаев у одного страхователя (0 — страховой случай не наступил).

0	1
p	q

$$\Rightarrow \mathrm{E}X = p; \quad \mathrm{Var}X = pq.$$

Случайная величина $X = x_1, \ldots, x_n$ — количество страховых случаев у страхователей имеет математическое ожидание $\mathrm{E}X = np$ и $\mathrm{Var}X = npq$.

В силу центральной предельной теоремы X — нормально распределенная случайная величина.

В среднем страховая компания должна будет выплатить npS страховых возмещений, т. е. если с каждого брать по pS страхового взноса, то у компании будет нулевой баланс. Реальная страховая ставка: $\tilde{p} > p$.

Таким образом, нам нужно исследовать поведение тех или иных объектов или явлений. А оно осуществляется на основе изучения статистических данных — наблюдений и измерений.

7.3. Вариационные ряды и их характеристики

Установление статистических закономерностей, присущих массовым случайным явлениям, основано на изучении статистических данных — сведений о том, какие значения принял в результате наблюдений интересующий нас признак X.

Рассмотрим X — числовую характеристику совокупности объектов.

Пример 7.4. Необходимо изучить распределение размера обуви, проданной в интересующем магазине, с целью обеспечить нужное количество обуви каждого размера. Получены следующие данные о размере проданной в магазине за сутки обуви (женской): 35, 35, 36, 36, …, 42, 42. Всего продано 100 штук.

Рассмотрение и осмысление данных, представленных в таком виде, практически невозможно из-за обилия числовой информации. Поэтому проводят группировку представленной совокупности чисел.

Различные значения признака X, наблюдавшиеся у объектов, называются *вариантами*, а их количество — *частотами*.

Сгруппированный ряд представляют в виде таблицы.

Пример 7.5. За смену продано 100 пар обуви.

x_i-варианты	36	37	38	39	40	41	42
n_i-частоты	2	6	13	20	25	21	13
m_i-частности	0,02	0,06	0,13	0,2	0,25	0,21	0,13

Здесь $\sum_{i=1}^{N} m_i = 1;\ \sum_{i=1}^{N} n_i = N;\ m_i = \frac{n_i}{N},\ N = 100.$

Частоты показывают, сколько раз встречаются наблюдения, у которых значение признака X равно данной варианте.

Вариационным рядом называется ранжированный в порядке возрастания или убывания ряд вариант с соответствующими весами (частотами или частностями).

Вариационный ряд можно определить для дискретных и непрерывных величин X. В последнем случае проводят интервальную группировку ряда. Число интервалов рекомендуется брать по формуле Стерджеса: $n = 1 + 3{,}322 \cdot \lg N$,

$h = \frac{x_{\max} - x_{\min}}{1 + 3{,}3221 \cdot \lg N}$, где $x_{\max} - x_{\min}$ — разность между наибольшим и наименьшим

значением признака. За начало первого интервала рекомендуют брать величину $x_{\text{нач}} = x_{\min} - h/2$. Частота показывает число членов совокупности, у которых признак X принимает значения в границах интервалов.

Пример 7.6. Имеется таблица данных:

Урожайность, ц/га	(8–12)	(12–16)	(16–20)	(20–25)	(25–40)
Количество хозяйств	10	14	30	26	20
Частности	0,1	0,14	0,3	0,26	0,2

Здесь $N = 100$.

В этом случае n_i — плотность распределения, а m_i — относительная плотность распределения вариационного ряда.

Полученный вариационный ряд позволяет выявить закономерности изменчивости признака, закономерности распределения обуви по размеру проданных пар и участков по урожайности, что сделать по первичным, несгруппированным данным оказалось затруднительно.

Наряду с понятием частот и относительных частот для описания вариационного ряда используются накопленные частоты и накопленные относительные частоты.

Графическое представление вариационных рядов

Представление вариационного ряда в виде таблицы не всегда удобно. Поэтому используют различные способы графического представления вариационных рядов:

- полигон частот — ломаная, соединяющая точки (x_i, n_i);
- полигон относительных частот — ломаная, соединяющая точки (x_i, m_i);
- в случае непрерывного признака X целесообразно строить различные гистограммы.

Гистограмма частот — содержит столбики с основанием — интервалом, высотой — плотностью вариационного ряда, деленной на величину соответствующего интервала. Площадь столбиков — число наблюдений N.

Гистограмма относительных частот — содержит столбики с основанием — интервалом, высотой — относительной плотностью вариационного ряда, деленной на величину соответствующего интервала. Площадь столбиков — 1.

Эмпирической функцией распределения называется функция $F_N(x)$, выражающая для каждого x долю значений параметра общего объема N, для которых рассматриваемый признак меньше x:

$$F_N(x) = \frac{n(x)}{N}, \quad \text{где } n(x) = \sum_{x_0 < x} n_i.$$

Вариационный ряд является статистическим аналогом (реализацией) распределения признака (случайной величины X). В этом смысле полигон или гистограмма аналогичен кривой распределения, а эмпирическая функция распределения — функции распределения случайной величины X.

Вариационный ряд содержит достаточно полную информацию об изменчивости признака X. Однако на практике часто оказывается знание лишь некоторых сводных характеристик вариационных рядов: средних или характеристик центральной тенденции и характеристик изменчивости (показателей вариации), расчет которых представляет собой следующий после группировки этап обработки данных наблюдений.

Средние величины характеризуют значения признака, вокруг которого концентрируются наблюдения. Наиболее распространенной среди средних величин является средняя арифметическая.

Средняя арифметическая вариационного ряда и ее свойства

Имеем среднюю арифметическую вариационного ряда:

$$\overline{x} = \frac{\sum_{i=1}^{n} n_i x_i}{N} = \sum_{i=1}^{n} m_i x_i,$$

где x_i — i-я варианта, если признак X — дискретный, и середина i-го интервала, если X — непрерывный признак.

С в о й с т в а средней арифметической:

1) $\overline{x \pm c} = \overline{x} \pm c$;

2) $\overline{cx} = c\overline{x}$;

3) $\dfrac{\sum_{i=1}^{n} n_i \left(x_i - \overline{x}\right)}{N} = \sum_{i=1}^{n} m_i \left(x_i - \overline{x}\right) = 0$;

4) $\overline{(x \pm y)} = \overline{x} \pm \overline{y}$;

5) если признак X разбит на группы, т. е. групповая и общая средние составляют:

Середина интервала	Q_1	…	Q_j	…	Q_m
x_1	s_{11}		s_{1j}		s_{1m}
…					
x_i	s_{i1}		s_{ij}		s_{im}
…					
x_n	s_{n1}		s_{nj}		s_{nm}
Итого	N_1		N_j		N_m
	$\overline{x}_1$		$\overline{x}_j$		$\overline{x}_m$

Здесь s_{ij} — частота появления i-го наблюдения в j-й группе;

$N_j = \sum_{i=1}^{n} s_{ij}$ — число наблюдений в j-й группе $\left(\sum_{j=1}^{m} N_j = N \right)$;

$\overline{x}_j = \frac{\sum_{i=1}^{n} s_{ij} x_i}{N_j}$ — средняя j-й группы;

$\overline{x} = \frac{\sum_{j=1}^{m} N_j \overline{x}_j}{N}$ — общая средняя.

Кроме средней арифметической также вычисляют медиану и моду.

Медианой вариационного ряда называется значение признака, приходящегося на середину ранжированного ряда наблюдений.

На медиану не влияют изменения крайних членов вариационного ряда. Медиана как показатель среднего предпочтительнее средней арифметической для ряда, у которого крайние варианты по сравнению с большими оказались чрезмерно большими или малыми.

Модой вариационного ряда называется варианта, которой соответствует наибольшая частота (модальный интервал).

Особенность моды как показателя среднего заключается в том, что она не изменяется при изменении крайних членов ряда, т. е. обладает устойчивостью к вариации признака.

Кроме средних величин для вариационного ряда рассчитывают еще показатели вариации.

Выборочная дисперсия и ее свойства

Дисперсией вариационного ряда называется величина

$$\mathrm{d}^2(X) = \frac{\sum_{i=1}^{n} n_i \left(x_i - \overline{x} \right)^2}{N} = \sum_{i=1}^{n} m_i \left(x_i - \overline{x} \right)^2 = \frac{\sum_{i=1}^{n} n_i x_i^2}{N} - \overline{x}^2,$$

где x_i — i-я варианта, если признак X — дискретный, и середина i-го интервала, если X — непрерывный признак.

$\sqrt{\mathrm{d}^2(X)}$ — среднеквадратичное отклонение вариационного ряда.

Свойства выборочной дисперсии:

1) $\mathrm{d}^2(cX) = c^2\mathrm{d}^2(X)$;

2) $\mathrm{d}^2(X \pm c) = \mathrm{d}^2(X)$.

Для вариационного ряда рассчитывают начальные и центральные моменты, частными случаями которых являются средняя арифметическая и дисперсия. В число таких показателей входят асимметрия и эксцесс.

7.4. Оценивание распределения случайных величин

Пусть имеется некоторая случайная величина (одномерная или многомерная), распределение которой неизвестно или известно с точностью до некоторого параметра θ. Имеется некоторая реализация этой случайной величины $X_1, X_2, \ldots, X_N$ (выборка). На основании имеющейся выборки мы должны оценить распределение случайной величины.

Для нахождения оценки характеристики используют два подхода.

1. При проведении эксперимента исследователь уже может что-то предполагать о распределении случайной величины. Например, число клиентов некоторой фирмы за час имеет распределение Пуассона с неизвестной интенсивностью λ, а возраст людей, страдающих головной болью,— нормальное распределение с неизвестным средним и дисперсией. Таким образом, в большинстве случаев распределение случайной величины известно с точностью до некоторого параметра $f_X(x, \theta)$. Тогда, оценив по выборке значение этого параметра, получим оценку распределения $\hat{f}_X(x,\theta) = f_X(x,\hat{\theta})$. Такой подход называется параметрическим.

Пример 7.7. Имеем:

а) $X \sim \mathrm{N}(m, \sigma)$,

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-m)^2}{2\sigma^2}}, \text{ тогда } \theta = \begin{pmatrix} m \\ \sigma^2 \end{pmatrix} \text{ — неизвестен;}$$

б) $X \sim \exp(\lambda)$, $\theta = \begin{pmatrix} \theta_1 \\ \ldots \\ \theta_k \end{pmatrix}$ — неизвестен;

в) $X \sim b_i(p)$,

$p(x = k) = C_n^k p^k q^{n-k} \Rightarrow \theta = p$ — неизвестен;

г) $X \sim U(a, b)$,

$$f_x(x) = \frac{1}{b-a}, \ \theta = \begin{pmatrix} a \\ b \end{pmatrix} \text{ — неизвестен.}$$

Таким образом, имеем $f_x(x, \theta)$, где θ — неизвестный параметр. Необходимо этот неизвестный параметр оценивать, т. е. получать оценку (приближенное значение) этого параметра, которая обозначается $\hat{\theta}$. Тогда имеем известную $\hat{f}_X(x,\theta)$. Стоит отметить, что подобных оценок можно получить несколько.

2. Непараметрический подход применяется в случае, когда вид функции распределения нам неизвестен. В такой ситуации необходимо составить алгоритм расчета значения функции в каждой точке.

Далее продолжим описание первого подхода.

Пусть функция плотности $f_X(x, \theta)$ известна, но параметр θ нет. В общем случае $\theta = \begin{pmatrix} \theta_1 \\ \dots \\ \theta_k \end{pmatrix}$ — вектор неизвестных параметров (их k неизвестных).

Как оценить эти k параметров? Существуют различные виды статистического оценивания параметров.

Виды статистического оценивания параметров:

1) точечное, т. е. получение оценки $\hat{\theta} = \varphi(x_1, x_2, \dots, x_n)$. Вообще говоря, $\hat{\theta} \neq \theta$, но необходимы методы, позволяющие находить $\hat{\theta}$ такое, что $\hat{\theta} \approx \theta$;

2) интервальное, т. е. построение интервала, который с заданной вероятностью γ покрывает истинное значение оцениваемого параметра; γ называют уровнем надежности (заданная вероятность накрытия интервалом), т. е. $p\Big(\theta \in (\hat{\theta}_1; \hat{\theta}_2)\Big) = \gamma$.

Точечное статистическое оценивание неизвестного параметра

Рассмотрим основные методы получения точечной оценки неизвестного параметра.

1. Метод математических ожиданий (метод аналогий).

В выборочном методе всегда имеется выборка, по которой можно посчитать выборочную среднюю $\overline{x} = \frac{\sum x_i}{n}$.

При этом $\theta = \mathrm{E}x$ — неизвестная величина. Тогда в качестве оценки математического ожидания можно взять известную оценку $\hat{\theta} = \frac{\sum x_i}{n}$, а в качестве неизвестной величины $\theta = \mathrm{E}X^2$ рассмотреть известную оценку $\hat{\theta} = \frac{\sum x_i^2}{n}$.

2. Метод моментов (метод Пирсона).

Пусть $\theta = \begin{pmatrix} \theta_1 \\ \dots \\ \theta_k \end{pmatrix}$ — вектор неизвестных параметров. В данном методе предлагается для нахождения оценки неизвестного параметра использовать выборочные моменты вместо истинных.

Таким образом, имеем $\nu_n = \mathrm{E}X^m$ — начальный момент m порядка.

$$\begin{bmatrix} Ex = g_1(\theta) \\ Ex^2 = g_2(\theta) \\ \dots \\ Ex^m = g_m(\theta) \end{bmatrix}$$ — система условий на момент распределения.

Тогда согласно избранному подходу имеем следующие оценки:

$$\begin{bmatrix} \overline{x} = \dfrac{\sum x_i}{n} = g_1(\theta) \\ \dfrac{\sum x_i^{\ 2}}{n} = g_2(\theta) \\ \dots \\ \dfrac{\sum x_i^{\ m}}{n} = g_m(\theta) \end{bmatrix}$$

Пример 7.8. Пусть:

а) $X \sim U(0; 0)$.

Методом моментов найдем оценку параметров θ.

Нам известно из теории вероятностей $E_x = \dfrac{\theta}{2};$ по выборке найдем среднее $\overline{x} = \dfrac{\theta}{2},$ следовательно, имеем: $\hat{\theta}_{MM} = 2\overline{x}.$ Стоит отметить, что индекс у оценки обозначает метод, по которому она получена.

Заметим, что сколько неизвестных параметров, столько моментов мы используем;

б) $X \sim \exp(\lambda)$, $\theta = \lambda$.

