И. Ю. АЛИБЕКОВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА В СРЕДЕ MATLAB

УЧЕБНОЕ ПОСОБИЕ

ИЗДАНИЕ ВТОРОЕ, СТЕРЕОТИПНОЕ

ЛАНЬ

САНКТ-ПЕТЕРБУРГ • МОСКВА • КРАСНОДАР
2021

УДК 519.2
ББК 22.171я73

А 50 **Алибеков И. Ю.** Теория вероятностей и математическая статистика в среде MATLAB : учебное пособие для вузов / И. Ю. Алибеков. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 184 с. — Текст : непосредственный.

ISBN 978-5-8114-6865-2

Данное пособие содержит основные сведения из теории вероятностей, математической статистики. Приводятся примеры решения задач по указанной тематике. Задачи решались традиционным способом и с помощью системы MATLAB. Даются рекомендации по использованию пакета символьной математики MuPAD и примеры его применения.

Автор постарался не перегружать книгу второстепенными деталями, чтобы помочь читателю быстро разобраться в материале и приступить к решению задач.

Пособие предназначено для студентов инженерных и экономических специальностей различных форм обучения.

УДК 519.2
ББК 22.171я73

Рецензенты:
А. И. ГУСЕВА — доктор технических наук,
профессор кафедры экономики и менеджмента
в промышленности Национального исследовательского
ядерного университета «МИФИ»;
Л. В. ЛАБУНЕЦ — доктор технических наук,
профессор кафедры автономных информационных
и управляющих систем Московского государственного
технического университета им. Н. Э. Баумана.

Обложка
Е. А. ВЛАСОВА

ОГЛАВЛЕНИЕ

Вероятностное знание —
вот предел человеческого разумения.
Цицерон

ВВЕДЕНИЕ

Методы теории вероятностей, математической статистики и эконометрики характеризуется рядом особенностей, создающих ряд трудностей в процессе преподавания.

Во-первых, это большой объем материала с большим числом дидактических единиц.

Во-вторых, данные предметы изучают вчерашние школьники, которые еще слабо подготовлены к восприятию большого массива абстрактно-логических понятий.

Особенно трудно дается освоение этих курсов студентами заочной формы обучения, которые физически не могут освоить сотни страниц учебника за краткий период сессии.

Все это диктует необходимость издания учебных пособий кратких по объему и с большим количеством иллюстраций, пособий, в которых бы в полной мере реализовались такие принципы, как наглядность и доступность.

Вместе с тем весьма актуальной остается задача совмещения изучаемых предметов с современными системами компьютерной математики, такими как MATLAB, Mathcad, Mathematica и др.

Компьютерная система MATLAB, в частности, представляет собой инструмент, ориентированный на широкий круг пользователей, которые не являются ни профессионалами в математике, ни программистами, воспитанными на узкоспециальных, малопонятных большинству конечных пользователей компьютерных языках. Она сочетает в себе простоту реализации численных методов математики с символьными преобразованиями (дробно-рациональные преобразования, подстановки, упрощения, решение уравнений, интегрирование, дифференцирование и т. п.), когда результат получается не в числовом виде, а в виде формулы.

Нельзя не отметить и графические возможности системы MATLAB, которая позволяет без труда представить в наглядной форме результаты расчетов и, при необходимости, получить анимацию графических изображений в пределах заданного изменения параметров.

К настоящему времени выпущено много руководств и справочников по MATLAB [5–9], содержащих подробное описание его обширной системы команд (их общее число давно перевалило за тысячу), расширений и приложений. Большинство из этих руководств неудобны для начинающего пользователя из-за обилия представленного материала.

Существующие учебные материалы, где рассматривается применение компьютерной математики для решения задач по указанной выше тематике, либо носят фрагментарный характер и не охватывают всех задач базовых курсов, либо это объемные учебники, перегруженные несущественными деталями в части программироания [3].

Данное учебное пособие задумано как недостающее связующее звено между первым и вторым типами учебников. Оно посвящено применению системы MATLAB при решении задач теории вероятностей, математической статистикии эконометрики. Авторы стремились найти наиболее простую программную реализацию решаемых задач. Решение задач компьютерным способом дублировалось традиционными способами решения, и проводилось сравнение результатов. Такое изложение материала позволило наглядно продемонстрировать преимущества и возможности, которые получае пользователь, применяя систему MATLAB при решении поставленных перед ним задач.

В данной работе подробно рассмотрен пакет MuPAD системы MATLAB, имеющий собственное ядро символьной математики, и его применение для аналитических вычислений и графического отображения информации. На примере решения конкретных практических задач показаны преимущества пакета MuPAD по сравнению с традиционно используемым пакетом символьных вычислений Symbolic Math системы MATLAB. В этом данное учебное пособие не имеет аналогов, за исключением работы [1].

Конечно, освоение системы MATLAB потребует труда, который мы постарались облегчить написанием данного пособия. Но этот труд можно рассматривать как важную инвестицию, которая многократно окупится в дальнейшем.

Для лучшего усвоения материала и в конце разделов приводятся задания по самостоятельной работе и контрольные вопросы.

Авторы выражают признательность за критические замечания, которые будут высказаны в адрес учебного пособия и выражают надежду, что данное пособие окажется полезной аспирантам и студентам при изучении курсов «теория вероятности и математическая статистика» и «эконометрика».

1.
ТЕОРИЯ ВЕРОЯТНОСТЕЙ В СРЕДЕ MATLAB

1.1.
КЛАССИЧЕСКОЕ ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ

1.1.1.
Случайные события

Определение 1.1.1. Случайные события — любые события или факты, относящиеся к результату эксперимента, которые могут происходить или не происходить.

Пример 1.1.1. Эксперимент — подбрасывание монеты. Случайное событие — выпадение «решки» или «орла». Исходы испытания — монета упала «орлом» или «решкой».

Отдельные случайные события в теории вероятностей обозначают прописными латинскими буквами, например, A, B, C и т. д.

Не все случайные явления можно изучать методами теории вероятностей, а лишь те, которые могут быть воспроизведены в одних и тех же условиях.

Основной числовой характеристикой случайного события является его вероятность.

Пусть производится серия из n испытаний, в каждом из которых может появиться или не появиться событие A. Если в результате испытания наблюдалось (появилось) событие A, то такой исход испытания называется благоприятным исходом.

Определение 1.1.2. Элементарное событие — событие или каждый отдельный возможный результат испытания.

Определение 1.1.3. Набор элементарных событий — набор всех возможных отдельных результатов испытаний.

Пример 1.1.2. Правильная игральная кость при бросании с равными шансами падает на любую из граней 1–6. В случае бросания двух костей сумма выпавших чисел заключена между 2 и 12. Как 9, так и 10 из чисел 1, 2, ..., 6 можно получить двумя разными способами: $9 = 3 + 6$ или $9 = 4 + 5$ и $10 = 4 + 6$ или $10 = 5 + 5$. Почему

9 появляется чаще, когда бросают две кости, а 10 — когда бросают три?

Р е ш е н и е.

В случае двух костей 9 и 10 могут получиться в результате следующих исходов: $9 = 3 + 6$ или $9 = 6 + 3$, или $9 = 4 + 5$, или $9 = 5 + 4$, $10 = 4 + 6$, или $10 = 6 + 4$, или $10 = 5 + 5$. Это значит, что при двух костях 9 можно «выбросить» четырьмя способами, а 10 — лишь тремя. Следовательно, здесь шансы получить 9 предпочтительней.

В случае трех костей ситуация меняется на противоположную: 9 можно «выбросить» 25 способами, а 10 — уже 26 способами. Потому 10 получается чаще, чем 9.

1.1.2.
Классификация событий. Вероятность события

Различные события различают по степени возможности их проявления и бывают взаимно связаны.

Типы событий: случайное, достоверное, невозможное.

Определение 1.1.3. Достоверное событие — событие, которое в результате опыта обязательно должно произойти.

Например, выпадение не менее одного очка при бросании игральной кости.

Определение 1.1.4. Невозможное событие — событие, которое не может иметь место в данном опыте.

Например, выпадение более 6 очков при бросании шестигранной игральной кости.

Если событие в данном опыте невозможно, то говорят, что вероятность его равна нулю. Математически это записывается так: P, $A = 0$. Если событие достоверно, то его вероятность равна единице: $P(A) = 1P(A) = 1$. Чем ближе вероятность события к 1, тем больше объективная возможность его появления его в опыте.

Определение 1.1.5. Два или несколько событий называются равновозможными, если нет оснований утверждать, что одно из них имеет больше данных появиться в итоге опыта по сравнению с другими, например выпадение герба и цифры при однократном бросании монеты.

Равная возможность исходов опыта — основная предпосылка классического определения вероятностей.

По характеру совместной связи события подразделяются на совместные и несовместные.

Определение 1.1.6. События называются несовместными, если появление какого-нибудь одного из них в данном опыте исключает возможность появления других событий.

Например, выпадение 3 и 5 вместе при однократном бросании игральной кости.

Определение 1.1.7. События называются совместными, если появление одного из них в данном опыте не исключает возможность появления других.

Например, выпадение 3 и 5 вместе при двукратном бросании игральной кости.

Определение 1.1.8. Полная группа событий — группа событий, из которых хотя бы одно непременно должно произойти в данном опыте.

Например, попадание или непопадание в мишень при одном выстреле.

Определение 1.1.9. Противоположные события — два единственно возможных и несовместных события AA и BB, для которых справедливо, что событие AA наступает, когда не наступает событие BB, и наоборот.

Противоположные события — частный случай событий, образующих полную группу.

Определение 1.1.10. События, вероятности которых малы или очень велики, называются практически невозможными или практически достоверными.

Определение 1.1.11. Вероятность события — численная мера, принимающая значения между 0 и 1 и характеризующая степень возможности появления события в данном опыте.

Обозначается $P(A)$, где AA — случайное событие.

1.1.3.
Классическое определение вероятности

Классическое определение вероятности дал еще Лаплас, но тогда ее приложение не выходило за сферу азартных игр.

Пьер-Симон Лаплас (1749–1827) — французский математик; один из создателей теории вероятностей.

Определение 1.1.12. (классическое определение вероятности события). Вероятность случайного события A равна числу элементарных событий (исходов опыта или случаев), благоприятствующих появлению события A, деленному на число всех возможных элементарных событий.

$$P(A) = \frac{m}{n}, \quad P(A) = \frac{m}{n}, \tag{1.1.1}$$

где m — число исходов опыта, благоприятствующих появлению события A; n — число всех возможных исходов опыта.

Формула (1.1.1) справедлива, когда мы имеем дело с так назаваемой схемой случаев, т. е. с полной группой несовместных равновозможных исходов опыта (случаев), и число этих случаев конечно.

Пример 1.1.3. Имеется партия из 30 изделий, среди которых 4 бракованных. Наугад выбираются 4 изделия. Какова вероятность, что все они окажутся бракованными.

Р е ш е н и е. Так как изделия выбираются наугад, у нас нет оснований считать, что один вариант отбора окажется предпочтительней другого. Поэтому все варианты отбора равновозможны. Кроме того, очевидно, что число вариантов отбора конечно. Таким образом, данная задача сводится к схеме случаев, и, следовательно, можно применить классическое определение вероятности (1.1.1). Извлечение 4-х бракованных деталей возможно только в одном случае, т. е. $m = 1$, а общее число возможных случаев равно $n = C_{30}^4 = \dfrac{30!}{4!(30-4)!}$. Вероятность искомого события: $P = \dfrac{m}{n} = \dfrac{1}{C_{30}^4}$.

Для расчета числа сочетаний C_{30}^4 в системе MATLAB можно применить функцию *nchoosec*(k, r), где k и r — неотрицательные целые числа. Команда возвращает вычисленное значение $C_k^r = \dfrac{k!}{r!(k-r)!}$.

1.1.4. Контрольные вопросы

1. Что изучает теория вероятностей?

2. Что понимается под термином «элементарное событие»?

3. Что является основной числовой характеристикой случайного события?

4. Дайте определения случайного, достоверного и невозможного событий.

5. При каких ограничениях действует классическое определение вероятности?

6. Как подразделяются события по характеру совместной связи?

7. Приведите примеры полной группы событий.

1.2. ГЕОМЕТРИЧЕСКОЕ И СТАТИСТИЧЕСКОЕ ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ. УСЛОВНАЯ ВЕРОЯТНОСТЬ СОБЫТИЯ

1.2.1. Геометрическое определение вероятности

Геометрическое определение вероятности является обобщением классического определения вероятности на эксперимент с бесконечным числом равновозможных случайных исходов, изображаемых точками, прямой, плоскостью, пространством и т. д.

Пусть плоская фигура g составляет часть плоской фигуры G. На фигуру G наудачу бросается точка. Это означает, что все точки области G «равноправны» в отношении попадания туда брошенной случайной точки. Фигуру g называют благоприятствующей событию A.

Определение 1.2.1. Геометрическая вероятность события A — отношение меры области, благоприятствующей появлению события A, к мере всей области.

П р и м е р 1.2.1. В квадрат со стороной 4 см «бросают» точку. Какова вероятность, что расстояние от этой точки до ближайшей стороны квадрата будет меньше 1 см?

Р е ш е н и е. Закрасим в квадрате множество точек, удаленных от ближайшей стороны меньше, чем на 1 см (рис. 1.2.1).

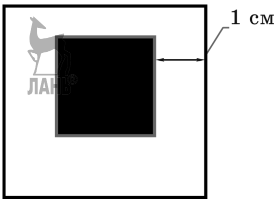

Рис. 1.2.1

Попадание случайной точки в заданную область

Площадь незакрашенной части квадрата $S(A) = 16$ см^2 – 4 см^2 = = 12 см^2. Площадь большого квадрата $S(\Omega) = 16$ см^2. Тогда, в соответствии с определением геометрической вероятности, получим $P(A) = \dfrac{S(A)}{S(\Omega)} = \dfrac{12}{16} = \dfrac{3}{4}$.

1.2.2.
Статистическая вероятность. Закон больших чисел

Результаты испытаний не всегда можно представить в виде совокупности элементарных событий. Введем понятие статистической вероятности. Если производить многократно повторение одного и того же опыта, то, согласно закону больших чисел, относительное число появлений данного события во всей серии опытов, или относительная частота его появления, будет близко к значению его вероятности. Оказывается, при большом числе испытаний n, относительная частота появления события A в различных сериях опытов отличается друг от друга мало и это отличие тем меньше, чем больше испытаний в сериях.

Определение 1.2.2. Абсолютной частотой случайного события A в серии из N случайных опытов называется число N_A, которое показывает, сколько раз произошло событие A в этой серии.

Определение 1.2.3. Относительной частотой случайного события называют отношение числа появлений этого события к общему числу проведенных экспериментов.

Например, при небольшом количестве опытов с подбрасыванием монеты относительная частота появлений герба может существенно отличаться от объективной вероятности, равной 0,5. Если же увеличивать число опытов, то относительная частота появлений герба будет все ближе к теоретической вероятности его появления. Такие опыты проводились Бюффоном (Франция) и Пирсоном (Англия), при этом были получены следующие результаты.

Число бросаний	**Относительная частота появления герба**
4040	0,50693
12 000	0,5016
24 000	0,5005

Полученные результаты являются одним из проявлений закона больших чисел, строгое доказательство которого приводится, например, в работе [1].

Определение 1.2.4. Статистической вероятностью события называют относительную частоту события при большом числе испытаний или число близкое к ней.

Пример 1.2.2. Наблюдения показывают, что в среднем среди 1000 новорожденных детей 515 мальчиков. Какова относительная частота рождения мальчика в такой серии наблюдений?

Р е ш е н и е. $W(A) = \dfrac{515}{1000} = 0{,}515.$

Согласно закону больших чисел, относительная частота $W(A)$ появления некоторого события A обладает определенной устойчивостью. То есть ее значения, изменяясь, колеблются около объективной вероятности $P(A)$ события, к которой она стремится при неограниченном возрастании числа испытаний n:

$$\lim_{n\to\infty} W(A) = P(A).$$

1.2.3. Условная вероятность события

Рассмотрим два последовательных случайных события. Ставится вопрос: какова вероятность наступления второго события, если первое событие уже произошло?

П р и м е р 1.2.3. Пусть в урне было 5 шаров (2 белых и 3 черных). Последовательно, один за другим, извлекаются два шара (без возврата в урну предыдущего шара). Найти вероятность извлечь белый шар во втором испытании.

Р е ш е н и е. После извлечения первого шара в ней останется 4 шара и один белый в их числе (если извлекли первым белый шар), или 2 белых (если в первый раз извлечен черный шар).

Задача определения вероятности появления определенного шара сводится к схеме случаев, поэтому можно применить классическое определение вероятности. В первом случае вероятность извлечь вторым белый шар будет 1/4, во втором 1/2. Таким образом, вероятность извлечь белый шар во втором испытании зависит от результата первого испытания.

Понятия условной вероятности и независимости введены А. Муавром в 1718 г.

Определение 1.2.5. Условная вероятность — это вероятность одного события, вычисленная в предположении, что другое событие произошло.

Вероятность события A_1 в предположении, что произошло событие A_2, обозначается как $P(A_1/A_2)$.

Определение 1.2.6. Два или несколько событий называются независимыми, если вероятность появления каждого из них не зависит от того, имели ли место другие события.

Определение 1.2.7. Два или несколько событий называются зависимыми, если появление одного из них влияет на вероятность наступления других.

Из этих определений следует математическая запись условия независимости двух событий:

$P(A_1/A_2) = P(A_1)$ — событие A_1 не зависит от события A_2.

$P(A_2/A_1) = P(A_2)$ — событие A_2 не зависит от события A_1.

Если выполняются оба эти условия, то события A_1 и A_2 называются взаимно-независимыми событиями.

Пример 1.2.4. Событие A — извлечение из колоды туза, событие B — вторая вынутая из колоды карта — туз. Тогда если после первого раза карта была возвращена в колоду, то вероятность вынуть вторично туз не меняется: $P(B) = P(A) = \frac{4}{36} = \frac{1}{9}$.

Если же первая карта в колоду не возвращается, то осуществление события A приводит к тому, что в колоде осталась 31 карта, в числе которых имеются только 3 туза. Поэтому $P(B/A) = \frac{3}{35}$.

1.2.4. Контрольные вопросы

1. Дайте определения геометрической и статистической вероятностей.
2. В чем отличие геометрического определения от классического определения вероятности?
3. В чем разница между абсолютной и относительной частотами?

1.2.5. Задачи для самостоятельно решения

1. Брошена игральная кость. Найти вероятность того, что выйдет четное число очков.

2. В ящике имеется 50 одинаковых деталей, из них 5 окрашенных. Наудачу вынимают одну деталь. Найти вероятность того, что извлеченная деталь окажется окрашенной.

3. Участники жеребьевки тянут из ящика жетоны с номерами от 1 до 100. Найти вероятность того, что номер первого, наудачу извлеченного жетона, не содержит цифры 5.

4. Куб, все грани которого окрашены, распилен на тысячу кубиков одинакового размера, которые затем тщательно перемешаны. Найти вероятность того, что наудачу извлеченный кубик будет иметь окрашенных граней: а) одну; б) две; в) три. (Ответ: а) 0,384; б) 0,096; в) 0,008.)

5. В замке на общей оси пять дисков, каждый из которых разделен на шесть секторов с различными написанными на них буквами. Замок открывается только в том случае, если каждый диск занимает одно определенное положение относительно корпуса замка. Найти вероятность того, что при произвольной установке дисков замок можно будет открыть.

6. В партии из 100 деталей отдел технического контроля обнаружил 6 нестандартных деталей. Чему равна относительная частота появления нестандартных деталей?

7. При стрельбе из винтовки относительная частота попадания в цель оказалась равной 0,85. Требуется найти число попаданий, если всего было произведено 120 выстрелов.

1.3.
АЛГЕБРА СОБЫТИЙ

Алгебра событий позволяет представить некоторое сложное событие в виде композиции более простых событий, вероятности которых либо известны, либо их можно определить по классической схеме. Композиция простых событий может осуществляться с помощью операций произведения, сложения и отрицания (инверсии) события.

1.3.1.
Произведение событий

Определение 1.3.1. Произведением нескольких событий называется событие, состоящие в совместном наступлении всех этих событий в результате испытания. Обозначают: AB, $A \cdot B$, $A \cap B$, A и B.

Замечание. Произведение означает связку «и» (ABC, это означает, что наступило и событие A, и событие B, и событие C).

П р и м е р 1.3.1. Событие A — «из колоды карт вынута дама», событие B — «из колоды карт вынута карта пиковой масти». Тогда событие $A \cdot B$ означает «вынута дама пиковой масти».

П р и м е р 1.3.2. Событие A — «число выпавших очков меньше 5», событие B — «число выпавших очков больше 2», C — «число выпавших очков четное». Тогда событие ABC означает, что «выпало 4 очка».

Теорема 1.3.1. Вероятность произведения взаимно-независимых событий равна произведению их вероятностей:

$$P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2). \tag{1.3.1}$$

Теорема 1.3.2. Вероятность произведения зависимых событий равна произведению вероятности одного события на условную вероятность другого события, вычисленную в предположении, что первое случайное событие уже произошло:

$$P(A_1 A_2) = P(A_1) \cdot P(A_2/A_1). \tag{1.3.2}$$

1.3.2.
Операции над событиями

Определение 1.3.2. Суммой двух событий A_1 и A_2 называется событие, состоящее в появлении хотя бы одного из этих событий. Обозначения суммы событий: $A_1 + A_2$, $A_1 \bigcup A_2$, A_1 или A_2.

Знаком плюс обозначается связка «или».

Теорема 1.3.3. Вероятность суммы двух несовместных событий равняется сумме их вероятностей:

$$P(A_1 + A_2) = P(A_1) + P(A_2). \tag{1.3.3}$$

Сформулированная теорема справедлива для любого числа несовместных событий.

Следствие теоремы сложения. Сумма вероятностей полной группы несовместных событий равна единице.

Теорема 1.3.4. Вероятность суммы двух совместных событий равняется сумме их вероятностей, уменьшенной на вероятность произведения этих событий:

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2). \tag{1.3.4}$$

Перечислим свойства операций сложения и умножения событий.

1. $A + B = B + A$ — коммутативность сложения событий.
2. $A + (B + C) = (A + B) + C$ — ассоциативность сложения.

3. $A \cdot B = B \cdot A$ — коммутативность умножения событий.

4. $A \cdot (B \cdot C) = (A \cdot B) \cdot C$ — ассоциативность умножения.

5. $A \cdot (B + C) = A \cdot B + A \cdot C$ — закон дистрибутивности.

Определение 1.3.3. Отрицанием некоторого события A называется событие, состоящее в том, что событие A не произойдет. Такое событие называется противоположным событию A и обозначается как $\overline{A}$.

События A и $\overline{A}$ несовместны и образуют полную группу событий, т. е.

$$P(A) + P(\overline{A}) = 1. \qquad (1.3.5)$$

1.3.3. Вероятность появления хотя бы одного события

Теорема 1.3.5. Вероятность появления хотя бы одного из событий $A_1, A_2, ..., A_n$, независимых в совокупности, равна разности между единицей и произведением вероятностей противоположных событий:

$$P\ (A_1 + A_2 + \ldots + A_n) = 1 - P\ (\overline{A}_1) \cdot P\ (\overline{A}_2) \cdot \ldots \cdot P(\overline{A}_n). \qquad (1.3.6)$$

Следствие. При производимых n одинаковых независимых испытаниях, в каждом из которых событие A появляется с одной и той же вероятностью, равной p, вероятность, что событие A появится хотя бы один раз, равна

$$P(\text{хотя бы один раз появится событие } A) = 1 - (1 - p)^n. \qquad (1.3.7)$$

Рассмотрим примеры вычисления вероятностей сложных событий, которые можно разложить на простые события с помощью алгебры событий. Вероятности простых событий, в свою очередь, вычисляются по классической формуле определения вероятностей.

Пример 1.3.3. В урне 30 шаров: 10 красных, 5 синих и 15 белых. Наугад извлекается один шар. Найти вероятность появления цветного шара.

Р е ш е н и е. Появление цветного шара означает появление либо красного, либо синего шара.

Вероятность появления красного шара (событие A): $P(A) = \frac{10}{30} = \frac{1}{3}.$

Вероятность появления синего шара (событие B): $P(B) = \frac{5}{30} = \frac{1}{6}.$

События A и B в данном опыте несовместны (появление шара одного цвета исключает появление шара другого цвета), поэтому искомую вероятность определяем по формуле (1.3.3):

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

П р и м е р 1.3.4. На стеллаже в библиотеке стоит 15 учебников, причем 5 из них в переплете. Библиотекарь случайным образом выбирает три учебника. Найти вероятность того, что хотя бы один из взятых учебников окажется в переплете.

Р е ш е н и е. Пусть события A — хотя бы один учебник в переплете. Событие $\overline{A}$ — ни один из взятых учебников не имеет переплета. Так как события $\overline{A}$ и A противоположные, то они образуют полную группу событий и по формуле (1.3.5) находим

$$P(A)=1-P(\overline{A})=1-\frac{C_{10}^3}{C_{15}^3}=1-\frac{24}{91}=\frac{67}{91}.$$

П р и м е р 1.3.5. Студент разыскивает нужную ему формулу в трех справочниках. Вероятности того, что формула содержится в первом, втором и третьем справочниках равны соответственно 0,6; 0,7 и 0,8. Найти вероятности того, что формула содержится:

1) только в одном справочнике;

2) только в двух справочниках.

Р е ш е н и е.

A — формула содержится в первом справочнике;

B — формула содержится во втором справочнике;

C — формула содержится в третьем справочнике.

Воспользуемся теоремами сложения и умножения вероятностей. Вероятность того, что формула содержится только в одном справочнике из трех:

$$P(A\overline{B}\,\overline{C}+\overline{A}\,B\,\overline{C}+\overline{A}\,\overline{B}C)=$$
$$=0{,}6\cdot 0{,}3\cdot 0{,}2+0{,}4\cdot 0{,}7\cdot 0{,}2+0{,}4\cdot 0{,}3\cdot 0{,}8=0{,}188.$$

Вероятность того, что формула содержится только в двух справочниках:

$$P(AB\overline{C}+A\overline{B}C+\overline{A}BC)=$$
$$=0{,}6\cdot 0{,}7\cdot 0{,}2+0{,}6\cdot 0{,}3\cdot 0{,}8+0{,}4\cdot 0{,}7\cdot 0{,}8=0{,}452.$$

1.3.4.
Принцип практической невозможности события

При решении многих практических задач приходится иметь дело с событиями, вероятность которых весьма мала, т. е. близка к нулю. Можно ли считать, что маловероятное событие A в единичном испытании не произойдет? Такого заключения сделать нельзя, так как не исключено, хотя и маловероятно, что событие A наступит. Казалось бы, появление или непоявление маловероятного события в единичном испытании предсказать невозможно. Однако длительный опыт показывает, что маловероятное событие в единичном испытании в подавляющем большинстве случаев не наступает.

На основании этого факта принимают следующий «принцип практической невозможности маловероятных событий»: если случайное событие имеет очень малую вероятность, то практически можно считать, что в единичном испытании это событие не наступит.

Естественно возникает вопрос: насколько малой должна быть вероятность события, чтобы можно было считать невозможным его появление в одном испытании? На этот вопрос нельзя ответить однозначно. Для задач, различных по существу, ответы будут разными. Например, если вероятность того, что парашют при прыжке не раскроется, равна 0,01, то было бы недопустимым применять такие парашюты, по крайней мере для масового применения. Если же вероятность того, что поезд дальнего следования прибудет с опозданием, равна 0,01, то можно практически быть уверенным, что поезд прибудет вовремя.

Определение 1.3.4. Достаточно малую вероятность, при которой событие можно считать практически невозможным, называют уровнем значимости.

На практике обычно принимают уровни значимости, заключенные между 0,01 и 0,05.

Подчеркнем, что рассмотренный здесь принцип позволяет делать предсказания не только о событиях, имеющих малую вероятность, но и о событиях, вероятность которых близка к единице. Действительно, если событие A имеет вероятность близкую к нулю, то вероятность противоположного события близка к единице. С другой стороны, непоявление события A означает наступление противоположного события $\overline{A}$.

Таким образом, из принципа невозможности маловероятных событий вытекает следующее важное для приложений следствие: если случайное событие имеет вероятность очень близкую к единице, то практически можно считать, что в единичном испытании это событие наступит.

Разумеется, и здесь ответ на вопрос о том, какую вероятность считать близкой к единице, зависит от существа задачи.

1.3.5.
Контрольные вопросы

1. Назовите действия над событиями.
2. Виды случайных событий.
3. Дайте статистическое определение вероятности.
4. Чему равна вероятность суммы двух несовместных событий?
5. Дайте определение произведения двух событий.
6. Как определяется вероятность появления хотя бы одного события?
7. Как определяется условная вероятность?
8. Приведите формулу для вычисления вероятностей совместных событий.

1.3.6.
Задачи для самостоятельного решения

1. В партии из 10 деталей 8 стандартных. Найти вероятность того, что среди наудачу извлеченных 2-х деталей есть хотя бы одна стандартная.

2. События A, B, C и D образуют полную группу. Вероятности событий таковы: $P(A) = 0{,}1$, $P(B) = 0{,}4$, $P(C) = 0{,}3$. Чему равна вероятность события D?

3. По статистическим данным ремонтной мастерской, в среднем, на 20 остановок токарного станка приходится: для смены резца — 10, из-за неисправности привода — 3, из-за несвоевременной подачи заготовок — 2. Остальные остановки происходят по другим причинам. Найти вероятность остановки станка по другим причинам.

4. Вероятность того, что стрелок при одном выстреле попадает в мишень, равна 0,9. Стрелок произвел 3 выстрела. Найти вероятность того, что все 3 выстрела дали попадание.

5. Брошены монета и игральная кость. Найти вероятность совмещения событий: «появился герб», «появилось 6 очков».

6. В двух ящиках находятся детали: в первом — 10 (из них 3 стандартных), во втором — 15 (из них 6 стандартных). Из каждого

ящика наудачу вынимают по одной детали. Найти вероятность того, что обе детали окажутся стандартными.

7. В студии телевидения 3 телевизионные камеры. Для каждой камеры вероятность того, что она включена в данный момент, равна 0,6. Найти вероятность того, что в данный момент включена хотя бы одна камера.

8. Чему равна вероятность того, что при бросании трех игральных костей 6 очков появится хотя бы на одной из костей.

1.4.
ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ СОБЫТИЯ И ФОРМУЛА БАЙЕСА

1.4.1.
Формула полной вероятности события

Важными следствиями двух теорем теории вероятностей — теоремы сложения и умножения — являются формула полной вероятности события, зависящего от нескольких несовместных гипотез, и формула Байеса. Выше мы рассматривали схему случаев, т. е. полную группу несовместных равновозможных событий. Если же события полной группы необязательно равновозможны, то они называются гипотезами. Гипотезы прнято обозначать латинской буквой H.

Определение 1.4.1. Гипотезы H_1, H_2, ..., H_n — это несовместные, необязательно равновозможные события, в условиях которых только и может появиться некоторое событие A.

Из определения следует: $\sum_{i=1}^{n} H_i = 1$.

Теорема 1.4.1. Полная вероятность события A равна сумме произведений вероятностей гипотез на условные вероятности событий, вычисленные соответственно при каждой из гипотез.

Таким образом:

$$P(A) = \sum_{i=1}^{n} P(H_i) \cdot P(A / H_i), \qquad (1.4.1)$$

где H_1, H_2, ..., H_n — совокупность первоначально выдвинутых гипотез, образующих полную группу событий.

$P(H_1)$, $P(H_2)$, ..., $P(H_n)$ — априорные (доопытные) вероятности гипотез.

П р и м е р 1.4.1. Имеются три одинаковые урны с шарами. В первой из них 3 белых и 4 черных шара, во второй — 2 белых и 5 черных, в третьей — 10 черных шаров. Из случайно выбранной урны наудачу вынут один шар. Найти вероятность того, что он белый.

Р е ш е н и е. Будем считать априорными гипотезами H_1, H_2 и H_3 — выбор урны с соответствующим номером. Так как по условию задачи все гипотезы равновозможны, то $P(H_1) = P(H_2) = P(H_3) = 1/3$.

Найдем условные вероятности появления события А при реализации каждой гипотезы:

$$P(A/H_1) = \frac{3}{7}; \quad P(A/H_2) = \frac{2}{3}; \quad P(A/H_3) = 0.$$

Тогда по формуле (1.4.1):

$$P(A) = \frac{1}{3} \cdot \frac{3}{7} + \frac{1}{3} \cdot \frac{2}{7} + \frac{1}{3} \cdot 0 \approx 0,238.$$

1.4.2. Формула Байеса

Формула Байеса — одна из основных теорем элементарной теории вероятностей, которая определяет вероятность подтверждения одной из первоначально выдвинутых гипотез на основе произведенного опыта (наблюдения).

Полученную по формуле вероятность можно далее уточнять, принимая во внимание данные новых наблюдений. Найденные по формуле Байеса вероятности называются апостериорными (послеопытными) вероятностями.

Формула Байеса применяется, когда событие A, которое может появиться только с одной из гипотез H_1, H_2, ..., H_n, образующих полную группу событий, произошло, и необходимо произвести количественную переоценку априорных вероятностей этих гипотез, т. е. найти апостериорные условные вероятности.

Рассмотрим полную группу несовместных событий, вероятности которых $P(H_1)$, $P(H_2)$, ..., $P(H_n)$ известны. Событие A может наступить только вместе с каким-либо из этих событий H_1, H_2, ..., H_n. Вероятность появления события A вычисляется по формуле полной вероятности.

Пусть событие A произошло, тогда это изменит вероятности первоначально выдвинутых гипотез $H_1, H_2, ..., H_n$.

Определим теперь условные вероятности осуществления этих гипотез в предположении, что событие A произошло, т. е. определим вероятности $P(H_1/A)$, $P(H_2/A)$, ..., $P(H_n/A)$. Учитывая коммутативное свойство произведения событий и формулу (1.3.2), можно записать: $P(AH_1) = P(A)P(H_1/A) = P(H_1)P(A/H_1)$. Откуда следует выражение для апостериорной вероятности гипотезы H_1:

$$P(H_1 / A) = \frac{P(H_1)P(A / H_1)}{P(A)}.$$

Аналогично можно найти апостериорные вероятности остальных гипотез, используя общую формулу:

$$P(H_k / A) = \frac{P(H_k)P(A / H_k)}{P(A)}, \qquad (1.4.2)$$

где $k = 1, 2, ..., n$; n — число первоначально выдвинутых гипотез, образующих полную группу событий; $P(H_k)$ — априорные вероятности гипотез; $P(A)$ — полная вероятность события A, вычисленная по формуле полной вероятноти.

Формула (1.4.2) называется формулой Байеса. Значение формулы Байеса состоит в том, что при наступлении события A, т. е. по мере получения новой информации, мы можем проверять и корректировать выдвинутые до испытания гипотезы. Это дает возможность корректировать управленческие решения в различных практических задачах.

Пример 1.4.2. После двух выстрелов двух стрелков, вероятности попаданий каждого из которых равны $p_1 = 0,6$ и $p_2 = 0,7$, в мишени оказалась одна пробоина. Найти вероятность того, что попал первый стрелок.

Решение. Пусть событие A — одно попадание при двух выстрелах, гипотезы: H_1 — первый стрелок попал в мишень, а второй промахнулся, H_2 — первый стрелок промахнулся, а второй попал, H_3 — оба стрелка попали, H_4 — оба промахнулись. Гипотезы H_1, H_2, H_3, H_4 образуют полную группу событий.

Найдем априорные вероятности этих гипотез:

$$p(H_1) = 0,6 \cdot 0,3 = 0,18;$$
$$p(H_2) = 0,4 \cdot 0,7 = 0,28;$$
$$p(H_3) = 0,6 \cdot 0,7 = 0,42;$$
$$p(H_4) = 0,4 \cdot 0,3 = 0,12.$$

Найдем условные вероятности появления события A в условиях каждой из выдвинутых гипотез:

$$P(A/H_1) = P(A/H_2) = 1, \quad P(A/H_3) = P(A/H_4) = 0.$$

Следовательно, полная вероятность, вычисленная по формуле (1.4.1), равна

$$P(A) = 0{,}18 \cdot 1 + 0{,}28 \cdot 1 + 0{,}42 \cdot 0 + 0{,}12 \cdot 0 = 0{,}46.$$

Вычислим апостериорную вероятность гипотезы H_1 с учетом того факта, что после выстрелов в мишени оказалась одна пробоина (событие A произошло). Применяя формулу Байеса, получим

$$P(H_1 / A) = \frac{P(H_1)P(A/H_1)}{P(A)} = \frac{0{,}18 \cdot 1}{0{,}46} \approx 0{,}39.$$

Аналогично пересчитываются вероятности остальных гипотез.

1.4.3.
Контрольные вопросы

1. При каких условиях применяется формула Байеса?
2. В каких случаях применяется формула полной вероятности? Каким требованиям должны удовлетворять гипотезы?
3. Что такое априорные и апостериорные вероятности?
4. Если все априорные вероятности гипотез одинаковы, то остаются ли их апостериорные вероятности также всегда одинаковыми?

1.5.
ПОВТОРЕНИЕ ОПЫТОВ

1.5.1.
Частная задача о повторении опытов (схема Бернулли)

Схемой Бернулли называется последовательность независимых опытов, в каждом из которых возможны лишь два исхода: появление события A или непоявление события A. Вероятность появления события A в каждом опыте постоянна и равна p, а вероятность его непоявления равна $q = 1 - p$.

Ставится задача определения вероятности того, что событие A в серии из n опытов произойдет ровно m раз.

Рассмотрим событие B_m, состоящее в том, что событие A появится в n опытах ровно m раз. Это событие может осуществиться различными способами. Если событие A появится m раз подряд, начиная с первого опыта, то событие B_m можно записать в виде

$$B_m = A_1 \cdot A_2 \cdot \ldots \cdot A_m \cdot \overline{A}_{m+1} \cdot \ldots \cdot \overline{A}_n,$$

где $\overline{A}_{m+1}$, …, $\overline{A}_n$ — непоявление соответствующих событий в опытах, начиная с номера $m + 1$.

Поскольку в схеме Бернулли опыты независимы, то и события, являющиеся результатами этих опытов также независимы. Поэтому

$$P(B_m) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_m) \cdot P(\overline{A}_{m+1}) \cdot \ldots \cdot P(\overline{A}_n) = p^m q^{n-m}.$$

Событие A может наступить m раз в n опытах в различной последовательности. При этом число возможных сочетаний появлений m событий A в серии из n опытов обозначается как C_n^m. Следовательно, искомая вероятность будет равна

$$P_n(m) = \mathrm{C}_n^m p^m q^{n-m}. \tag{1.5.1}$$

Эта формула называется формулой Бернулли.

Заметим, что правая часть формулы (1.5.1) является одним из членов разложения бинома Ньютона:

$$(p+q)^n = \sum_{i=1}^{n} C_n^i p^i q^{n-i}, \tag{1.5.2}$$

когда $i = m$.

То есть все вероятности $P_n(m)$ для значений m от 0 до 1 являются членами разложения выражения $(p + q)^n$.

П р и м е р 1.5.1. Игральная кость выбрасывается 15 раз. Найти вероятность того, что в данной серии опытов выпадет ровно 10 троек.

Р е ш е н и е. Условие задачи удовлетворяет условиям схемы Бернулли. Здесь $n = 15$, $m = 10$, вероятность выпадения трех очков в каждом опыте равна $p = \frac{1}{6}$, вероятность невыпадения трех очков в каждом опыте равна $q = (1 - p) = \frac{5}{6}$. Тогда:

$$P_{15}(10) = \mathrm{C}_{15}^{10} \cdot \left(\frac{1}{6}\right)^{10} \cdot \left(\frac{5}{6}\right)^{15-10}.$$

Применяя теорему сложения вероятностей несовместных событий, можно получить следующую формулу для вероятности появления события A не менее k раз в серии из n опытов по схеме Бернулли

$$P(m \ge k) = \sum_{m=k}^{n} C_n^m \cdot p^m \cdot (1-p)^{n-m}. \qquad (1.5.3)$$

Вычисление вероятностей по формулам (1.5.1) и (1.5.3) будет рассмотрено ниже.

Формулу (1.5.3) можно, в частности, применить для нахождения вероятности появления события A не менее одного раза в серии из n опытов, или, другими словами, хотя бы один раз. Но проще найти эту вероятность, исходя из противоположного события $\overline{A}$, которое состоит в том, что событие A не появится ни разу в n опытах. Найдем вероятность $P(\overline{A})$ по формуле Бернулли, полагая, что $m = 0$:

$$P(\overline{A}) = P_n(0) = C_n^o \cdot p^0 \cdot (1-p)^{n-0} = \frac{n!}{0! \cdot (n-0)!} \cdot 1 \cdot (1-p)^n = (1-p)^n.$$

Учитывая соотношение между вероятностями противоположных событий, имеем:

$$P(A) = P(m \ge k) = 1 - (1-p)^n. \qquad (1.5.4)$$

Формула (1.5.4) вычисляет вероятность появления хотя бы одного события в серии из n опытов при условии, что в каждом опыте это событие может появиться с одной и той же вероятностью p.

Пример 1.5.2. Игральная кость выбрасывается 15 раз. Найти вероятность того, что в данной серии опытов выпадет хотя бы одна тройка.

Решение. Условие задачи удовлетворяет условиям схемы Бернулли. Здесь $n = 15$, $m = 10$, вероятность выпадения трех очков в каждом опыте равна $p = \frac{1}{6}$. Тогда по формуле (1.5.4) получим

$$P(m \ge k) = 1 - (1-p)^n = 1 - \left(1 - \frac{1}{6}\right)^{15}.$$

1.5.2.
Общая задача о повторении опытов

Общая задача о повторении опытов отличается от схемы Бернулли тем, что вероятность появления некоторого события A меняется от опыта к опыту. Это может происходить, когда опыты производятся в неодинаковых условиях.

Пусть вероятность появления события A в i-м опыте равна p_i, а вероятность его непоявления $q_i = 1 - p_i$, $i = 1, 2, ..., n$. Тогда искомая вероятность $P_n(m)$ того, что событие A в n опытах появится m раз и не появится $n - m$ раз, равна сумме всех возможных произведений, в которых буква p с разными индексами входит m раз, а буква q с разными индексами входит $n - m$ раз. Так как вероятности p_i изменяются от опыта к опыту и, соответственно, изменяются вероятности q_i, то для нахождения вероятности $P_n(m)$ нужно разложить по степеням не выражение $(p + q)^n$, как в формуле (1.5.2), а произведение n биномов вида $(p_i + q_i)$. Сложив затем все слагаемые, содержащие m сомножителей p с разными индексами и $n - m$ сомножителей q с разными индексами, мы получим искомую вероятность $P_n(m)$ в общей задаче о повторении опытов.

Формализовать решение этой задачи можно с помощью функции

$$\varphi(z) = \prod_{i=1}^{n}(p_i z + q_i), \tag{1.5.5}$$

где z — произвольный параметр.

Искомая вероятность $P_n(m)$ будет равна коэффициенту при z^m в разложении по произведения (1.5.5) по степеням z.

Таким образом, решение общей задачи о повторении опытов для различных значений $m = 0, 1, 2, ..., n$ можно выразить формулой

$$\prod_{i=1}^{n}(p_i z + q_i) = \sum_{m=0}^{n} P_n(m) \cdot z^m. \tag{1.5.6}$$

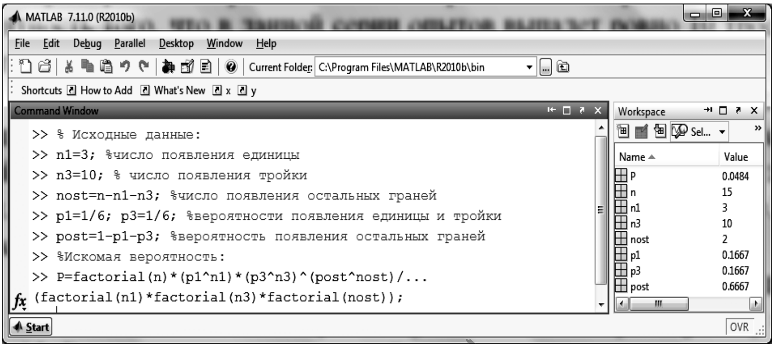

Рис. 1.5.1

Решение примера 1.5.2 в среде MATLAB

П р и м е р 1.5.3. Изделия поставляются пятью независимыми поставщиками. Вероятность годного изделия от первого поставщика равна $p_1 = 0{,}7$, от второго — $p_2 = 0{,}8$, от третьего — $p_3 = 0{,}75$, от четвертого — $p_4 = 0{,}65$ и от пятого — $p_5 = 0{,}85$. Найти вероятность того, что три поставщика из пяти поставят годные детали.

Р е ш е н и е. Решение этой задачи в среде MATLAB представлено на рисунке 1.5.2.

Результаты расчета показывают, что искомая вероятность равна коэффициенту при z^3 в разложении произведения $\varphi(z) = \prod_{i=1}^{5}(p_i z + q_i)$, т. е. $P_5(3) = 0{,}2673$.

Вычисление коэффициентов многочлена при степенях z выполняется командой *sym2poly*. При этом необязательно предварительно приводить многочлен к стандартному виду. Если многочлен задан не в форме полинома, то функция *sym2poly* приводит его к полиномиальной форме и возвращает вектор-строку коэффициентов, начиная с коэффициента при старшей степени аргумента.

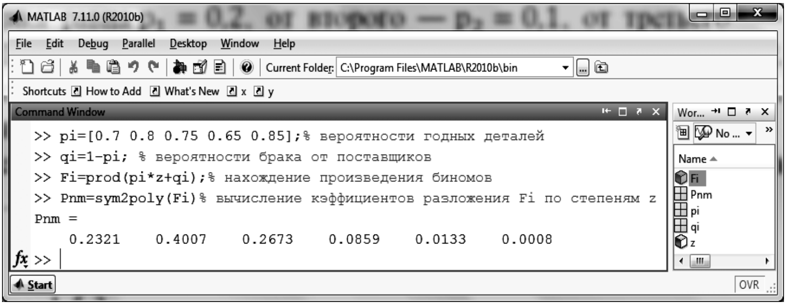

Рис. 1.5.2

Решение задачи на обобщенную теорему о повторении опытов

1.5.3.
Независимые испытания с несколькими исходами

Пусть в серии, состоящей из n независимых опытов, в результате каждого опыта произойдет не одно из двух событий, как это было в схеме Бернулли, а одно из m событий с вероятностями $p_1, p_2, ..., p_m$, причем $p_1 + p_2 + ...+ p_m = 1$. Найдем вероятность $P_{n_1, n_2, \ldots, n_m}$ того, что событие с номером 1 появится n_1 раз, событие с номером 2 появится n_2 раз, ..., событие с номером m появится n_m раз. При этом $n_1 + n_2 + ... + n_m = n$.

Рассмотрим один элементарный исход, когда после проведения серии из n опытов в первых n_1 опытах наступило событие с номером 1, в следующих n_2 опытах наступило событие с номером 2, ..., в последних n_m опытах наступило событие с номером m:

$$(\underbrace{1, 1, \ldots, 1}_{n_1}, \underbrace{2, 2, \ldots, 2}_{n_2}, \ldots, \underbrace{m, m, \ldots, m}_{n_m}).$$

Вероятность такого исхода серии n независимых опытов равна $p_1^{n_1} \cdot p_2^{n_2} \cdot \ldots \cdot p_m^{n_m}$. Остальные благоприятствующие исходы отличаются лишь расположением чисел на n местах. Подсчитаем число благоприятствующих исходов, применяя основное правило комбинаторики и формулу для подсчета числа сочетаний. Очевидно, что n_1 единиц можно разместить на n местах $C_n^{n_1}$ различными способами. Каждому из этих вариантов соответствует размещение n_2 двоек на оставшихся $n - n_1$ местах и т. д.

В результате получим следующее количество вариантов размещения:

$$C_n^{n_1} \cdot C_{n-n_1}^{n_2} \cdot C_{n-n_1-n_2}^{n_3} \cdot \ldots \cdot C_{n-n_1-n_2-\ldots-n_{m-1}}^{n_m} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_m!}.$$

Тогда искомая вероятность рассчитывается по формуле

$$P_n(n_1, n_2, \ldots, n_m) = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_m!} \cdot p^{n_1} \cdot p^{n_2} \cdot \ldots \cdot p^{n_m}. \qquad (1.5.7)$$

П р и м е р 1.5.4. Игральная кость выбрасывается 15 раз. Найти вероятность того, что в данной серии опытов выпадет ровно 10 троек и три единицы.

Р е ш е н и е. Здесь каждое испытание имеет три, а не два исхода: выпадение тройки, выпадение единицы, выпадение остальных граней. Поэтому данная серия опытов не соответствует схеме Бернулли.

Итак, в данной задаче количество исходов в одном опыте $m = 3$, число опытов в серии $n = 15$. Число появления тройки $n_1 = 10$, число появление единицы $n_2 = 3$, число остальных случаев $n_3 = 2$, так, что $n_1 + n_2 + n_3 = n = 15$. Вероятность появления тройки и единицы в одном опыте серии равны, соответственно, $p_1 = p_2 = \frac{1}{6}$, вероятность появления остальных граней в одном опыте равна $p_3 = \frac{4}{6} = \frac{2}{3}$.

Тогда искомая вероятность найдется подстановкой в формулу (1.5.5) соответствующих значений:

$$P_{15}(10, 3, 2) = \frac{15!}{10! \cdot 3! \cdot 2!} \cdot \left(\frac{1}{6}\right)^{10} \cdot \left(\frac{1}{6}\right)^{3} \cdot \left(\frac{2}{3}\right)^{2}.$$

Решим эту задачу в среде MATLAB (рис. 1.5.1).

Найденная вероятность $P = 0{,}0484$ отображена в окне Workspace рабочей области MATLAB.

1.6. ФОРМУЛЫ МУАВРА — ЛАПЛАСА

1.6.1. Локальная теорема Муавра — Лапласа. Функция Гаусса

Теорема 1.6.1 (*локальная теорема Муавра — Лапласа*). Пусть произведено n повторных независимых испытаний, в каждом из которых некоторое событие A наступает с вероятностью p, причем число испытаний достаточно велико ($n > 100$). Тогда вероятность $P_{m,\,n}$ того, что в этих n испытаниях событие A наступит m раз, вычисляется по следующей приближенной формуле:

$$P_n(m) \approx \frac{\varphi(x)}{\sqrt{npq}}, \qquad (1.6.1)$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ — нормированная функция Гаусса; $x = \frac{m - np}{\sqrt{npq}}$; $q = 1 - p$.

Формула (1.6.1) называется формулой Муавра — Лапласа. Эта формула дает удовлетворительную для практики точность при условии $npq > 20$.

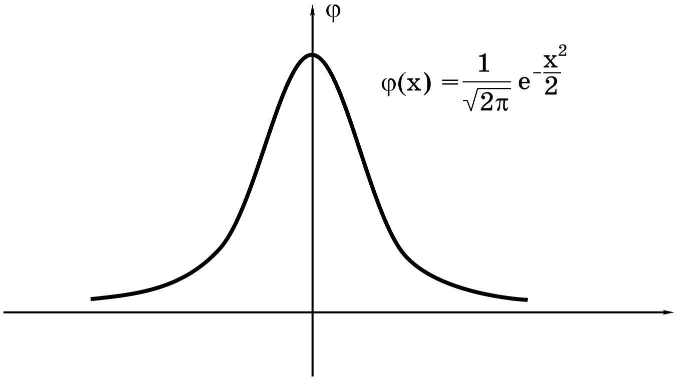

Рис. 1.6.1
Нормированная функция Гаусса

Свойства нормированной функции Гаусса:

1) функция Гаусса четна: $\varphi(-x) = f(x)$, поэтому ее график симметричен относительно оси OY;

2) $\varphi(x) > 0$ при всех x, т. е. график $y = \varphi(x)$ расположен строго выше оси OX;

3) $\lim\limits_{x\to+\infty} \varphi(x) = \lim\limits_{x\to-\infty} \varphi(x) = 0$, т. е. ось OX является горизонтальной асимптотой графика этой функции; на практике полагаем $\varphi(x) = 0$ при $x > 4$.

Схематично график функции Гаусса изображен на рисунке 1.6.1.

Значения функции Гаусса для различных значений x приводятся в приложении 4.

Пример 1.6.1. Имеется партия деталей, состоящая из 1000 шт. В среднем, среди деталей такого вида стандартные детали составляют 90%. Найти вероятность того, что число стандартных деталей в данной партии окажется равным 890.

Решение. Число испытаний в данном случае достаточно велико: $n = 1000 > 10$, поэтому локальная теорема Муавра — Лапласа применима. Из условия следует, что вероятность быть стандартной для произвольной детали данной партии равна $p = \frac{90}{100} = 0{,}9$, $q = 1 - p = 0{,}1$, $m = 890$.

Тогда

$$x = \frac{m - np}{\sqrt{npq}} = \frac{890 - 1000 \cdot 0{,}9}{\sqrt{1000 \cdot 0{,}9 \cdot 0{,}1}} = -1{,}05.$$

По формуле Муавра — Лапласа:

$$P_{1000}(890) = \frac{\varphi(-1{,}05)}{\sqrt{1000 \cdot 0{,}9 \cdot 0{,}1}}.$$

Учитывая, что нормированная функция Гаусса четная, по таблице в приложении 5 находим $\varphi(-1{,}05) = \varphi(1{,}05) \approx 0{,}2299$. Окончательно получаем

$$P_{1000}(890) = \frac{0{,}2299}{\sqrt{1000 \cdot 0{,}9 \cdot 0{,}1}} \approx 0{,}0242.$$

На рисунке 1.6.2 показано решение той же задачи в среде MATLAB.

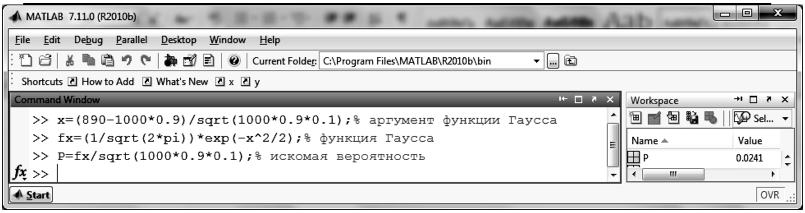

Рис. 1.6.2

Решение задачи примера 1.6.1 в среде MATLAB

Вычисленное значение искомой вероятности $p = 0{,}0241$ отображается в окне Workspace рабочей области MATLAB.

Значения ранее вычисленной вероятности $P_{1000}(890) \approx 0{,}0242$ и вероятности p, полученной с помощью MATLAB, практически совпадают.

Заметим, что значения нормированной функции Гаусса $\varphi(x)$ в среде MATLAB в данном примере можно также вычислить с помощью команды *normpdf*(x, 1,1).

1.6.2. Интегральная теорема Муавра — Лапласа

Т е о р е м а 1.6.2 (*интегральная теорема Муавра — Лапласа*). Пусть произведено n повторных независимых испытаний, в каждом из которых некоторое событие A наступает с вероятностью p, причем число испытаний достаточно велико ($n > 100$). Тогда вероятность того, что число m наступлений события A в этих n испытаниях будет заключено в границах от m_1 до m_2, вычисляется по следующей приближенной формуле:

$$P(m_1 \le m \le m_2) = \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right), \qquad (1.6.2)$$

где $\Phi(u) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^u e^{-\frac{x^2}{2}} dx$ — функция Лапласа.

Формула (1.6.2) называется интегральной формулой Муавра — Лапласа. Она дает удовлетворительную для практики точность при условии $npq > 20$.

Перечислим свойства функции Лапласа.

1. $\Phi(0) = 0$.

2. $\Phi(-\infty) = -0{,}5$.

3. $\Phi(+\infty) = 0{,}5$.

4. $\Phi(u)$ — неубывающая функция.

5. $\Phi(-u) = -\Phi(u)$ — нечетная функция.

Практически можно считать, что уже при $x > 4\Phi(u) \approx 0{,}5$.

Вместо функции Лапласа $\Phi(u)$ можно использовать функцию $N(u) = \dfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{u} e^{-x^2/2} dx$ (приложение 1), которая связана с $\Phi(u)$ соотношением:

$$N(u) = \Phi(u) + 0{,}5. \tag{1.6.3}$$

Тогда интегральную формулу Муавра — Лапласа (1.6.2) можно переписать в виде

$$P(m_1 \le m \le m_2) = N\left(\frac{m_2 - np}{\sqrt{npq}}\right) - N\left(\frac{m_1 - np}{\sqrt{npq}}\right). \tag{1.6.4}$$

При этом следует помнить следующие свойства функции $N(u)$.

1. $N(0) = 0{,}5$.

2. $N(-\infty) = 0$.

3. $N(+\infty) = 1$.

4. $N(u)$ — неубывающая функция.

5. $N(-u) = 1 - F(u)$ — функция общего вида.

Практически можно считать, что уже при $x > 4N(u) \approx 1$, в чем можно убедиться по таблице значений функции $N(u)$, которая приводится в приложении 1.

Пример 1.6.2. Каждая из 1000 деталей партии стандартна с вероятностью 0,9. Найти вероятность того, что число стандартных деталей этой партии будет не меньше 880.

Решение. Число n повторных независимых испытаний в данном случае равно числу деталей в партии (каждая из деталей партии будет проверяться на предмет качества, а в этой проверке и состоит испытание). По условию задачи $n = 1000 > 100$, поэтому интегральная теорема Муавра — Лапласа применима. Неравенство $m \ge 880$, где m — число стандартных деталей в партии, здесь равносильно условию $880 \le m \le 1000$, поэтому $m_1 = 880$, $m_2 = 1000$, $p = 0{,}9$, $q = 1 - 0{,}9 = 0{,}1$, $np = 1000 \cdot 90 = 900$, $npq = 1000 \cdot 90 \cdot 0{,}1 = 90$. Тогда по интегральной формуле Муавра — Лапласа находим

$$P(880 \le m \le 1000) = N\left(\frac{1000 - 900}{\sqrt{90}}\right) - N\left(\frac{880 - 900}{\sqrt{90}}\right) =$$

$$= N(10{,}5) - N(-2{,}11).$$

По свойствам функции $N(u)$:

$$N(10{,}5) \approx 1; \quad N(-2{,}11) = 1 - N(2{,}11).$$

По таблице приложения 1 находим $N(2{,}11) \approx 0{,}9826$.

Тогда $N(-2{,}11) = 1 - 0{,}9826 \approx 0{,}0174$.

Окончательно:

$$P(880 \le m \le 1000) = 1 - 0{,}0174 \approx 0{,}9826.$$

Решение той же задачи в среде MATLAB представлено на рисунке 1.6.3.