Тогда $E_x = \dfrac{1}{\lambda};$ $\overline{x} = \dfrac{1}{\lambda} \Rightarrow \hat{\lambda} = \dfrac{1}{\overline{x}};$

в) $X \sim N(m; \sigma^2); \theta = \dfrac{m}{\sigma^2};$

$E_x = m, E_{x^2} = m^2 + \sigma^2,$ так как $\mathrm{Var}(X) = E(X - EX)^2 = EX^2 - (EX)^2;$

$\sigma^2 = E_{x^2} - (E_x)^2 \Rightarrow E_{x^2} = \sigma^2 + (E_x)^2 = \sigma^2 + m^2.$

Таким образом, $\overline{x} = \hat{m},$ учитывая $\dfrac{\sum x_i^2}{n} = m^2 + \sigma^2,$ получим $\hat{\sigma}^2 = \dfrac{\sum x_i^2}{n} - \overline{x}^2;$

$\dfrac{\sum\left(x_i - \overline{x}\right)^2}{n} = \dfrac{\sum x_i^2}{n} - \overline{x}^2$ — выборочная дисперсия.

3. Метод минимальных расстояний.

В данном методе предлагается x_i сопоставить с так называемой функцией цели $C_i(\theta)$.

Составим выражение $x_i - C_i(\theta)$ для любого $i=\overline{1,n},$ тогда можно рассмотреть выражение $C(\theta)=\sum\limits_{i=1}^{n}\left|x_i - C_i(\theta)\right|$ и найти $\hat{\theta}_{MD}=\arg\min C(\theta).$ Данный метод называется методом наименьших модулей.

Если рассмотреть $C(\theta)=\sum\limits_{i=1}^{n}\left(x_i - C_i(\theta)^2\right)$ и найти $\hat{\theta}_{MD}=\arg\min C(\theta),$ то получим уже другой метод, носящий название «метод наименьших квадратов» (также используются сокращения МНК или MLS от англ. *least square method*). Данный метод имеет активное применение в эконометрических исследованиях.

Пример 7.9. Неизвестный параметр — математическое ожидание, т. е. $\theta = \mathrm{E}X$.

Составим $C_i(\theta)=\mathrm{E}X=\theta.$ Согласно методу наименьших квадратов необходимо найти минимум: $\sum\limits_{i=1}^{n}\left(x_i-\theta\right)^2.$

Находим необходимое условие существования экстремума:

$$C'_\theta=\frac{\partial C}{\partial\theta}=-2\sum\left(x_i-\theta\right)=\theta.$$

Тогда имеем: $\hat{\theta}\approx\theta.$

4. Метод максимального правдоподобия (*Maximum Likelihood*).

Этот метод среди всех возможных параметров выбирается, так как вероятность получить данную выборку максимальна.

Основу метода правдоподобия составляет функция правдоподобия:

$$L(\theta)=f_{X_1X_2\ldots X_n}\left(x_1,x_2,\ldots,x_n,\theta\right)=\prod f_{x_i}\left(x_i,\theta\right).$$

Далее ищут максимум функции правдоподобия $L(\theta)$ и полагают, что $\hat{\theta}_{ML}=\arg\max L(\theta).$

Пример 7.10. Дано дискретное распределение:

X	–1	0	1
p	θ	1,2–2θ	θ–0,2

Необходимо найти θ.

Выборка $x_1 = 0$; $x_2 = 0$ независима.

$$L(\theta)=p\left((x_1=0)(x_2=1)\right)=p(x_1=0)\cdot p(x_2=1)=(1{,}2-2\theta)(\theta-0{,}2)\to\max.$$

Необходимые условия экстремума таковы:

$$L'_\theta = 1{,}2\theta - 2\theta^2 + 0{,}4\theta - 0{,}24 = 0;$$

$$\Rightarrow 1{,}6 - 4\theta = 0 \Rightarrow \hat{\theta}_{ML} = 0{,}4;$$

$$L''_\theta = -4 < 0 \Rightarrow \hat{\theta}_{ML} = 0{,}4.$$

Таким образом, получили оценку методом максимального правдоподобия, а теперь для сравнения применим метод моментов:

$$\mathrm{E}X = -\theta + \theta - 0{,}2 = -0{,}2;$$

$$\mathrm{E}X^2 = \theta + \theta - 0{,}2 = 2\theta - 0{,}2;$$

$$\frac{\sum x_i^2}{n} = 2\theta - 0{,}2;$$

$$\hat{\theta}_{MM} = \frac{\sum x_i^2}{2n} + 0{,}1;$$

$$n = 2 \Rightarrow \hat{\theta}_{MM} = \frac{0^2 + 1^2}{4} + 0{,}1 = 0{,}25 + 0{,}1 = 0{,}35.$$

Оценки неизвестного параметра методом максимального правдоподобия и методом моментов оказалась разные.

Нахождение оценки $\hat{\theta}_n$ упрощается, если максимизировать не саму функцию L, а $\ln L$, поскольку максимум обеих функций достигается при одном и том же значении θ.

Кратко опишем механизм метода максимального правдоподобия для двух этих случаев.

$$L(\theta) = f_{X_1 X_2 \ldots X_N}\left(x_1, \ldots, x_N, \theta\right) = \prod_{i=1}^{N} f\left(x_i, \theta\right),$$

$$L(\theta) \to \max,$$

$$\hat{\theta}_{ML} = \arg\max L(\theta).$$

Необходимое условие:

$$L'(\theta) = 0.$$

Рассмотрим:

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f_x\left(x_i, \theta\right),$$

$$\hat{\theta}_{ML} = \arg\max L(\theta),$$

$$l'_\theta = \frac{\partial l(\theta)}{\partial \theta} = 0,$$

а затем надо отобрать то решение, которое обращает функцию $\ln L$ в max, т. е. воспользоваться достаточным условием существования экстремума (в двумерном случае с использованием матрицы Гессе или дифференциала второго порядка).

Пример 7.11. Найти оценку методом максимального правдоподобия, если известно следующее:

$$x \sim N\left(m, \sigma^2\right);\ \theta = \begin{pmatrix} m \\ \sigma^2 \end{pmatrix};\ f_x(x) = \frac{1}{\sqrt{\sigma^2}\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Прологарифмируем для удобства функцию плотности нормального распределения:

$$\ln f_x(x) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln(2\pi) - \frac{\left(x-m\right)^2}{2\sigma^2}.$$

Составим логарифм функции правдоподобия:

$$l(\theta) = \sum_{i=1}^{n} \ln f_{x_i}\left(x_i, \theta\right) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{\sum\left(x_i - m\right)^2}{2\sigma^2}.$$

Далее воспользуемся необходимым условием существования экстремума, т. е. найдем частные производные по неизвестным параметрам и приравняем их к нулю, получим:

$$l'_m = 2\frac{\sum(x_i - m)}{2\sigma^2} = 0;$$

$$l'_{\sigma^2} = -\frac{m}{2\sigma^2} + \frac{\sum\left(x_i - m\right)^2}{2\sigma^4} = 0.$$

Из первого уравнения следует $\sum(x_i - m) = 0,$ раскрыв знак суммы, имеем: $\sum x_i - \sum_{i=1}^{n} m = 0,$ учитывая, что $\sum_{i=1}^{n} m = nm,$ получим: $\sum x_i - nm = 0,$ откуда $\hat{m}^2_{ML} = \frac{\sum x_i}{n}.$

Аналогично из второго уравнения имеем: $\dfrac{m}{2\sigma^2} = \dfrac{\sum\left(x_i - \overline{x}\right)^2}{2\sigma^4}$, следовательно, $\dfrac{n\sigma^2}{2\sigma^4} = \dfrac{\sum\left(x_i - \overline{x}\right)^2}{2\sigma^4}$, откуда очевидно получаем: $\hat{\sigma}^2_{ML} = \dfrac{\sum\left(x_i - \overline{x}\right)^2}{n}$.

Стоит отметить, что мы преждевременно обозначили данные выражения за оценки неизвестных параметров, так как необходимо еще проверить достаточное условие существования экстремумов. Для этого найдем вначале частные производные второго порядка:

$$l''_{mm} = -\frac{n}{\sigma^2} < 0; \quad l''_{\sigma^2\sigma^2} = \frac{n}{2\sigma^4} - \frac{2\sum\left(x_i - m\right)^2}{2\sigma^6};$$

$$l''_{m\sigma^2} = -\frac{\sum\left(x_i - m\right)}{\sigma^4} = 0,$$

так как

$$\sum\left(x_i - \hat{m}\right) = \sum\left(x_i - \overline{x}\right) = \sum x_i - \sum \overline{x} = \sum x_i - n\overline{x} = \sum x_i - \sum x_i = 0.$$

Проверим знак $l''_{\sigma^2\sigma^2}$, для этого заменим σ^2 на его оценку, получим:

$$l''_{\sigma^2\sigma^2} = \frac{n}{2\left[\dfrac{\sum\left(x_i - \overline{x}\right)^2}{n}\right]} - \frac{\sum\left(x_i - \overline{x}\right)^2}{\left[\dfrac{\sum\left(x_i - \overline{x}\right)^2}{n}\right]^3} =$$

$$= \frac{n^3}{2\left(\sum\left(x_i - \overline{x}\right)^2\right)^2} - \frac{n^3}{\left(\sum\left(x_i - \overline{x}\right)^2\right)^2} = -\frac{n^3}{2\left(\sum\left(x_i - \overline{x}\right)^2\right)^2} < 0.$$

Составим матрицу Гессе:

$$H = \begin{pmatrix} -\dfrac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\dfrac{n^3}{\left(\sum\left(x_i - \overline{x}\right)^2\right)^2} \end{pmatrix}.$$

Исследовав матрицу Гессе по критерию Сильвестра, получим, что она знакоотрицательная, а значит, действительно найденная точка является точкой максимума и полученная оценка является оценкой максимального правдоподобия.

7.5. Свойства статистических оценок

Свойства статистических оценок различают на больших и конечных выборках.

1) Свойства оценок на больших выборках (при $n \to \infty$):

- состоятельность;
- асимптотическая нормальность;
- асимптотическая несмещенность;
- асимптотическая эффективность.

2) Свойства оценок на конечных выборках:

- асимптотическая несмещенность;
- асимптотическая эффективность.

Здесь при дальнейших рассмотрениях имеем θ и $\hat{\theta}$, где $\hat{\theta} \neq \theta$, при этом в общем случае надеемся, что $\hat{\theta} \approx \theta$.

Заметим, что θ — число, $\hat{\theta}$ — случайная величина.

1. *Состоятельность.* Оценка $\hat{\theta}$ параметра θ называется *состоятельной*, если она удовлетворяет закону больших чисел, т. е. сходится по вероятности к оцениваемому параметру:

$$\lim_{x \to \infty} p\left(\left|\hat{\theta} - \theta\right| < E\right) = 1 \quad \text{или} \quad p\lim_{x \to \infty} \hat{\theta} = \theta.$$

То есть, если оценка состоятельна, то практически достоверно, что при достаточно большом n верно, что $\hat{\theta} = \theta$.

Отметим, что оценки метода моментов и метода максимального правдоподобия состоятельны, а метода наименьших квадратов — нет.

2. *Несмещенность.* Оценка $\hat{\theta}$ параметра θ называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру, т. е. $\mathrm{E}\hat{\theta} = \theta$.

Если $\mathrm{E}\hat{\theta} \neq \theta$, то смещенная оценка систематически завышает/занижает истинное значение оцениваемого параметра. В этом случае можно ввести выражение $bias(\hat{\theta}) = \theta - \mathrm{E}\hat{\theta}$ — величина смещения.

Пусть $x_i \sim iid(m, \sigma^2)$, $\mathrm{E}X = m$. Рассмотрим следующее выражение:

$$\mathrm{E}\overline{x} = \mathrm{E}\left(\frac{\sum x_i}{n}\right) = \frac{\mathrm{E}\sum x}{n} = \frac{\sum \mathrm{E}x_i}{n} = \frac{\sum m}{n} = \frac{nm}{n} = m = \mathrm{E}X.$$

Таким образом, можно сделать вывод, что выборочное среднее есть несмещенная оценка.

Отметим также следующие факты:

$d_x^2 = \dfrac{\sum\left(x_i - \overline{x}\right)^2}{n}$ — смещенная, но асимптотически несмещенная оценка;

$S_x^2 = \dfrac{\sum\left(x_i - \overline{x}\right)^2}{n-1}$ — несмещенная оценка, которая в данном случае называется *исправленной выборочной дисперсией*.

Отметим, что из двух несмещенных оценок выбирают ту, дисперсия у которой меньше.

Рассмотренные нами методы оценки не гарантируют несмещенности.

3. *Эффективность*. Эффективной в классе несмещенных оценок называется оценка, обладающая минимальной дисперсией среди всех несмещенных оценок:

$$\hat{\theta}_1, \hat{\theta}_2; \quad \mathrm{E}_{\hat{\theta}_1} = \theta; \quad \mathrm{E}_{\hat{\theta}_2} = \theta;$$

$Var(\hat{\theta}_1) > Var(\hat{\theta}_2) \Rightarrow$ оценка $\hat{\theta}_2$ более эффективна, чем $\hat{\theta}_1$.

Обращаем внимание читателей, что по дисперсии сравниваются только несмещенные оценки.

$$\overline{x} = \widehat{\mathrm{E}x}; \quad Var\overline{x} = \frac{VarX}{n}; \quad Var\overline{x} = \frac{1}{n^2}\sum_1^n Var(x_i) = \frac{1}{n^2}\sum_1^n \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} = \frac{VarX}{n}.$$

Информацией Фишера о неизвестном параметре $\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}$ называется

$$I(\theta) = -\mathrm{E}\left(\frac{\partial^2 \ln f\left(x,\theta\right)}{\partial\theta\partial\theta'}\right).$$

Если $\theta = (\theta)$, то $I(\theta) = -\mathrm{E}\left(\dfrac{\partial^2 \ln f\left(x,\theta\right)}{(\partial\theta)^2}\right)$.

Теорема 7.1 (*Рао — Фреше — Крамера*). Неравенство Рао — Крамера.

Пусть выполняется некое условие регулярности, а именно:

а) $I(\theta)$ существует и положительно;

б) область определения случайной величины X не зависит от θ.

Тогда матрица $[n \cdot I(\theta)]^{-1}$ дает нижнюю границу дисперсий:

- состоятельных оценок;
- асимптотически нормальных оценок θ.

В частности, если θ — скаляр, то $Var(\hat{\theta}) \geq \dfrac{1}{n \cdot I(\theta)}$.

Если дисперсия несмещенной оценки достигает границы, указанной неравенством Рао — Крамера, значит, она эффективна, т. е. обладает наименьшей дисперсией в классе несмещенных оценок.