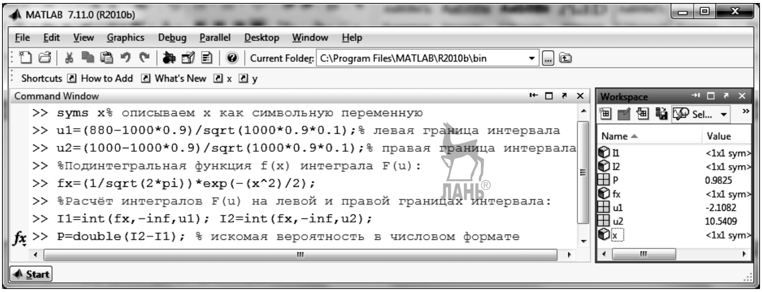

Рис. 1.6.3

Решение задачи примера 1.6.2

Вычисленная вероятность $P = 0{,}9825$ отображена в поле Workspace рабочего окна MATLAB. Функция double переводит результат вычисления в формат числа с плавающей точкой двойной точности. С подробностями можно познакомиться с помощью справочной системы MATLAB. Для этого надо набрать команду в окне MATLAB, установить курсор на этой команде и нажать клавишу F1 на клавиатуре компьютера.

Следует отметить, что результат, полученный в среде MATLAB, более точен, поскольку табличные значения интегралов, найденные по таблице, приводятся не для всех значений аргументов. В то время как система MATLAB вычисляет определенные интегралы численным методом, а точнее — методом Симпсона, в основе которого лежит параболическая интерполяция.

Следствия из интегральной теоремы Муавра — Лапласа.

Следствие 1. Вероятность того, что число m наступлений события A в n повторных независимых испытаниях будет отличаться от величины np не более чем на некоторую величину ε (по абсолютной величине), вычисляется по формуле

$$P(|m-np|\le\varepsilon)=2\cdot\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right). \qquad (1.6.5)$$

Следствие 2. Вероятность того, что относительная частота m/n появления события A в n повторных независимых испытаниях будет отличаться от вероятности p наступления этого события в одном испытании не более чем на ε (по абсолютной величине), вычисляется по формуле

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=2\cdot\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right). \qquad (1.6.6)$$

П р и м е р 1.6.3. Подлежат исследованию 1000 проб руды. Вероятность промышленного содержания металла в каждой пробе равна 0,15. Найти границы, в которых с вероятностью 0,9973 будет заключено число проб руды с промышленным содержанием металла.

Р е ш е н и е. Перепишем формулу (1.6.5) в виде

$$P(np-\varepsilon\le m\le np+\varepsilon)=2\cdot\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right).$$

Очевидно, искомые границы для числа m проб руды с промышленным содержанием металла из из общего числа $n = 1000$ проб определяются величинами $np - \varepsilon$ и $np + \varepsilon$, где $np = 1000 \cdot 0{,}15 = 150$. Определим величину ε. Из следствия 1 интегральной теоремы Муавра — Лапласа и условия задачи следует

$$\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right)=\frac{0{,}9973}{2}\approx 0{,}4987.$$

Найдем такое значение аргумента $u=\dfrac{\varepsilon}{\sqrt{npq}}$, что $\Phi(u) = 0{,}4987$.

По формуле (1.6.3) находим

$$N(u) = \Phi(u) + 0{,}5 = 0{,}4987 + 0{,}5 \approx 0{,}9987.$$

Из таблицы приложения 1 для $N(u) = 0{,}9987$ получим искомое значение аргумента $u = 3$.

Тогда

$$\frac{\varepsilon}{\sqrt{npq}}=3 \text{ и } \varepsilon=3\sqrt{npq}=3\cdot\sqrt{1000\cdot 0{,}15\cdot 0{,}85}\approx 34.$$

Окончательно находим искомые границы:

$$np - \varepsilon = 1000 \cdot 0{,}15 - 34 = 116;$$
$$np + \varepsilon = 1000 \cdot 0{,}15 + 34 = 184.$$

т. е. с вероятностью 0,9973 число проб руды с промышленным содержанием металла (из данных 1000 проб) попадет в интервал (116; 184).

Решение этого примера в среде MATLAB дано на рисунке 1.6.4.

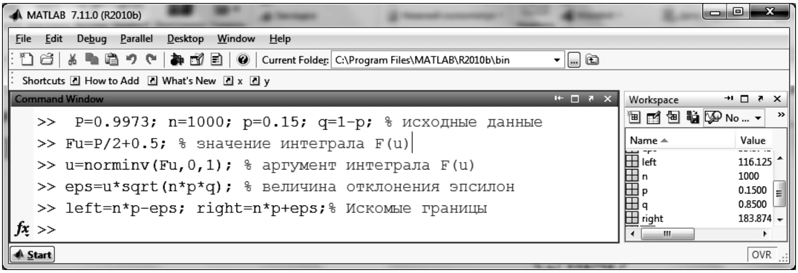

Рис. 1.6.4

Решение примера 1.6.3 в среде MATLAB

Результаты расчета левой (left) и правой (right) границ искомого интервала отображены в окне Workspace рабочей среды MATLAB. Если округлить эти величины, то получим для левой границы значение 116 и для правой границы значение 184, что совпадает с результатами, полученными без применения MATLAB. Команда *norminv*(*Fu*, 0,1) вычисляет аргумент u по заданному значению интеграла $F(u, m, \sigma) = \; = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{u} e^{-(x-m)^2/2\sigma^2} dx$ с параметрами $m = 0$, $\sigma = 1$,

т. е. вычисляет обратное значение функции $N(u) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{u} e^{-x^2/2} dx$:

$u = N^{-1}(u)$.

П р и м е р 1.6.4. В лесхозе приживается в среднем 80% саженцев. Сколько саженцев надо посадить, чтобы с вероятностью 0,9981 можно было утверждать, что доля прижившихся саженцев будет находиться в границах от 0,75 до 0,85?

Р е ш е н и е. $p = 80/100 = 0{,}8$ — вероятность прижиться для каждого из саженцев, $q = 1 - p = 1 - 0{,}8 = 0{,}2$. Пусть n — необходимое число саженцев (искомая величина данной задачи), m — число прижившихся из них, тогда m/n — доля прижившихся саженцев. По условию, $P\left(0,75 \le \frac{m}{n} \le 0,85\right) = 0,9981.$

Данные границы для доли m/n симметричны относительно величины $p = 0,8$, поэтому неравенство $0,75 \le m/n \le 0,85$ равносильно неравенству $|m/n - 0,8| \le 0,05.$

Следовательно, вероятность 0,9981 — это та самая вероятность, которая вычисляется по следствию 2 интегральной теоремы Муавра — Лапласа при $\Delta = 0,05$, $p = 0,8$, $q = 0,2$:

$$P\left(\left|\frac{m}{n} - 0,8\right| \le 0,05\right) = 2 \cdot \Phi\left(\frac{0,05\sqrt{n}}{\sqrt{0,8 \cdot 0,2}}\right) \approx 0,9981.$$

Тогда

$$\Phi\left(\frac{0,05\sqrt{n}}{\sqrt{0,8 \cdot 0,2}}\right) = \frac{0,9981}{2} = 0,4991.$$

Найдем такое значение $u = \frac{0,05\sqrt{n}}{\sqrt{0,8 \cdot 0,2}}$, что $\Phi(u) = 0,4991$. Для нахождения u применим тот же прием, что и в примере (1.6.3): $N(u) = \Phi(u) + 0,5 = 0,9991$. По таблице значений $N(u)$ находим искомое значение: $u \approx 3,12$. Тогда

$$\frac{0,05\sqrt{n}}{\sqrt{0,8 \cdot 0,2}} \approx 3,12; \quad \sqrt{n} = \frac{3,1\sqrt{0,8 \cdot 0,2}}{0,05}; \quad n = \frac{3,12^2 \cdot 0,8 \cdot 0,2}{0,05^2} \approx 624.$$

Заметим, что значение n округлено до целых в большую сторону, чтобы обеспечить, как говорят, «запас по вероятности». Кроме того, видно, что полученное значение n достаточно велико (более 100), поэтому применение интегральной теоремы Муавра — Лапласа для решения данной задачи оправдано.

В рабочей среде MATLAB данная задача решается аналогично тому, что было показано в примере (1.6.3). Рекомендуется проделать это самостоятельно.

1.7. ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

1.7.1. Классификация случайных величин

Случайные события — это качественная характеристика случайного результата опыта, но случайный результат можно характеризовать и количественно, если каждому элементарному исходу опыта поставить в соответствие вещественное число.

Определение 1.7.1. Случайная величина — это величина, которая в результате опыта может принять то или иное значение, но неизвестно заранее какое именно.

Принятие случайной величиной конкретного значения тоже представляет собой событие, поэтому все теоремы, рассмотренные ранее для случайных событий, можно применять и для случайных величин.

Примеры случайных величин: число выпадений герба при бросании монеты, ошибка при измерении, число бракованных деталей при выпуске продукции.

Названия случайных величин обозначают заглавными буквами, предпочтительно из конца латинского алфавита, например X, Y, Z, а их всевозможные значения соответственно малыми буквами x, y, z.

Например, если случайная величина X — число попаданий при трех выстрелах, то ее возможные значения: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$.

Среди случайных величин можно выделить два основных класса: дискретные случайные величины и непрерывные случайные величины.

Определение 1.7.2. Дискретная случайная величина — это величина, принимающая счетное множество значений. Ее значения отделимы друг от друга.

Определение 1.7.3. Непрерывная случайная величина — это величина, возможные значения которой непрерывно заполняют некоторый интервал.

Примеры дискретных случайных величин: число появлений герба при трех бросаниях монеты (возможные значения 0, 1, 2, 3); относительная частота появления герба в том же опыте (возможные значения $0, \frac{1}{3}, \frac{2}{3}, 1$).

Примеры непрерывных случайных величин: абсцисса (ордината) точки попадания при выстреле; расстояние от точки попадания до центра мишени; время прихода поезда метро на платформу.

1.7.2. Ряд распределения дискретной случайной величины

Рассмотрим дискретную случайную величину X с возможными значениями $x_1, x_2, ..., x_n$. Каждое из этих значений возможно, но не достоверно, и величина X может принять каждое из них с некоторой вероятностью. В результате опыта величина X примет одно из этих значений, т. е. произойдет одно из полной группы несовместных событий:

$$\begin{cases} X = x_1, \\ X = x_2, \\ \quad ... \\ X = x_n. \end{cases}$$

Обозначим вероятности этих событий буквой p с соответствующими индексами:

$$P(X = x_1) = p_1,\ P(X = x_2) = p_2,\ ...,\ P(X = x_n) = p_n.$$

Так, перечисленные выше события несовместны и образуют полную группу, то

$$\sum_{i=1}^{n} p_i = 1, \tag{1.7.1}$$

т. е. сумма вероятностей всех возможных значений случайной величины равна единице.

Условие (1.7.1) называется условием нормировки для дискретной случайной величины.

Случайная величина будет полностью описана с вероятностной точки зрения, если мы зададим это распределение, т. е. в точности укажем, какова вероятность каждого из событий. Этим мы установим так называемый закон распределения случайной величины.

Определение 1.7.4. Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Про случайную величину мы будем говорить, что она подчинена данному закону распределения.

Закон распределения может быть представлен в виде таблицы, аналитической зависимости или графика.

Простейшей формой задания закона распределения является таблица, в которой перечислены возможные значения случайной величины и соответствующие им вероятности (табл. 1.7.1).

Такую таблицу мы будем называть рядом распределения случайной величины X.

Для ряда распределения должно соблюдаться условие нормировки (1.7.1).

Таблица 1.7.1

Ряд распределения вероятностей

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Чтобы придать ряду распределения более наглядный вид, часто прибегают к его графическому изображению: по оси абсцисс откладываются возможные значения случайной величины, по оси ординат — вероятности этих значений. Для наглядности полученные точки соединяются отрезками прямых линий. Такая фигура называется многоугольником распределения. На рисунке 1.7.1 показан пример многоугольника распределения.

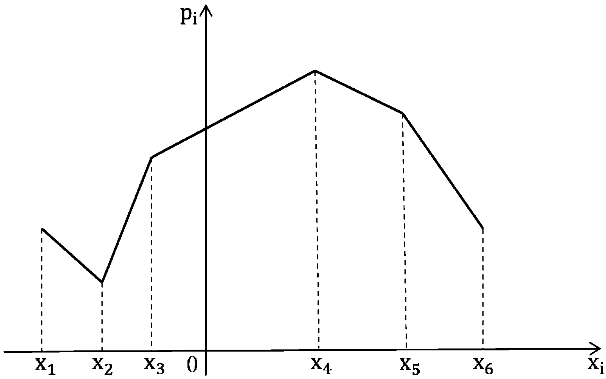

Рис. 1.7.1

Многоугольник распределения

Многоугольник распределения, так же как и ряд распределения, полностью характеризует случайную величину; он также является одной из форм закона распределения.

Замечание 1. Сумма всех ординат многоугольника распределения равна единице.

Замечание 2. При построении многоугольника распределения надо помнить, что соединение полученных точек носит условный характер. В промежутках между значениями случайной величины вероятность не принимает никакого значения. Точки соединены только для наглядности.

П р и м е р 1.7.1. Два стрелка делают по одному выстрелу по мишени. Вероятности их попадания при одном выстреле равны соответственно $p_1 = 0{,}6$ и $p_2 = 0{,}7$. Требуется составить ряд распределения случайной величины X — числа попаданий после двух выстрелов.

Р е ш е н и е. Очевидно, что X может принимать три значения: 0, 1 и 2. Их вероятности:

$$P(X = 0) = (1 - p_1)(1 - p_2) = 0{,}4 \cdot 0{,}3 = 0{,}12;$$
$$P(X = 1) = p_1(1 - p_2) + (1 - p_1)p_2 = 0{,}6 \cdot 0{,}3 + 0{,}4 \cdot 0{,}7 = 0{,}46;$$
$$P(X = 2) = p_1 \cdot p_2 = 0{,}6 \cdot 0{,}7 = 0{,}42.$$

Следовательно, ряд распределения в данном случае имеет вид:

X	0	1	2
P	0,12	0,46	0,42

1.7.3. Функция распределения

В предыдущем параграфе мы ввели в рассмотрение ряд распределения как закон распределения дискретной случайной величины. Однако эта характеристика не является универсальной; она существует только для дискретных случайных величин. Нетрудно убедиться, что для непрерывной случайной величины такой характеристики построить нельзя. Действительно, непрерывная случайная величина имеет бесчисленное множество возможных значений, сплошь заполняющих некоторый промежуток (так называемое «несчетное множество»). Составить таблицу, в которой были бы перечислены все возможные значения такой случайной величины, невозможно.

Для количественной характеристики этого распределения вероятностей удобно воспользоваться не вероятностью события $X = x$, а вероятностью события $X < x$, где x — некоторая текущая переменная. Вероятность этого события, очевидно, зависит от x и является некоторой функцией от x. Эта функция называется функцией распределения случайной величины X и обозначается $F(x)$:

$$F(x) = P(X < x). \qquad (1.7.2)$$

Функцию распределения $F(x)$ иногда называют интегральной функцией распределения или интегральным законом распределения.

Функция распределения — самая универсальная характеристика случайной величины. Она существует как для дискретных, так и непрерывных случайных величин. Функция распределения полностью характеризует случайную величину с вероятностной точки зрения, т. е. является одной из форм закона распределения.

Сформулируем некоторые общие свойства функции распределения.

1. Функция распределения $F(x)$ есть неубывающая функция своего аргумента, т. е. при $x_2 > x_1$, $F(x_2) > F(x_1)$.

2. На минус бесконечности функция распределения равна нулю: $F(-\infty) = 0$.

3. На плюс бесконечности функция распределения равна единице: $F(+\infty) = l$.

4. Вероятность того, что случайная величина примет значение из интервала $[a; b]$ равна разности значений функции распределения на концах интервала:

$$P(a \le X \le b) = F(b) - F(a). \tag{1.7.3}$$

Эта формула следует из определения интегральной функции распределения (1.7.2).

Не давая строгого доказательства этих свойств, проиллюстрируем их с помощью наглядной геометрической интерпретации. Для этого будем рассматривать случайную величину X как случайную точку на оси OX, которая в результате опыта может занять то или иное положение. Тогда функция распределения $F(x)$ есть вероятность того, что случайная точка X в результате опыта попадет левее точки x. Будем увеличивать x, т. е. перемещать точку x вправо по оси абсцисс. Очевидно, при этом вероятность того, что случайная точка X попадет левее x, не может уменьшиться; следовательно, функция распределения $F(x)$ с возрастанием x убывать не может. Чтобы убедиться в том, что $F(-\infty) = 0$, будем неограниченно перемещать точку x влево по оси абсцисс. При этом попадание случайной точки X левее x в пределе становится невозможным событием; естественно полагать, что вероятность этого события стремится к нулю, т. е. $F(-\infty) = 0$.

Аналогичным образом, неограниченно перемещая точку x вправо, убеждаемся, что $F(+\infty) = l$, так как событие $X < x$ становится в пределе достоверным.

Зная ряд распределения дискретной случайной величины, можно легко построить функцию распределения этой величины. Действительно, $F(x) = P(X < x) = \sum\limits_{x_i} P(\mathrm{X} < x_i)$, где неравенство $X < x_i$ под знаком суммы указывает, что суммирование распространяется на все те значения x_i, которые меньше x.

Когда текущая переменная x проходит через одно из возможных значений дискретной величины X, функция распределения меняется скачкообразно, причем величина скачка равна вероятности этого значения.

Пример 1.7.2. Производится один опыт, в котором может появиться или не появиться событие A. Известно, что вероятность появления события A равна 0,7. Пусть случайная величина X — число появлений события A в опыте. Построить ее функцию распределения.

Решение. Ряд распределения величины X имеет вид:

X	0	1
P	0,3	0,7

Построим функцию распределения величины X.

1. При $x = 0$, $F(0) = P(X < 0) = 0$, так как данная случайная величина не имеет значений меньших нуля, а потому событие $(X < 0)$ для нее является невозможным. Таким образом, в интервале $x \le 0$ интегральная функция распределения $F(x)$ везде постоянна и равна 0.

2. При $x = 1$, $F(1) = P(X < 1) = P(X = 0) = 0{,}3$, так как случайная величина по условию задачи может принимать только одно значение меньшее единицы, которое равно нулю, с вероятностью 0,3. Таким образом, в интервале $0 < x \le 1$ интегральная функция распределения $F(x)$ везде постоянна и равна 0,3.

3. При $x > 1$, например, $x = 2$.

$$F(2) = P(X < 2) = P(X = 1) + P(X = 0) = 0{,}7 + 0{,}3 = 1,$$

так как случайная величина по условию задачи может принимать только два значения меньше 2: либо $X = 0$, либо $X = 1$.

Аналогично, при любом значении переменной x, которое больше единицы, будем иметь $F(x) = 1$.

Окончательно имеем

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le 0; \\ 0{,}3 & \text{при} \quad 0 < x \le 1; \\ 1 & \text{при} \quad x > 1. \end{cases}$$

График найденной функции распределения изображен на рисунке 1.7.2.

В качестве самостоятельной работы рекомендуется решить следующую задачу.

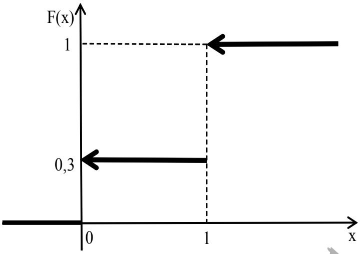

Рис. 1.7.2

График интегральной функции распределения случайной величины в примере 1.7.2

Пусть X — случайное число очков, выпавшее при одном выбрасывании игральной кости. Найти интегральную функцию распределения случайной величины X.

Функция распределения любой дискретной случайной величины всегда есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям этих значений. Сумма всех скачков функции $F(x)$ равна единице.

По мере увеличения числа возможных значений случайной величины и уменьшения интервалов между ними число скачков становится больше, а сами скачки — меньше; ступенчатая кривая становится все более плавной.

На рисунке 1.7.3 показан характерный вид интегральной функции распределения дискретной случайной величины.

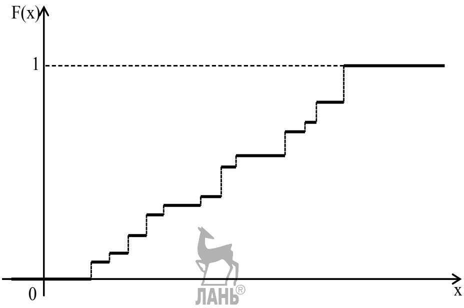

Рис. 1.7.3

Характерный вид интегральной функции распределения дискретной случайной величины

1.8.
ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

В тех случаях, когда невозможно найти закон распределения или этого не требуется, можно ограничиться нахождением числовых характеристик случайной величины. Они выражают наиболее существенные особенности того или иного распределения. В теории вероятностей применяется большое количество различных числовых характеристик, характеризующих те или иные особенности закона распределения. Из них мы рассмотрим только некоторые, наиболее часто применяемые.

1.8.1.
Характеристики положения

Характеристики положения указывают некоторое среднее, ориентировочное значение, около которого группируются все возможные значения случайной величины. К ним относятся:

— математическое ожидание;

— мода;

— медиана.

Определение 1.8.1. Математическим ожиданием дискретной случайной величины называется сумма произведений всех возможных значений случайной величины на соответствующие им вероятности.

Для математического ожидания случайной величины X используются следующие обозначения: $M(X)$ или m_x.

Согласно данному определению:

$$M(X) = \sum_{i=1}^{n} x_i p_i. \tag{1.8.1}$$

При этом $\sum_{i}^{n} p_i = 1$.

В дальнейшем мы иногда будем пользоваться понятием центрированной случайно величины: $\dot{X} = X - m_X$.

Покажем, что математическое ожидание центрированной случайной величины равно нулю:

$$M(\dot{X}) = M[X - m_x] = \sum_{i=1}^{n}(x_i - m_x)\cdot p_i = \sum_{i=1}^{n} x_i p_i - \sum_{i=1}^{n} m_x p_i =$$

$$= m_x - m_x \sum_{i=1}^{n} p_i = m_x - m_x \cdot 1 = 0.$$

Перечислим некоторые свойства математического ожидания.

1. Математическое ожидание дискретной случайной величины есть неслучайная (постоянная) величина.

2. Математическое ожидание постоянной равно самой постоянной:

$$M(C) = C.$$

3. Постоянный множитель можно выносить за знак математического ожидания:

$$M(C \cdot x) = C \cdot M(x).$$

П р и м е р 1.8.1. Производится три независимых выстрела по мишени; вероятность попадания при каждом выстреле равна 0,4. Случайная величина X — число попаданий. Определить математическое ожидание величины X.

Р е ш е н и е.

1. Пользуясь схемой Бернулли (п. 1.5.1), строим ряд распределения случайной величины X:

x_i	0	1	2	3
p_i	0,216	0,432	0,288	0,064

2. По ряду распределения и формуле (1.8.1) вычисляем искомое математическое ожидание:

$$m_x = 0 \cdot 0{,}216 + 1 \cdot 0{,}432 + 2 \cdot 0{,}288 + 3 \cdot 0{,}064 = 1{,}2.$$

Определение 1.8.2. Мода — наиболее вероятное значение случайной величины X. Будем обозначать ее символом $\mathcal{M}$.

Так, на рисунке 1.7.1 $\mathcal{M} = x_4$.

Определение 1.8.3. Медиана (**Me**) — это такое значение случайной величины, что выполняется условие

$$P(X < \mathbf{Me}) = P(X > \mathbf{Me}),$$

т. е. одинаково вероятно, окажется ли случайная величина меньше или больше **Me**.

Замечание. Если распределение симметричное и модальное (имеет одну моду), то математическое ожидание, медиана и мода характеризуются одним положением и совпадают.

П р и м е р 1.8.2. Если ряд распределения дискретной случайной величины X имеет вид:

x_i	1	2	3	4
p_i	0,1	0,7	0,15	0,05

то мода $\mathcal{M} = 2$.

П р и м е р 1.8.3. Пусть две случайные величины X и Y заданы рядами распределения:

x_i	49	50	51
p_i	0,1	0,8	0,1

y_i	0	100
p_i	0,5	0,5

Найдем математические ожидания дискретных этих случайных величин:

$$M(X) = 49 \cdot 0{,}1 + 50 \cdot 0{,}8 + 51 \cdot 0{,}1 = 50;$$
$$M(Y) = 0 \cdot 0{,}5 + 100 \cdot 0{,}5 = 50.$$

Получили $M(X) = M(Y)$, т. е. для различных случайных величин получены одинаковые математические ожидания. Следовательно, наряду с математическим ожиданием желательно знать, насколько значения случайной величины отклоняются от него. Для этого вводятся понятия дисперсии и среднеквадратического отклонения случайной величины как характеристики ее рассеивания относительно математического ожидания.

1.8.2. Характеристики рассеивания

Значения наблюдаемых в практике случайных величин всегда колеблются около среднего значения (математического ожидания). Это явление называется рассеиванием величины около ее среднего значения. Числовые характеристики, описывающие степень рассеивания случайной величины, называются характеристиками рассеивания, и основные из них — дисперсия и среднеквадратическое отклонение. Само слово дисперсия означает «рассеивание».

Отклонения противоположных знаков в среднем взаимно погашаются. Поэтому в качестве меры рассеивания принимают математическое ожидание квадрата отклонения случайной величины от ее математического ожидания или математическое ожидание квадрата центрированной случайной величины.

Определение 1.8.4. Дисперсией называется математическое ожидание квадрата разности случайной величины и ее математического ожидания.

Для дисперсии случайной величины мы будем использовать следующие обозначения: $D(X)$ или D_x.

Согласно данному определению:

$$D(X) = M[(X - m_x)^2]\sum_{i=1}^{n}(x_i - m_x)^2 \cdot p_i. \qquad (1.8.2)$$

Чем большие отклонения в обе стороны от среднего значения m_x данной случайной величины и чем больше вероятности таких отклонений, тем больше дисперсия случайной величины. В частном случае, когда среднее значение случайной величины равно нулю, дисперсия характеризует разброс значений случайной величины в обе стороны от нуля.

Дисперсия, как и математическое ожидание, являются величиной не случайной.

В ряде случаев для вычисления дисперсии удобней использовать формулу

$$D(X) = M(X^2) - M^2(X). \qquad (1.8.3)$$

Это означает, что дисперсия равна разности математического ожидания квадрата случайной величины и квадрата ее математического ожидания.

Перечислим некоторые свойства дисперсии.

1. Постоянный множитель можно выносить за знак дисперсии, возведя его в квадрат:

$$D(cX) = M[(cX - M(cX))^2] = M[(cX - cM(X))^2] =$$
$$= c^2M[(X - M(X))^2] = c^2D(X).$$

2. Дисперсия неслучайной (детерминированной) величины равна нулю:

$$D(C) = M(C - M(C))^2 = M(C - C)^2 = M(0) = 0.$$

3. Дисперсия принимает только неотрицательные значения. Это свойство следует из определения дисперсии.

Из определения дисперсии ясно, что дисперсия случайной величины имеет размерность квадрата случайной величины.

Для наглядности удобнее пользоваться характеристикой рассеивания, размерность которой совпадает с размерностью случайной величины. Для этого из дисперсии извлекают квадратный корень. Полученная величина называется среднеквадратическим отклонением случайной величины (сокращенно с. к. о.). Среднеквадратическое отклонение будем обозначать $\sigma[X]$ или σ_X. Таким образом, с. к. о. вычисляется по формуле

$$\sigma[X] = \sqrt{D(X)}. \qquad (1.8.4)$$

Математическое ожидание и с. к. о. — наиболее часто применяемые характеристики случайной величины. Они характеризуют наиболее важные черты распределения: его положение и степень разбросанности. Так, например, если случайная величина X есть доходность некоторого актива, то $M[X]$ — средняя (прогнозная) доходность актива, $D[X]$ — мера отклонения (колебания) доходности от ожидаемого среднего значения, т. е. риск актива.

П р и м е р 1.8.4. Для двух случайных величин X и Y, заданных в примере 1.8.3, найдем их дисперсии и среднеквадратические отклонения.

Р е ш е н и е. Из предыдущего примера: $M(X) = 50$, $M(Y) = 50$.

Вычисляем искомые дисперсии и среднеквадратические отклонения по формулам (1.8.2) и (1.8.4):

$$D_x = (49-50)^2 \cdot 0{,}1 + (50-50)^2 \cdot 0{,}8 + (51-50)^2 \cdot 0{,}1 = 0{,}2;$$
$$\sigma_X = \sqrt{0{,}2} \approx 0{,}447;$$
$$D_Y = (0-50)^2 \cdot 0{,}5 + (100-50)^2 \cdot 0{,}5 = 1250;$$
$$\sigma_Y = \sqrt{1250} \approx 35{,}355.$$

Вычисления вышеописанных числовых характеристик в системе MATLAB реализуется с помощью функций суммирования (sum), произведения (prod), операторов возведения в степень (^) и деления (/). Их описание приводится в работе [2]. Кроме того, с синтаксисом операторов можно ознакомиться в справочной системе MATLAB следующим образом: Help — Function Browser — MATLAB — Data Analysis — Basic Operations.

Определение 1.8.5. Начальным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени этой величины:

$$a_k = M[X^k] = \sum_{i=1}^{n} x_i^k \cdot p_i. \qquad (1.8.5)$$

Определение 1.8.6. Центральным моментом k-го порядка случайной величины X называется математическое ожидание k-й степени центрированной случайной величины $\dot{X}$:

$$\mu_k = M[\dot{X}^k] = M[(X - m_x)^k]. \qquad (1.8.6)$$

Замечание. Начальный момент 1-го порядка равен математическому ожиданию: $\alpha_1 = M[X]$.

Центральный момент 2-го порядка равен дисперсии: $\mu_2 = D[X]$.

1.8.3.
Контрольные вопросы

1. Дайте определение дискретной случайной величины.

2. Какими способами можно задать дискретную случайную величину?

3. Дайте определение математического ожидания дискретной случайной величины. Назовите свойства математического ожидания.