Пример 7.12. Имеем: $f_x(x,\theta)=\frac{1}{\theta}e^{-\frac{1}{\theta}x}$, $x\geq 0.$ Требуется найти оценку методом максимального правдоподобия; проверить состоятельность, несмещенность и эффективность.

Решение. Составим функцию правдоподобия $l(\theta)=\prod_{i=1}^{n}\frac{1}{\theta}e^{-\frac{1}{\theta}x_i}\to\max.$

Тогда:

$$\ln f_x(x,\theta)=-\ln\theta-\frac{1}{\theta}x_i;$$

$$\sum\ln f_x(x,\theta)=\sum\left(-\ln\theta-\frac{1}{\theta}x_i\right);$$

$$l(\theta)=-n\ln\theta-\frac{1}{\theta}\sum x_i\to\max.$$

Применим необходимое условие существования экстремума:

$$l'_\theta=-\frac{n}{\theta}+\frac{1}{\theta^2}\sum x_i=0.$$

Проверим:

$$\frac{-n\theta}{\theta^2}+\frac{1}{\theta^2}\sum x_i\cdot\theta^2\Rightarrow -n\theta+\sum x_i=0\Rightarrow\hat{\theta}_{ML}=\frac{\sum x_i}{n}=\overline{x}.$$

При этом нужно еще доказать, что это точка максимума, т. е. проверить достаточность условия существования экстремума, но мы упустим этот момент, оставляя его для самостоятельной работы читателя.

Таким образом, оценка неизвестного параметра найдена. Теперь проверим свойства найденной оценки.

Проверим состоятельность: $p\lim_{x\to\infty}\hat{\theta}=p\lim_{x\to\infty}\overline{x}=\mathrm{E}_x=\theta;\ x_i\sim iid.$

$$\mathrm{E}_x=\int_0^{+\infty}x^2\frac{1}{\theta}e^{\frac{-1}{\theta}x}\mathrm{d}x=\theta;$$

$$VarX=\int_0^{+\infty}x^2\frac{1}{\theta}e^{\frac{-1}{\theta}x}\mathrm{d}x-\left(\mathrm{E}_x\right)^2=\theta^2.$$

Откуда и следует, что рассматриваемая оценка состоятельна.

Проверим несмещенность:

$$\mathrm{E}\hat{\theta}=\mathrm{E}\overline{x}=\mathrm{E}\left[\frac{\sum x_i}{n}\right]=\frac{\sum \mathrm{E}x_i}{n}=\frac{\sum \mathrm{E}x}{n}=\frac{n\mathrm{E}x}{n}=\mathrm{E}x=\theta,$$

откуда следует, что оценка несмещенная.

Проверим эффективность. Для этого вначале найдем дисперсию оценки:

$$Var\left(\hat{\theta}_{ML}\right) = Var\left[\frac{\sum x_i}{n}\right] = \frac{1}{n^2} Var\left(\sum x_i\right) = \frac{1}{n^2} \sum Var(x_i) = \frac{1}{n^2}\theta^2 = \frac{n\theta^2}{n^2} = \theta^2 / n \Rightarrow$$

$$\Rightarrow Var\left(\hat{\theta}_{ML}\right) = \frac{\theta^2}{n}.$$

Далее посмотрим, совпадет ли дисперсия оценки с нижней границей, полученной из неравенства Рао — Фреше — Крамера.

Ищем нижнюю границу из теоремы Рао — Фреше — Крамера, для этого находим $I(\theta)$ — информацию Фишера, для чего ищем вторую производную:

$$\frac{\partial \ln f_x\left(x, \theta\right)}{\partial \theta} = -\frac{1}{\theta} + \frac{1}{\theta^2} x_i \text{ — производная от } \ln\left(\frac{1}{\theta} \mathrm{e}^{-\frac{1}{\theta}x}\right) = -\ln\theta - \frac{1}{\theta} x_i;$$

$$\frac{\partial^2 \ln f_x\left(x, \theta\right)}{\partial \theta^2} = \frac{1}{\theta^2} + \frac{2}{\theta^3} x_i;$$

$$\ln\left(\frac{1}{\theta} \mathrm{e}^{-\frac{1}{\theta}x}\right) = -\ln\theta - \frac{1}{\theta} x_i;$$

$$I(\theta) = -\mathrm{E}\left(\frac{1}{\theta^2} + \frac{2}{\theta^3} x_i\right) = -\left(\frac{1}{\theta^2} + \frac{2}{\theta^3} \mathrm{E}_x\right) = -\left(\frac{1}{\theta^2} + \frac{2}{\theta^2}\right) = -\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2} = I(\theta).$$

Из неравенства Рао — Крамера:

$$Var(\hat{\theta}) \geq \frac{1}{n \cdot I(\theta)}.$$

А у нас:

$$\frac{1}{n \cdot I(\theta)} = \frac{1}{n \cdot \frac{1}{\theta^2}} = \frac{\theta^2}{n} \Rightarrow Var(\hat{\theta}),$$

следовательно, дисперсия оценки совпадает с нижней границей Рао — Фреше — Крамера, значит, наша оценка эффективна.

Пример 7.13. Предприимчивый Семен Семенович хочет проверить надпись на рекламном баннере, которая гласит, что в лотерее «СтопЖирЛото» почти треть всех билетов выигрышные. Чтобы реализовать свою задумку, он попросил n своих друзей купить по 10 лотерейных билетов. Пусть X_i — число выигрышных билетов друга i и p — вероятность выигрыша одного билета.

а) Какое распределение имеет величина X_i?

б) Записать функцию правдоподобия $L(p)$ для выборки $X_1, \ldots, X_n$.

в) Методом максимального правдоподобия найти оценку p.

г) Найти информацию Фишера для одного наблюдения $i(p)$.

д) Будет ли оценка параметра p, полученная по методу максимального правдоподобия, эффективной?

е) Найти оценку максимального правдоподобия математического ожидания и дисперсии выигранных произвольным другом билетов.

ж) Дана реализация случайной выборки пяти друзей Семена Семеновича. Число выигрышных билетов оказалось равно (3, 4, 0, 2, 6). Найти значение точечной оценки вероятности выигрыша p. Как вы думаете, похоже ли утверждение организаторов на правду?

Решение. а) Очевидно, что X_i будет иметь биномиальное распределение с неизвестным параметром p, а именно: $C_{10}^{x_i} \cdot p^{x_i} \cdot (1-p)^{10-x_i}$, $x_i = 0,1,\ldots,10$.

б) Составим функцию правдоподобия согласно определению, получим:

$$L(p) = \prod_{i=1}^{n} C_{10}^{x_i} p^{x_i} \left(1-p\right)^{10-x_i}.$$

в) Для того, чтобы найти оценку параметра p методом максимального правдоподобия, в данном случае проще рассмотреть логарифм функции правдоподобия, т. е.:

$$l(p) = \ln\left(\prod_{i=1}^{n} C_{10}^{x_i} \cdot p^{x_i}\left(1-p\right)^{10-x_i}\right) = \sum_{i=1}^{n} \ln\left(C_{10}^{x_i} p^{x_i}\left(10-p\right)^{10-x_i}\right) =$$

$$= \sum_{i=1}^{n}\left(\ln C_{10}^{x_i} + x_i \ln p + \left(10-x_i\right)\ln\left(10-p\right)\right) =$$

$$= \sum_{i=1}^{n} \ln C_{10}^{x_i} + \ln p \sum_{i=1}^{n} x_i + \ln\left(10-p\right)\sum_{i=1}^{n}\left(10-x_i\right).$$

Далее ищем точку максимума полученной функции. Вначале пользуемся необходимым условием существования экстремума, т. е. находим производную по переменной p:

$$l'(p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n}(10-x_i)}{1-p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{10n - \sum_{i=1}^{n} x_i}{1-p} =$$

$$= \frac{\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i - 10np + p\sum_{i=1}^{n} x_i}{p(1-p)} = \frac{\sum_{i=1}^{n} x_i - 10np}{p(1-p)}.$$

Далее, чтобы найти стационарную точку, приравняем найденную производную к нулю:

$$\frac{\sum_{i=1}^{n} x_i - 10np}{p(1-p)} = 0.$$

Откуда найдем:

$$\hat{p}_{ML} = \frac{\sum_{i=1}^{n} x_i}{10n} = 0,1\overline{x}.$$

И вновь мы преждевременно взяли данное выражение за оценку параметра p, так как найденная стационарная точка может не являться точкой максимума. Чтобы проверить это, воспользуемся достаточным условием существования экстремума, т. е. проверим знак второй производной в найденной стационарной точке:

$$l''(p) = -\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{\sum_{i=1}^{n} (10 - x_i)}{(1-p)^2}.$$

Вторая производная представляет собой разность двух дробей, знаменатели и числители которых неотрицательны (в силу того, что согласно условию $x_i =$ 0, 1, …, 10). Таким образом, выражение — отрицательно, т. е. полученная точка является точкой максимума, значит, $\hat{p}_{ML} = \frac{\sum_{i=1}^{n} x_i}{10n} = 0,1\overline{x}$ является оценкой параметра p, полученной по методу максимального правдоподобия.

г) Найдем информацию Фишера для общего количества наблюдений:

$$I\left(\hat{p}_{ML}\right) = -\mathrm{E}\left(\frac{\partial^2 l\left(\hat{p}_{ML}\right)}{\partial\left(\hat{p}_{ML}\right)^2}\right) = -\mathrm{E}\left(-\frac{\sum_{1}^{n} x_i}{\left(\hat{p}_{ML}\right)^2} - \frac{10n - \sum_{i=1}^{n}(10 - x_i)}{\left(1-\hat{p}_{ML}\right)^2}\right) =$$

$$= \frac{1}{\left(\hat{p}_{ML}\right)^2}\sum_{i=1}^{n} x_i + \frac{10n - 10\sum_{i=1}^{n} x_i}{\left(1-\hat{p}_{ML}\right)^2}.$$

Учитывая, что $\hat{p}_{ML} = \frac{\sum_{i=1}^{n} x_i}{10n}$, получим $\sum_{i=1}^{n} x_i = 10n\hat{p}_{ML}$, тогда имеем:

$$I\left(\hat{p}_{ML}\right) = \frac{10\hat{p}_{ML}n}{\left(\hat{p}_{ML}\right)^2} - \frac{10n - 10n\hat{p}_{ML}}{\left(1-\hat{p}_{ML}\right)^2} = \frac{10n}{\hat{p}_{ML}} + \frac{10n\left(1-\hat{p}_{ML}\right)}{\left(1-\hat{p}_{ML}\right)^2} =$$

$$= \frac{10n - 10n\hat{p}_{ML} + 10n\hat{p}_{ML}}{p(1-p)} = \frac{10n}{\hat{p}_{ML}\left(1-\hat{p}_{ML}\right)}.$$

Откуда получаем информацию Фишера для одного наблюдения:

$$i\left(\hat{p}_{ML}\right) = \frac{I\left(\hat{p}_{ML}\right)}{n} = \frac{10}{\hat{p}_{ML}\left(1-\hat{p}_{ML}\right)}.$$

д) Чтобы проверить найденную по методу максимального правдоподобия оценку на эффективность, нужно сначала проверить ее на несмещенность, затем найти ее дисперсию и сравнить ее с оценкой в неравенстве Рао — Фреше — Крамера. Проверку на несмещенность оставим читателю для самостоятельной работы и сразу найдем дисперсию:

$$Var\left(\hat{p}_{ML}\right) = Var(0{,}1\overline{x}) = \left(\frac{0{,}1}{n}\right)^2 Var\sum_{i=1}^{n} x_i = \frac{1}{100n^2}\cdot n\cdot 10p\left(1-p\right) = \frac{p\left(1-p\right)}{10n}.$$

Таким образом, дисперсия совпадает с нижней оценкой из неравенства Рао — Фреше — Крамера, что позволяет сделать вывод, что полученная нами оценка эффективна.

е) Известно из теории, что $\mathrm{E}(x_i) = 10p$, значит, $\hat{\mathrm{E}}(x_i) = 10\hat{p}_{ML} = \overline{x}$.

Аналогично известно из теории, что $Var(x_i) = 10p(1-p)$, значит, $\widehat{Var}(x_i) = \overline{x}\left(1 - \frac{\overline{x}}{10}\right)$.

ж) $\hat{p}_{ML} = \frac{3+4+0+2+6}{10\cdot 5} = 0{,}3.$

Теорема 7.2. Среднее арифметическое $\overline{x}$ является состоятельной, несмещенной и эффективной оценкой математического ожидания случайной величины.

Утверждение. 1) Метод математических ожиданий: оценка, полученная этим методом, состоятельная и асимптотически нормальная.

2) Метод моментов: оценка состоятельная, асимптотически нормальная, но не гарантирует несмещенность.

3) Метод минимальных расстояний: выполнений свойств не гарантируется, все свойства необходимо проверять для каждой оценки индивидуально.

4) Метод максимального правдоподобия: оценка состоятельная, асимптотически нормальная, асимптотически несмещенная, асимптотически эффективная (в пределе этот метод самый лучший).

$$\sqrt{n}\left(\hat{\theta}_{ML}-\theta\right)\xrightarrow[n\to\infty]{d}N\left(0;I^{-1}(\theta)\right).$$

У читателя может возникнуть вопрос: как сравнивать смещенные и несмещенные оценки? Ответ на этот вопрос прост. Рассмотрим следующее выражение:

$$MSE(\hat{\theta})=\mathrm{E}\left(\theta-\mathrm{E}\hat{\theta}\right)^2=Var\hat{\theta}+\left(\theta-\mathrm{E}\hat{\theta}\right)^2=Var\hat{\theta}+bias^2(\hat{\theta}).$$

Та оценка, у которой MSE (*mean square error*) меньше, — точнее!

С теорией статистического оценивания параметров тесно связана проверка статистических гипотез. Она используется всякий раз, когда необходимо сделать обоснованный вывод о преимуществах того или иного способа инвестиций, об эффективности метода управления, об уровне доходности ценных бумаг, о значимости математической модели.

7.6. Общая схема проверки статистических гипотез

Статистическая гипотеза — гипотеза о виде распределения случайной величины или о параметрах распределения.

Имеем: X — случайная величина; $f_x\left(x,\theta\right);\theta$ — неизвестный параметр.

Выдвинутую гипотезу называют *нулевой* и обозначают H_0.

Альтернативная гипотеза — это гипотеза, противоречащая нулевой. Ее обозначают H_1 или H_a.

Нулевую гипотезу формулируют так: «нет различий», «=».

Альтернативную так: «различия», «≠».

Для проверки гипотезы используется выборка. Поскольку выборка действительность отражает не полностью, то при принятии решений возможны ошибки.