4. Определение дисперсии дискретной случайной величины. Формулы для вычисления дисперсии. Свойства дисперсии.

5. Дайте определение начальных и центральных моментов k-го порядка.

1.9.
ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

1.9.1.
Биномиальное распределение (закон Бернулли)

Пусть производится n независимых испытаний, в каждом из которых случайное событие A может наступить с вероятностью p и не наступить с вероятностью $q = 1 - p$. Вероятность $P_n(m)$ того, что в n независимых опытах событие A наступит m раз находится по формуле (1.5.1). Это задача о повторении опытов или схема Бернулли. Ее решение рассматривалось в п. 1.5.1.

Дискретная случайная величина, которая может принимать только целочисленные значения с вероятностью

$$P_n(m) = \mathrm{C}_n^m p^m (1-p)^{n-m}, \tag{1.9.1}$$

называется распределенной по биноминальному закону, или закону распределения Бернулли.

Величины n и p являются параметрами биномиального распределения. Биномиальный коэффициент C_n^m есть число сочетаний из n элементов по m элементов ($m \le n$).

Ряд распределения случайной величины, подчиненной биномиальному закону, представлен таблицей 1.9.1.

Таблица 1.9.1

Ряд биномиального распределения

m	0	1	2	...	n
$P_n(m)$	$C_n^0 \cdot p^0 \cdot (1-p)^n$	$C_n^1 \cdot p^1 \cdot (1-p)^{n-1}$	$C_n^2 \cdot p^2 \cdot (1-p)^{n-2}$		$C_n^n \cdot p^n \cdot (1-p)^0$

При этом условие нормировки (1.7.1) соблюдается:

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

Равенство получено на основании формулы бинома Ньютона.

Рассмотрим числовые характеристики биномиального распределения. В соответствии с формулой (1.8.1) математическое ожидание дискретной случайной величины, распределенной по биномиальному закону, определяется выражением

$$M[X] = \sum_{m=0}^{n} m \cdot P_n(m) = np. \qquad (1.9.2)$$

Дисперсия биномиального распределения (1.8.2) вычисляется по формуле

$$D[X] = \sum_{m=0}^{n} (m-np)^2 P_n(m) = npq, \qquad (1.9.3)$$

где $q = (1-p)$ есть вероятность ненаступления события A в отдельном опыте.

Доказательство справедливости этих формул будет дано в п. 1.14.2.

Пример 1.9.1. Составить ряд распределения дискретной случайной величины X — числа попаданий при 5 выстрелах, если вероятность попадания при одном выстреле равна $p = 0{,}8$. Найти математическое ожидание, дисперсию и среднеквадратическое отклонение случайной величины X. Построить многоугольник распределения случайной величины X и ее функцию распределения.

Решение.

$$P(X=0) = P_5(0) = C_5^0 \cdot p^0 \cdot (1-p)^5 = 0{,}2^5 \approx 0{,}0003;$$

$$P(X=1) = P_5(1) = C_5^1 \cdot p^1 \cdot (1-p)^4 = 5 \cdot 0{,}8 \cdot 0{,}2^4 \approx 0{,}0064;$$

$$P(X=2) = P_5(2) = C_5^2 \cdot p^2 \cdot (1-p)^3 = 10 \cdot 0{,}8^2 0{,}2^3 \approx 0{,}0512;$$

$$P(X=3) = P_5(3) = C_5^3 \cdot p^3 \cdot (1-p)^2 = 10 \cdot 0{,}8^3 0{,}2^2 \approx 0{,}2048;$$

$$P(X=4) = P_5(4) = C_5^4 \cdot p^4 \cdot (1-p)^1 = 5 \cdot 0{,}8^4 0{,}2^1 \approx 0{,}4096;$$

$$P(X=5) = P_5(5) = C_5^5 \cdot p^5 \cdot (1-p)^0 = 1 \cdot 0{,}8^5 0{,}2^0 \approx 0{,}3277.$$

Ряд распределения:

X	0	1	2	3	4	5
$P_n(m)$	0,0003	0,0064	0,0512	0,2048	0,4096	0,3277

Математическое ожидание:

$$M[X] = np = 5 \cdot 0{,}8 = 4.$$

Дисперсия:

$$D[X] = npq = 5 \cdot 0{,}8 \cdot 0{,}2 = 0{,}8.$$

Среднеквадратическое отклонение:

$$\sigma_X = \sqrt{D_X} = \sqrt{0{,}8} = 0{,}89.$$

Вычислим значения функции распределения.

$$F(0) = P(X < 0) = 0;$$
$$F(1) = P(X < 1) = P(X = 0) \approx 0{,}0003;$$
$$F(2) = P(X < 2) = P(X = 0) + P(X = 1) \approx 0{,}0067;$$
$$F(3) = P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) \approx 0{,}0579;$$
$$\ldots$$
$$\ldots$$
$$F(6) = P(X < 6) = 1.$$

На рисунке 1.9.1 показано решение этой задачи в среде MATLAB.

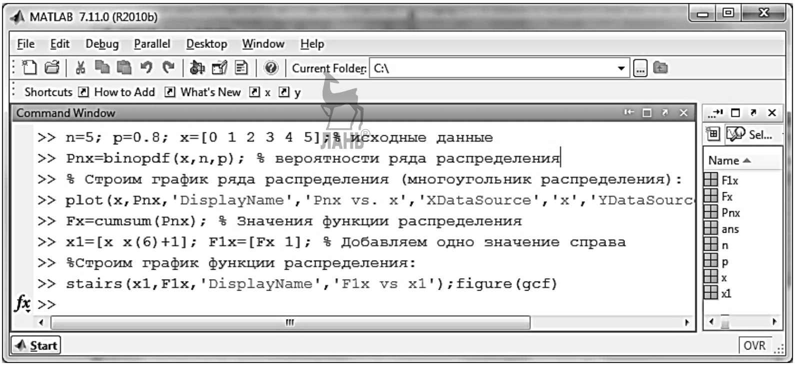

Рис. 1.9.1

Расчет ряда и функции распределения биномиального распределения

Команда $Pnx = binopdf(x, n, p)$ возвращает значения вероятностей $P_n(x) = C_n^x \cdot p^x (1-p)^{n-x}$ для всех значений вектора $x = (0, 1, 2, 3, 4, 5)$ и присваивает их вектору *Pnx*. Таким образом, с помощью функции *binopdf* в MATLAB реализуется вычисление биномиального ряда распределения.

Функция *plot*(...) появляется в командном окне автоматически при выполнении следующих действий. Сначала, удерживая клавишу ⟨*ctrl*⟩, кликаем в окне *Workspace* иконки векторов *x* и *Pnx* в указанной последовательности (поскольку строится зависимость *Pnx* от *x*). Затем на инструментальной панельке того же окна кликаем меню *plot*. Из выпадающего списка выбираем вид графика. В нашем случае он обозначен как *plot*(*x*, *Pnx*). Полученный график можно редактировать, используя возможности меню *Edit* инструментальной панели в окне рисунка.

Полученный таким образом многоугольник распределения показан на рисунке 1.9.2.

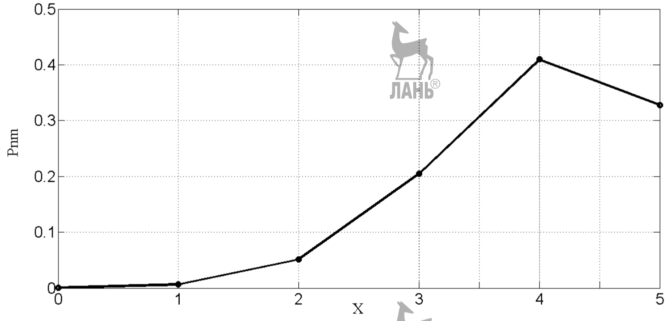

Рис. 1.9.2

Многоугольник распределения к примеру 1.9.1

Команда $Fx = cumsum(Pnx)$ возвращает накопленные (кумулятивные) суммы элементов вектора *Pnx* и присваивает их вектору *Fx*, элементами которого являются значения функции распределения $F(x)$. Заметим, что для вычисления функции биномиального распределения $F(x)$ можно вместо функции *cumsum* использовать функцию *binocdf*(x, n, p), где x — вектор-строка, которая содержит все возможные значения случайной величины X, n и p — параметры биномиального распределения.

График функии распределения изображает зависимость $F1x$ от $x1$ и строится аналогично тому, как строился график ряда распределения. Только учитывая ступенчатый характер графика функции распределения дискретной случайной величины, выбираем в графическом меню окна Worcspace лестничный тип графика — *stairs*($x1$, $F1x$). На рисунке 1.9.3 показан график функции распределения $F(x)$, заданной в примере 1.9.1 случайной величины X.

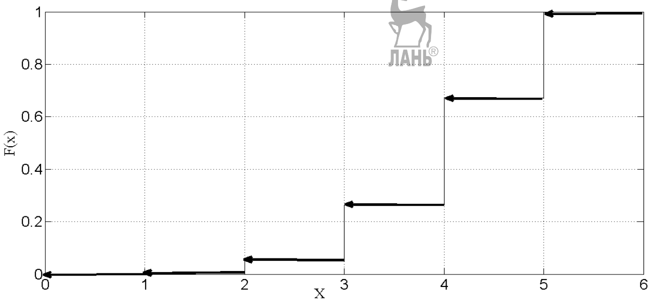

Рис. 1.9.3

Функция распределения в примере 1.9.1

П р и м е р 1.9.1. В условиях примера 1.9.1 найти вероятность того что число попаданий в мишень при пяти выстрелах будет не меньше трех.

Решение задачи в среде MATLAB показано на рисунке 1.9.4.

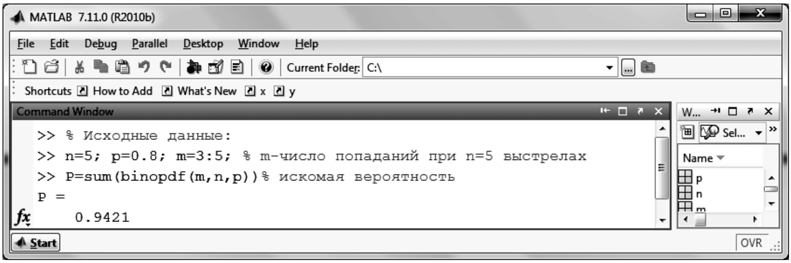

Рис. 1.9.4

Расчет вероятности $P(m \geq k)$

1.9.2. Закон Пуассона

При большом числе испытаний n и малой вероятности p формулой Бернулли пользоваться неудобно. В этом случае для вычисления вероятности того, что в n опытах событие произойдет m раз, используют формулу Пуассона.

$$P_m = \frac{\lambda^m}{m!} \cdot e^{-\lambda}, \quad (m = 0, 1, \ \ldots), \tag{1.9.4}$$

где λ — параметр закона Пуассона.

Рассмотрим дискретную случайную величину X, которая может принимать только целые, неотрицательные значения: 0, 1, 2, ..., m, ... Причем последовательность этих значений теоретически не ограничена. Говорят, что случайная величина X распределена по закону Пуассона, если вероятность того, что она примет определенное значение m, выражается формулой Пуассона (1.9.4).

Ряд распределения случайной величины X, распределенной по закону Пуассона представлен таблицей 1.9.2.

Таблица 1.9.2

Ряд распределения Пуассона

	0	1	2	...	m	...
P_m	$e^{-\lambda}$	$\frac{\lambda}{1!}\cdot e^{-\lambda}$	$\frac{\lambda^2}{2!}\cdot e^{-\lambda}$	...	$\frac{\lambda^m}{m!}\cdot e^{-\lambda}$	...

Математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона, равны параметру распределения λ:

$$M[X] = D[X] = \lambda.$$

Закон Пуассона является предельным для биномиального распределения, если одновременно устремлять число опытов n к бесконечности, а вероятность p появления события в каждом опыте к нулю, так, что их произведение сохраняет постоянное значение $np = \lambda$:

$$\lim_{n\to\infty} C_n^m p^m (1-p)^{n-m} = \frac{\lambda^m}{m!}\cdot e^{-\lambda}.$$

Это предельное свойство биномиального распределения находит практическое применение. Так, при $p \le 0{,}1$ и $np \le 10$ вычислить вероятность $P_{m,\,n}$ того, что событие A произойдет m раз в n опытах, можно по приближенной формуле:

$$P_n(m) \approx \frac{(np)^m}{m!} e^{-np}, \tag{1.9.5}$$

где $np = \lambda$ — параметр того закона Пуассона, которым приближенно заменяется биномиальное распределение.

П р и м е р 1.9.2. Автоматическая телефонная станция производит в среднем 2000 соединений в час. Вероятность ошибочного соединения $p = 0{,}001$. Найти вероятности того, что за час будет:

а) ровно три ошибочных соединения $m = 3$;

б) менее трех ошибочных соединений $m < 3$;

в) более трех ошибочных соединений $m > 3$.

Решение. По смыслу задачи рассматриваемые события (ошибочные соединения) независимы, поэтому имеет место биномиальное распределение числа ошибочных соединений. Число опытов велико ($n = 2000$), вероятность ошибочного соединения в одном опыте мала ($p = 0{,}001$). Поэтому биномиальное распределение можно приближенно заменить распределением Пуассона (1.9.4) с параметром $\lambda = np = 2000 \cdot 0{,}001 = 2$.

Вычисляем искомые вероятности:

а) вероятность ровно трех ошибочных соединений равна

$$P_{2000}(3) \approx \frac{(2)^3}{3!} e^{-2} \approx 0{,}1804;$$

б) вероятность менее трех ошибочных соединений:

$$P_{2000}(<3) = P_{2000}(0) + P_{2000}(1) + P_{2000}(2) = e^{-2} + \frac{2^1}{1!} \cdot e^{-2} + \frac{2^2}{2!} \cdot e^{-2} \approx 0{,}6768;$$

в) более трех ошибочных соединений:

$$P_{2000}(>3) = 1 - (P_{2000}(0) + P_{2000}(1) + P_{2000}(2) + P_{2000}(3)) \approx 0{,}1429.$$

Решение той же задачи в рабочей среде MATLAB показано на рисунке 1.9.5.

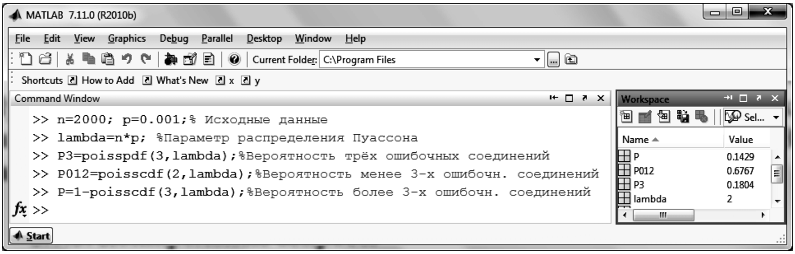

Рис. 1.9.5

Решение задачи примера 1.9.2

Результаты расчета отображены в окне Workspace рабочей области MATLAB.

Функция *poisspdf*(x, λ) вычисляет вероятность того, что случайная величина, распределенная по закону Пуассона с параметром λ, примет некоторое значение x. Функция *poisscdf*(x, λ) вычисляет вероятность того, что случайная величина примет значение не больше заданного значения x. Причем x может принимать целые положительные значения, включая ноль.

В заключение заметим, что применение локальной теоремы Муавра — Лапласа для решения этой задачи будет неоправданно, так как не выполняется одно из условий ее применения: $npq > 20$. В нашем примере $npq = 2000 \cdot 0{,}001 \cdot 0{,}999 \approx 2$.

1.9.3. Контрольные вопросы

1. Какое распределение называется биноминальным? Записать формулы для вычисления числовых характеристик дискретной случайной величины, распределенной по биномиальному закону.
2. Какое распределение называется распределением Пуассона?
3. В каком случае применяется закон распределения Пуассона и в чем состоит его особенность?
4. Какая формула используется для вычисления вероятности того, что в n испытаниях событие A появится ровно m раз при малом числе испытаний?
5. Какая формула используется для вычисления вероятности того, что в n испытаниях событие A появится ровно m раз при большом числе испытаний и малой вероятности p?

1.9.4. Задачи для самостоятельного решения

1. Определить вероятность того, что в семье, имеющей 5 детей, будет не больше трех девочек. Вероятности рождения мальчика и девочки предполагаются одинаковыми.
2. Завод выпускает 96% изделий первого сорта и 4% изделий второго сорта. Наугад выбирают 1000 изделий. Пусть X — число изделий первого сорта в данной выборке. Найти закон распределения, математическое ожидание и дисперсию случайной величины X.

1.10. НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Определение 1.10.1. Случайная величина называется непрерывной, если ее функция распределения непрерывна на всей числовой прямой и дифференцируема при всех x, за исключением, быть может, отдельных значений.

1.10.1.
Интегральный закон распределения

Непрерывную случайную величину, так же как и дискретную, можно задать с помощью функции распределения. Определяется она так же, как и для дискретной случайной величины, формулой (1.7.2).

Все свойства функции распределения, перечисленные в п. 1.7.3, справедливы и для функций распределения непрерывных случайных величин.

На рисунке 1.10.1 приведен характерный вид функции распределения непрерывной случайной величины.

Как видим, функция распределения непрерывной случайной величины имеет непрерывный характер.

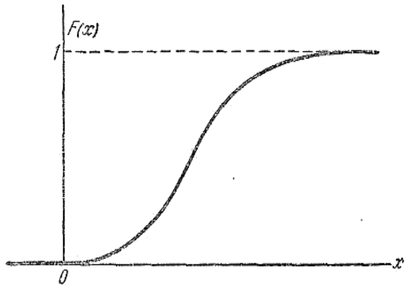

Рис. 1.10.1
Характерный вид функции распределения непрерывной случайной величины

1.10.2.
Плотность распределения

Плотность распределения, так же как и ряд распределения и функция распределения, есть одна из форм закона распределения, но, в отличие от них, плотность распределения используется для описания только непрерывных случайных величин.

Плотность распределения $f(x)$ характеризует плотность, с которой распределяются значения случайной величины в окрестности данной точки x. Пример графического изображения плотности распределения показан на рисунке 1.10.2.

Рассмотрим непрерывную случайную величину X с плотностью распределения $f(X)$ и элементарный участок dx, примыкающий к точке x (рис. 1.10.2).

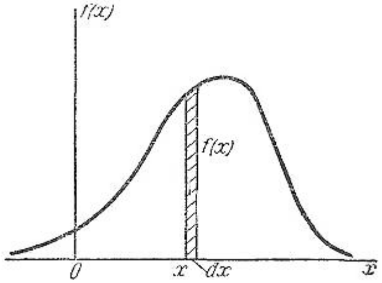

Рис. 1.10.2

Кривая плотности распределения

Вероятность попадания случайной величины X на этот элементарный участок (с точностью до бесконечно малых высшего порядка) равна $f(x)dx$. Величина $f(x)dx$ называется элементом вероятности. Геометрически это есть площадь элементарного прямоугольника, опирающегося на элементарный отрезок dx.

Выразим вероятность попадания непрерывной случайной величины X на отрезок (α, β) через плотность распределения.

Очевидно, она равна сумме элементов вероятности на всем этом участке, т. е. интегралу:

$$P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx. \tag{1.10.1}$$

Геометрически вероятность попадания величины X на участок (α, β) равна площади кривой распределения, опирающейся на этот участок (рис. 1.10.3).

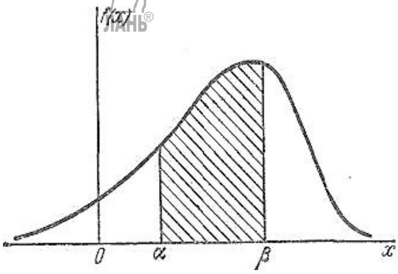

Рис. 1.10.3

Геометрическое представление вероятности попадания случайной величины на заданный интервал

Формула

$$f(x) = F'(x) \qquad (1.10.2)$$

выражает плотность распределения через функцию распределения.

Зададимся обратной задачей: выразить функцию распределения через плотность.

По определению

$$F(x) = P(X < x) = P(-\infty < X < x),$$

Следовательно, учитывая, что $P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx,$ имеем

$$F(x) = \int_{-\infty}^{x} f(x)dx. \qquad (1.10.3)$$

Таким образом, геометрически $F(x)$ есть не что иное, как площадь под кривой плотности распределения, лежащая левее точки x.

Укажем основные свойства плотности распределения.

1. Плотность распределения есть неотрицательная функция:

$$f(x) > 0.$$

Это свойство непосредственно вытекает из того, что функция распределения $F(x)$ есть неубывающая функция.

2. Интеграл в бесконечных пределах от плотности распределения равен единице (условие нормировки):

$$\int_{-\infty}^{\infty} f(x)dx = 1. \qquad (1.10.4)$$

Это следует из формулы (1.10.3) и из того, что $F(+\infty) = 1$.

Геометрически основные свойства плотности распределения означают, что:

1) вся кривая распределения лежит не ниже оси абсцисс;

2) полная площадь, ограниченная кривой плотности распределения и осью абсцисс, равна единице.

Функция распределения $F(x)$, как всякая вероятность, есть величина безразмерная. Размерность плотности распределения $f(x)$, как следует из ее определения, обратна размерности случайной величины.

Пример 1.10.1. Функция распределения непрерывной случайной величины X задана выражением

$$F(x)=\begin{cases}0 & \text{при } x\le 0;\\ ax^2 & \text{при } 0<x\le 1;\\ 1 & \text{при } x>1.\end{cases}$$

Требуется:

а) найти коэффициент a;

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

в) найти вероятность попадания величины X на участок от 0,25 до 0,5.

Р е ш е н и е.

1. Так как функция распределения величины X непрерывна, то $F(x) = a\mathrm{x}^2 = 1$, когда $x = 1$, следовательно, $a = 1$.

2. Плотность распределения величины X, согласно (1.10.2), выражается формулой

$$f(x)=\begin{cases}0 & \text{при } x\le 0;\\ 2x & \text{при } 0<x\le 1;\\ 0 & \text{при } x>1.\end{cases}$$

3. По формуле (1.7.3) имеем

$$P(0{,}25 < X < 0{,}5) = F(0{,}5) - F(0{,}25) = 0{,}52 - 0{,}252 = 0{,}1875.$$

Решим ту же задачу, используя возможности пакета символьной математики MuPAD системы MATLAB.

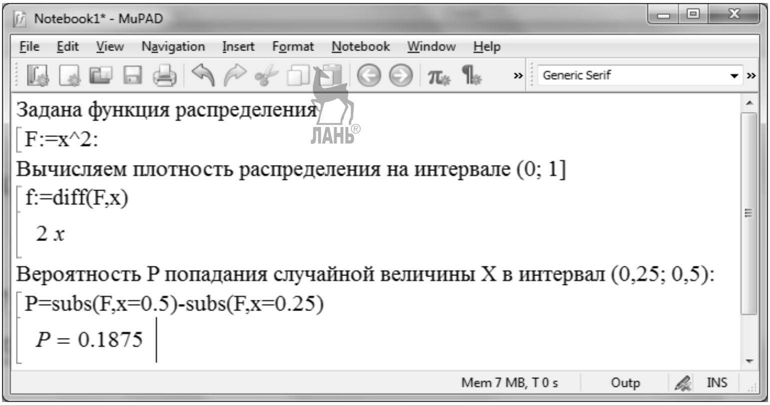

Рис. 1.10.4

Нахождение плотности распределения в символьной форме

Откроем рабочую тетрадь Notebook-MuPAD, набрав в командной строке MATLAB команду mupad. Вычисление неизвестного коэффициента функции рапределения выполняется с помощью уже известных арифметических операторов MATLAB.

Интерес представляет та часть задачи, которая связана с символьными преобразованиями.

Процедура получения символьного выражения искомой плотности распределения представленная на рисунке 1.10.4.

Более подробно возможности пакета MuPAD можно найти в работе [2].

П р и м е р 1.10.2. Случайная величина *X* подчинена закону распределения с плотностью:

$$f(x)=\begin{cases} a\cdot\cos x & \text{при } -\pi/2\le x\le\pi/2; \\ 0 & \text{при } x<\pi/2 \text{ или } x>\pi/2. \end{cases}$$

Требуется найти:

1) коэффициент *a*;

2) функцию распределения *F*(*x*);

3) вероятность попадания величины *X* на участок от 0 до π/4.

Р е ш е н и е.

1. Для определения коэффициента *a* воспользуемся свойством плотности распределения:

$$\int_{-\infty}^{\infty} f(x)dx=\int_{-\pi/2}^{\pi/2} a\cdot\cos x\cdot dx=2a=1,$$

откуда следует $a=\dfrac{1}{2}.$

2. Получим выражение функции распределения:

при $x<\dfrac{\pi}{2}$:

$$F(x)=\int_{-\infty}^{x} 0\cdot dx=0;$$

при $-\dfrac{\pi}{2}\le x\le\dfrac{\pi}{2}$:

$$F(x)=\int_{-\pi/2}^{x}\frac{1}{2}\cos x\cdot dx=\frac{1}{2}\left(\sin x-\sin\left(-\frac{\pi}{2}\right)\right)=\frac{1}{2}(\sin x+1);$$

при $x > \dfrac{\pi}{2}$:

$$F(x) = \int_{-\infty}^{x} 0 \cdot dx + \int_{-\pi/2}^{\pi/2} \frac{1}{2}\cos x \cdot dx + \int_{\pi/2}^{x} 0 \cdot dx = \frac{1}{2}\left(\sin\frac{\pi}{2} - \sin\left(-\frac{\pi}{2}\right)\right) = 1.$$

Полученный результат можно записать в виде:

$$F(x) = \begin{cases} 0 & \text{при } x < -\dfrac{\pi}{2}; \\ \dfrac{1}{2}(\sin x + 1) & \text{при } -\dfrac{\pi}{2} \le x \le \dfrac{\pi}{2}; \\ 1 & \text{при } x > \dfrac{\pi}{2}. \end{cases}$$

3. $P\left(0 < x < \dfrac{\pi}{4}\right) = F\left(\dfrac{\pi}{4}\right) - F(0) = \dfrac{1}{2}\left(\left(\sin\dfrac{\pi}{4} + 1\right) - (\sin 0 + 1)\right) = \dfrac{\sqrt{2}}{4}.$

Получение символьного выражения функции распределения с помощью пакета MuPAD показано на рисунке 1.10.5.

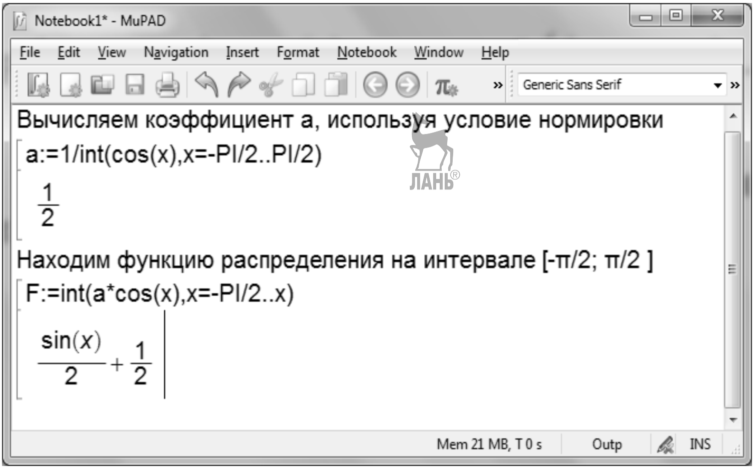

Рис. 1.10.5

Символьное вычисление функции распределения

1.11. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

1.11.1. Характеристики положения

Математическое ожидание непрерывной случайной величины определяется выражением

$$M[X] = \int_{-\infty}^{\infty} x \cdot f(x)dx, \tag{1.11.1}$$

где $f(x)$ — плотность распределения случайной величины x.

Эта формула получается из формулы (1.8.1) математического ожидания дискретной случайной величины, если в ней заменить отдельные значения x_i непрерывно изменяющейся переменной x, соответствующие вероятности p_i — элементом вероятности $f(x)dx$, конечную сумму — интегралом.

В дальнейшем мы будем пользоваться таким способом распространения формул, выведенных для дискретных случайных величин, на случай непрерывных величин.

В ряде случаев математическое ожидание $M[X]$ удобнее обозначать как m_X. Эти обозначения для математического ожидания будут в дальнейшем применяться параллельно в зависимости от удобства той или иной записи формул.

Модой непрерывной случайной величины является то значение случайной величины, при котором плотность вероятности максимальна. Условимся, как и для дискретной случайной величины, обозначать моду буквой $\mathcal{M}$.

Медианой случайной величины X называется такое ее значение **Me**, для которого $P(X < \mathbf{Me}) = P(X > \mathbf{Me})$, т. е. одинаково вероятно, окажется ли случайная величина меньше или больше **Me**.

Геометрическая медиана — это абсцисса точки, в которой площадь, ограниченная кривой распределения, делится пополам.

1.11.2. Характеристики рассеивания. Моменты распределения

Дисперсия непрерывной случайной величины определяется выражением

$$D[X]=\int_{-\infty}^{\infty}(x-m_x)^2dx. \qquad (1.11.2)$$

Среднеквадратическое отклонение непрерывной случайной величины определяется так же, как и для дискретной случайной величины (п. 1.8.2).

Вычисление числовых характеристик непрерывных случайных величин сводится к вычислению в общем случае несобственных интегралов. Синтаксис команд вычисления определенных интегралов в командном окне MATLAB и в рабочей тетради Notebook-MuPAD и соответствующие примеры приводятся в работе [2].

Моменты распределения.

Чаще всего применяются на практике моменты двух видов: начальные и центральные.