H_0	Принимаем H_0	Не принимаем H_0
Верная	ОК	Ошибка I рода (α)
Неверная	Ошибка II рода (β)	ОК

$$p\left(\text{I рода}\right)=p\left(H_a\mid H_0\ \text{верна}\right)=\alpha\text{ — уровень значимости.}$$

Пример 7.14. Пусть α — вероятность вынесения судом обвинительного приговора при том, что подсудимый невиновен.

$$p(\text{II рода}) = p(H_0 | H_a \text{ верна}) = \beta;$$

β — вероятность вынесения оправдательного приговора при том, что подсудимый виновен.

$$\mu = 1 - \beta \text{ — мощность критерия.}$$

При этом чем больше α, тем меньше β, и наоборот.

Поэтому вероятность ошибки I рода выбираем сами.

Установим $\alpha = 0{,}05$, $\beta \to \min$.

Для проверки гипотез используется специальным образом подобранная случайная величина, называемая *статистическим критерием*.

Закон распределения критерия при справедливости нулевой гипотезы известен точно или приблизительно (асимптотически).

Наиболее часто используемые случайные величины:

- Стандартное нормальное распределение.
- Распределение Стьюдента.
- Распределение Фишера.
- Распределение χ^2.

Все возможные значения критерия разбиваются на две непересекающиеся области: критическую область и область принятия гипотезы.

Критическая область — это область таких значений критерия, попадание в которую критерием маловероятно ($p = \alpha$) при справедливости нулевой гипотезы.

Область принятия гипотезы — это область, вероятность попадания в которую критерием вероятна при справедливости нулевой гипотезы.

Если наблюдаемое значение критерия попало в критическую область, значит, гипотеза неверна и мы ее отвергаем.

Если наблюдаемое значение критерия не попало в критическую область (попало в область принятия гипотезы), мы говорим, что нет основания отвергнуть нулевую гипотезу.

Разделяют критическую область и область принятия гипотезы критические точки.

$k_{\text{набл}} = k(\text{выборка})$;

$w_{\text{кр}}$ — критическая область, k — критерий;

$p(k \in w_{\text{кр}} | H_0) = \alpha$ — вероятность попасть в критическую область при выполнении нулевой гипотезы.

Вывод: $k \in w_{\text{кр}}$ — H_0 отвергаем; $k \notin w_{\text{кр}}$ — нет оснований отвергнуть H_0 (так как гипотеза — это не установленный факт, а лишь как достаточно правдоподобное, не противоречащее выборке утверждение).

7.7. Проверка гипотезы о виде распределения случайной величины

Имеем некоторую случайную величину X. Предполагаем, что она имеет распределение A. То есть формулируем гипотезы:

$$H_0: X \sim A;$$

$$H_a: X \nsim A.$$

Для проверки гипотезы H_0 выбирают некоторую случайную величину u, характеризующую степень расхождения теоретического и эмпирического распределения. Проверку проводят с помощью *критерия согласия Пирсона*.

Условия применения критерия Пирсона:

1. Объем выборки большой.
2. Данные должны быть сгруппированы.
3. Частота каждого интервала (каждой варианты) должна быть достаточно большой.

Частоты интервала — эмпирические частоты. Теоретическая частота — это частота, которую бы имел в среднем данный интервал (варианта), если бы случайная величина X имела предполагаемое распределение.

Обозначим: n_i — эмпирическая частота; n_i' — теоретическая частота.

Тогда $n_i' = p_i n$,

где $p_i = p\left(X = x_i | H_0\right)$, если X — дискретная;

$p_i = p\left(x_i \in \left(x_{i-1}; x_i\right) | H_0\right)$, если X — непрерывная.

Иногда распределение A параметрическое, и чтобы посчитать p_i, нам нужно оценить неизвестные параметры:

если $n_i \approx n_i' \Rightarrow H_0$;

если $n_i \not\approx n_i' \Rightarrow$ отвергаем H_0;

$u = \sum \dfrac{\left(n_i - n_i'\right)^2}{n_i'} \sim \chi^2\left(m - r - 1\right)$, где u называют статистикой Пирсона.

При справедливости H_0 u имеет распределение χ^2 с количеством степеней свободы $m - r - 1$, где r — число параметров распределения A, которые мы оцениваем для того, чтобы посчитать p_i; m — число интервалов.

Таким образом, проверка сводится к следующему:

1. Считаем $u_{\text{набл}}$ — наблюдаемое значение критерия.

2. Для выбранного уровня значимости по таблице χ^2 (см. Приложение, табл. 4) ищем критическое значение $u_{\text{кр}}$.

3. Делаем вывод.

Рассматриваем $w_{\text{кр}} = \{u > u_{\text{кр}}\}$ (рис. 7.1) при $p(u > u_{\text{кр}} | H_0) = \alpha$ — уровень значимости, тогда $u_{\text{кр}} = \chi^2_{\text{кр}}(\alpha, m - r - 1)$, $m - r - 1$ — количество степеней свободы.

Вывод: $u_{\text{набл}} > u_{\text{кр}} \Rightarrow$ отвергаем H_0, $u_{\text{набл}} < u_{\text{кр}} \Rightarrow$ Н.О.О.H_0 (нет оснований отвергнуть H_0).

Пример 7.15. Дана таблица частот:

	n_i	n'_i
Орел	40	50
Решка	60	50

$$u_{\text{набл}} = \frac{(40-50)^2}{50} + \frac{(60-50)^2}{50} = \frac{2 \cdot 100}{50} = 4;$$

$$u_{\text{кр}} = 3{,}84\Big(u_{\text{кр}}\big(0{,}05; 2-0-1\big)\Big) = u_{\text{кр}}\big(0{,}05; 1\big) = 3{,}84.$$

Вывод: $u_{\text{набл}} = 4 > u_{\text{кр}} = 3{,}84 \Rightarrow$ отвергаем H_0.

Особое внимание стоит уделить *тестированию нормальности*:

$$H_0\colon x \sim N;$$

$$H_a\colon x \nsim N.$$

Критерий Харке — Бера применяется для проверки нормальности распределения. Он проверяет не совпадение всего распределения с нормальным, а только моментов распределения. Как известно, для нормального распределения асимметрия равна нулю, а эксцесс — трем. Критерий проверяет незна-

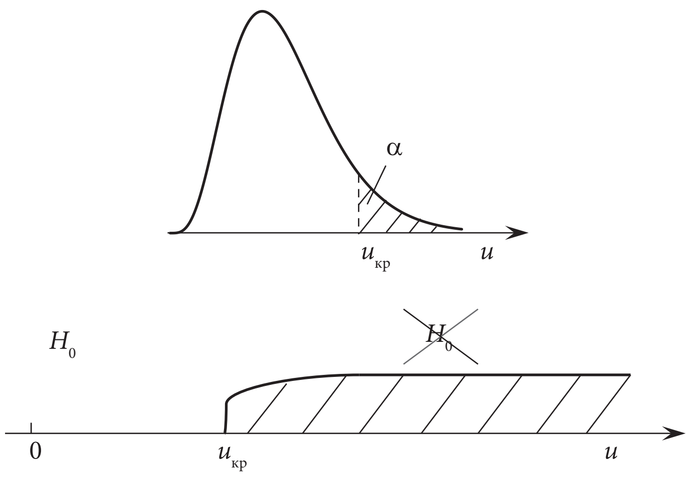


Рис. 7.1. Геометрическая иллюстрация к критерию согласия Пирсона

чимость отличия выборочных моментов распределения от теоретических для нормального распределения:

$$H_0: As = 0,\ K = 3;$$

$$H_a: As \neq 0 \quad \text{или} \quad K \neq 3.$$

Для проверки гипотезы используется следующий статистический критерий:

$$JB = \frac{N}{6}\left(As^2 + \frac{1}{4}\left(K-3\right)^2\right) \sim \chi_2^2.$$

В случае справедливости нулевой гипотезы данный критерий имеет асимптотическое распределение χ^2 с двумя степенями свободы. Мы отвергаем нулевую гипотезу, если наблюдаемое значение критерия больше критической точки распределения χ^2 с двумя степенями свободы для выбранного уровня значимости: $JB_{\text{набл}} > \chi^2\left(\alpha, 2\right) \Rightarrow$ отвергаем H_0.

Критерий Харке — Бера обладает малой мощностью, т. е. он часто распределения, не являющиеся нормальными, относит к нормальному распределению.

7.8. Проверка нормальности из графического анализа гистограмм

Проверка гипотезы о равенстве математического ожидания нормально распределенной случайной величины некоторому числу

$$X \sim N\left(m_x, \sigma_x\right).$$

H_0: $m_x = m_0$, m_0 — некоторое известное число.

1. σ_x — известно.

Имеем x_i: $x_1, x_2, \ldots, x_n \to \overline{X}$.

Вспомним факты относительно выборочного среднего X:

1) $\overline{x} = \dfrac{\sum x_i}{n}$, $\overline{x} \sim N$;

2) $\mathrm{E}_{\overline{x}} = \mathrm{E}_x = m_0$, если верно H_0;

3) $Var\,\overline{X} = \dfrac{Var\,X}{n} = \dfrac{\sigma_x^2}{n} \Rightarrow \sigma_{\overline{x}} = \dfrac{\sigma_x}{\sqrt{n}}$.

Тогда при нулевой гипотезе имеем:

$$\overline{x} \sim N\left(m_0, \frac{\sigma_x^2}{n}\right);$$

$$\bar{x} \sim N\left(m_0, \frac{\sigma_x}{\sqrt{n}}\right);$$

$$\frac{\bar{x} - m_0}{\sigma_x / \sqrt{n}} \sim N(0,1).$$

Рассмотрим:

$$z = \frac{(\bar{x} - m_0)\sqrt{n}}{\sigma_x} \sim N(0,1).$$

Данную величину называют z-статистикой.

а) H_a: $m_x \neq m_0$.

Рассматриваем $w_{\text{кр}} = \{|z| > z_{\text{кр}}\}$ при $P\left(|z| > z_{\text{кр}} | H_0\right) = \alpha$ (рис. 7.2).

Ищем $z_{\text{кр}}$: $\Phi\left(z_{\text{кр}}\right) = \frac{1}{2} - \frac{\alpha}{2}$.

Вывод: если $|z_{\text{набл}}| > z_{\text{кр}} \Rightarrow H_0$ отвергаем; если $|z_{\text{набл}}| < z_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

Пример 7.16. Проверить, что средний рост равен 170 см.

$$H_0: m_x = 170;$$

$$H_a: m_x \neq 170;$$

$$\alpha = 0{,}05 \Rightarrow z_{\text{кр}} = 1{,}96;$$

$$\sigma_x = 6 \text{ см}, n = 100, \bar{x} = 169 \text{ см};$$

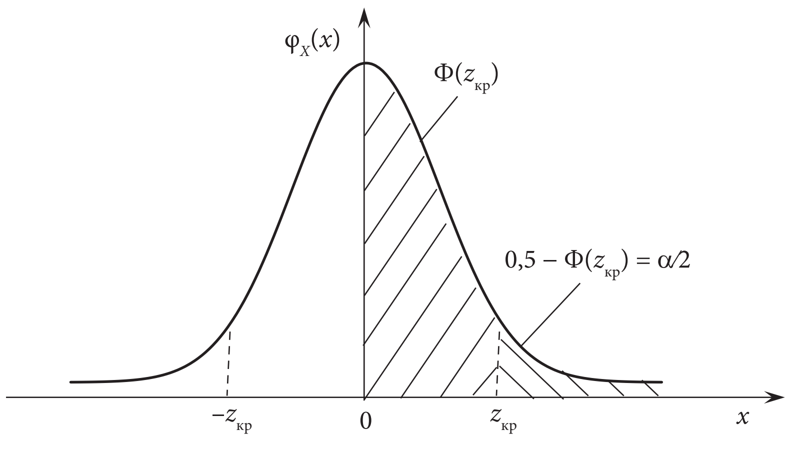


Рис. 7.2. Геометрическая иллюстрация к проверке гипотезы о равенстве математического ожидания

$$z=\frac{\left(169-170\right)\sqrt{100}}{6}=-1,67.$$

Вывод: $\left|z_{\text{набл}}\right|<z_{\text{кр}}\Rightarrow$ нет оснований отвергнуть H_0.

Здесь можно задать закономерный вопрос: какие m_0 не отвергнутся нашей проверкой?

$$\left|z_{\text{набл}}\right|<z_{\text{кр}};$$

$$z=\frac{\left(\overline{x}-m_0\right)\sqrt{n}}{\sigma_x}<z_{\text{кр}}$$ — решим неравенство относительно m_0:

$$\left|\overline{x}-m_0\right|\sqrt{n}<z_{\text{кр}}\sigma_x;$$

$$\left|\overline{x}-m_0\right|<\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};$$

$$-\overline{x}-\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}}\le -m_0\le -\overline{x}+\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};$$

$$\overline{x}-\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}}\le m_0\le \overline{x}+\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};$$

$$\left(\overline{x}-\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};\,\overline{x}+\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}}\right);$$

$$CI\left(\gamma,m_x\right)=\left(\overline{x}-\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}};\,\overline{x}+\frac{z_{\text{кр}}\sigma_x}{\sqrt{n}}\right),$$

где $\gamma = 1 - \alpha$ — уровень надежности; *CI* — доверительный интервал (*Confidence interval*) — интервал, который с заданной надежностью γ накрывает истинное значение математического ожидания случайной величины *X*.

б) H_a: $m_x > m_0$.

$$w_{\text{кр}}=\left\{z>z_{\text{кр}}\right\};$$

$$\mathrm{P}\left(z>z_{\text{кр}}\mid H_0\right)=\alpha;$$

$$z_{\text{кр}}:\Phi\left(z_{\text{кр}}\right)=\frac{1}{2}-\alpha;$$

$$\alpha=0,05,\ z_{\text{кр}}^{0,05}=1,65.$$

Вывод: если $z_{\text{набл}} > z_{\text{кр}} \Rightarrow H_0$ отвергаем; если $z_{\text{набл}} < z_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

в) H_a: $m_x < m_0$.

$$w_{\text{кр}} = \left\{ z < z_{\text{кр}} \right\};$$

$$\mathrm{P}\left(z < z_{\text{кр}} \mid H_0 \right) = \alpha;$$

$$z_{\text{кр}}: \Phi\left(z_{\text{кр}} \right) = \frac{1}{2} - \alpha;$$

$$\alpha = 0{,}05,\ z_{\text{кр}}^{0{,}05} = -1{,}65.$$

Вывод: если $z_{\text{набл}} < z_{\text{кр}} \Rightarrow H_0$ отвергаем; если $z_{\text{набл}} > z_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

2. σ_x — неизвестно.

$$x_i: x_1, x_2, \ldots, x_n \to \overline{X}, s_x.$$

Отметим факты, которые справедливы, только если $x \sim N$:

1) $\overline{x} \sim N\left(m_0, \dfrac{\sigma_x^2}{n} \right)$;

2) $\dfrac{s_x^2\left(n-1\right)}{\sigma_x^2} \sim \chi^2\left(n-1\right)$;

3) $\overline{x}$ и $\overline{s}_X$ независимы.

Рассмотрим распределение Стьюдента:

$$t_{\text{набл}} = \frac{N\left(0,1\right)}{\sqrt{\dfrac{\chi^2\left(m\right)}{m}}} = \frac{\dfrac{\left(\overline{x} - m_0\right)\sqrt{n}}{\sigma_x}}{\sqrt{\dfrac{s_x^2\left(n-1\right)}{\sigma_x^2\left(n-1\right)}}} = \frac{\left(\overline{x} - m_0\right)\sqrt{n}}{s_x} \sim t\left(n-1\right);$$

$$t_{\text{набл}} = \frac{\left(\overline{x} - m_0\right)\sqrt{n}}{s_x} \sim t\left(n-1\right).$$

а) H_a: $m_x \neq m_0$.

$$w_{\text{кр}} = \left\{ \left| t \right| > t_{\text{кр}} \right\}.$$

Критическое значение распределения Стьюдента $t_{\text{кр}}^{\text{дв}}(\alpha, n-1)$ находим из таблиц распределения критических точек распределения Стьюдента для выбранного уровня значимости α, числа степеней свободы $n - 1$ и двусторонней критической области. Также данное значение можно получить посредством программы Microsoft Excel, используя предопределенную функцию СТЬЮДРАСПОБР.

После того, как найдем наблюдаемое и критическое значения критерия, делаем **вывод**: если $|t_{\text{набл}}| > t_{\text{кр}} \Rightarrow H_0$ отвергаем; если $|t_{\text{набл}}| < t_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

Пример 7.17. Рассмотрим вес людей, больных гипертонией. Можно ли утверждать, что средний вес людей, больных гипертонией, равен 90 кг, если известно $n = 50$, $\overline{x} = 88$ кг, $s_x = 6$ кг?

Решение.

$$H_0 : m_x = m_0 = 90;$$

$$H_a : m_x \neq 90;$$

$$t = \frac{(88 - 90)\sqrt{50}}{6} = -2,36;$$

$$t_{\text{кр}}^{\text{дв}} = (0,05;\, 49) = 2,01;$$

$$|t_{\text{н}}| = 2,36 > t_{\text{кр}} = 2,01 \Rightarrow \text{отвергаем } H_0.$$

Делаем вывод, что нельзя утверждать, что средний вес больных гипертонией равен 90 кг.

Найдем m_0, которые не отвергнутся нашей проверкой:

$$|t| < t_{\text{кр}};$$

$$\frac{|\overline{x} - m_0|\sqrt{n}}{s_x} < t_{\text{кр}};$$

$$|\overline{x} - m_0| < \frac{t_{\text{кр}} s_x}{\sqrt{n}};$$

$$-\frac{t_{\text{кр}} s_x}{\sqrt{n}} - \overline{x} < -m_0 < \frac{t_{\text{кр}} s_x}{\sqrt{n}} - \overline{x};$$

$$\overline{x} - \frac{t_{\text{кр}} s_x}{\sqrt{n}} < m_0 < \frac{t_{\text{кр}} s_x}{\sqrt{n}} + \overline{x};$$

$$CI(\gamma, m_x) = \left(\overline{x} - \frac{t_{\text{кр}} s_x}{\sqrt{n}};\, \overline{x} + \frac{t_{\text{кр}} s_x}{\sqrt{n}} \right).$$

Таким образом, мы построили доверительный интервал для m_0 нормально распределенной случайной величины с уровнем надежности γ в случае, когда дисперсия неизвестна.

б) H_a: $m_x > m_0$.

Тогда рассматриваем $w_{кр} = \{t > t_{кр}\}$ при $P\left(t > t_{кр} \mid H_0\right) = \alpha$ (рис. 7.3).

Величину $t_{кр}^{одн}(\alpha; n-1)$ находим в таблицах критических точек распределения Стьюдента (см. Приложение, табл. 5), с уровнем значимости α, числом степеней свободы n – 1 и односторонней критической областью или с помощью функции СТЬЮДЕНТ.ОБР.ПХ в программе Microsoft Excel, используя равенство

$$t_{кр}^{одн}\left(\alpha; n-1\right) = t_{кр}^{дв}\left(2\alpha; n-1\right).$$

После нахождения критического и наблюдаемого значений критерия делаем **вывод**: если $t_{набл} > t_{кр} \Rightarrow H_0$ отвергаем; если $t_{набл} < t_{кр} \Rightarrow$ нет оснований отвергнуть H_0.

в) H_a: $m_x < m_0$.

Рассмотрим $w_{кр} = \{t < t_{кр}\}$ при $P(t < t_{кр} \mid H_0) = \alpha$, тогда $t_{кр} = -t_{кр}^{одн}(\alpha; n-1)$.

Вывод: если $t_{набл} < t_{кр} \Rightarrow H_0$ отвергаем; если $t_{набл} > t_{кр} \Rightarrow$ нет оснований отвергнуть H_0.

Проверка гипотезы о равенстве математических ожиданий двух нормально распределенных случайных величин

Пусть $X \sim N(m_x; \sigma_x^2)$; $Y \sim N(m_y; \sigma_y^2)$.

Предполагаем, что $\sigma_x = \sigma_y$. Тогда:

$$H_0: m_x = m_y;$$

$$H_a: m_x \neq m_y.$$

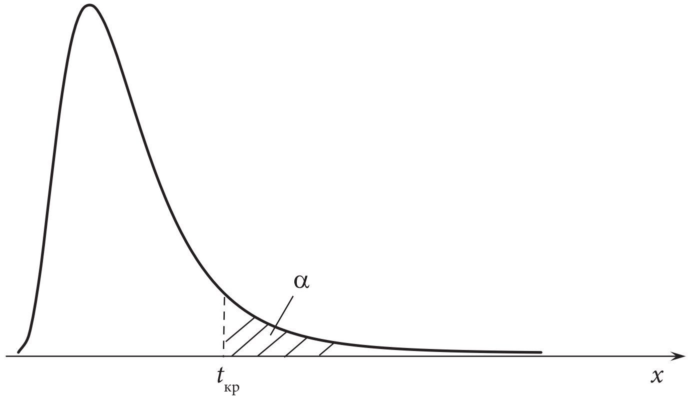


Рис. 7.3. Геометрическая иллюстрация проверки правосторонней гипотезы

Имеем:

$$X\colon x_1, x_2, \ldots, x_{N_x} \to \text{считаем } \overline{x},\ s_x^2 \text{ (исправленная дисперсия);}$$

$$Y\colon y_1, y_2, \ldots, y_{N_y} \to \text{считаем } \overline{y}, s_y^2 \text{ (исправленная дисперсия).}$$

Тогда:

$$t = \frac{\overline{x} - \overline{y}}{\sqrt{s_x^2\left(n_x - 1\right) + s_y^2\left(n_y - 1\right)}} \sqrt{\frac{n_x n_y\left(n_x + n_y - 2\right)}{n_x + n_y}} \sim t\left(n_x + n_y - 2\right);$$

$$w_{\text{кр}} = \left\{\left|t\right| > t_{\text{кр}}\right\};$$

$$\mathrm{P}\left(\left|t\right| > t_{\text{кр}} \mid H_0\right) = \alpha.$$

Величину $t_{\text{кр}}^{\text{дв}}(\alpha; n_x + n_y - 2)$ ищем в таблице критических точек распределения Стьюдента (см. Приложение, табл. 5) со степенью свободы $n_x + n_y - 2$ и уровнем значимости α.

Далее делаем **вывод**: если $|t_{\text{набл}}| > t_{\text{кр}} \Rightarrow H_0$ отвергаем; если $|t_{\text{набл}}| < t_{\text{кр}} \Rightarrow$ нет оснований отвергнуть H_0.

Проверка гипотезы о равенстве дисперсий двух нормально распределенных совокупностей

Пусть $X \sim N(m_x; \sigma_x)$; $Y \sim N(m_y; \sigma_y)$.

$$H_0\colon \sigma_x^2 = \sigma_y^2;$$

$$H_a\colon \sigma_{\text{больш}}^2 > \sigma_{\text{меньш}}^2,$$

где $\sigma_{\text{больш}}^2$ — это X или Y в зависимости от того, чья исправленная дисперсия больше.

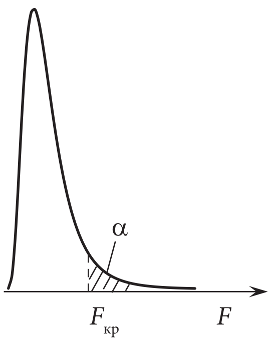
$$F_{\text{набл}} = \frac{s_{\text{больш}}^2}{s_{\text{меньш}}^2} \sim F\left(n_{\text{больш}} - 1,\ n_{\text{меньш}} - 1\right);$$

$$n_{\text{больш}} - 1 \text{ — степень свободы числителя;}$$

$$n_{\text{меньш}} - 1 \text{ — степень свободы знаменателя,}$$

где $n_{\text{больш}}$ — объем выборки, по которой вычислена большая из исправленных дисперсий.

Рассмотрим $w_{\text{кр}} = \{F > F_{\text{кр}}\}$ при $\mathrm{P}\left(F > F_{\text{кр}} \mid H_0\right) = \alpha$ (рис. 7.4), тогда $F_{\text{кр}}(n_{\text{больш}} - 1,$ $n_{\text{меньш}} - 1)$.



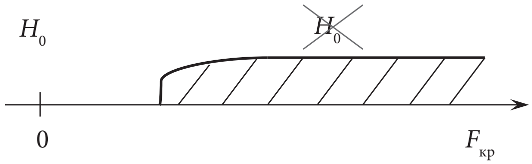


Рис. 7.4. Геометрическая иллюстрация к проверке гипотезы о равенстве дисперсий

Величину $F_{кр}$ находим в таблицах критических точек распределения Фишера (см. Приложение, табл. 6) для выбранного уровня значимости α и числа степеней свободы числителя: $n_{больш}-1$ и знаменателя: $n_{меньш}-1$ или с помощью функции FРАСПОБР в программе Microsoft Excel.

Вывод: если $F_{набл} < F_{кр} \Rightarrow H_0$ отвергаем; если $F_{набл} > F_{кр} \Rightarrow$ нет оснований отвергнуть H_0.

Пример 7.18. Можно ли достоверно утверждать, что средний рост мужчин и женщин одинаковый?

$$\overline{x} = 168 \text{ см}, s_x^2 = 36 \text{ см}^2, n_x = 100;$$

$$\overline{y} = 171 \text{ см}, s_y^2 = 40 \text{ см}^2, n_y = 50.$$

Решение. Для начала проверим равенство дисперсий по Фишеру, чтобы можно было проверить равенство математических ожиданий по Стьюденту.

$$H_0: \sigma_x^2 = \sigma_y^2.$$

$$H_a: \sigma_y^2 > \sigma_x^2, \text{ так как } s_y^2 > s_x^2;$$

$$F_{набл} = \frac{40}{36} = 1{,}11; \quad F_{кр}\left(0{,}05; 99; 49\right) = 1{,}47;$$

$$F_{набл} = 1{,}11 < F_{кр} = 1{,}47 \Rightarrow$$

нет оснований отвергнуть H_0, а значит, можем сделать проверку гипотезы с помощью критерия Стьюдента. Если бы отвергли, то тогда использовали бы критерий Стьюдента для неравных дисперсий, но в данном пособии таких критериев мы рассматривать не будем.

$$H_0: m_x = m_y;$$

$$H_a: m_x \neq m_y;$$

$$t_{\text{набл}} = \frac{168-171}{\sqrt{36\cdot 99 - 40\cdot 49}} \sqrt{\frac{100\cdot 50\cdot (100+50-2)}{100+50}} = -2{,}84;$$

$$t_{\text{кр}}^{\text{дв}}(0{,}05; 148) = 1{,}99.$$

Вывод: $|t_{\text{набл}}| = 2{,}84 > t_{\text{кр}} = 1{,}99 \Rightarrow H_0$ отвергаем, т. е. нельзя сказать, что средний рост одинаковый.

Проверка гипотез о равенстве долей признака

Сравнение долей признака в двух совокупностях — достаточно часто встречающаяся на практике задача. Например, если доля признака в одной группе отличается от такой же доли в другой, то указывает ли это на то, что наличие признака в одной совокупности действительно вероятнее или полученное расхождение долей является случайным.

Оценку разности между двумя долями признака в независимых выборках можно осуществить двумя способами. Вначале рассмотрим процедуру, в которой наблюдаемое значение z-статистики аппроксимируется стандартизованным нормальным распределением. Позже опишем процедуру, в которой используется тестовая χ^2-статистика, аппроксимированная χ^2-распределением с одной степенью свободы. Сразу отметим, что эти два критерия эквивалентны.

Для оценки различий между двумя генеральными совокупностями на основе независимых выборок можно применять z-критерий. На основе разности между двумя выборочными долями признака $P_{s_1} - P_{s_2}$ вычисляется z-статистика, используемая для оценки разности между двумя долями признака в генеральных совокупностях. Если объем выборок достаточно велик, эта тестовая статистика имеет стандартизованное нормальное распределение. z-критерий для оценки разности между двумя долями имеет вид:

$$z_{\text{набл}} = \frac{p_{s_1} - p_{s_2}}{\sqrt{\overline{p}(1-\overline{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad \overline{p} = \frac{m_1 + m_2}{n_1 + n_2}, \quad p_{s_1} = \frac{m_1}{n_1}, \quad p_{s_2} = \frac{m_2}{n_1},$$

где p_{s_1} — доля успехов в первой выборке; m_1 — количество успехов в первой выборке; n_1 — объем выборки из первой генеральной совокупности; p_{s_1} — доля успехов в первой генеральной совокупности; p_{s_2} — доля успехов во второй выборке; m_2 — количество успехов во второй выборке; n_2 — объем выборки из второй генеральной совокупности; p_2 — доля успехов во второй генеральной совокупности; $\overline{p}$ — оценка доли успехов в объединенной генеральной совокупности.

При достаточно большом объеме выборок z-статистика подчиняется стандартизованному нормальному распределению.