Определение 1.11.1. Начальным моментом k-го порядка непрерывной случайной величины X называется интеграл вида

$$\alpha_k=\int_{-\infty}^{\infty}x^k f(x)dx. \qquad (1.11.3)$$

Нетрудно убедиться, что математическое ожидание представляет собой не что иное, как начальный момент первого порядка случайной величины X: $\alpha_1[X]=M[X]$.

В п. 1.8.2 дается общее определение начального момента k-го порядка, справедливое и для непрерывных случайных величин: $\alpha_k[X]=M[X^k]$. То есть начальным моментом k-го порядка случайной величины X является математическое ожидание k-й степени этой случайной величины.

Для определения центрального момента используем понятие центрированной случайной величины, рассмотренное в п. 1.8.2: $\dot{\mathrm{X}}=X-m_X$. Центрирование случайной величины равносильно переносу начала координат в точку, абсцисса которой равна математическому ожиданию, поэтому математическое ожидание центрированной случайной величины равно нулю: $M[\dot{\mathrm{X}}]=0$.

Моменты центрированной случайной величины носят название центральных моментов.

Для непрерывной случайной величины k-й центральный момент (центральный момент k-го порядка) выражается интегралом

$$\mu_k=\int_{-\infty}^{\infty}(x-m_x)^k f(x)dx. \qquad (1.11.4)$$

Очевидно, для любой случайной величины центральный момент первого порядка равен нулю: $\mu_1 = M[\dot{X}] = M[X - m_x] = 0$, так как математическое ожидание центрированной случайной величины всегда равно нулю.

Для более подробного описания закона распределения с помощью числовых характеристик применяются моменты высших порядков.

Третий центральный момент служит для характеристики асимметрии распределения. Если распределение симметрично относительно математического, то все моменты нечетного порядка (если они существуют) равны нулю. Действительно, интеграл

$$\mu_k = \int_{-\infty}^{\infty} (x - m_x)^k f(x)dx$$

при нечетном значении k равен нулю как интеграл в симметричных пределах от нечетной функции.

Естественно поэтому в качестве характеристики асимметрии распределения выбрать какой-либо из нечетных центральных моментов. Простейший из них есть третий центральный момент. Он имеет размерность куба случайной величины. Чтобы получить безразмерную характеристику, третий момент μ_3 делят на куб среднего квадратичного отклонения.

Полученная величина носит название «коэффициента асимметрии» или просто «асимметрии»; мы обозначим ее Sk:

$$Sk = \frac{\mu_3}{\sigma^3}. \quad (1.11.5)$$

На рисунке 1.11.1 показаны два асимметричных распределения.

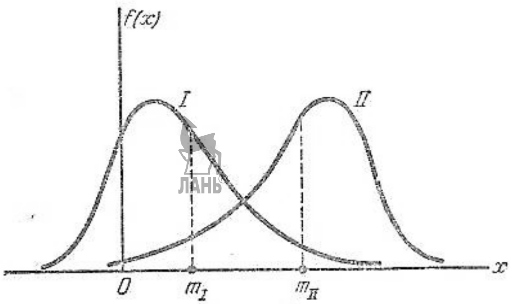

Рис. 1.11.1

Ассиметричные распределения

Одно них (кривая *I*) имеет положительную асимметрию ($Sk > 0$); другое (кривая *II*) — отрицательную ($Sk < 0$).

Четвертый центральный момент μ_4 служит для характеристики так называемой островершинности или плосковершинности распределения.

Эти свойства распределения описываются с помощью эксцесса. Эксцессом случайной величины X называется величина

$$\mathrm{Ex} = \frac{\mu_4}{\sigma^4} - 3. \qquad (1.11.6)$$

Число 3 вычитается из отношения $\frac{\mu_4}{\sigma^4}$, потому что для весьма важного и широко распространенного в природе нормального закона распределения (с которым мы подробно познакомимся ниже) — $\frac{\mu_4}{\sigma^4} = 3$.

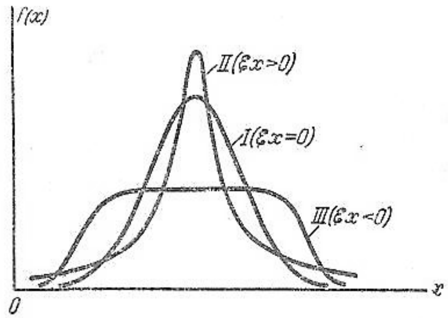

Рис. 1.11.2

Связь эксцесса и формы кривой распределения

Таким образом, для нормального распределения эксцесс равен нулю. Кривые, более островершинные по сравнению с нормальной кривой распределения, обладают положительным эксцессом; более плосковершинные кривые — отрицательным эксцессом.

На рисунке 1.11.2 изображены: нормальное распределение (кривая *I*), распределение с положительным эксцессом (кривая *II*) и распределение с отрицательным эксцессом (кривая *III*).

Часто числовыми характеристиками пользуются для приближенной замены одного распределения другим, причем обычно стремятся произвести эту замену так, чтобы сохранились неизменными несколько важнейших моментов.

П р и м е р 1.11.1. Непрерывная случайная величина X подчинена закону распределения с плотностью: $f(x) = A \cdot e^{-|x|}$.

Найти коэффициент A. Определить математическое ожидание, дисперсию, с. к. о., асимметрию, эксцесс случайной величины X.

Р е ш е н и е. Для определения A воспользуемся свойством плотности распределения (условие нормировки):

$$\int_{-\infty}^{\infty} f(x)dx = 2A\int_{0}^{\infty} e^{-x}dx = 2A = 1.$$

Отсюда $A = \frac{1}{2}$.

Математическое ожидание $m_x = \frac{1}{2}\int_{-\infty}^{\infty} x \cdot e^{-|x|}dx = 0$, так как функция $x \cdot e^{-|x|}$ нечетная.

Дисперсия и с. к. о. равны соответственно:

$$D_x = 2 \cdot \frac{1}{2}\int_{0}^{\infty} x^2 e^{-x}dx = 2; \quad \sigma_x = \sqrt{D_x} = \sqrt{2}.$$

Так как распределение симметрично, то коэффициент асимметрии $Sk = 0$.

Для вычисления эсцесса находим μ_4:

$$\mu_4 = 2 \cdot \frac{1}{2}\int_{0}^{\infty} x^4 e^{-x}dx = 24.$$

Отсюда $\mathrm{Ex} = \frac{\mu_4}{\sigma_x^4} - 3 = 3$.

П р и м е р 1.11.2. Случайная величина X подчинена закону распределения, плотность которого задана графически на рисунке 1.11.3.

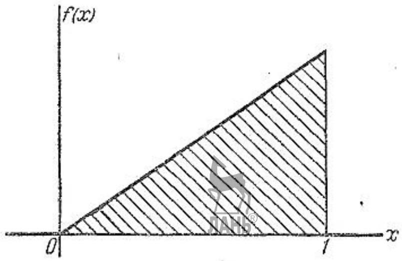

Рис. 1.11.3

График плотности распределения к примеру 1.11.2

Написать выражение плотности распределения. Найти математическое ожидание, дисперсию, с. к. о. и асимметрию распределения.

Р е ш е н и е. Выражение плотности распределения имеет вид:

$$f(x)=\begin{cases} a\cdot x & \text{при}\ \ 0<x<1; \\ 0 & \text{при}\ \ x<0\ \ \text{или}\ \ x>0. \end{cases}$$

Пользуясь свойством плотности распределения (условие нормировки), находим: $a = 2$.

Математическое ожидание случайной величины X:

$$m_x = 2\int_0^1 2x^2 dx = \frac{2}{3}.$$

Дисперсию найдем через второй начальный момент:

$$\alpha_2 = 2\int_0^1 x^2 dx = \frac{1}{2}; \qquad D_x = \alpha_2 - m_x^2 = \frac{1}{18}.$$

Отсюда

$$\sigma_x = \frac{1}{3\sqrt{2}}.$$

Для нахождения коэффициента асимметрии найдем μ_3:

$$\mu_3 = 2\int_0^1 \left(x-\frac{2}{3}\right)^3 x dx = -\frac{1}{135}.$$

Отсюда

$$Sk = \frac{\mu_3}{\sigma_x^3} = -\frac{2}{5}\sqrt{2}.$$

Решение этой задачи в MATLAB осуществляется с помощью функций интегрирования *int*, синтаксис которой приводится в работе [2].

1.11.3.
Контрольные вопросы

1. Сформулировать определение непрерывной случайной величины.
2. Что такое плотность распределения вероятностей?
3. Каким свойством обладает плотность распределения вероятностей?
4. Какими свойствами обладает функция распределения непрерывной случайной величины?

5. Как найти интегральную функцию, зная плотность распределения и наоборот?

6. Перечислить свойства интегральной функции распределения.

7. Дать определения числовым характеристикам непрерывной случайной величины.

1.12. ЗАКОНЫ РАСПРЕДЕЛЕНИЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Рассмотрим типовые законы распределения для непрерывных случайных величин. Ограниченный объем данного учебного пособия не позволяет рассмотреть все законы распределения, находящие практическое применение, поэтому остановимся на двух из них: равномерном и нормальном распределениях.

1.12.1. Равномерное распределение

Определение 1.12.1. Распределение непрерывной случайной величины X называется равномерным на отрезке $[\alpha, \beta]$, если ее плотность распределения на этом отрезке постоянна и имеет вид

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha} & \text{при } \alpha \le x \le \beta;\\[2mm] 0 & \text{при } x<\alpha \text{ или } x>\beta,\end{cases} \tag{1.12.1}$$

где α и β — параметры равномерного распределения.

Величина плотности распределения в формуле (1.12.1) определяется условием нормировки (1.10.4).

Функция распределения выражается площадью под плотностью распределения, лежащей левее точки x. Следовательно, в данном случае

$$F(x)=\begin{cases}0 & \text{при } x<\alpha;\\[1mm] \dfrac{x-\propto}{\beta-\alpha} & \text{при } \alpha<x<\beta;\\[2mm] 1 & \text{при } x>\beta.\end{cases} \tag{1.12.2}$$

Математическое ожидание равномерно распределенной случайной величины X равно

$$m_x=\int_{\alpha}^{\beta}\frac{1}{\beta-\alpha}\cdot x dx=\frac{\alpha+\beta}{2}. \tag{1.12.3}$$

В силу симметричности функции плотности распределения относительно математического ожидания медиана случайной величины X также равна $\frac{\alpha+\beta}{2}$.

Моды закон равномерной плотности не имеет.

Дисперсия равномерно распределенной случайной величины X:

$$D_x = \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta}\left(x-\frac{\alpha+\beta}{2}\right)^2 dx = \frac{(\beta-\alpha)^2}{12}, \qquad (1.12.4)$$

откуда среднее квадратичное отклонение:

$$\sigma_x = \sqrt{D_x} = \frac{\beta-\alpha}{2\sqrt{3}}. \qquad (1.12.5)$$

В силу симметричности функции плотности распределения относительно математического ожидания асимметрия закона распределения равна нулю:

$$Sk = \frac{\mu_3}{\sigma_x^3} = 0.$$

Для определения эксцесса находим четвертый центральный момент:

$$\mu_4 = \frac{1}{\beta-\alpha}\int_{\alpha}^{\beta}\left(x-\frac{\alpha+\beta}{2}\right)^4 dx = \frac{(\beta-\alpha)^4}{80}.$$

Отсюда

$$\mathrm{Ex} = \frac{\mu_4}{\sigma_x^4} - 3 = -1{,}2.$$

Вероятность попадания случайной величины X, распределенной по равномерному закону на участок $[a,\ b]$, представляющий собой часть отрезка $[\alpha,\ \beta]$, равна:

$$P(a < X < B) = \int_a^b \frac{1}{\beta-\alpha}dx = \frac{b-a}{\beta-\alpha}. \qquad (1.12.6)$$

П р и м е р 1.12.1. Поезда метрополитена идут с интервалом 2 мин. Пассажир выходит на платформу в некоторый момент времени. Время T, в течение которого ему придется ждать поезд, представляет собой случайную величину, распределенную с равномерной плотностью на интервале [0; 2] мин. Найти вероятность того, что пришедшему на остановку пассажиру придется ожидать поезд не более 1 мин.

Р е ш е н и е. Плотность распределения времени t ожидания поезда:

$$f(t)=\begin{cases}\dfrac{1}{2} & \text{при} \quad 0 \le x \le 2; \\ 0 & \text{при} \quad x<0 \quad \text{или} \quad x>2.\end{cases}$$

Тогда

$$P(0 \le T \le 1)=\int_0^1 \frac{1}{2}dx=\frac{1}{2}=0,5 \text{ мин.}$$

Числовые характеристики этого закона распределения определяются по формулам, приведенным выше для равномерного закона распределения вероятности:

— среднее время ожидания поезда на платформе $m_T = 1$ мин;

— с. к. о. времени ожидания $\sigma_T \approx 0,57$ мин.

Перечислим некоторые функции MATLAB для работы с равномерным распределением непрерывной случайной величины X.

1. *unifpdf*(x, a, b) возвращает численное значение равномерной плотности распределения в точке x.

Замечание. Здесь и далее: x — значение случайной величины; a и b — левая и правая границы интервала, в котором плотность распределени не равна 0. В общем случае x, a и b могут быть векторами, матрицами или многомерными массивами.

2. *unifcdf*(x, a, b) возвращает численное значение функции распределения в точке x.

3. [*mx*, *Dx*] = *unifstat*(a, b) возвращает математическое ожидание *mx* и дисперсию *Dx* случайной величины X, равномерно распределенной на интервале $[a, b]$.

1.12.2.
Нормальное распределение

Главная особенность, выделяющая нормальный закон среди других законов распределения, состоит в том, что он является предельным законом, к которому приближаются другие законы распределения при весьма часто встречающихся типичных условиях.

Вместе с тем, закон распределения суммы случайных величин, каждая из которых может иметь различные распределения, приближается к нормальному распределению с ростом числа слагаемых. Точная формулировка этих утверждений является содержанием группы теорем, называемых центральной предельной теоремой. Эти научно установленные факты широко используются, в частности при решении задач математической статистики.

Определение 1.12.2. Распределение непрерывной случайной величины X называется нормальным, если ее плотность распределения имеет вид

$$f(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{(x-m_X)^2}{2\sigma_X^2}}, \qquad (1.12.7)$$

где m_X — математическое ожидание, σ_X — среднеквадратическое отклонение случайной величины X.

Величины m_X и σ_X являются парметрами нормального распределения.

Математическое ожидание m_X является центром рассеивания случайной величины X. Если изменять m_X, то кривая нормального распределения будет смещаться вдоль оси абсцисс, не изменяя своей формы (рис. 1.12.1).

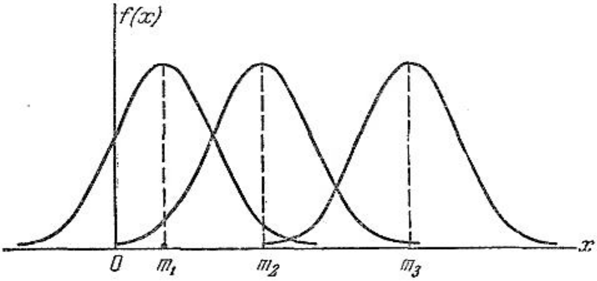

Рис. 1.12.1

Влияние параметра m_X на кривую нормального распределения

Среднее квадратичное отклонение σ_X характеризует форму кривой распределения. Это есть характеристика рассеивания. Так как площадь под кривой распределения, согласно условию нормировки, всегда должна оставаться равной единице, то при увеличении σ_X кривая распределения становится более плоской, растягиваясь вдоль оси абсцисс. Напротив, при уменьшении σ_X кривая распределения вытягивается вверх, одновременно сжимаясь с боков, и становится более иглообразной. На рисунке 1.12.2 показаны три нормальные кривые (*I*, *II*, *III*) при $m_X = 0$; из них кривая *I* соответствует самому большому σ_X, а кривая *III* — самому малому значению σ_X.

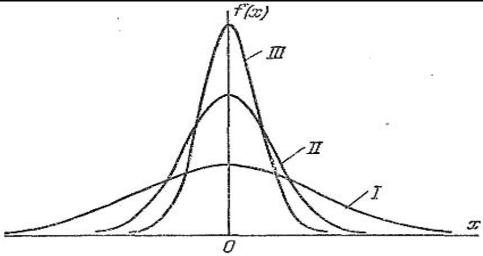

Рис. 1.12.2

Влияние параметра σ_X на форму кривой нормального распределения

Асимметрия нормального распределения равна нулю ($Sk = 0$), так как кривая плотности нормального распределения симметрична относительно m_X.

Эксцесс также равен нулю ($Ex = 0$). Об этом уже говорилось в п. 1.11.2.

Найдем функцию распределения $F(x)$ случайной величины X, распределенной по нормальному закону с параметрами m_x и σ_x:

$$F(x) = \int_{-\infty}^{x} f(x)dx = \frac{1}{\sigma_x\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-m_x)^2}{2\sigma_x^2}} dx. \qquad (1.12.8)$$

Сделаем замену переменной:

$$\frac{x - m_x}{\sigma_x} = u.$$

Тогда

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-m_x}{\sigma_x}} e^{-\frac{u^2}{2}} dt. \qquad (1.12.9)$$

Этот интеграл не выражается через элементарные функции, но его можно вычислить приближенно по таблицам, построенным для специальной функции вида:

$$N(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{x^2}{2}} dx. \qquad (1.12.10)$$

Функция (1.12.10) называется стандартной (нормированной) функцией нормального распределения. Она представляет собой интегральную функцию распределения для нормально распределенной случайной величины с параметрами: $m_X = 0$, $\sigma_X = 1$. Таблица значений функции стандартного нормального распределения (1.12.10) дается в приложении 1.

Перечислим свойства функции стандартного нормального распеделения $N(u)$.

1. $N(0) = 0{,}5$.
2. $N(-\infty) = 0$.
3. $N(\infty) = 1$.
4. $N(u)$ — неубывающая функция.
5. $N(-u) = 1 - N(u)$.

Интегральная функция $F(x)$ нормально распределенной случайной величины X с заданными параметрами m_x и σ_x выражается через функцию $N(u)$ стандартного нормального распределения следующим образом:

$$F(x) = N\left(\frac{x - m_x}{\sigma_x}\right), \qquad (1.12.11)$$

где $\dfrac{x - m_x}{\sigma_x} = u$ — нормированное отклонение значения случайной величины от математического ожидания m_x.

Вероятность попадания на участок от a до b, случайной величины X, распределенной по нормальному закону с любыми заданными параметрами m_x, σ_x, с учетом выражений (1.7.3), (1.12.11), можно вычислить по формуле

$$P(a < X < b) = N\left(\frac{b - m_x}{\sigma_x}\right) - N\left(\frac{a - m_x}{\sigma_x}\right). \qquad (1.12.12)$$

По формуле (1.12.12) можно определить, что нормально распределенная случайная величина X принимает значения, находящиеся в пределах $\pm 3\sigma_X$ относительно математического ожидания, с вероятностью:

$$P(|X - m_X|) \le 3\sigma_X = 0{,}9973. \qquad (1.12.13)$$

На основании полученного результата формулируется «правило трех сигм»: значения нормальной случайной величины практически достоверно отклоняются от ее математического ожидания не более чем на 3σ.

Замечание. Функция $N(u) = \dfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{u} e^{-\frac{x^2}{2}} dx$ является лишь одной из форм так называемого интеграла вероятностей. Для решения прикладных задач, в частности для вычисления вероятности попадания случайной величины в заданный интервал, симметричный отнсительно начала координат, очень часто пользуются и другими формами интеграла вероятностей, например функцией Лапласа (п. 1.6.2). Выбор функции $N(u)$ в данной работе продиктован возможностями MATLAB.

П р и м е р 1.12.2. Случайная величина X, распределенная по нормальному закону, представляет собой ошибку измерения некоторого расстояния. При измерении допускается систематическая ошибка в сторону завышения на 1,2 м; среднеквадратическое отклонение ошибки измерения равно 0,8 м. Найти вероятность того, что отклонение измеренного значения от истинного значения не превзойдет по абсолютной величине 1,6 м.

Р е ш е н и е. Согласно условию, ошибка измерения есть случайная величина X, подчиненная нормальному закону с параметрами $m_x = 1{,}2$ и $\sigma_X = 0{,}8$. Вероятность попадания этой величины на участок от $a = -1{,}6$ до $b = +\,1{,}6$ вычисляется с помощью функции стандартного нормального распределения:

$$P(-1{,}6 < X < 1{,}6) = N\left(\frac{1{,}6-1{,}2}{0{,}8}\right) - N\left(\frac{-1{,}6-1{,}2}{0{,}8}\right) =$$

$$= N(0{,}5) - N(-3{,}5) = N(0{,}5) - (1 - N(-3{,}5)) \approx 0{,}6915 - 0{,}0002 \approx 0{,}6913.$$

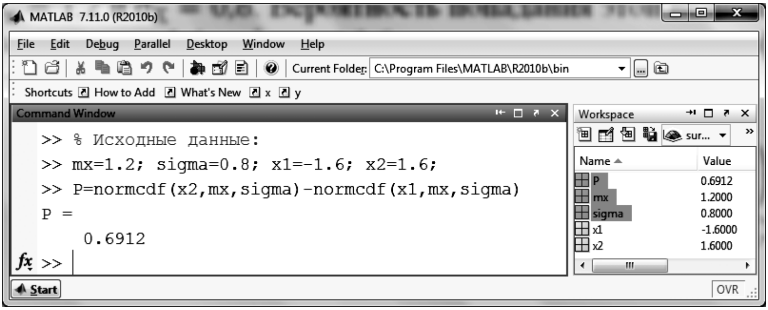

Рис. 1.12.3

Вычисление вероятности P попадания нормально распределенной случайной величины в заданный интервал

Перечислим некоторые функции MATLAB для работы с нормальным распределением.

1. *normpdf*(x, mx, $sigma$) возвращает численное значение плотности распределения в точке x.

Замечание. Здесь и далее *x* — значение случайной величины; *mx* и *sigma* — математическое ожидание и среднеквадратическое отклонение случайной величины *X*. В общем случае формальные параметры *x*, *mx* и *sigma* могут быть векторами, матрицами или многомерными массивами.

2. *normcdf*(*x*, *mx*, *sigma*) возвращает численное значение функции распределения в точке *x*.

3. *norminv* (*P*, *mx*, *sigma*) возвращает значение *x*, соответствующее заданному значению *P* функции распределения, т. е. вычисляет значения функции, обратной по отношению к заданной функции распределения.

Решим задачу, сформулированную в примере 1.12.2, с помощью MATLAB (рис. 1.12.3).

1.12.3.
Задачи для самостоятельного решения

1. Плотность $f(x)$ равномерного распределения в интервале $[a, b]$ имеет некоторое постоянное значение, равное C; вне этого интервала $f(x) = 0$. Найти значение постоянного параметра C.

2. Закон равномерного распределения задан плотностью вероятности $f(x) = 1/(b - a)$ в интервале $[a, b]$; вне этого интервала $f(x) = 0$. Найти функцию распределения F(x).

3. Найти математическое ожидание случайной величин X, распределенной равномерно в интервале [2, 8].

4. Найти дисперсию и стандартное отклонение случайной величины X, распределенной равномерно в интервале [2, 8].

5. Равномерно распределенная случайная величина X задана плотностью распределения $f(x) = ½$ в интервале $[a - 1, a + 1]$; вне этого интервала $f(x) = 0$. Найти математическое ожидание и дисперсию X.

1.13.
МНОГОМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Часто результат опыта описывается не одной, а двумя или более случайными величинами, которые каким-то образом влияют друг на друга.

Систему таких случайных величин можно рассматривать, в общем случае, как многомерную случайную величину.

Свойства многомерной случайной величины не исчерпываются свойствами отдельных величин, ее составляющих. Помимо этого, они включают в себя также взаимные связи между одномерными случайными величинами, образующими систему.

В данном разделе в основном будут рассмотрены двумерные случайные величины и их характеристики.

Мы будем рассматривать как полные, исчерпывающие вероятностные характеристики — законы распределения, так и их числовые характеристики.

При рассмотрении вопросов, связанных с системами случайных величин, удобно пользоваться геометрической интерпретацией системы. Например, систему двух случайных величин (X, Y) можно изображать случайной точкой или случайным вектором на плоскости с координатами x и y. Система трех случайных величин (X, Y, Z) может быть изображена, соответственно, случайной точкой (случайным вектором) в трехмерном пространстве и т. д.

Одним из достоинстсв системы MATLAB является то, что все величины, с которыми она оперирует, можно рассматривать в общем случае как многомерные векторы. Это позволяет без особого труда решать многомерные задачи. В их числе вычисление числовых характеристик и законов распределения многомерных случайных величин, а также нахождение многомерных функций от многомерных случайных величин и их законов распределения. Эта тема достаточно обширна и требует отдельного внимания.

В данной работе мы ограничимся рассмотрением нескольких частных случаев, имеющих широкое практическое применение. Начнем с нахождения законов распределения непрерывной двумерной случайной величины и ее числовых характеристик.

1.13.1. Законы распределения непрерывной двумерной случайной величины

Определение 1.13.1. Функцией распределения системы двух случайных величин (X, Y) называется вероятность совместного выполнения двух неравенств $X < x$ и $Y < y$:

$$F(x, y) = P((X < x)(Y < y)). \tag{1.13.1}$$

Геометрически функция распределения системы двух случайных величин (X, Y) есть вероятность попадания случайной точки (X, Y) в бесконечный квадрат с вершиной в точке (x, y), лежащий левее и ниже ее (рис. 1.13.1).

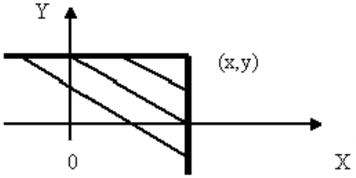

Рис. 1.13.1

Геометрический смысл двумерной функции распределения

Сформулируем свойства функции распределения системы двух случайных величин.

1. Функция распределения $F(x,\ y)$ есть неубывающая функция своих аргументов: при $x_2 > x_1 F(x_2, y) \geq F(x_1, y)$; при $y_2 > y_1$ $F(x, y_2) \geq F(x, y_1)$.

2. Везде на $-\infty$ функция распределения равна нулю:

$$F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0.$$

3. При одном из аргументов, стремящемся к $+\infty$, функция распределения системы превращается в функцию распределения случайной величины, соответствующей другому аргументу:

$$\lim_{y\to\infty} F(x, y) = F_1(x), \ \ \lim_{x\to\infty} F(x, y) = F_2(y), \qquad (1.13.2)$$

где $F_1(x)$; $F_2(y)$ — соответственно функции распределения случайных величин X и Y.

4. Если оба аргумента стремятся к $+\infty$, функция распределения системы равна единице: $F(+\infty, +\infty) = 1$.

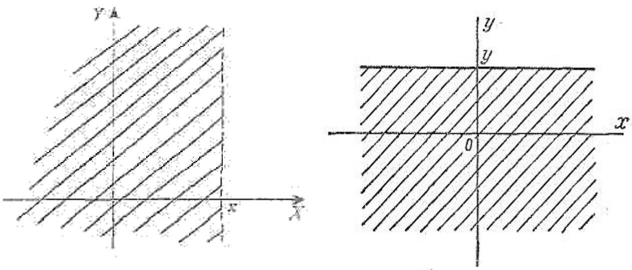

Рис. 1.13.2

Геометрический смысл функций распределения случайных величин, входящих в систему: $F_1(x)$ (слева) и $F_2(y)$ (справа)

Функция распределения $F_1(x)$ представляет собой вероятность попадания случайной точки в полуплоскость, ограниченную справа абсциссой *x*. Соответственно, $F_2(x)$ представляет собой вероятность попадания случайной точки в полуплоскость, ограниченную сверху ординатой *y*. Наглядно это показано на рисунке 1.13.2.

Определение 1.13.2. Плотностью распределения $f(x, y)$ непрерывной двумерной случайной величины называется вторая смешанная частная производная от двумерной функции распределения:

$$f(x, y) = F''_{xy}(x, y). \tag{1.13.3}$$

График двумерной плотности распределения $f(x, y)$ называется поверхностью распределения.

Рассматривая в п. 1.10.2 одномерную плотность распределения $f(\mathrm{x})$, мы ввели понятие «элемента вероятности»: $f(x)dx$. Аналогичное понятие «элемента вероятности» вводится и для двумерной случайной величины.

Элементом вероятности двумерной случайной величины называется выражение $f(x, y)dxdy$. Геометрически оно означает вероятность попадания случайной точки в элементарный прямоугольник со сторонами $d\mathrm{x}$, dy, примыкающий к точке с координатами (x, y).

Из формулы (1.13.3) и определения двойного интеграла следует, что вероятнсть попадания случайной точки (x, y) в некоторую область D есть двойной интеграл от двумерной плотности распределения по этой области:

$$P((X, Y) \subset D) = \iint\limits_{(D)} f(x, y)dxdy. \tag{1.13.4}$$

Вероятность попадания случайной точки в область D геометрически изображается объемом цилиндрического тела, опирающегося на область D и ограниченного сверху поверхностью распределения $f(x, y)$.

Если в формуле (1.13.4) в качестве области интегрирования D принять область, показанную на рисунке 1.13.1, то вероятность попадания в нее случайной точки по определению 1.13.1 есть двумерная функция распределения системы случайных величин (X, Y):

$$F(x, y) = \iint\limits_{-\infty}^{x,y} f(x, y)dxdy. \tag{1.13.5}$$

Перечислим свойства плотности распределения двумерной непрерывной случайной величины.

1. Плотность распределения системы есть функция неотрицательная:

$$f(x, y) \geq 0.$$

2. Двойной интеграл в бесконечных пределах от плотности распределения системы равен единице:

$$\iint\limits_{-\infty}^{+\infty} f(x, y)dxdy = 1. \qquad (1.13.6)$$

Это следует из того, что вероятность попадания случайной точки на всю плоскость xOy есть вероятность достоверного события.