Нулевая гипотеза заключается в том, что доли признака в двух генеральных совокупностях одинаковы. Следовательно, проверку равенства долей признака в двух генеральных совокупностях можно свести к оценке доли признака в объединенной генеральной совокупности. Оценка объединенной доли равна результату деления количества успехов в обеих выборках $m_1 + m_2$ на сумму объемов выборок $n_1 + n_2$.

С помощью z-критерия можно определить, существуют ли различия между долями успеха в двух группах (двусторонний тест), а также установить, превышает ли доля успехов в одной группе долю успехов в другой (односторонний критерий).

Чтобы проверить нулевую и альтернативные гипотезы H_0: $p_1 = p_2$, H_1: $p_1 \neq p_2$, следует использовать наблюдаемое значение z-статистики. При заданном уровне значимости α нулевая гипотеза отклоняется, если вычисленная z-статистика больше верхнего или меньше нижнего критического значения стандартизованного нормального распределения.

Пример 7.19. Неврологи решили провести исследование по выявлению факторов риска головной боли. Для исследования были собраны данные двух групп: исследуемой (у всех опрошенных присутствует головная боль) в количестве 1 035 человек и контрольной (у всех опрошенных отсутствует головная боль) в количестве 565 человек. В первую очередь проверке подвергся такой фактор, как повышенное артериальное давление. Согласно опросу в исследуемой группе у 650 респондентов наблюдалось повышенное артериальное давление, а в контрольной — только у 150 человек. На уровне значимости 0,05 проверить гипотезу об отсутствии существенных различий между наличием артериальной гипертензии в обеих группах.

Решение. Имеем нулевую гипотезу о том, что доли страдающих повышенным давлением исследуемой и контрольной групп равны. В качестве альтернативной возьмем гипотезу, что доли не равны. Считаем наблюдаемое значение z-критерия:

$$z_{\text{набл}} = \frac{p_{s_1} - p_{s_2}}{\sqrt{\overline{p}\left(1-\overline{p}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} = \frac{\dfrac{650}{1\,035} + \dfrac{150}{565}}{\sqrt{\dfrac{650+150}{1\,035+565}\left(1-\dfrac{650+150}{1\,035+565}\right)\left(\dfrac{1}{1\,035}+\dfrac{1}{565}\right)}} \approx 13,4.$$

Критическое значение находим в таблицах критических точек стандартного *нормального* распределения (значение функции Лапласа) для выбранного уровня значимости $\alpha = 0,05$, и оно равно 1,96, что меньше наблюдаемого значения z-критерия. Следовательно, нет оснований отвергнуть нулевую гипотезу, иными словами, полученные данные не противоречат гипотезе об одинаковых долях наличия повышенного давления в обеих группах.

Далее рассмотрим способ, основанный на сравнении количества успехов в двух группах, а не их долей в генеральных совокупностях. В этой процедуре наблюдаемое значение χ^2-статистики аппроксимируется χ^2-распределением с одной степенью свободы. Результат, полученный с помощью χ^2-критерия, эквивалентен результату применения z-критерия.

Для сравнения количества успехов в двух независимых группах необходимо заполнить таблицу перекрестной классификации с двумя входами, содержащую количество успехов и неудач в каждой из групп. Такую таблицу часто называют таблицей сопряженности признаков или факторной.

Строки	Столбцы (группы)		Итого
	1	2	
Успехи	m_1	m_2	m
Неудачи	$n_1 - m_1$	$n_2 - m_2$	$n - m$
Всего	n_1	n_2	n

Здесь использованы следующие обозначения: m_1 — количество успехов в первой группе; m_2 — количество успехов во второй группе; $n_1 - m_1$ — количество неудач в первой группе; $n_2 - m_2$ — количество неудач во второй группе; $m = m_1 + m_2$ — общее количество успехов; $n - m = (n_1 - m_1) + (n_2 - m_2)$ — общее количество неудач; n_1 — объем первой выборки; n_2 — объем второй выборки; $n = n_1 + n_2$ — суммарный объем выборок. Представленная таблица имеет две строки и два столбца, поэтому она называется факторной таблицей 2×2. Ячейки, образованные пересечением каждой строки и столбца, содержат количество успехов или неудач.

Чтобы проверить нулевую и альтернативные гипотезы H_0: $p_1 = p_2$; H_1: $p_1 \neq p_2$, используем тестовую χ^2-статистику.

Критерий «хи-квадрат» для сравнения двух долей

Наблюдаемое значение χ^2-статистики равно сумме квадратов разностей между наблюдаемым и ожидаемым количеством успехов, деленных на ожидаемое количество успехов в каждой ячейке таблицы:

$$\chi^2 = \sum_{\text{по всем ячейкам}} \frac{\left(f_0 - f_e\right)^2}{f_c},$$

где f_0 — наблюдаемое количество успехов или неудач в конкретной ячейке таблицы сопряженности признаков; f_e — теоретическое, или ожидаемое, количество успехов или неудач в конкретной ячейке таблицы сопряженности признаков при условии, что нулевая гипотеза является истинной.

Наблюдаемое значение χ^2-статистики аппроксимируется χ^2-распределением с одной степенью свободы.

Чтобы вычислить ожидаемое количество успехов или неудач в каждой ячейке таблицы сопряженности признаков, необходимо понимать их смысл. Если нулевая гипотеза является истинной, т. е. доли успехов в двух генеральных совокупностях равны, выборочные доли, вычисленные для каждой из двух групп, могут отличаться друг от друга лишь по случайным причинам, причем обе доли являются оценкой общего параметра генеральной совокупности p. В этой ситуации статистика, объединяющая обе доли в одной общей (средней) оценке параметра p, содержит больше информации, чем каждая из них в отдельности. Эта статистика, обозначаемая символом $\overline{p}$, представляет собой общую долю успехов в объединенных группах (т. е. равна общему количеству успехов, деленному на суммарный объем выборок). Ее дополнение, $1-\overline{p}$, представляет собой общую долю неудач в объединенных группах. Используя обозначения, смысл которых описан выше в таблице, можно вывести формулу для вычисления параметра $\overline{p}=\dfrac{X_1+X_2}{n_1+n_2}=\dfrac{X}{n}$, где $\overline{p}$ — средняя доля признака.

Чтобы вычислить ожидаемое количество успехов f_e (т. е. содержимое первой строки таблицы сопряженности признаков), необходимо умножить объем выборки на параметр $\overline{p}$. Чтобы вычислить ожидаемое количество неудач f_e (т. е. содержимое второй строки таблицы сопряженности признаков), необходимо умножить объем выборки на параметр $1-\overline{p}$.

Наблюдаемое значение статистики аппроксимируется χ^2-распределением с одной степенью свободы. При заданном уровне значимости α нулевая гипотеза отклоняется, если вычисленная χ^2-статистика больше $\chi_U^{\ 2}$, верхнего критического значения χ^2-распределения с одной степенью свободы. Таким образом, решающее правило выглядит следующим образом: гипотеза H_0 отклоняется, если $\chi^2 > \chi_U^{\ 2}$, в противном случае гипотеза H_0 не отклоняется.

Если нулевая гипотеза является истинной, вычисленная χ^2-статистика близка к нулю, поскольку квадрат разности между наблюдаемой f_0 и ожидаемой f_e величинами в каждой ячейке очень мал. С другой стороны, если нулевая гипотеза H_0 является ложной и между долями успехов в генеральных совокупностях существует значимая разница, вычисленная χ^2-статистика должна быть большой. Это объясняется разностью между наблюдаемым и ожидаемым количеством успехов или неудач в каждой ячейке, которая увеличивается при возведении в квадрат. Однако вклады разностей между ожидаемыми и наблюдаемыми величинами в общую χ^2-статистику могут быть неодинаковыми. Одна и та же фактическая разность между f_0 и f_e может оказать большее влияние на χ^2-статистику, если в ячейке содержатся результаты небольшого количества наблюдений, чем разность, соответствующая большему количеству наблюдений.

Проверка предположений, касающихся факторной таблицы 2×2

Для получения точных результатов на основе данных, приведенных в таблице 2×2, необходимо, чтобы количество успехов или неудач превышало число 5. Если это условие не выполняется, следует применять точный *критерий Фишера*.

При сравнении долей критерии z и χ^2 приводят к одинаковым результатам. Это можно объяснить существованием тесной связи между стандартизованным нормальным распределением и χ^2-распределением с одной степенью свободы. В этом случае χ^2-статистика всегда является квадратом z-статистики.

Таким образом, можно утверждать, что при проверке нулевой и альтернативной гипотез H_0: $p_1 = p_2$; H_1: $p_1 \neq p_2$ критерии z и χ^2 являются эквивалентными. Однако, если необходимо не просто обнаружить различия, но и определить, какая доля больше ($p_1 > p_2$), следует применять z-критерий с одной критической областью, ограниченной хвостом стандартизованного нормального распределения.

Стоит заметить, что критерий «*хи*-квадрат» можно распространить на более общий случай и применять для проверки гипотезы о равенстве нескольких долей признака.

Задачи для самостоятельного решения

7.1. Дан ряд распределения случайной величины:

X	-1	0	2
P	θ	$1{,}1 - 2\theta$	$\Theta - 0{,}1$

Имеется независимая выборка $X_1 = 0$, $X_2 = -1$. Найти оценку неизвестного параметра методом максимального правдоподобия, методом моментов, если возможно, проверить состоятельность и несмещенность. Найти дисперсию оценки метода моментов.

7.2. Асимметричная монетка подбрасывается N раз, k раз выпадает «орел». Оценить вероятность выпадения «орла», если возможно, проверить состоятельность и несмещенность.

7.3. Методом моментов построить оценку параметра следующего распределения: $f(x, \theta) = \dfrac{2x}{\theta^2}$, $x \in [0, \theta]$. Проверить состоятельность, несмещенность, найти дисперсию оценки. Можно ли здесь применить неравенство Рао — Фреше — Крамера для проверки эффективности оценки?

7.4. Методом моментов и методом максимального правдоподобия построить оценку параметра следующего распределения: $f(x,\theta)=\theta x^{\theta-1},\ x>0$. Проверить состоятельность, несмещенность, найти дисперсию оценки. Можно ли здесь применить неравенство Рао — Фреше — Крамера для проверки эффективности оценки?

7.5. Пусть p_0, p_1, и p_2 — доли семей студентов, у которых 0, 1 и два ребенка соответственно (в нашей популяции нет семей студентов с тремя и более детьми). Известно, что $p_1=3p_2$. У четырех случайно опрошенных семей оказалось следующее количество детей: 1, 2, 0, 1. Найти оценки максимального правдоподобия и метода моментов величин p_0, p_1 и p_2.

7.6. Закон распределения случайной величины задан таблицей:

X	-1	0	1
p	θ	2θ	$1-3\theta$

а) Имеется три наблюдения случайной величины: $X_1=0$, $X_2=1$, $X_3=0$. Найти *ML*- и *MM*-оценки неизвестного параметра.

б) Решить эту задачу для произвольной случайной выборки $X_1, \ldots, X_n$.

в) Для метода моментов проверить состоятельность и несмещенность, найти дисперсию оценки.

7.7. Построить там, где возможно, оценки *ML* и *MM* неизвестных параметров по случайной выборке $X_1, \ldots, X_n$ и везде, где можно, проверить состоятельность, несмещенность, найти информацию Фишера о неизвестном параметре и дисперсию оценок:

а) неизвестного параметра p геометрического распределения;

б) неизвестного параметра p биномиального распределения;

в) $f(x,\theta)=\dfrac{2x}{\theta^2}$ при $x\in(0,\theta)$;

г) $f(x,\theta)=\dfrac{\theta\left(\ln^{\theta-1}x\right)}{x}$ при $x\in(1,e)$.

7.8. По случайной выборке $X_1, \ldots, X_n$ из нормального распределения $N(\mu_1,\mu_2-\mu_1^2)$ методом моментов оценить параметры μ_1, μ_2. Дать определения несмещенности и состоятельности и проверить выполнение этих свойств для оценки μ_1.

7.9. По случайной выборке $X_1, \ldots, X_n$ из нормального распределения $N(\theta, 1)$ методом максимального правдоподобия оценить параметр θ. Будет ли найденная оценка эффективной (ответ обосновать)?

7.10. Имеется квадрат со стороной a, нам необходимо оценить площадь этого квадрата. Для этого проводим n независимых замеров стороны квадрата и получаем выборку длин сторон $X_1, \ldots, X_n$, $X_i \sim N(a, \sigma^2)$. Предлагаются две оценки площади квадрата:

а) $\dfrac{\sum X_i^2}{N}$ — средний квадрат;

б) $\overline{X}^2$ — квадрат среднего.

Какая оценка предпочтительнее?

7.11. Имеются две оценки дисперсии случайной величины σ^2, построенные по независимой выборке: выборочная дисперсия d_X^2 и исправленная выборочная дисперсия s_X^2:

а) показать, что d_X^2 смещена, а s_X^2 нет;

б) но $Var(d_X^2) < Var(s_X^2)$;

в) сравнить $MSE(d_X^2)$ и $MSE(s_X^2)$.

Являются ли эти оценки состоятельными?

7.12. Пусть $X = (X_1, \ldots, X_n)$ — случайная выборка из нормального распределения с нулевым математическим ожиданием и дисперсией θ.

а) Используя второй начальный момент, найти оценку параметра θ методом моментов.

б) Сформулировать определение несмещенности оценки и проверить выполнение данного свойства для оценки, найденной в пункте «а».

в) Сформулировать определение состоятельности оценки и проверить выполнение данного свойства для оценки, найденной в пункте «а».

г) Найти оценку параметра θ методом максимального правдоподобия.

д) Вычислить информацию Фишера о параметре θ, заключенную в n наблюдениях случайной выборки.

е) Сформулировать неравенство Рао — Фреше — Крамера.

ж) Сформулировать определение эффективности оценки и проверить выполнение данного свойства для оценки, найденной в пункте «г».

7.13. Транспортная компания утверждает, что 10 % пассажиров, купивших билет, не являются на рейс. В случайной выборке из шести рейсов автобуса, имеющего 180 посадочных мест, число неявившихся оказалось: 5, 10, 25, 0, 17, 30. Пусть число пассажиров X, неявившихся на рейс, хорошо описывается распределением Пуассона. При помощи метода максимального правдоподобия найти:

а) оценку $E(X)$ и ее числовое значение по выборке;

б) оценку дисперсии X и ее числовое значение по выборке;

в) оценку стандартного отклонения X и ее числовое значение по выборке;

г) оценку вероятности того, что на рейс явятся все пассажиры, а также найти ее числовое значение по выборке;

д) используя асимптотические свойства оценок максимального правдоподобия, построить 95 %-ный доверительный интервал для $E(X)$.

7.14. По выборке $X_1, \ldots, X_n$ из равномерного распределения $R\,(0, \theta)$ с неизвестным параметром $\theta > 0$ требуется оценить θ. Будут ли оценки $T_1 = 2\bar{X}$, $T_2 = \dfrac{n+1}{n}\bar{X}$ несмещенными? Какая из них является более точной (эффективной)? Являются ли эти оценки состоятельными?

7.15. Пусть X — случайная величина с неизвестным математическим ожиданием m и дисперсией σ^2. Для оценки параметра m были сделаны две независимые выборки. По одной было посчитано выборочное среднее $\bar{X}_1$, объем выборки составил N_1, по второй — $\bar{X}_2$, объем выборки N_2. На основании этих двух оценок предложены два способа получить оценку m:

$$\hat{m}_1 = \frac{\bar{X}_2 + \bar{X}_2}{2};$$

$$\hat{m}_2 = \frac{N_1\bar{X}_2 + N_2\bar{X}_2}{N_1 + N_2}.$$

Какой способ лучше?

7.16. У 20 студентов, сдававших экзамен по теории вероятностей, сердце билось в среднем со скоростью 96 уд. в минуту. Стандартное отклонение выборки было равно 5 уд. в минуту. Найти 95 %-ный доверительный интервал для генерального среднего.

7.17. Средний доход фирмы в день составлял 1 020 ед. После реорганизации выборочный средний доход в день за 30 рабочих дней составил 1 070 ед. с выборочным средним квадратичным отклонением (исправленным) 90 ед. Можно ли утверждать (на уровне значимости 5 %), что реорганизация привела к увеличению среднего дохода?

7.18. В книжном магазине проведено исследование продаж фантастического романа писателя Бурьяненко «Танцы в пустоте». В течение 25 рабочих дней роман продавался ежедневно в среднем по 64 экз. со средним квадратичным отклонением 10 экз. Можно ли утверждать на уровне значимости 5 %, что этот роман расходится хуже, чем предыдущий роман автора «Звездная жуть», если тот расходился в среднем по 70 экз. в день?

7.19. Инвестиция 1 рассчитана на 12 лет, дисперсия ежегодных прибылей составляет 20 %. Инвестиция 2 рассчитана на 10 лет, дисперсия ежегодных прибылей составляет 30 %. Предполагается, что распределение ежегодных прибылей на инвестиции подчиняется нормальному закону распределения. Равны ли риски 1 и 2? Доверительная вероятность 95 %.

7.20. Пробег автомобиля до капитального ремонта является нормальной случайной величиной с неизвестным математическим ожиданием и неизвестной дисперсией. По выборке из 100 наблюдений найти 99 %-ный доверительный интервал для математического ожидания пробега, если известно, что интервал (92,16; 107,84) соответствует 95 %-ному доверительному интервалу.

7.21. Опрос домохозяйств в регионе X и Y выявил следующие результаты:

$$\sum_i X_i = 540, \quad \sum_i Y_i = 480, \quad \sum_i \left(X_i - 36\right)^2 = 410{,}26;$$

$$\sum_i \left(Y_i - 40\right)^2 = 358{,}3, \quad \sum_i \left(X_i - 40\right)^2 = 426{,}26, \quad \sum_i \left(Y_i - 36\right)^2 = 376{,}3.$$

Построить 90 %-ный доверительный интервал для математического ожидания региона Y. На 5 %-ном уровне значимости проверить гипотезу о том, что средний доход в H не превышает средний доход в X, предполагая, что распределение доходов нормально.

7.22. Предприимчивый Семен Семенович решил заключить контракт с новой фирмой-поставщиком, менеджер которой утверждает, что средний срок безотказной работы предлагаемого изделия — 2 900 ч. Однако Семен Семенович не верит простым словам и все перепроверяет. И в этот раз при помощи своих знакомых-бизнесменов он насобирал выборку из 50 изделий, по которой средний срок безотказной работы оказался равным 2 720 ч с исправленным выборочным средним квадратичным отклонением 700 ч. Помогите Семену Семеновичу при 5 %-ном уровне значимости проверить, что средний срок не окажется меньше, чем гарантирует это фирма-поставщик.

7.23. На основании сделанного прогноза средняя дебиторская задолженность однотипных предприятий региона должна составить $m_0 = 120$ ден. ед. Выборочная проверка 10 предприятий дала среднюю задолженность $\overline{x} = 135$ ден. ед., а исправленное среднее квадратичное отклонение задолженности $s = 20$ ден. ед. На уровне значимости 0,05 выяснить, можно ли принять данный прогноз.

7.24. Продавщица тетя Галя утверждает, что все три вида кофе «Невкусе», «Чиго» и «ДжанаМанах» одинаково популярны. Менеджер решил проверить, верно ли это утверждение. За неделю менеджер насчитал, что 45, 41 и 50 человек покупали эти марки кофе соответственно. Правдоподобно ли утверждение тети Гали?

7.25. Для некоторой отрасли проведено исследование об оплате труда мужчин и женщин. Их зарплаты в тысячах рублей в месяц составляют для мужчин: 50, 40, 45, 45, 35, для женщин: 60, 30, 30, 35, 30.

а) Считая, что распределение заработных плат мужчин хорошо описывается нормальным распределением, построить 99 %-ный доверительный интервал для математического ожидания заработной платы мужчин и 90 %-ный доверительный интервал для стандартного отклонения заработной платы женщин.

б) Сформулировать предпосылки, необходимые для построения доверительного интервала для разности математических ожиданий заработных плат мужчин и женщин. Считая предпосылки выполненными, построить 90 %-ный доверительный интервал для разности математических ожиданий заработных плат мужчин и женщин. Можно ли считать зарплаты мужчин и женщин одинаковыми?

7.26. Контрольную работу по теории вероятностей по индивидуальным вариантам выполняли студенты двух групп высшей школы экономики и менеджмента УрФУ. В первой группе было предложено 105 задач, из которых верно решено 60, во второй группе из 140 предложенных задач верно решено 69. На уровне значимости 0,02 проверить гипотезу об отсутствии существенных различий в усвоении учебного материала студентами обеих групп.

7.27. Проводились испытания нового лекарства. В эксперименте участвовали 3 000 мужчин и 3 500 женщин. У 50 мужчин и 110 женщин наблюдались побочные эффекты. Можно ли утверждать, что побочные эффекты от нового лекарства у женщин возникают чаще, чем у мужчин? Доверительная вероятность $\gamma = 95\,\%$.

7.28. Руководство фирмы, владеющей тремя магазинами, решило выяснить, посещают ли покупатели все три магазина одинаково охотно либо имеется некоторое различие. Для проверки была собрана информация о количестве покупателей, сделавших покупки в течение недели. Оказалось, что в первом магазине это число составляет 160 человек, во втором — 225, в третьем — 215. Каков будет результат?

7.29. Наблюдение за 1 000 автомобилями на скоростной дороге с четырьмя полосами движения показало, что первую полосу предпочли 294 водителя, вторую — 276, третью — 238, а остальные выбрали четвертую. Можно ли на основании этих данных утверждать, что равное количество водителей выбирает каждую из полос?

7.30. На основании сделанного прогноза средняя дебиторская задолженность однотипных предприятий региона должна составить $a_0 = 120$ ден. ед. Выборочная проверка 10 предприятий дала среднюю задолженность 135 ден. ед., а среднее квадратичное отклонение задолженности $s = 20$ ден. ед. На уровне значимости 0,05 выяснить, можно ли принять данный прогноз.

7.31. В рекламе утверждается, что месячный доход по акциям A превышает доход по акциям B более чем на 0,3 % (или на 0,003). В течение годичного периода средний месячный доход по акциям B составил 0,5 %, а по акциям A — 0,65 %, его средние квадратичные отклонения соответственно 1,9 и 2,0 %.

Полагая распределения доходности по каждой акции нормальными, на уровне значимости 0,05 проверить утверждение, содержащееся в рекламе.

7.32. Поступление страховых полисов в 130 филиалах страховых компаний в регионе A составило $26 \cdot 10^4$ у. е., в регионе B на 100 филиалов пришлось $18 \cdot 10^4$ у. е. Дисперсия величины страховых взносов в регионе A равна $39 \cdot 10^8$ (у. е.)^2, в регионе B — $25 \cdot 10^8$ (у. е.)^2. На уровне значимости $\alpha = 0,05$ определить, существенно ли различается средняя величина поступления страховых взносов в регионах A и B из расчета на один филиал.

7.33. Производительность двух моторных заводов, выпускающих дизельные двигатели, характеризуется следующими данными:

1-й завод	72	84	69	74	82	67	75	86	68	61
2-й завод	55	65	73	66	58	71	77	68	68	59

Можно ли считать одинаковыми производительности дизельных двигателей на обоих заводах при уровне значимости 0,05?

7.34. Проведено исследование розничного товарооборота продовольственных магазинов в двух районах области (по 50 магазинов в каждом). Априори известны средние значения товарооборота 78,9 и 78,68 тыс. руб. Полученные в результате оценки среднеквадратичных отклонений в первом и во втором районах области соответственно равны 7,22 и 7,79 тыс. руб. Можно ли считать, что разброс розничного товарооборота магазинов в районах неодинаков при уровне значимости 0,05?

7.35. Менеджер банка утверждает, что размер кредита, выдаваемого клиентам, составляет в среднем 4 800$ со стандартным отклонением 800$. В выборке из 25 клиентов, бравших кредиты, средний размер кредита оказался равен 4 235$. При $\alpha = 0,10$ есть ли достаточные основания опровергать утверждение менеджера?

7.36. В таблице представлены результаты наблюдений: X — число сделок на фондовой бирже за квартал, $N = 400$ (инвесторов).

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	146	97	73	34	23	10	6	3	4	2	2

Найти: выборочное среднее, выборочную дисперсию, выборочное среднеквадратичное отклонение, исправленную выборочную дисперсию, исправленное выборочное среднеквадратичное отклонение, моду, медиану, асимметрию, эксцесс.

7.37. В таблице представлены результаты наблюдений: X — месячный доход жителя региона (в руб.), $N = 1\,000$ (жителей).

x_i	Менее 500	500–1 000	1 000–1 500	1 500–2 000	2 000–2 500	Свыше 2 500
n_i	58	96	239	328	147	132

Найти: выборочное среднее, выборочную дисперсию, выборочное среднеквадратичное отклонение, исправленную выборочную дисперсию, исправленное выборочное среднеквадратичное отклонение, моду, медиану, асимметрию, эксцесс.

7.38. Имеется монетка. Мы предполагаем, что монетка «жульническая» и «орел» выпадает в три раза чаще, чем «решка». Для проверки этой гипотезы предложена следующая процедура: бросаем монетку 3 раза и считаем, что монетка жульническая, если «орел» выпал 3 раза. Найти вероятность ошибок 1-го и 2-го рода.

7.39. В таблице представлены данные о годовых доходах и расходах на личное потребление (в долларах) для 10 семей. Найти выборочную ковариацию.

Годовой доход	Расходы на личное потребление
2 508	2 406
2 572	2 464
2 408	2 336
2 522	2 281
2 700	2 641
2 531	2 385
2 390	2 297
2 595	2 416
2 524	2 460
2 685	2 448

7.40. Имеется равномерное распределение с неизвестным параметром k:

$$f(x)=\begin{cases}\dfrac{1}{k}, & x\in[0;k],\\ 0, & x\notin[0;k].\end{cases}$$

Нужно проверить нулевую гипотезу H_0: $k=1$ при альтернативной гипотезе H_a: $k=2$.

Для проверки имеется лишь одно наблюдение x_1. Предложены два способа:

а) H_0 отвергается при $x_1\geq\dfrac{1}{2}$;

б) H_0 отвергается при $1\leq x_1\leq 2$.

Найти вероятность ошибок 1-го и 2-го рода.

7.41. Из генеральной совокупности извлечена выборка объема $n = 12$:

Варианта x_i	–0,5	–0,4	–0,2	0	0,2	0,6	0,8	1	1,2	1,5
Частота n_i	1	2	1	1	1	1	1	1	2	1

Оценить с надежностью 0,95 математическое ожидание m нормально распределенного признака генеральной совокупности с помощью доверительного интервала. Решить задачу для надежности 0,9 и 0,99.

7.42. За последние 5 лет годовой рост цены актива A составлял в среднем 20 % со средним квадратичным отклонением (исправленным) 5 %. Построить доверительный интервал с вероятностью 95 % для цены актива в конце следующего года, если в начале года она равна 100 ден. ед.

7.43. Из генеральной совокупности извлечена выборка объема $n = 10$:

Варианта x_i	–2	1	2	3	4	5
Частота n_i	2	1	2	2	2	1

Оценить с надежностью 0,95 математическое ожидание m нормально распределенного признака генеральной совокупности по выборочному среднему при помощи доверительного интервала.

7.44. За последние 9 лет годовой рост цены актива A составлял в среднем 22 % со средним квадратичным отклонением (исправленным) 6 %. Построить доверительный интервал с вероятностью 90 % для средней цены актива в конце следующего года, если в начале она была равна 200 ден. ед.

7.45. Для отрасли, включающей 1 200 фирм, составлена случайная выборка из 19 фирм. По выборке оказалось, что в фирме в среднем работают $\overline{X} = 77{,}5$ человека при среднем квадратичном отклонении $S_x = 25$ человек. Пользуясь 95 %-ным доверительным интервалом, оценить среднее число работающих в фирме по всей отрасли и общее число работающих в отрасли. Предполагается, что количество работников фирмы имеет нормальное распределение.

7.46. Пробег в тысячах километров автомобиля «Лада Калина» до капитального ремонта двигателя является нормальной случайной величиной с неизвестным математическим ожиданием μ и известной дисперсией 49. По выборке из 20 автомобилей найти значение доверительного интервала для математического ожидания пробега с уровнем доверия 0,95.

7.47. Вес упаковки с лекарством является нормальной случайной величиной с неизвестными математическим ожиданием μ и дисперсией σ^2. Контрольное взвешивание 15 упаковок показало, что выборочное среднее равно 60, а несмещенная оценка дисперсии равна 50. Построить доверительные интервалы для математического ожидания и дисперсии веса упаковки (для дисперсии односторонний с нижней границей).

7.48. В ходе анкетирования 20 сотрудников банка «Альфа» ответили на вопрос о том, сколько времени они проводят на работе ежедневно. Среднее выборочное оказалось равно 9,5 ч при стандартном отклонении 0,5 ч. Аналогичные показатели для 25 сотрудников банка «Бета» составили 9,8 и 0,6 ч соответственно. Проверить гипотезу о равенстве дисперсий времени, проводимого на работе, сотрудниками банков «Альфа» и «Бета». Проверить гипотезу о том, что сотрудники банка «Альфа» проводят на работе столько же времени, что и сотрудники банка «Бета».