Равенство (1.13.6) является условием нормировки.

Чтобы получить плотность распределения одной из непрерывных случайных величин, входящих в систему, нужно двумерную плотность распределения проинтегрировать в бесконечных пределах по аргументу, соответствующему другой случайной величине:

$$f_1(x) = \int\limits_{-\infty}^{\infty} f(x, y)dy, \qquad f_2(y) = \int\limits_{-\infty}^{\infty} f(x, y)dx. \qquad (1.13.7)$$

Зная плотности распределения $f_1(x)$ и $f_2(y)$ каждой из случайных величин, входящих в систему (X, Y), можно рассчитать их числовые характеристики по формулам (1.11.1) и (1.11.2).

Пример 1.13.1. Система двух непрерывных случайных величин (X, Y) подчинена закону распределения с плотностью

$$f(x, y) = \frac{1}{\pi^2(1+x^2)(1+y^2)}.$$

Определить:

1) плотности $f_1(x)$ и $f_2(y)$ распределения каждой из случайных величин X и Y, входящих в систему;

2) функцию распределения $F(x, y)$;

3) функции распределения $F_1(x)$ и $F_2(y)$ случайных величин X и Y;

4) вероятность попадания случайной точки в прямоугольную область, ограниченную по осям X и Y координатами [0; 1];

5) построить графики функций $f(x, y)$ и $F(x, y)$.

Поскольку речь идет о символьных преобразованиях, то в данном случае удобно воспользоваться возможностями пакета символьной математики MuPAD системы MATLAB. Решение задачи в рабочей тетради Notebook — MuPad приведено на рисунках 1.13.3–1.13.5.

Функции MuPAD, использованнные при решении этой задачи, подробно описаны в работе [2].

Построим поверхность распределения $f(x, y)$ и график найденной нами функции распределения.

Дадим некоторые рекомендации для построения графиков в рабочей тетради Notebook — MuPAD.

Устанавливаем курсор в начало области ввода команд (обозначена квадратной скобкой. Из меню View открываем панель Command Bar (если она не открыта). Для построения трехмерных графиков кликаем соответствующую «иконку» внизу панели Command Bar. В области ввода команд появится шаблон вида plot(#f,#3D). Далее вставляем требуемые параметры функции вместо символов #. Переход от одного символа # к другому осуществляется с клавиатуры кнопкой табуляции ⟨*Tab*⟩. Если нужно отредактировать график, то кликаем левой кнопкой мыши по полю графика. Справа открывается окно Object Browsers. Кликаем ссылку Coordinate system2d. В окне Axes Title Font устанавливаем требуемый размер шрифта обозначения осей. В окне Tick Marks устанавливаем размер шрифта цифр разметки осей (Tick Label Font). В окне Object Browsers кликаем раздел Function 2d.

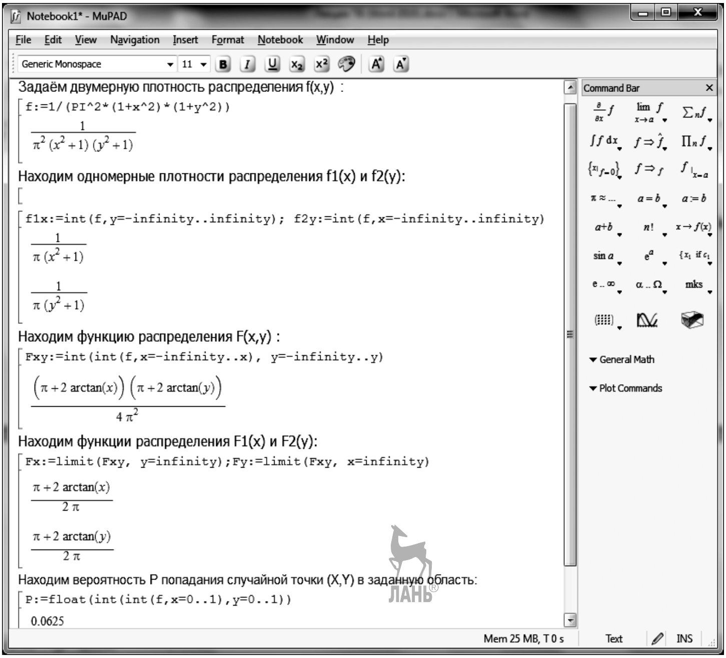

Рис. 1.13.3

Расчет законов распределения системы случайных величин

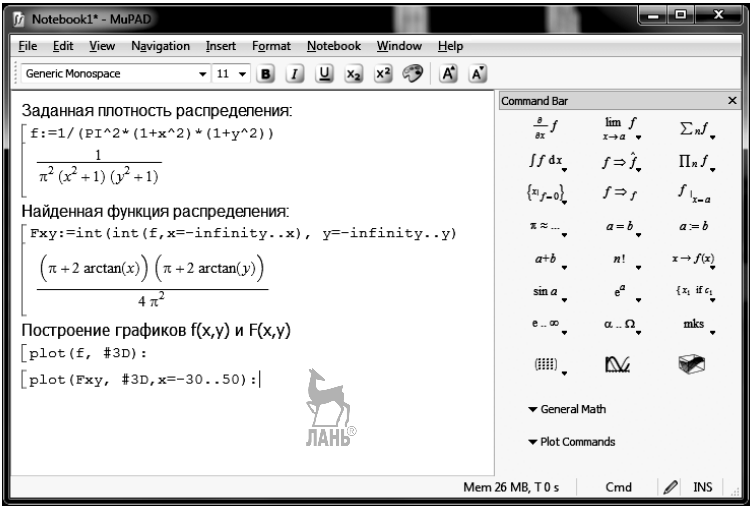

Рис. 1.13.4

Построение графиков законов распределения двумерной непрерывной случайной величины

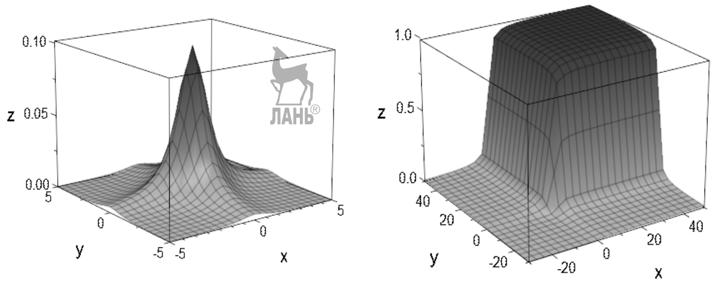

Рис. 1.13.5

Графики плотности f(x, y) (слева) и функции F(x, y) распределения
двумерной непрерывной случайной величины

В подразделе Style устанавливаем цвет (Line color) и толщину (Line width) линий графика. Кликаем отформатированный таким образом график правой кнопкой мыши. В открывшемся списке выбираем опцию Export Graphics. Далее кликаем кнопку Next. В открывшемся окне Image Format в верхнем поле выбираем нужное расширение. Затем кликаем кнопку Choose. В открывшемся окне в поле имени файла присваиваем файлу рисунка имя и указываем нужное расширение, например Рисунок1.tif, и назначаем куда экспортировать файл, например Рабочий стол. Далее кликаем кнопку Сохранить. В открывшемся окне кликаем кнопку Next, затем Export.

1.13.2. Условные законы распределения непрерывных случайных величин

В п. 1.13.1 показано, как, зная закон распределения системы, заданный в форме плотности распределения или в форме функции распределения, найти законы распределения отдельных величин, входящих в систему. Естественно, возникает вопрос об обратной задаче: нельзя ли по законам распределения отдельных величин, входящих в систему, восстановить закон распределения системы? Оказывается, что в общем случае этого сделать нельзя: зная только законы распределения отдельных величин, входящих в систему, не всегда можно найти закон распределения системы. Для того чтобы исчерпывающим образом охарактеризовать систему, недостаточно знать распределение каждой из величин, входящих в систему; нужно знать зависимость между величинами. Эта зависимость может быть охарактеризована с помощью так называемых условных законов распределения.

Определение 1.13.3. Закон распределения одной из случайных величин, входящих в систему при фиксированном значении другой случайной величины, называется условным законом распределения.

Условный закон распределения можно задавать как условной функцией распределения, так и условной плотностью распределения. Условная функция распределения обозначается $F(x/y)$ или $F(y/x)$, условная плотность распределения — $f(x/y)$ или $f(y/x)$.

Чтобы наглядно пояснить понятие условного закона распределения, рассмотрим пример. Система случайных величин X и Y представляет собой рост и вес человека. Пусть нас интересует рост X человека безотносительно к его весу Y. Случайная величина X подчинена закону распределения с плотностью $f_1(x)$. Этот закон распределения мы можем исследовать, рассматривая всех без исключения людей и оценивая их только по росту, т. е. $f_1(x)$ есть безусловный закон распределения роста человека.

Для определения закона распределения роста человека, имеющего вполне определенный вес, например 80 кг, мы будем исследовать не всех людей, а только данную весовую группу, и получим условный закон распределения роста человека при весе 80 кг с плотностью $f(x/y)$ при $y = 80$. Этот условный закон распределения в общем случае отличается от безусловного закона $f_1(x)$; очевидно, люди с большим весом должны, в среднем, быть более высокими. Следовательно, условный закон распределения роста человека X существенно зависит от его веса Y.

Приведем без доказательства формулы для условных законов распределения:

$$f(x/y) = \frac{f(x,y)}{f_2(y)}; \qquad f(y/x) = \frac{f(x,y)}{f_1(x)}. \qquad (1.13.8)$$

Условные функции распределения $F(x/y)$ и $F(y/x)$ определяются аналогично.

Наглядное представление об условных законах распределения можно получить из графиков на рисунке 1.13.5. Кривые условных плотностей и функций распределения изображены тонким линиями сечений соответствующих поверхностей распределения вертикальными плоскостями, параллельными плоскостям $z0x$ и $z0y$.

Основными числовыми характеристиками условных законов распределения являются условное математическое ожидание $m_{X/y}$ ($m_{Y/x}$) и условная дисперсия $D_{X/y}$ ($D_{Y/x}$).

Под условными числовыми характеристиками случайной величины понимаются соответствующие числовые характеристики одной случайной величины при условии, что другая случайная величина приняла какое-либо фиксированное значение.

Для системы непрерывных случайных величин (X, Y) с условными плотностями распределения $f(x/y)$ и $f(y/x)$, условные математические ожидания $m_{x/y}$ и $m_{y/x}$ определяются формулами:

$$m_{X/y} = \int_{-\infty}^{+\infty} x \cdot f(x/y)dx; \qquad m_{Y/x} = \int_{-\infty}^{+\infty} y \cdot f(y/x)dy, \qquad (1.13.9)$$

а условные дисперсии:

$$D_{X/y} = \int_{-\infty}^{+\infty} (x - m_{x/y})^2 f(x/y)dx; \qquad D_{Y/x} = \int_{-\infty}^{+\infty} (y - m_{y/x})^2 f(y/x)dy.$$

$$(1.13.10)$$

Определение 1.13.4. Случайные величины X и Y называются независимыми, если закон распределения каждой из них не зависит от того, какое значение приняла другая случайная величина. В противном случае величины X и Y называются зависимыми.

Для непрерывных случайных величин условие независимости, скажем, Y от X может быть записано в виде

$$f(y/x) = f_2(y)$$

при любом значении y.

Напротив, в случае, если Y зависит от X, то

$$f(y/x) \neq f_2(y).$$

Можно доказать, что если величина Y не зависит от X, то и величина X не зависит от Y.

Для независимых непрерывных случайных величин:

$$f(x, y) = f_1(y) \cdot f_2(y). \qquad (1.13.11)$$

Остановимся несколько подробнее на важных понятиях о «зависимости» и «независимости» случайных величин. Понятие «зависимости» случайных величин, которым мы пользуемся в теории вероятностей, несколько отличается от обычного понятия «зависимости» величин, которым мы оперируем в математике. Действительно, обычно под «зависимостью» величин подразумевают только один тип зависимости — полную, жесткую, так называемую функциональную зависимость. Величины X и Y называются функционально зависимыми, если, зная значение одной из них, можно однозначно указать значение другой.

В теории вероятностей мы встречаемся с другим, более общим типом зависимости — с вероятностной, или стохастической, зависимостью. Если величина Y связана с величиной X вероятностной зависимостью, то, зная значение X, нельзя с уверенностью определить значение Y, а можно указать только ее закон распределения, зависящий от того, какое значение приняла величина X.

Вероятностная зависимость может быть более или менее тесной; по мере увеличения «тесноты» вероятностной зависимости она все более приближается к функциональной зависимости. Таким образом, функциональную зависимость можно рассматривать как крайний, предельный случай наиболее тесной вероятностной зависимости. Другой крайний случай — полная независимость случайных величин. Между этими двумя крайними случаями лежат все градации вероятностной зависимости — от самой сильной до самой слабой.

1.13.3. Законы распределения дискретной двумерной случайной величины

Закон распределения системы дискретных случайных величин (X, Y) задается таблицей распределения (табл. 1.13.1), которая является двумерным аналогом ряда распределения.

Таблица 1.13.1

Таблица распределения двумерной дискретной случайной величины

$p(x, y)$	y_1	y_2	...	y_j	...	y_m
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	...	$p(x_1, y_j)$	...	$p(x_1, y_m)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	...	$p(x_2, y_j)$	...	$p(x_2, y_m)$
...	...	...	...	...	...	...
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	...	$p(x_i, y_j)$	...	$p(x_i, y_m)$
...	...	...	...	...	...	...
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	...	$p(x_n, y_j)$	...	$p(x_n, y_m)$

По аналогии со свойством (1.7.1) закона распределения одномерной случайной величины, справедливо равенство

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} = 1, \qquad (1.13.12)$$

где

$$p_{ij} = p(x_i, y_j) = P((X = x_i), (Y = y_j)).$$

Зная закон распределения $p(x, y)$ двумерной дискретной случайной величины (X, Y), можно найти законы распределения $p(x)$ и $p(y)$ ее составляющих X и Y в виде соответствующих рядов:

$$P(X = x_i) = p(x_i, y_1) + p(x_i, y_2) + ... + p(x_i, y_m);$$
$$P(Y = y_j) = p(x_1, y_j) + p(x_2, y_j) + ... + p(x_n, y_j),$$

где $i = 1, ..., n$; $j = 1, ..., m$.

П р и м е р 1.13.2. Задан закон распределения двумерной дискретной случайной величины (X, Y):

p(x, y)		**Y**		
		–2	**3**	**6**
X	–0,8	0,10	0,30	0,10
	–0,5	0,15	0,25	0,10

Требуется найти законы распределения составляющих.

Р е ш е н и е. Складывая стоящие в таблице вероятности «по строкам», получим ряд распределение для отдельной случайной величины X, входящей в систему(X, Y):

X	–0,8	–0,5
$p(x)$	–0,50	0,50

Складывая стоящие в таблице вероятности «по столбцам», получим ряд распределение для отдельной случайной величины Y, входящей в систему (X, Y):

Y	–2	3	6
$p(y)$	0,25	0,55	0,20

Отсюда нетрудно найти математические ожидания и дисперсии отдельно для каждой из величин X и Y по формулам, представленным в п. 1.8. Рекомендуется проделать это в качестве самостоятельной работы.

Заметим, что числовые характеристики для X и Y можно получить непосредственно из двумерного закона распределения, заданного в условии задачи. Об этом будет сказано ниже, при рассмотрении числовых характеристик двумерных случайных величин.

Для системы дискретных случайных величин (X, Y), заданной таблицей распределения $p(x, y)$, условные математические ожидания определяются формулами:

$$m_{X/y_j} = \sum_{i=1}^{n} x_i p_{ij}; \quad m_{Y/x_i} = \sum_{j=1}^{m} y_j p_{ij}, \tag{1.13.13}$$

а условные дисперсии:

$$D_{X/y_j} = \sum_{i=1}^{n} (x_i - m_{x/y_j})^2 p_{ij}; \quad D_{Y/x_i} = \sum_{j=1}^{m} (y_j - m_{Y/x_i})^2 p_{ij}. \tag{1.13.14}$$

П р и м е р 1.13.3. Вычислить условные математические ожидания и условные дисперсии случайных величин X и Y, если совместный закон распределения задан таблицей примера 1.13.2.

Р е ш е н и е.

$$m_{X/-2} = -0{,}8 \cdot 0{,}10 - 0{,}5 \cdot 0{,}15 = -0{,}155;$$
$$m_{X/3} = -0{,}8 \cdot 0{,}30 - 0{,}5 \cdot 0{,}25 = -0{,}365;$$
$$m_{X/6} = -0{,}8 \cdot 0{,}10 - 0{,}5 \cdot 0{,}10 = 0{,}13;$$

$$m_{Y/-0,8} = -2 \cdot 0,10 + 3 \cdot 0,30 + 6 \cdot 0,10 = 1,3;$$
$$m_{Y/-0,5} = -2 \cdot 0,15 + 3 \cdot 0,25 + 6 \cdot 0,10 = 1,05;$$
$$D_{X/-2} = (-0,8 + 0,155)^2 \cdot 0,10 + (-0,5 + 0,155)^2 \cdot 0,15 \approx 0,059;$$
$$D_{X/3} = (-0,8 + 0,365)^2 \cdot 0,30 + (-0,5 + 0,365)^2 \cdot 0,25 \approx 0,139;$$
$$D_{X/6} = (-0,8 + 0,13)^2 \cdot 0,10 + (-0,5 + 0,13)^2 \cdot 0,10 \approx 0,059;$$
$$D_{Y/-0,8} = (-2 - 0,13)^2 \cdot 0,10 + (3 - 0,13)^2 \cdot 0,30 + (6 - 0,13)^2 \cdot 0,10 \approx 6,37;$$
$$D_{Y/-0,5} = (-2 - 1,05)^2 \cdot 0,15 + (3 - 1,05)^2 \cdot 0,25 + (6 - 1,05)^2 \cdot 0,10 \approx 4.8.$$

Как и в одномерном случае, для описания двумерной дискретной случайной величины (X, Y) помимо таблицы распределения может применяться двумерная функция распределения $F(x_i, y_j)$, которая имеет тот же смысл и свойства, что и функция распределения для непрерывной случайной величины (п. 1.13.1):

$$F(x_i, y_j) = P(X < x_i, Y < y_j).$$

Разница в том, что переменные x и y индексируются, а график самой функции $F(x_i, y_j)$ носит ступенчатый характер.

Отметим, что стандартной функции для рисования функции распределения двумерной дискретной случайной величины в системе MATLAB нет. Изобразить ее можно с помощью стандартной функции *fill*3. Она применяется для рисования произвольно ориентированных в пространстве многоугольников [3]. Но эта тема достаточно обширна и требует отдельного рассмотрения.

1.13.4. Числовые характеристики системы двух случайных величин

В пп. 1.8, 1.10 мы рассмотрели числовые характеристики одномерных дискретных и непрерывных случайных величин. Напомним, что в качестве таких характеристик используются начальные и центральные моменты различных порядков. Из этих характеристик важнейшими являются две: математическое ожидание и дисперсия. Аналогичные числовые характеристики — начальные и центральные моменты различных порядков — можно ввести и для системы двух случайных величин.

Определение.1.13.5. Начальным моментом порядка k, s системы случайных величин (X, Y) называется математическое ожидание произведения X^k и Y^s:

$$\alpha_{k,s} = M[X^k \cdot Y^s]. \qquad (1.13.15)$$

Чаще всего используются начальные моменты порядков 1,0 и 0,1 и центральные моменты порядков 2,0 и 0,2, которые имеют смысл математических ожиданий m_x, m_y и дисперсий D_x, D_y. Ниже приводятся формулы расчета этих моментов для дискретных и для непрерывных случайных величин:

$$m_X = \alpha_{1,0} = \sum_i \sum_j x_i p_{ij}; \quad m_Y = \alpha_{0,1} = \sum_i \sum_j y_j p_{ij}; \tag{1.13.16}$$

$$D_X = \mu_{2,0} = \sum_i \sum_j (x_i - m_x)^2 p_{ij}; \quad D_Y = \mu_{0,2} = \sum_i \sum_j (y_j - m_y)^2 p_{ij}; \tag{1.13.17}$$

$$m_X = \alpha_{1,0} = \iint\limits_{-\infty}^{+\infty} xf(x,y)dxdy; \quad m_Y = \alpha_{0,1} = \iint\limits_{-\infty}^{+\infty} yf(x,y)dxdy; \tag{1.13.18}$$

$$\begin{aligned} D_X &= \mu_{2,0} = \iint\limits_{-\infty}^{+\infty} (x - m_X)^2 f(x,y)dxdy; \\ D_X &= \mu_{2,0} = \iint\limits_{-\infty}^{+\infty} (y - m_Y)^2 f(x,y)\ dxdy. \end{aligned} \tag{1.13.19}$$

Если известны одномерные знаки распределения $f_1(x)$ и $f_2(y)$, то рассчитать математические ожидания и дисперсии проще по формулам:

$$m_X = \int\limits_{-\infty}^{\infty} x \cdot f_1(x)dx; \quad m_Y = \int\limits_{-\infty}^{\infty} y \cdot f_2(y)dy; \tag{1.13.20}$$

$$D_X = \int\limits_{-\infty}^{\infty} (x - m_X)^2 \cdot f_1(x)dx; \quad D_Y = \int\limits_{-\infty}^{\infty} (y - m_Y)^2 \cdot f_2(y)dy. \tag{1.13.21}$$

Определение.1.13.6. Центральным моментом порядка k, s системы случайных величин (X, Y) называется математическое ожидание произведения k-й и s-й степени соответствующих центрированных величин:

$$\mu_{k,\,s} = M[(X - m_x)^k \cdot (Y - m_y)^s]. \tag{1.13.22}$$

Особую роль в качестве характеристики системы двух случайных величин, как дискретных, так и непрерывных, играет второй смешанный центральный момент:

$$\mu_{11} = M[(X - m_X)(Y - m_Y)] = K_{XY}, \tag{1.13.23}$$

представляющий собой математическое ожидание произведения центрированных случайных величин.

Характеристика K_{XY} называется корреляционным моментом или ковариацией случайных величин X и Y и является характеристикой связи между ними.

Для дискретных случайных величин фомула для расчета корреляционного момента принимает вид

$$K_{XY} = \mu_{1,1} = \sum_i \sum_j (x_i - m_X)(y_j - m_Y), \qquad (1.13.24)$$

где суммирование распространяется на все возможные значения i, j.

Для непрерывных случайных величин:

$$K_{XY} = \mu_{1,1} = \iint\limits_{-\infty}^{+\infty} (x - m_X)(y - m_Y) f(x, y) dxdy. \qquad (1.13.25)$$

П р и м е р 1.13.4. Найти числовые характеристики m_X, m_Y, D_X, D_Y.

K_{XY} двумерного закона распределения, заданного таблицей примера 1.13.2.

Р е ш е н и е.

$$m_X = \sum_{i=1}^{2} \sum_{j=1}^{3} x_i p_{ij} = x_1(p_{11} + p_{12} + p_{13}) + x_2(p_{21} + p_{22} + p_{23}) =$$

$$= –0,8 \cdot (0,10 + 0,30 + 0,10) + (–0,5) \cdot (0,15 + 0,25 + 0,10) = –0,65;$$

$$m_Y = \sum_{i=1}^{2} \sum_{j=1}^{3} y_j p_{ij} = \sum_{j=1}^{3} \sum_{i=1}^{2} y_j p_{ij} =$$

$$= y_1(p_{11} + p_{21}) + y_2(p_{12} + p_{22}) + y_3(p_{13} + p_{23}) =$$

$$= –2 \cdot (0,10 + 0,15) + 3 \cdot (0,30 + 0,25) + 6 \cdot (0,10 + 0,10) = 2,35;$$

$$D_X = \sum_{i=1}^{2} \sum_{j=1}^{3} (x_i - m_x)^2 p_{ij} = (x_1 - m_x)^2 (p_{11} + p_{12} + p_{13}) =$$

$$= + (x_2 - m_x)^2(p_{21} + p_{22} + p_{23}) = (–0,8 + 0,65)^2(0,10 + 0,30 + 0,10) +$$

$$+ (–0,5 + 0,65)^2(0,15 + 0,25 + 0,10) = 0,025;$$

$$D_Y = \sum_{j=1}^{3} \sum_{i=1}^{2} (y_j - m_y)^2 p_{ij} = (y_1 - m_y)^2 (p_{11} + p_{21}) +$$

$$+ (y_2 - m_y)^2(p_{12} + p_{22}) + (y_3 - m_y)^2(p_{13} + p_{23}) =$$

$$= (–2 – 2,35)^2(0,10 + 0,15) + (3 – 2,35)^2(0,30 + 0,25) +$$

$$+ (6 – 2,35)^2(0,10 + 0.10) \approx 7,63;$$

$$K_{XY} = \sum_{i=1}^{2} \sum_{j=1}^{3} (x_i - m_x)(y_j - m_y) p_{ij} = (x_1 - m_x)(y_1 - m_y) p_{11} +$$

$$+ (x_1 - m_x)(y_2 - m_y)p_{12} + (x_1 - m_x)(y_3 - m_y)p_{13} +$$

$$+ (x_2 - m_x)(y_1 - m_y)p_{21} + (x_2 - m_x)(y_2 - m_y)p_{22} +$$

$$+ (x_2 - m_x)(y_3 - m_y)p_{23} =$$

$$= (-0{,}8 + 0{,}65)(-2 - 2{,}35) \cdot 0{,}1 + (-0{,}8 + 0{,}65)(3 - 2{,}35) \cdot 0{,}3 +$$
$$+ (-0{,}8 + 0{,}65)(6 - 2{,}35) \cdot 0{,}1 + (-0{,}5 + 0{,}65)(-2 - 2{,}35) \cdot 0{,}15 +$$
$$+ (-0{,}5 + 0{,}65)(3 - 2{,}35) \cdot 0{,}25 + (-0{,}5 + 0{,}65)(6 - 2{,}35) \cdot 0{,}1 \approx -0{,}037.$$

Корреляционный момент K_{XY} (ковариацию) часто удобно вычислять по формуле

$$K_{XY} = M[X \cdot Y] - M[X] \cdot M[Y]. \quad (1.13.26)$$

Из формул (1.13.24), (1.13.25) следует, что корреляционный момент K_{XY} описывает не только связь между случайными величинами X и Y, но степень их рассеивания. Действительно, если, например, одна из величин (X, Y) весьма мало отклоняется от своего математического ожидания (почти не случайна), то корреляционный момент будет мал, какой бы тесной зависимостью ни были связаны величины (X, Y). Поэтому для характеристики связи между случайными величинами переходят от момента K_{XY} к безразмерной характеристике:

$$r_{XY} = \frac{K_{XY}}{\sigma_X \sigma_Y}, \quad (1.13.27)$$

где σ_X, σ_Y — среднеквадратические отклонения величин X, Y.

Эта характеристика называется коэффициентом корреляции величин X и Y. Очевидно, коэффициент корреляции обращается в нуль одновременно с корреляционным моментом. Следовательно, для независимых случайных величин коэффициент корреляции равен нулю.

Случайные величины, для которых корреляционный момент (а значит, и коэффициент корреляции) равен нулю, называются некоррелированными.

Заметим, что равенство нулю коэффициента корреляции — необходимое, но недостаточное условие независимости случайных величин, т. е. из некоррелированности величин еще не следует их независимость. Условие независимости случайных величин — более жесткое, чем условие некоррелированности.

Вместе с тем, коэффициент корреляции характеризует не всякую зависимость, а только так называемую линейную зависимость. Линейная вероятностная зависимость случайных величин заключается в том, что при возрастании одной случайной величины другая имеет тенденцию возрастать (или же убывать) по линейному закону.

В общем случае коэффициент корреляции изменяется в пределах $-1 \le r_{XY} \le 1$. В случае $r_{XY} > 0$ говорят о положительной корреляции величин X и Y, в случае $r_{XY} < 0$ — об отрицательной корреляции.

Если случайные величины X и Y связаны точной линейной функциональной зависимостью $Y = aX + b$, то $rXY = \pm 1$, причем знак «плюс» или «минус» берется в зависимости от того, положителен или отрицателен коэффициент a.

П р и м е р 1.13.5. Задана плотность распределения двумерной случайной величины (X, Y): $f(x, y) = 0{,}5\sin(x + y)$, причем $X, Y \in [0; \pi]$. Вычислить ковариацию и коэффициент корреляции в заданной системе случайных величин.

Решение задачи, выполненное в рабочей тетради Notebook — MuPAD системы MATLAB, представлено на рисунке 1.13.6.

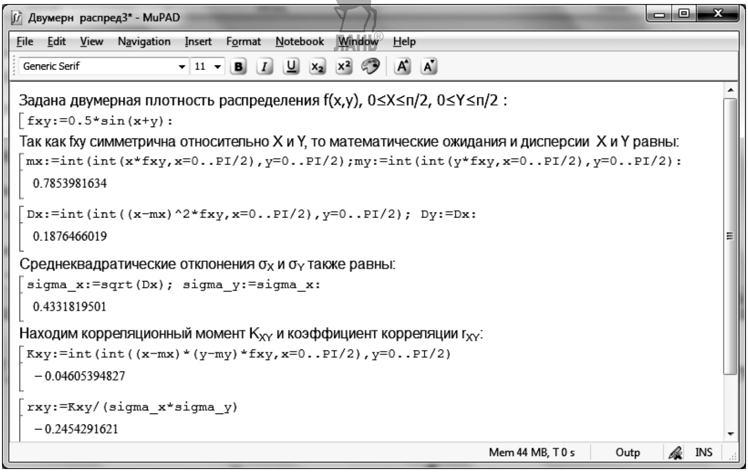

Рис. 1.13.6

Расчет числовых характеристик
двумерной непрерывной случайной величины

Поскольку в данной задаче не требовалось находить одномерные законы распределения $f_1(x)$ и $f_2(y)$, то математические ожидания и дисперсии рассчитывались сразу по заданной двумерной плотности вероятности $f(x, y)$ с помощью формул (1.13.18), (1.13.19).

Стоит обратить внимание на существенную разницу полученных значений K_{XY} и r_{XY}. Если судить по вычисленному корреляционному моменту K_{XY}, то вероятностная зависимость между случайными величинами X и Y весьма слабая. В то же время величина коэффициента корреляции r_{XY} указывает на более сильную вероятностную зависимость между ними. По причинам, о которых говорилось выше, коэффициент корреляции r_{XY} следует считать более объективным показателем степени вероятностной зависимости между случайными величинами, чем корреляционный момент K_{XY}.

1.14.
ФУНКЦИИ СЛУЧАЙНЫХ ВЕЛИЧИН

1.14.1.
Числовые характеристики функций случайных величин

Пусть случайная величина X задана законом распределения; другая случайная величина Y связана с X функциональной зависимостью

$$Y = \varphi(X).$$

Требуется, не находя закона распределения величины Y, определить ее числовые характеристики: математическое ожидание $m_Y = M[\varphi(x)]$ и дисперсию $D_Y = D[\varphi(x)]$.

Рассмотрим сначала случай, когда X есть дискретная случайная величина с рядом распределения:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Выпишем возможные значения величины Y и вероятности этих значений:

$y_i = \varphi(x_i)$	$\varphi(x_1)$	$\varphi(x_2)$	...	$\varphi(x_n)$
p_i	p_1	p_2	...	p_n

Полученная таблица не является в строгом смысле слова рядом распределения случайной величины Y, так как в общем случае некоторые из значений $\varphi(x_1)$, $\varphi(x_2)$, ..., $\varphi(x_n)$ могут совпадать между собой; к тому же эти значения в верхней строке таблицы не обязательно идут в возрастающем порядке. Для того чтобы перейти к ряду распределения величины Y, нужно было бы расположить значения $\varphi(x_1)$, $\varphi(x_2)$, ..., $\varphi(x_n)$ в порядке возрастания, объединить столбцы с равными значениями Y и сложить соответствующие вероятности. Но в данном случае нас не интересует закон распределения величины Y как таковой; для решения нашей задачи достаточно такой «неупорядоченной» формы ряда распределения. Математическое ожидание и дисперсию случайной величины $Y = \varphi(x)$ можно определить по формулам:

$$m_Y = M[\varphi(x)] = \sum_{i=1}^{n} \varphi(x_i) p_i; \qquad (1.14.1)$$

$$D_Y = D[\varphi(x)] = \sum_{i=1}^{n} [\varphi(x_i) - m_Y]^2 p_i. \qquad (1.14.2)$$

Заменяя в этих формулах сумму интегралом, а вероятность p_i — элементом вероятности $f(x)dx$, получим аналогичные формулы для числовых характеристик функции $y = \varphi(x)$ непрерывной случайной величины:

$$m_Y = M[\varphi(x)] = \int_{-\infty}^{+\infty} \varphi(x) f(x) dx; \tag{1.14.3}$$

$$D_Y = D[\varphi(x)] = \int_{-\infty}^{+\infty} [\varphi(x) - m_Y]^2 f(x) dx, \tag{1.14.4}$$

где $f(x)$ — плотность распределения величины X.

Если одномерная величина Z есть функция $Z = \varphi(X, Y)$ двумерной дискретной случайной величины (X, Y), то математическое ожидание и дисперсия дискретной случайной величины Z вычисляются по формулам:

$$m_Z = M[\varphi(X, Y)] = \sum_i \sum_j \varphi(x_I, y_j) p_{ij}; \tag{1.14.5}$$

$$D_Z = D[\varphi(X, Y)] = \sum_i \sum_j [\varphi(x_i, y_j) - m_Z]^2 p_{ij}, \tag{1.14.6}$$

где p_{ij} — вероятность того, что система (X, Y) примет значения (x_i, y_j).

Для функции $Z = \varphi(X, Y)$ непрерывной двумерной случайной величины (X, Y) математическое ожидание и дисперсия непрерывной случайной величины Z вычисляются по формулам:

$$m_Z = M[\varphi(X, Y)] = \iint_{-\infty}^{+\infty} \varphi(x, y) f(x, y) dxdy; \tag{1.14.7}$$

$$D_Z = D[\varphi(X, Y)] = \int_{-\infty}^{+\infty} [\varphi(x, y) - m_Z]^2 f(x, y) dxdy. \tag{1.14.8}$$

П р и м е р 1.14.1. Для закона распределения двумерной дискретной случайной величины (X, Y), заданного таблицей распределения (см. пример 1.13.2 в п. 1.13.3), найти числовые характеристики m_Z и D_z функции $Z = 2X - Y$.

Р е ш е н и е.

$$m_z = M[2X - Y] = \sum_{i=1}^{2} \sum_{j=1}^{3} z_{ij} p_{ij} = \sum_{i=1}^{2} \sum_{j=1}^{3} (2x_i - y_j) p_{ij} = (2x_1 - y_1) p_{11} +$$

$$+ (2x_1 - y_2)p_{12} + (2x_1 - y_3)p_{13} + (2x_2 - y_1)p_{21} + (2x_2 - y_2)p_{22} +$$
$$+ (2x_2 - y_3)p_{23} = (2 \cdot (-0{,}8) - (-2)) \cdot 0{,}1 + (2 \cdot (-0{,}8) - 3) \cdot 0{,}3 +$$

$$+ (2 \cdot (-0{,}8) - 6) \cdot 0{,}1 + (2 \cdot (-0{,}5) - (-0{,}2)) \cdot 0{,}15 +$$
$$+ (2 \cdot (-0{,}5) - 3) \cdot 0{,}25 + (2 \cdot (-0{,}5) - 6) \cdot 0{,}1 =$$
$$= 0{,}04 - 1{,}38 - 0{,}76 - 0{,}12 - 1 - 0{,}7 = -3{,}92;$$
$$D_z = \sum_{i=1}^{2}\sum_{j=1}^{3}(Z_{ij} - m_Z)^2 p_{ij} = \sum_{i=1}^{2}\sum_{j=1}^{3}(2x_i - y_j - m_Z)^2 p_{ij} =$$
$$= (2x_1 - y_1 - m_z)^2 p_{11} + (2x_1 - y_2 - m_z)^2 p_{12} + (2x_1 - y_3 - m_z)^2 p_{13} +$$
$$+ (2x_2 - y_1 - m_z)^2 p_{21} + (2x_2 - y_2 - m_z)^2 p_{22} + (2x_2 - y_3 - m_z)^2 p_{23} +$$
$$+ (2 \cdot (-0{,}8) - 6)^2 \cdot 0{,}1 + (2 \cdot (-0{,}5) - (-2))^2 \cdot 0{,}15 +$$
$$+ (2 \cdot (-0{,}5) - 3)^2 \cdot 0{,}25 + (2 \cdot (-0{,}5) - 6)^2 \cdot 0{,}1 = 21{,}19.$$

1.14.2. Теоремы о числовых характеристиках функций случайных величин

В п. 1.14.1 мы привели ряд формул, позволяющих находить числовые характеристики функций по известным законам распределения аргументов. Во многих случаях для нахождения числовых характеристик функций достаточно знать только числовые характеристики аргументов. Нахождение числовых характеристик функций по заданным числовым характеристикам аргументов широко применяется в теории вероятностей и позволяет значительно упрощать решение ряда практических задач. Такие упрощенные методы относятся преимущественно к линейным функциям, но и некоторые элементарные нелинейные функции также допускают подобный подход.

Сформулируем некоторые теоремы о числовых характеристиках функций случайных величин.

Теоремы о математическом ожидании функций случайных величин.

1. Математическое ожидание суммы двух случайных величин равно сумме математических ожиданий этих случайных величин:

$$M(X + Y) = M(X) + M(Y). \qquad (1.14.9)$$

2. Математическое ожидание произведения двух случайных величин равно произведению математических ожиданий этих случайных величин плюс корреляционный момент:

$$M(XY) = M(X)M(Y) + K_{XY}, \qquad (1.14.10)$$

3. Пусть $X_1, X_2, ..., X_n$ — такие случайные величины, математические ожидания которых равны между собой, т. е. $M(X_i) = a$, где $i = 1, 2, ..., n$ и a — некоторое число. Тогда среднее арифметическое

этих случайных величин равно их общему математическому ожиданию, т. е. $M\left(\frac{X_1+X_2+...+X_n}{n}\right)=a$.

Теоремы о дисперсиях.

1. Дисперсия суммы двух случайных величин равна сумме дисперсий этих случайных величин плюс удвоенный корреляционный момент:

$$D(X+Y)=D(X)+D(Y)+2K_{XY}. \qquad (1.14.11)$$

В частном случае, когда случайные величины X и Y независимы, $K_{XY}=0$ и формула (1.14.11) принимает вид

$$D(X+Y)=D(X)+D(Y). \qquad (1.14.12)$$

2. Дисперсия произведения двух независимых случайных величин определяется по формуле

$$D(XY)=D(X)\cdot D(Y)+m_X^2\cdot D(Y)+m_Y^2\cdot D(X). \qquad (1.14.13)$$

3. Пусть случайные величины X_1, X_2, ..., X_n независимы и $D[X_i]=\sigma^2$, где $i=1, 2, ..., n$. Тогда

$$D\left(\frac{X_1+X_2+...+X_n}{n}\right)=\frac{\sigma^2}{n}. \qquad (1.14.14)$$

Эти теоремы справедливы как для дискретных, так и для непрерывных случайных величин. Доказательство их можно провести, опираясь на определения числовых характеристик случайных величин, рассмотренные в предыдущих разделах.

В качестве примера рссмотрим задачу нахождения числовых характеристик дискретной случайной величины X, распределенной по биномиальному закону. Пусть производится n опытов, в каждом из которых может появиться или не появиться некоторое событие A. Вероятность появления события A в i-м опыте равна p_i. Найти математическое ожидание m_x, дисперсию D_x и среднеквадратическое отклонение σ_x случайной величины X — числа появлений события A в n опытах.

Представим дискретную случайную величину X в виде

$$X=X_1+X_2+\ldots+X_n=\sum_{i=1}^{n}X_i,$$

где X_1 — число появлений события в первом опыте; X_2 — число появлений события во втором опыте; X_n — число появлений события в n-м опыте.

Причем каждая из величин X_i есть дискретная случайная величина с двумя возможными значениями: 0 или 1. Ряд распределения каждой из величин X_i, $i = 1, 2, ..., n$ имеет вид:

0	1
q_i	p_i

где $q_i = 1 - p_i$ — вероятность того, что событие A в i-м опыте не появится.

По теореме сложения математических ожиданий:

$$M[X] = M[X_1 + \ldots + X_n] = M[X_1] + \ldots + M[X_n] = \sum_{i=1}^{n} M[X_i].$$

Математическое ожидание $M[X_i]$ случайной величины X_i:

$$M[X_i] = 0 \cdot q_i + 1 \cdot p_i = p_i.$$

Подставляя это выражение в предыдущую формулу, имеем

$$m_X = M[X] = \sum_{i=1}^{n} p_i,$$

т. е. математическое ожидание числа появлений события при нескольких опытах равно сумме вероятностей события в отдельных опытах.

В частности, когда условия опытов одинаковы, т. е. $p_1 = p_2 = ... = p_n = p$, формула математического ожидания принимает вид

$$m_X = M[X] = np.$$

Найдем дисперсию и среднеквадратическое отклонение числа появлений события в n опытах. В силу независимости опытов в схеме Бернулли, случайные величины X_1, X_2, ..., X_n также независимы, и к ним применима теорема сложения дисперсий (1.14.12):

$$D[X] = D[X_1 + \ldots + X_n] = D[X_1] + \ldots + D[X_n] = \sum_{i=1}^{n} D[X_i].$$

Найдем дисперсию $D[X_i]$ дискретной случайной величины X_i, заданной ее рядом распределения:

$$D[X_i] = (0 - p_i)^2 \cdot q_i + (1 - p_i)^2 \cdot p_i = p_i q_i.$$

Подставляя это выражение в предыдущую формулу, имеем

$$D_x = D[X] = \sum_{i=1}^{n} p_i q_i,$$

т. е. дисперсия числа появлений события при нескольких опытах равна сумме вероятностей появления и непоявления события в отдельных опытах.

Если $p_1 = p_2 = ... = p_n = p$, то формулы дисперсии и среднего квадратичного отклонения упрощаются и принимают вид:

$$D_x = D[X] = npq; \quad \sigma_x = \sqrt{npq}.$$

Полученный результат приводился без доказательства в п. 1.9.1.

1.15. ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ФУНКЦИЙ СЛУЧАЙНЫХ ВЕЛИЧИН

1.15.1. Закон распределения функции одного случайного аргумента

Задача нахождения закона распределения функции одного дискретного случайного аргумента по известному закону распределения самого аргумента решалась в п. 1.14.1 в связи с определением числовых характеристик.

Рассмотрим теперь задачу о нахождении закона распределения функции одного непрерывного случайного аргумента.

Пусть имеется непрерывная случайная величина X с заданной плотностью распределения $f(x)$. Другая случайная величина Y связана с ней известной функциональной зависимостью $Y = \varphi(X)$. Предполагаем, что функция $\varphi(X)$ непрерывна и дифференцируема.

Требуется найти плотность распределения $g(y)$ случайной величины Y.

Рассмотрим участок оси абсцисс $[a, b]$, на котором лежат все возможные значения величины X, т. е. $P(a \le X \le b) = 1$. В частном случае, когда область возможных значений X ничем не ограничена, $a = -\infty$, $b = \infty$.

Способ решения поставленной задачи зависит от поведения функции $\varphi(X)$ на участке $[a, b]$: возрастает ли она на этом участке, убывает или колеблется.

1. Функция $Y = \varphi(X)$ монотонно возрастает или убывает.

Пусть функция $Y = \varphi(X)$ монотонно возрастает. Найдем сначала функцию распределения величины Y: $G(y) = P(Y < y)$.

Проведем прямую AB, параллельную оси абсцисс, на расстоянии y от нее (рис. 1.15.1).

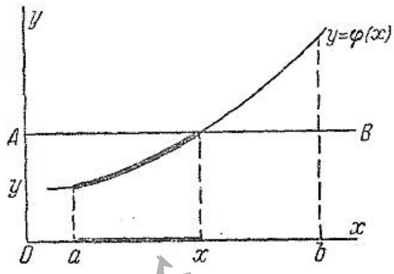

Рис. 1.15.1
Монотонно возрастающая функция непрерывной случайной величины

Чтобы выполнялось условие $Y < y$, случайная точка (X, Y) должна попасть на тот участок кривой, который лежит ниже прямой AB. Очевидно, это произойдет, когда случайная величина X попадет на участок оси абсцисс от a до x, где x — абсцисса точки пересечения кривой $y = \varphi(x)$ и прямой АВ. Тогда функция распределения $G(y)$ примет вид

$$G(y) = P(Y < y) = P(a < X < x) = \int_a^x f(x)dx, \qquad (1.15.1)$$

где $f(x)$ — плотность распределения случайной величины X.

Верхний предел x интеграла можно выразить через y:

$$x = \psi(y),$$

где ψ — функция, обратная функции φ. Тогда получим окончательное выражение для функции распределени случайной величины y:

$$G(y) = \int_a^{\psi(y)} f(x)dx.$$

Дифференцируя интеграл по переменной y, входящей в верхний предел, получим выражение для плотности распределения случайной величины y:

$$g(y) = G'(y) = f(\psi(y))\psi'(y). \qquad (1.15.2)$$

Аналогично можно показать, что если функция $Y = \varphi(X)$ на участке $[a, b]$ монотонно убывает, то

$$g(y) = -f(\psi(y))\psi'(y). \qquad (1.15.3)$$

Объединяем две полученные формулы (1.15.2), (1.15.3) в одну и получаем выражение для плотности распределения в случае, когда функция случайного аргумента монотонно возрастает или убывает:

$$g(y) = f(\psi(y)) \cdot |\psi'(y)|. \qquad (1.15.4)$$

П р и м е р 1.15.1. Случайная величина X подчинена нормальному закону с плотностью распределения:

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left(-\frac{(x-m_x)^2}{2\sigma_x^2}\right),$$

а случайная величина Y связана с ней линейной функциональной зависимостью: $Y = kX + b$, где k и b — неслучайные коэффициенты.

Требуется найти закон распределения случайной величины Y.

Р е ш е н и е. Так как функция $y = kx + b$ монотонна на участке от $-\infty$ до $+\infty$, то можно применить формулу (1.15.4). В нашем случае:

$$x = \psi(y) = \frac{y-b}{k}; \quad \psi'(y) = \frac{1}{k}; \quad |\psi'(y)| = \frac{1}{|k|}.$$

Тогда

$$g(y) = f(\psi(y)) \cdot |\psi'(y)| = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left(-\frac{\left(\frac{y-b}{k} - m_x\right)^2}{2\sigma_x^2}\right) \cdot \frac{1}{|k|}$$

или окончательно

$$g(y) = \frac{1}{|k|\,\sigma_x\sqrt{2\pi}} \exp\left(-\frac{[y-(km_x+b)]^2}{2\,|k|^2\,\sigma_x^2}\right).$$

Полученный закон есть не что иное, как нормальный закон распределения с параметрами: $m_y = km_x + b$, $\sigma_y = |k| \cdot \sigma_x$.

На этом примере мы убедились, что линейная функция случайного аргумента, подчиненного нормальному закону распределения, также подчинена нормальному закону

2. Функция $Y = \varphi(X)$ на участке $[a, b]$ не является монотонной.

Имеется непрерывная случайная величина X с плотностью распределения $f(x)$; другая величина Y связана с X функциональной зависимостью: $Y = \varphi(X)$, причем функция $y = \varphi(x)$ не монотонна на участке $[a, b]$ возможных значений аргумента (рис. 1.15.2).

Сначала найдем интегральную функцию распределения $G(y)$ случайной величины Y. Для этого проведем прямую AB параллельно оси абсцисс на расстоянии y от нее. Выделим те участки кривой

$y = \varphi(x)$, для которых выполняется условие $Y < y$. Пусть этим участкам соответствуют участки оси абсцисс: $\Delta_1(y)$, $\Delta_2(y)$, ...

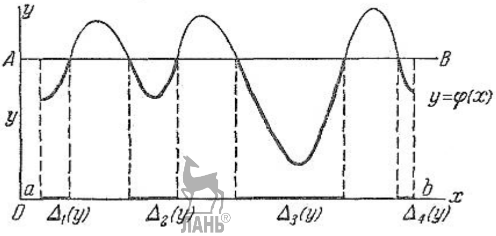

Рис. 1.15.2

Немонотонная функция непрерывной случайной величины

Событие $Y < y$ равносильно попаданию случайной величины X на один из участков $\Delta_1(y)$, $\Delta_2(y)$, ..., поэтому

$$G(y) = P(Y < y) = \sum_i P(X \subset \Delta_i(y)) = \sum_i \int_{\Delta_i(y)} f(x)dx. \qquad (1.15.5)$$

Границы интервалов $\Delta_i(y)$ зависят от величины y и при заданном конкретном виде функции $y = \varphi(x)$ могут быть выражены как явные функции от y. Дифференцируя $G(y)$ по y, получим плотность распределения случайной величины Y: $g(y) = G'(y)$.

Пример 1.15.2. Величина X подчинена закону равномерной плотности на участке от $-\frac{\pi}{2}$ до $+\frac{\pi}{2}$:

$$f(x) = \begin{cases} \frac{1}{\pi} & \text{при } |x| < \frac{\pi}{2}; \\ 0 & \text{при } |x| > \frac{\pi}{2}. \end{cases}$$

Найти закон распределения $g(y)$ величины $Y = \cos X$.

Решение. Строим график функции $y = \cos x$ (рис. 1.15.3).

Очевидно $a = -\frac{\pi}{2}$, $b = +\frac{\pi}{2}$, и в интервале $[a, b]$ функция $y = \cos x$ немонотонна. Применяя формулу (1.15.5), имеем

$$G(y) = \int_{-\pi/2}^{x_1} f(x)dx + \int_{x_2}^{\pi/2} f(x)dx.$$

Выразим пределы x_1 и x_2 через y: $x_1 = -\arccos y$; $x_2 = \arccos y$.

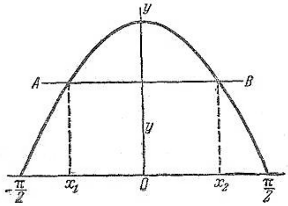

Рис. 1.15.3

График функции y = cosx на интервале $\left[-\frac{\pi}{2};\frac{\pi}{2}\right]$

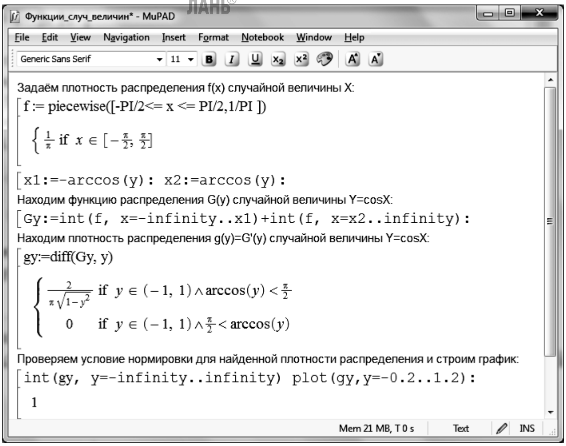

Рис. 1.15.4

Нахождение закона распределения функции одномерной случайной величины

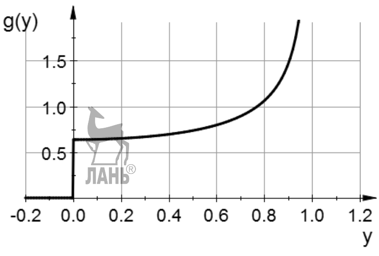

Рис. 1.15.5

График плотности распределения

Отсюда

$$G(y)=\int_{-\pi/2}^{-\arccos y} f(x)dx+\int_{\arccos y}^{\pi/2} f(x)dx.$$

Чтобы найти плотность $g(y)$, продифференцируем это выражение по переменной y, входящей в пределы интегралов:

$$g(y)=G'(y)=f(-\arccos y)\frac{1}{\sqrt{1-y^2}}+f(\arccos y)\frac{1}{\sqrt{1-y^2}}.$$

Имея в виду, что в заданной области $f(x)=\dfrac{1}{\pi}$, получим:

$$g(y)=\frac{2}{\pi}\frac{1}{\sqrt{1-y^2}}.$$

Замечание. Полученный закон распределения для Y определен на интервале от 0 до 1, т. е. в пределах изменения Y. Вне этих пределов плотность $g(y)$ равна нулю.

Решение той же задачи в рабочей тетради Notebook — MuPAD представлено на рисунках 1.15.4 и 1.15.5.

При задании плотности распределения $f(x)$ используется оператор *piecewise* (рис. 1.15.4), с помощью которого задается кусочная функция, т. е. функция, которая меняет свой вид в разных областях изменения аргумента. Для получения подробной справки по этому оператору достаточно набрать его в рабочей тетради Notebook — MuPAD, разместить курсор в поле команды и нажать на клавиатуре клавишу $\langle f1 \rangle$.

1.15.2.
Закон распределения функции двух случайных аргументов

Пусть X и Y — дискретные независимые случайные величины, от которых зависит другая случайная величина Z. Для нахождения закона распределения $Z=\varphi(X, Y)$ нужно найти все возможные значения Z и соответствующие им вероятности и построить для нее ряд распределения.

П р и м е р 1.15.3. Дискретные независимые случайные величины X и Y заданы законами распределения:

Y	0	1	2
p_j	0,2	0,5	0,3

X	–2	1	3
p_i	0,3	0,4	0,3

Требуется найти закон распределения дискретной случайной величины $Z = X + Y$.

Р е ш е н и е. Найдем возможные значения Z:

$$z_{11} = x_1 + y_1 = -2 + 0 = -2;\ p_{11} = 0{,}3 \cdot 0{,}2 = 0{,}06;$$
$$z_{12} = x_1 + y_2 = -2 + 1 = -1;\ p_{12} = 0{,}3 \cdot 0{,}5 = 0{,}15;$$
$$z_{13} = x_1 + y_3 = -2 + 2 = 0;\ p_{13} = 0{,}3 \cdot 0{,}3 = 0{,}09;$$
$$z_{21} = x_2 + y_1 = 1 + 0 = 1;\ p_{21} = 0{,}4 \cdot 0{,}2 = 0{,}08;$$
$$z_{22} = x_2 + y_2 = 1 + 1 = 2;\ p_{22} = 0{,}4 \cdot 0{,}5 = 0{,}20;$$
$$z_{23} = x_2 + y_3 = 1 + 2 = 3;\ p_{23} = 0{,}4 \cdot 0{,}3 = 0{,}12;$$
$$z_{31} = x_3 + y_1 = 3 + 0 = 3;\ p_{31} = 0{,}3 \cdot 0{,}2 = 0{,}06;$$
$$z_{32} = x_3 + y_2 = 3 + 1 = 4;\ p_{32} = 0{,}3 \cdot 0{,}5 = 0{,}15;$$
$$z_{33} = x_3 + y_3 = 3 + 2 = 5;\ p_{33} = 0{,}3 \cdot 0{,}3 = 0{,}09.$$

Сложив вероятности повторившегося дважды значения $Z = 3$, составим ряд распределения для Z:

Z	–2	–1	0	1	2	3	4	5
p	0,06	0,15	0,09	0,08	0,2	0,18	0,15	0,09

Перейдем к нахождению законов распределения функций двух непрерывных случайных величин.

Пусть система двух непрерывных случайных величин (X, Y) задана плотностью распределения $f(x, y)$. Случайная величина Z связана с X и Y функциональной зависимостью: $Z = \varphi(X, Y)$.

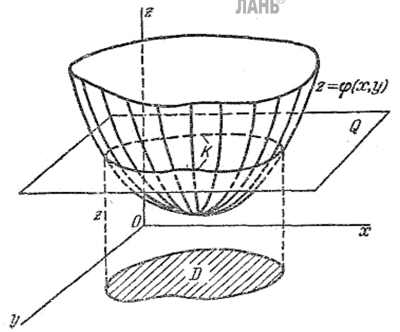

Рис. 1.15.6

Графическое изображение функции двух переменных

Требуется найти закон распределения случайной величины Z.

Для решения задачи воспользуемся геометрической интерпретацией, аналогичной той, которую мы применяли в случае одного непрерывного аргумента. Функциональная зависимость $Z = \varphi(X, Y)$ геометрически является некоторой поверхностью (рис. 1.15.6).

Сначала найдем интегральную функцию распределения

$G(z) = P(Z < z) = P(\varphi(X, Y) < z)$ величины Z.

Пусть Q — плоскость, параллельная плоскости xOy, на некотором расстоянии z от нее. Эта плоскость пересечет поверхность $Z = \varphi(X, Y)$ по некоторой кривой K. Спроектируем кривую K на плоскость xOy. Получим линию равного уровня $\varphi(X, Y) = z$, разделяющую плоскость xOy на две области. Для одной из них высота поверхности над плоскостью xOy будет меньше, для другой — больше z. Обозначим через D ту область, для которой эта высота меньше текущего значения z. Чтобы выполнялось неравенство $\varphi(X, Y) < z$, случайная точка (X, Y), очевидно, должна попасть в область D, следовательно,

$$G(z) = P((X, Y) \subset D) = \iint\limits_{(D)} f(x, y)dxdy. \qquad (1.15.6)$$

В это выражение величина z входит неявно, через пределы интегрирования.

Дифференцируя интегральную функцию распределения $G(z)$ по z, получим плотность распределения величины Z:

$$g(z) = G'(z). \qquad (1.15.7)$$

Зная конкретный вид функции $Z = \varphi(X, Y)$, можно выразить пределы интегрирования через z и написать выражение $g(z)$ в явном виде.

На практике, чтобы найти область интегрирования D, достаточно записать уравнение линии равного уровня $\varphi(X, Y) = z$, где z — параметр, и выяснить, при каком соотношении между X и Y выполняется условие $\varphi(X, Y) < z$.

Пример 1.15.4. Система случайных величин (X, Y) задана двумерной плотностью распределения:

$$f(x, y) = \begin{cases} 6 \cdot e^{-2x-3y} & \text{при } x > 0 \text{ и } y > 0; \\ 0 & \text{при остальных значениях } x \text{ и } y. \end{cases}$$

Случайная величина Z зависит от X и Y так, что $Z = Y/X$.

Требуется найти плотность распределения одномерной случайной величины Z.

Решение. Область интегрирования D есть пересечение области определения заданной плотности вероятности $f(x, y)$ и области, удовлетворяющей условию $Y/X < z$ или $Y < zX$, где z — неслучайный параметр. Областью определения функции $f(x, y)$ по условию является первая четверть плоскости xOy. Таким образом, область D расположена в первой четверти плоскости xOy ниже прямой $Y = zX$.