7.49. Даны независимые выборки доходов выпускников двух ведущих экономических вузов *A* и *B*, по 50 выпускников каждого вуза известно: $\overline{x}_A = 650$, $\overline{x}_B = 690$, $S_A^2 = 50$, $S_B^2 = 70$. Предполагая, что распределение доходов подчиняется нормальному закону, проверить гипотезу об отсутствии преимуществ выпускников вуза *B* на уровне значимости 0,05.

7.50. Бухгалтер компании решил предпринять выборочную проверку и выбрал 18 из 1 200 компонент, продававшихся в прошлом месяце. Стоимость отобранных компонент 82; 30; 98; 116; 80; 150; 200; 88; 70; 90; 160; 100; 86; 76; 90; 140; 76; 68 (ден. ед.). Найти оценку средней стоимости всех компонент и построить для нее доверительный интервал с надежностью 0,95.

ПРИЛОЖЕНИЕ

Таблица 1

Таблица значений функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290

Окончание табл. 1

x	0	1	2	3	4	5	6	7	8	9
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

Таблица 2

Таблица значений функции Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-t^2/2}dt$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000
4	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

4,1	0,499979	4,2	0,499987	4,3	0,499991	4,4	0,499995	4,5	0,4999966
4,6	0,4999979	4,7	0,4999987	4,8	0,4999992	4,9	0,4999995	5,0	0,4999997

Таблица 3

Таблица значений функции Пуассона

k | λ	0,1	0,2	0,3	0,4	0,5	0,6,	0,7	0,8	0.9
0	0.9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066
1	0.0905	0,1638	0,2222	0,2681	0,3033	0,3293	0,3476	0,3596	0,3639
2	0.0045	0,0164	0,0333	0,0536	0,0758	0,0983	0,1217	0,1438	0,1648
3	0.0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494
4	—	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111
5	—	—	—	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020
6	—	—	—	—	—	—	0,0001	0,0002	0,0003
k | λ	1,0	2,0	3,0	4,0	5,0	6,0,	7,0	8,0	9.0
0	0.3679	0,1383	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001
1	0.3679	0,2707	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011
2	0.1839	0,2707	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050
3	0.0613	0,1805	0,224	0,1984	0,1404	0,0892	0,0521	0,0286	0,0150
4	0.0153	0,0902	0,1680	0,1954	0,1755	0,1339	0,0912	0,0572	0,0337
5	0.0031	0,0361	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607
6	0.0005	0,0120	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911
7	0.0001	0,0034	0,0216	0,0595	0,1045	0,1377	0,1490	0,1396	0,1171
8	—	0,0009	0,0081	0,0298	0,0655	0,1033	0,1304	0,1396	0,1318
9	—	0,0002	0,0027	0,0132	0,0363	0,0689	0,1014	0,1241	0,1318
10	—	—	0,0008	0,0053	0,0181	0,0413	0,0710	0,0993	0,1186
11	—	—	0,0002	0,0019	0,0082	0,0255	0,0452	0,0722	0,0970,
12	—	—	0,0001	0,0006	0,0034	0,0113	0,0264	0,0481	0,0728
13	—	—	—	0,0002	0,0013	0,0052	0,0142	0,0296	0,0504
14	—	—	—	0,0001	0,0005	0,0022	0,0071	0,0169	0,0324
15	—	—	—	—	0,0002	0,0009	0,0033	0,0090	0,0194
16	—	—	—	—	—	0,0003	0,0015	0,0045	0,0109
17	—	—	—	—	—	0,0001	0,0006	0,0021	0,0058
18	—	—	—	—	—	—	0,0002	0,0009	0,0029
19	—	—	—	—	—	—	0,0001	0,0004	0,0014
20	—	—	—	—	—	—	—	0,0002	0,0006
21	—	—	—	—	—	—	—	0,0001	0,0003
22	—	—	—	—	—	—	—	—	0,0001

Таблица 4

Критические точки распределения χ^2

k	Уровень значимости α												
	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995
1	7,879	6,635	5,024	3,841	2,706	1,323	0,455	0,102	0,0158	0,0039	0,00098	0,00016	0,00004
2	10,597	9,210	7,378	5,991	4,605	2,773	1,386	0,575	0,2107	0,1026	0,05064	0,02010	0,01003
3	12,838	11,345	9,348	7,815	6,251	4,108	2,366	1,213	0,584	0,352	0,216	0,115	0,072
4	14,860	13,277	11,143	9,488	7,779	5,385	3,357	1,923	1,064	0,711	0,484	0,297	0,207
5	16,750	15,086	12,833	11,071	9,236	6,626	4,351	2,675	1,610	1,145	0,831	0,554	0,412
6	18,548	16,812	14,449	12,592	10,645	7,841	5,348	3,455	2,204	1,635	1,237	0,872	0,676
7	20,278	18,475	16,013	14,067	12,017	9,037	6,346	4,255	2,833	2,167	1,690	1,239	0,989
8	21,955	20,090	17,535	15,507	13,362	10,219	7,344	5,071	3,490	2,733	2,180	1,647	1,344
9	23,589	21,666	19,023	16,919	14,684	11,389	8,343	5,899	4,168	3,325	2,700	2,088	1,735
10	25,188	23,209	20,483	18,307	15,987	12,549	9,342	6,737	4,865	3,940	3,247	2,558	2,156
11	26,757	24,725	21,920	19,675	17,275	13,701	10,341	7,584	5,578	4,575	3,816	3,053	2,603
12	28,300	26,217	23,337	21,026	18,549	14,845	11,340	8,438	6,304	5,226	4,404	3,571	3,074
13	29,819	27,688	24,736	22,362	19,812	15,984	12,340	9,299	7,042	5,892	5,009	4,107	3,565
14	31,319	29,141	26,119	23,685	21,064	17,117	13,339	10,165	7,790	6,571	5,629	4,660	4,075
15	32,801	30,578	27,488	24,996	22,307	18,245	14,339	11,037	8,547	7,261	6,262	5,229	4,601
16	34,267	32,000	28,845	26,296	23,542	19,369	15,339	11,912	9,312	7,962	6,908	5,812	5,142
17	35,718	33,409	30,191	27,587	24,769	20,489	16,338	12,792	10,085	8,672	7,564	6,408	5,697
18	37,156	34,805	31,526	28,869	25,989	21,605	17,338	13,675	10,865	9,390	8,231	7,015	6,265
19	38,582	36,191	32,852	30,144	27,204	22,718	18,338	14,562	11,651	10,117	8,907	7,633	6,844
20	39,997	37,566	34,170	31,410	28,412	23,828	19,337	15,452	12,443	10,851	9,591	8,260	7,434

Окончание табл. 4

k	Уровень значимости α												
	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95	0,975	0,99	0,995
21	41,401	38,932	35,479	32,671	29,615	24,935	20,337	16,344	13,240	11,591	10,283	8,897	8,034
22	42,796	40,289	36,781	33,924	30,813	26,039	21,337	17,240	14,041	12,338	10,982	9,542	8,643
23	44,181	41,638	38,076	35,172	32,007	27,141	22,337	18,137	14,848	13,091	11,689	10,196	9,260
24	45,559	42,980	39,364	36,415	33,196	28,241	23,337	19,037	15,659	13,848	12,401	10,856	9,886
25	46,928	44,314	40,646	37,652	34,382	29,339	24,337	19,939	16,473	14,611	13,120	11,524	10,520
26	48,290	45,642	41,923	38,885	35,563	30,435	25,336	20,843	17,292	15,379	13,844	12,198	11,160
27	49,645	46,963	43,195	40,113	36,741	31,528	26,336	21,749	18,114	16,151	14,573	12,879	11,808
28	50,993	48,278	44,461	41,337	37,916	32,620	27,336	22,657	18,939	16,928	15,308	13,565	12,461
29	52,336	49,588	45,722	42,557	39,087	33,711	28,336	23,567	19,768	17,708	16,047	14,256	13,121
30	53,672	50,892	46,979	43,773	40,256	34,800	29,336	24,478	20,599	18,493	16,791	14,953	13,787

Таблица 5

Критические точки распределения Стьюдента

k	Уровень значимости α (двусторонняя критическая область)					
	0,1	0,05	0,02	0,01	0,002	0,001
1	6,3138	12,7062	31,8205	63,6567	318,3088	636,6192
2	2,9200	4,3027	6,9646	9,9248	22,3271	31,5991
3	2,3534	3,1824	4,5407	5,8409	10,2145	12,9240
4	2,1318	2,7764	3,7469	4,6041	7,1732	8,6103
5	2,0150	2,5706	3,3649	4,0321	5,8934	6,8688
6	1,9432	2,4469	3,1427	3,7074	5,2076	5,9588
7	1,8946	2,3646	2,9980	3,4995	4,7853	5,4079
8	1,8595	2,3060	2,8965	3,3554	4,5008	5,0413
9	1,8331	2,2622	2,8214	3,2498	4,2968	4,7809
10	1,8125	2,2281	2,7638	3,1693	4,1437	4,5869
11	1,7959	2,2010	2,7181	3,1058	4,0247	4,4370
12	1,7823	2,1788	2,6810	3,0545	3,9296	4,3178
13	1,7709	2,1604	2,6503	3,0123	3,8520	4,2208
14	1,7613	2,1448	2,6245	2,9768	3,7874	4,1405
15	1,7531	2,1314	2,6025	2,9467	3,7328	4,0728
16	1,7459	2,1199	2,5835	2,9208	3,6862	4,0150
17	1,7396	2,1098	2,5669	2,8982	3,6458	3,9651
18	1,7341	2,1009	2,5524	2,8784	3,6105	3,9216
19	1,7291	2,0930	2,5395	2,8609	3,5794	3,8834
20	1,7247	2,0860	2,5280	2,8453	3,5518	3,8495
21	1,7207	2,0796	2,5176	2,8314	3,5272	3,8193
22	1,7171	2,0739	2,5083	2,8188	3,5050	3,7921
23	1,7139	2,0687	2,4999	2,8073	3,4850	3,7676
24	1,7109	2,0639	2,4922	2,7969	3,4668	3,7454
25	1,7081	2,0595	2,4851	2,7874	3,4502	3,7251
26	1,7056	2,0555	2,4786	2,7787	3,4350	3,7066
27	1,7033	2,0518	2,4727	2,7707	3,4210	3,6896
28	1,7011	2,0484	2,4671	2,7633	3,4082	3,6739
29	1,6991	2,0452	2,4620	2,7564	3,3962	3,6594
30	1,6973	2,0423	2,4573	2,7500	3,3852	3,6460
40	1,6839	2,0211	2,4233	2,7045	3,3069	3,5510
50	1,6759	2,0086	2,4033	2,6778	3,2614	3,4960
60	1,6706	2,0003	2,3901	2,6603	3,2317	3,4602

k	Уровень значимости α (двусторонняя критическая область)					
	0,1	0,05	0,02	0,01	0,002	0,001
70	1,6669	1,9944	2,3808	2,6479	3,2108	3,4350
80	1,6641	1,9901	2,3739	2,6387	3,1953	3,4163
90	1,6620	1,9867	2,3685	2,6316	3,1833	3,4019
100	1,6602	1,9840	2,3642	2,6259	3,1737	3,3905
110	1,6588	1,9818	2,3607	2,6213	3,1660	3,3812
120	1,6577	1,9799	2,3578	2,6174	3,1595	3,3735
200	1,6525	1,9719	2,3451	2,6006	3,1315	3,3398
∞	1,64485	1,95996	2,32634	2,57582	3,0902	3,2905
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Таблица 6

Квантили распределения Фишера 95 %-ного уровня

k_2	k_1									
	1	2	3	4	5	6	7	8	9	10
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9
2	15,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,05
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,90	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38

k_2	k_1									
	1	2	3	4	5	6	7	8	9	10
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

k_2	k_1								
	12	15	20	24	30	40	60	120	∞
1	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,91	2,85	2,77	2,74	2,70	2,66	3,20	2,58	2,54
11	2,79	2,72,	2,65	2,61	2,57	2,53	3,00	2,45	2,40
12	2,69	2,62	2,54	2,51	2,47	2,43	2,85	2,34	2,30
13	2,60	2,53	2,46	2,42	2,38	2,34	2,72	2,25	2,21
14	2,53	2,46	2,39	2,35	2,31	2,27	2,61	2,18	2,13
15	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01

Окончание табл. 6

k_2	k_1								
	12	15	20	24	30	40	60	120	∞
17	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	2,23	2,15	2,07	2,03	1,98,	1,94	1,89	1,84	1,78
23	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	2,12	2,04	1,96	1,84	1,87	1,82	1,77	1,71	1,65
29	2,10	2,03	1,94	1,81	1,85	1,81	1,75	1,70	1,64
30	2,09	2,01	1,93	1,78	1,84	1,79	1,74	1,68	1,62
40	2,00	1,92	1,84	1,76	1,74	1,69	1,64	1,58	1,51
60	1,92	1,84	1,75	1,88	1,65	1,59	1,53	1,47	1,39
120	1,83	1,75	1,66	1,76	1,55	1,50	1,43	1,35	1,25
∞	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,	

Учебное издание

Шевалдина Ольга Яковлевна
Выходец Евгения Владимировна
Кузнецова Ольга Леонидовна
Трофимова Елена Александровна
Гилёв Денис Викторович
Кисляк Надежда Валерьевна

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА: РЕШЕНИЕ ЗАДАЧ

Учебное пособие

Заведующий редакцией *М. А. Овечкина*
Редактор *С. Г. Галинова*
Корректор *С. Г. Галинова*
Компьютерная верстка *В. К. Матвеев*

Подписано в печать 23.04.2021 г. Формат 70 × 100 1/16.
Бумага офсетная. Цифровая печать. Усл. печ. л. 17,74.
Уч.-изд. л. 13,50. Тираж 100 экз. Заказ 68.

Издательство Уральского университета
Редакционно-издательский отдел ИПЦ УрФУ
620083, Екатеринбург, ул. Тургенева, 4
Тел.: +7 (343) 389-94-79, 350-43-28
E-mail: rio.marina.ovechkina@mail.ru

Отпечатано в Издательско-полиграфическом центре УрФУ
620083, Екатеринбург, ул. Тургенева, 4
Тел.: +7 (343) 358-93-06, 350-58-20, 350-90-13
Факс: +7 (343) 358-93-06
http://print.urfu.ru

ЛИДЕР УНИВЕРСИТЕТСКОГО КНИГОИЗДАНИЯ

УНИВЕРСИТЕТСКАЯ КНИГА

2020

9785799631895