Запишем выражение для интегральной функции распределения в нашем случае:

$$G(z) = \iint\limits_{(D)} f(x, y)dxdy = \int\limits_0^\infty \int\limits_0^y 6e^{-2x-3y}dxdy = 6 \cdot \int\limits_0^\infty e^{-2x}\left(\int\limits_0^{zx} e^{-3y}dy\right)dx.$$

В данном примере ставится задача нахождения не функции распределения $G(z)$, а плотности распределения $g(z)$. Поэтому, не вычисляя полученный кратный интеграл, дифференцируем его по аргументу z функции распределения $G(z)$:

$$g(z) = G'(z) == 6\int\limits_0^\infty e^{-2x}e^{-3zx}(zx)'_z\,dx = 6\int\limits_0^\infty xe^{-(2+3z)x}dx = \frac{6}{(2+3z)^2}.$$

Нетрудно убедиться, что полученная функция обладает всеми свойствами плотности распределения случайной величины.

Решение той же задачи в рабочей тетради Notebook — MuPAD системы MATLAB представлено на рисунках 1.15.7, 1.15.8.

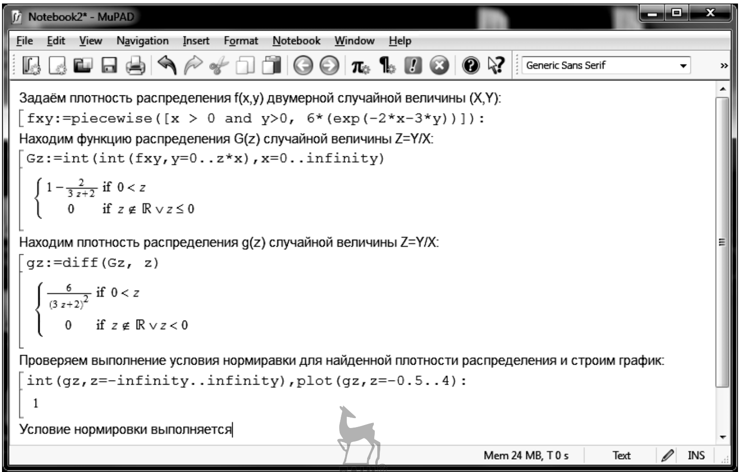

Рис. 1.15.7

Нахождение закона распределения функции двумерной случайной величины

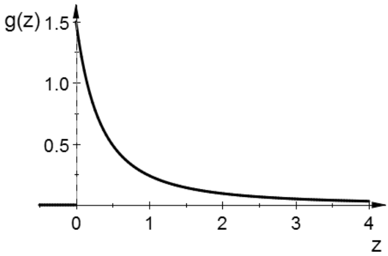

Рис. 1.15.8

График плотности распределения

Частным случаем изложенного выше метода является нахождение закона распределения суммы двух непрерывных случайных величин. Используя формулы (1.15.6), (1.15.7), нетрудно доказать, что в случае, когда $Z = X + Y$, плотность распределения $g(z)$ случайной величины Z определяется одним из интегралов:

$$g(z) = \int_{-\infty}^{\infty} f(z-y,\, y)dy; \qquad (1.15.8)$$

$$g(z) = \int_{-\infty}^{\infty} f(x,\, z-x)dx. \qquad (1.15.9)$$

И, наконец, в самом простом случае, когда случайные величины X и Y независимы, плотность распределения их суммы определяется одним из интегралов свертки плотностей рапределения $f_1(x)$ и $f_2(y)$:

$$g(z) = \int_{-\infty}^{\infty} f_1(z-y)\cdot f_2(y)dy; \qquad (1.15.10)$$

$$g(z) = \int_{-\infty}^{\infty} f_1(x)\cdot f_2(z-x)dx. \qquad (1.15.11)$$

Интегралы (1.15.10), (1.15.11) следуют из интегралов (1.15.8), (1.15.9), с учетом того, что для независимых случайных величин выполняется условие $f(x, y) = f_1(x) \cdot f_2(y)$.

1.15.3.
Задачи для самостоятельного решения

1. Случайная величина X равномерно распределена на отрезке $[0, 2]$. Найти плотность распределения случайной величины $Y = \sqrt{X+1}$.

2. Пусть двумерная случайная величина (X, Y) равномерно распределена внутри треугольника: $D = \{x, y\colon x > 0, y > 0, x + y < 2\}$. Вычислить вероятность неравенства $X > Y$.

3. В условиях предыдущей задачи определить, независимы ли случайные величины X и Y.

4. В условиях задачи 2 вычислить $M[X \cdot Y]$.

5. Пусть X и Y — независимые случайные величины, имеют одинаковую плотность распределения с параметром $\lambda = 2$:

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & \text{при } x \geq 0; \\ 0 & \text{при } x < 0; \end{cases}$$

$$f(y) = \begin{cases} \lambda \cdot e^{-\lambda y} & \text{при } y \geq 0; \\ 0 & \text{при } y < 0. \end{cases}$$

Найти плотность распределения суммы $Z = X + Y$.

2.
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА В СРЕДЕ MATLAB

Математическая статистика — это приложение теории вероятностей к практическим задачам обработки результатов эксперимента.

Основными задачами математической статистики являются:

- определение приближенных значений параметров распределения случайных величин;
- определение закона распределения случайной величины по результатам наблюдений;
- статистическая проверка гипотез относительно закона распределения случайной величины и его параметров;
- многомерный статистический анализ систем случайных величин;
- проведение имитационного моделирования случайных явлений как альтернативы реальному эксперименту;
- проведение экспертных оценок для принятия управленческих решений.

Изучение математической статистики мы начнем с рассмотрения основных понятий выборочного метода. При этом для решения задач будем применять пакеты расширения Statistics Toolbox и символьной математики MuPAD, входящие в систему MATLAB.

2.1.
ВЫБОРОЧНЫЙ МЕТОД. ОЦЕНКА ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

2.1.1.
Генеральная и выборочная совокупности

Пусть в результате наблюдений за некоторой случайной величиной получена определенная совокупность ее числовых значений (исходная выборка значений) и требуется оценить значение какого-либо параметра этой совокупности, например среднее значение при-были для малых предприятий некоторого региона или долю выборщиков, проголосовавших за данного кандидата на выборах.

Различают понятия генеральной и выборочной совокупностей.

Генеральная совокупность — это множество всех мыслимых результатов наблюдений, которые могут быть получены в данных условиях эксперимента.

Для изучения генеральной совокупности из нее выделяют часть — так называемую выборочную совокупность, которую для краткости часто называют выборкой.

Число элементов в исходной выборке, полученной в результате многократных наблюдений, называют объемом выборки.

По выборке находят оценку исследуемого параметра случайной величины — выборочное значение параметра.

На основании этих результатов делают вывод о значении исследуемого параметра для генеральной совокупности.

Нужно подчеркнуть, что полученное таким образом выборочное значение параметра является случайной величиной, поскольку оно яляется функцией результатов наблюдений, имеющих случайный характер.

Теоретической основой описанного выше выборочного метода является закон больших чисел [4]. Согласно этому закону, при неограниченном увеличении объема выборки, случайный выборочный параметр практически достоверно стремится к определенному параметру генеральной совокупности. Говорят: выборочный параметр сходится по вероятности к параметру генеральной совокупности.

Чтобы по исследованию выборки можно было сделать выводы об интересующем нас параметре, нужно, чтобы выборка правильно представляла пропорции генеральной совокупности, т. е. была репрезентативной (представительной).

Учитывая закон больших чисел, можно утверждать, что это условие выполняется, если каждый объект выбран случайно, причем для любого объекта вероятность попасть в выборку одинакова.

При оценке параметров случайной величины по выборочным данным мы будем также использовать термин «выборочная статистика» или просто «статистика».

Определение 2.1.1. Согласно ГОСТ Р50779.24-2005, k-й порядковой статистикой в выборке объема n называют k-е значение x_k в выборке из n значений, расположенных в неубывающем по величине порядке: $x_1 \le x_2 \le ... \le x_n$.

Примерами выборочных статистик являются также статистические оценки таких параметров, как математическое ожидание, дисперсия, мода, медиана случайной величины или оценки параметров закона распределения случайной величины. Термином «статистика» будем также обозначать некоторую функцию случайной величины. Нужный смысл этого термина будет ясен из контекста.

Пусть для изучения дискретной или непрерывной случайной величины X извлечена выборка x_1, x_2, ..., x_m объема n. Причем значение x_1 случайной величины X наблюдалось n_1 раз, значение x_2 наблюдалось n_2 раз, значение x_m наблюдалось n_m раз и $\sum_{i=1}^{m} n_i = n$.

Возможные значения $x_1, x_2, ..., x_m$ случайной величины X принято называть вариантами. Числа n_1, n_2, ..., n_m повторения вариантов называются частотами, числа $w_i = \dfrac{n_i}{n}$ — относительными частотами.

Перечень вариантов и соответствующих им частот (или относительных частот) называется вариационным или статистическим рядом.

Статистический (вариационный) ряд вида

x_i	x_1	x_2	…	x_m
n_i	n_1	n_2	…	n_m

является одной из форм представления исходной выборки. Во второй строке вместо частот n_i могут быть указаны относительные частоты w_i.

Замечание. В математической статистике под законом распределения понимают связь между наблюдаемыми вариантами и соответствующими им частотами, или их относительными частотами.

При исследовании непрерывных случайных величин значения вариантов группируют в интервалы (обычно одинаковой длины), которые будем называть частичными интервалами. В первой строке таблицы указываются частичные интервалы, во второй — число наблюдений случайной величины, попавших в данный интервал, или относительные частоты.

Полученный таким образом статистический ряд является еще одной формой представления исходной выборки. Такой статистический ряд называют интервальным статистическим рядом. Выбор числа и длины частичных интервалов не должен приводить к большой потере информации об исследуемой случайной величине. Пусть значения случайной величины X располагаются на отрезке $[a, b]$, объем выборки — n.

Для определения оптимальной длины частичного интервала можно использовать формулу Стерджесса:

$$\Delta = \frac{b-a}{1+\log_2 n}. \tag{2.1.1}$$

Здесь $k = 1 + \log_2 n$ — число интервалов (берется ближайшее целое число).

Первый частичный интервал начинается в точке

$$x_{\min} = a - \frac{\Delta}{2}.$$

Для определения числа интервалов можно использовать более простую формулу: $k = \sqrt{n}$.

П р и м е р 2.1.1. Пусть измерен рост 50 случайно выбранных человек с точностью до 1 см. В результате измерений получена следующая выборка объема $n = 50$: 175, 179, 170, 163, 159, 171, 170, 152, 168, 172, 160, 167, 165, 167, 156, 170, 181, 153, 163, 167, 179, 172, 170, 186, 180, 187, 178, 175, 168, 168, 171, 173, 178, 170, 183, 181, 180, 160, 165, 158, 173, 160, 167, 172, 180, 169, 168, 170, 188, 176.

Рост является непрерывной случайной величиной, но в силу ограниченной точности измерений любые значения этой величины будут принадлежать некоторому дискретному множеству. Значения роста в выборке изменяются от 152 до 188 см и принимают $m = 37$ различных значений, объем выборки $n = 50$ человек. Нахождение статистических параметров данной выборки в таком виде представляет заметные вычислительные трудности.

Упорядочим данные, входящие в исходную выборку по возрастанию (ранжируем выборку): 152, 153, 156, 158, 159, 160, 160, 160, 163, 163, 165, 165, 167, 167, 167, 167, 168, 168, 168, 168, 169, 170, 170, 170, 170, 170, 170, 171, 171, 172, 172, 172, 173, 173, 175, 175, 176, 178, 178, 179, 179, 180, 180, 180, 181, 181, 183, 186, 187, 188.

Построим интервальный статистический ряд.

Ширина частичного интервала:

$$\Delta = \frac{188-152}{1+\log_2 50} \approx \frac{36}{6,644} \approx 6.$$

Число частичных интервалов: $k \approx 6,644 \approx 7$.

Начало первого частичного интервала в точке $x_{\min} = 152 - \frac{6}{2} = 149.$

Строим интервальный статистический ряд:

X	[149; 155)	[155; 161)	[161; 167)	[167; 173)	[173; 179)	[179; 185)	[185; 191)
n_i	2	6	4	20	7	8	3
w_i	0,04	0,12	0,08	0,4	0,14	0,16	0,06

При изучении статистических рядов наряду с понятием частоты, используется понятие накопленной (кумулятивной) частоты n_i^Σ. Накопленная частота показывает количество наблюдаемых вариантов, значения которых меньше заданного значения x. Отношение накопленной частоты n_i^Σ к общему числу наблюдений n будем называть относительной накопленной (кумулятивной) частотой и обозначать как w_i^Σ. Накопленные частоты (относительные накопленные частоты) находятся последовательным суммированием частот (относительных частот) всех предшествующих частичных интервалов, включая данный интервал. Например, для $x = 173$ в рассмотренном примере накопленная частота: $n_i^\Sigma = 2+6+4+20=32$, т. е. 32 человека имели рост менее 173 см.

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Наиболее часто используют следующие виды графического представления вариационных рядов:

- полигон частот (относительных частот);
- гистограмма частот (относительных частот);
- гистограмма накопленных (кумулятивных) частот;
- эмпирическая (статистическая) функция распределения.

Все перечисленные способы представления вариационных рядов (табличные и графические) являются различными формами статистического закона распределения случайной величины. Выбор формы представления статистического закона распределения зависит от решаемой задачи.

Полигон и гистограмма позволяют выявить преобладающие значения случайной величины и характер распределения частот и относительных частот. Полигон частот (или относительных частот) служит для изображения статистического ряда дискретной случайной величины. Он представляет собой ломаную линию, отрезки которой соединяют точки $(x_1; n_1)$, $(x_2; n_2)$, ..., (x_i, n_i), ... или $(x_1; w_1)$, $(x_2; w_2)$, ..., (x_i, w_i), ...

На рисунке 2.1.1 приведен полигон частот статистического ряда:

X	1	2	3	4	5	6
Частоты (n_i)	22	16	31	24	21	13

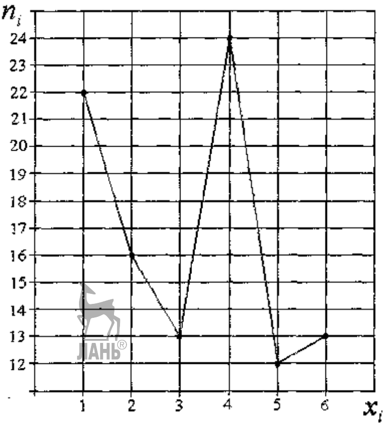

Рис. 2.1.1

Полигон частот

Гистограмма частот (относительных частот) служит для изображения только интервальных статистических рядов. Она представляет собой ступенчатую фигуру, состоящую из прямоугольников, основания которых равны частичным интервалам $\Delta x_i = x_{i+1} - x_i$, а высоты $h_i = \dfrac{n_i}{\Delta x_i}$ $\left(h_i = \dfrac{w_i}{\Delta x_i} \right)$. Площадь всей гистограммы частот равна n (объему выборки), а площадь всей гистограммы относительных частот равна 1.

На рисунке 2.1.2 приведена гистограмма частот интервального статистического ряда.

Над прямоугольниками гистограммы обозначены их площади. Таким образом, высота первого прямоугольника равна $\dfrac{2}{155-149} = \dfrac{1}{3}$, второго — $\dfrac{6}{161-155} = 1$ и т. д.

X	[149; 155)	[155; 161)	[161; 167)	[167; 173)	[173; 179)	[179; 185)	[185; 191)
n_i	2	6	4	20	7	8	3

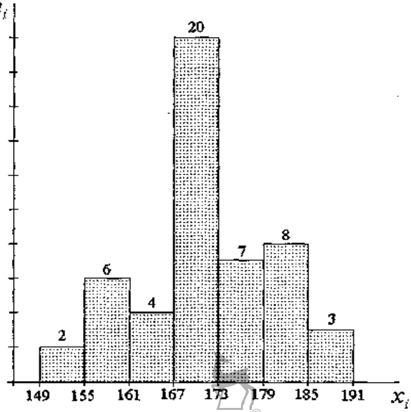

Рис. 2.1.2

Гистограмма частот

На рисунке 2.1.3 построена гистограмма кумулятивных (накопленных) относительных частот для интервального статистического ряда.

X	[149; 155)	[155; 161)	[161; 167)	[167; 173)	[173; 179)	[179; 185)	[185; 191)
w_i	0,04	0,12	0,08	0,4	0,14	0,16	0,06

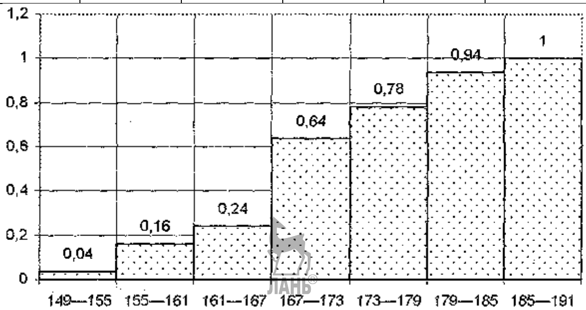

Рис. 2.1.3

Гистограмма кумулятивных частот

Перейдем к рассмотрению понятия эмпирической функции распределения. Пусть задано статистическое распределение случайной величины X. Обозначим через n_x число вариант, меньших заданного значения x, n — объем выборки. Относительная частота события $(X < x)$ равна n_x/n. При изменении x меняется и относительная частота, т. е. n_x/n есть функция от x. Поскольку эта функция строится по данным опыта, ее называют эмпирической.

Эмпирической функцией распределения (функцией распределения выборки) $F^*(x)$ называется относительная частота события $(X < x)$:

$$F^*(x) = \frac{n_x}{n}, \tag{2.1.2}$$

где n_x — число вариантов, меньших x; n — объем выборки.

Теоретической функцией распределения называется функция распределения $F(x) = P(X < x)$ случайной величины X, вычисленная по генеральной совокупности.

Заметим, что эмпирическая и теоретическая функции распределения определяются единообразно, но по различным совокупностям: эмпирическая — по выборочной, теоретическая — по генеральной совокупности. Поэтому свойства эмпирической функции распределения полностью аналогичны свойствам теоретической функции распределения случайной величины X.

Согласно закону больших чисел, при возрастании объема выборки n, различия между функциями $F^*(x)$ и $F(x)$ уменьшаются.

На рисунке 2.1.4 построена эмпирическая функция распределения для статистического ряда.

Для $x \le 1$ условие $(X < x)$ не может быть выполнено, т. е. $F^*(x) = 0$ при $x \le 1$. Для $1 < x \le 2$ условие $(X < x)$ выполняется в 22 случаях из 100, т. е. $F^*(x) = 0{,}22$ при $1 < x \le 2$. Продолжая, получаем:

$$F(x) = \begin{cases} 0 & x \le 1; \\ 0{,}22 & 1 < x \le 2; \\ 0{,}38 & 2 < x \le 3; \\ 0{,}51 & 3 < x \le 4; \\ 0{,}75 & 4 < x \le 5; \\ 0{,}87 & 5 < x \le 6; \\ 1 & x > 6. \end{cases}$$

X	1	2	3	4	5	6
Частоты (n_i)	22	16	13	24	12	13
Относительные частоты (w_i)	0,22	0,16	0,13	0,24	0,12	0,13

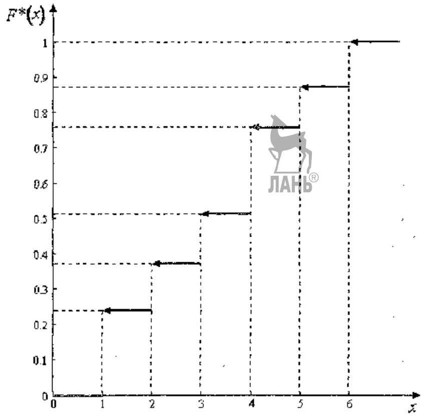

Рис. 2.1.4

Эмпирическая функция распределения

2.1.2. Точечные оценки параметров распределения

Основные точечные статистические оценки параметров случайной величины X и способы их вычисления сведены в таблице 2.1.1.

Таблица 2.1.1

Наименование	**Обозначение**	**Способ вычисления**
1	**2**	**3**
Выборочное математическое ожидание	$\overline{X}$	$\frac{1}{n}\sum_{i=1}^{m} x_i n_i$, где $n=\sum_{i=1}^{m} n_i$ (для вариационного ряда), или $\frac{1}{n}\sum_{i=1}^{n} x_i$ (для исходной выборки)

1	2	3
Исправленная выборочная дисперсия	s_x^2	$\frac{1}{n-1}\sum_{i=1}^{m}(x_i-\bar{X})^2\cdot n_i$ (для вариационного ряда), или $\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{X})^2$ (для исходной выборки)
Выборочное стандартное (среднеквадратичное) отклонение	s_x	$\sqrt{s_x^2}$
Выборочный коэффициент эксцесса	E_x	$\frac{1}{n-2}\cdot\frac{\sum_{i=1}^{m}(x_i-\bar{X})^4\cdot n_i}{s_x^4}-3$
Выборочный коэффициент ассиметрии	A_x	$\frac{1}{n-2}\cdot\frac{\sum_{i=1}^{m}(x_i-\bar{X})^3\cdot n_i}{s_x^3}$
Выборочная медиана	*Med*	По ГОСТ Р50779.24-2005
Размах выборки	R	$R = x_{\max} - x_{\min}$

Замечание. Выборочная медиана, согласно ГОСТ, указаному в таблице 2.1.1, находится путем нумерации элементов выборки, расположенной в порядке неубывания и выбора значения:

- $\frac{n+1}{2}$-й порядковой статистики (см. определение 2.1.1 в п. 2.1.1), если n нечетное;
- $\left(\frac{n}{2}+1\right)$-й порядковой статистики, если n нечетное.

К выборочным оценкам параметров случайной величины предъявляются следующие требования.

1. Состоятельность, которая означает, что с увеличением объема выборки n вероятность ошибки при данной оценке параметра распределения уменьшается.

2. Несмещенность. Требование несмещенности гарантирует отсутствие систематической ошибки при оценивании параметра.

3. Эффективность, которая означает, что выбранная несмещенная оценка параметра, по сравнению с другими возможными его оценками, обладала наименьшей дисперсией.

Совместить выполнение требований несмещенности и эффективности оценки параметра распределения удается не всегда.

Первостепенное значение придается условию несмещенности оценки. Это означает, что мы, в первую очередь, ограничиваемся рассмотрением только несмещенных оценок, а затем уже пытаемся минимизировать их дисперсию.

На практике не всегда удается удовлетворить всем этим требованиям. Например, может оказаться, что даже если эффективная оценка существует, формулы для ее вычисления оказываются слишком сложными, и приходится удовлетворяться другой оценкой, дисперсия которой несколько больше. Иногда, для упрощения расчетов, применяются незначительно смещенные оценки, обладающие большей дисперсией по сравнению с эффективными оценками.

2.1.3. Обработка первичной статистической информации в среде MATLAB

В этом параграфе мы научимся вводить данные в программу, проводить их предварительную сортировку, вычислять точечные оценки параметров генеральной совокупности и строить гистограмму распределения случайной величины.

Пусть выборка значений случайной величины X хранится в текстовом файле с заданным именем в виде либо одной вектор-строки, либо одного вектор-столбца, либо матрицы.

Для считывания исходных данных из файла удобно использовать вкладку Import Data... из основного меню File окна MATLAB. В раскрывшемся списке следует дважды кликнуть левой кнопкой мыши имя требуемого файла данных и далее следовать указаниям мастера импорта данных.

Пример 2.1.2. В этом примере для импорта данных в рабочую среду MATLAB будем использовать файл *xrayl.txt*, расположенный в папке с именем *bin* системы MATLAB. Импортированные данные отобразятся в окне Workspace рабочей среды MATLAB.

Для их просмотра достаточно дважды кликнуть левой кнопкой мыши по иконке файла данных. Откроется окно с массивом данных.

Если выборка представлена в виде двумерного массива данных, т. е. в виде матрицы, то после ввода преобразуем данные в вектор-столбец, т. е. в одномерный массив. Это можно осуществить с помощью команды *sort*(*x*(:)), которая также осуществляет сортировку данных, выстраивая их в неубывающем порядке. Перед выполнением этой команды значения, хранящиеся в импортированном файле данных нужно присвоить переменной x.

Объем выборки n можно определить с помощью команды *length*(x).

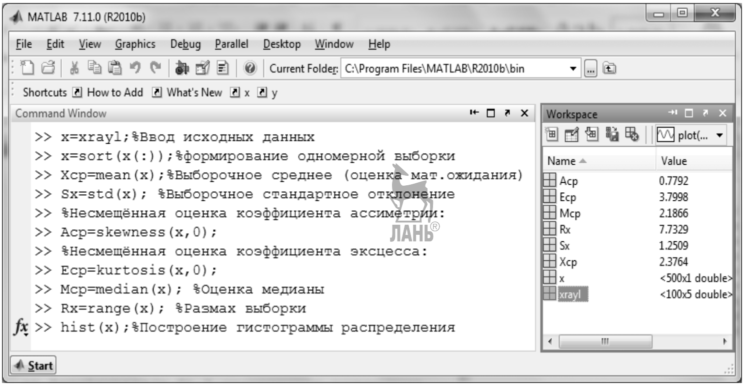

Рис. 2.1.5

Расчет точечных оценок параметров случайной величины X в рабочей среде MATLAB

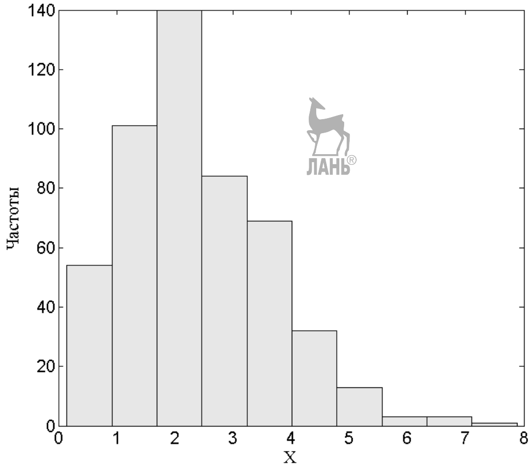

Рис. 2.1.6

Гистограмма распределения случайной величины X

Синтаксис команд, вычисляющих оценки параметров случайной величины, представленных в таблице 2.1.1, будет ясен из рисунка 2.1.5.

Все рассчитанные значения оценок параметров отображены в окне Workspace рабочей среды MATLAB (рис. 2.1.5).

С помощью функции *hist*(*x*) (заключительная команда на рисунке 2.1.5) в отдельном окне строится гистограмма распределения случайной величины X (рис. 2.1.6).

2.1.4. Интервальная оценка математического ожидания

Чтобы дать представление о точности и надежности оценки того или иного параметра случайной величины в математической статистике используются так называемая интервальная оценка каждого параметра и связанные с ней понятия доверительного интервала и доверительной вероятности.

Определение 2.1.2. Интервальной оценкой неизвестного параметра распределения называется числовой интервал, называемый доверительным интервалом, который с заданной вероятностью накрывает неизвестное значение этого параметра. Вероятность интервальной оценки называется доверительной вероятностью.

Таким образом, интервальная оценка дополняет точечную оценку $\hat{\theta}$ некоторого параметра θ информацией о ее случайной погрешности Δ, т. е. характеризуется доверительным интервалом, в пределах которого с заданной доверительной вероятностью может находиться оцениваемый параметр. Иными словами, интервальная оценка характеризует так называемую предельную ошибку выборки как модуль разности между истинным и выборочным значениями параметра: $\Delta = |\theta - \hat{\theta}|$. Возможные значения Δ с заданной доверительной вероятностью β находятся внутри доверительного интервала $(\varepsilon_1;\ \varepsilon_2)$. Величины ε_1 и ε_2 есть левая и правая границы доверительного интервала.

Так как для несмещенных оценок параметров математическое ожидание предельной ошибки выборки равно 0, т. е. $M[\Delta] = 0$, то

$$P(\varepsilon_1 \le \Delta \le \varepsilon_2) = \beta,$$

где β — заданная доверительная вероятность.

Проанализируем результаты интервальной оценки математического ожидания, полученные тремя способами: метод интервальной оценки при больших объемах выборки, метод, основанный на распределении Стьюдента, с помощью MATLAB.

Первый метод оценки математического ожидания при больших объемах выборки ($n > 30$) основан на предположении о нормальном распределении выборочного среднего значения $\bar{X}$. Справедливость этого предположения вытекает из центральной предельной теоремы теории вероятностей [4]. Тогда можно утверждать: вероятность того, что отклонение выборочного среднего значения $\bar{X}$ от математического ожидания m_X не превзойдет (по абсолютной величине) число $\Delta > 0$, равна:

$$P(|\bar{X} - m_x| \le \Delta) = \Phi(u) = \beta, \tag{2.1.3}$$

где

$$u = \frac{\Delta}{\sigma_{\bar{X}}} \approx \frac{\Delta}{s_{\bar{X}}} = \frac{\Delta}{s_X / \sqrt{n}},$$

β — доверительная вероятность; $\Delta = u \dfrac{s_X}{\sqrt{n}}$ — предельная ошибка выборки,

$$\Phi(u) = \frac{2}{\sqrt{2\pi}} \int_0^u e^{-x^2/2} dx$$

— интеграл Лапласа.

Из (2.1.3) следует, что интервальная оценка (доверительный интервал) для математического ожидания m_X определяется неравенством

$$\bar{X} - \Delta \le m_X \le \bar{X} + \Delta \tag{2.1.4}$$

или

$$\bar{X} - u\frac{s_X}{\sqrt{n}} \le m_X \le \bar{X} + u\frac{s_X}{\sqrt{n}}. \tag{2.1.5}$$

Параметр u может быть найден по заданной доверительной вероятности β из равенства

$$\Phi(u) = 2N(u) - 1 = \beta, \tag{2.1.6}$$

где $N(u)$ — стандартная функция нормального распределения (п. 1.12.2).

Тогда $N(u) = \dfrac{1+\beta}{2}$. Далее вычисляем $N(u)$ и по таблице приложения 1 находим u. Затем по формуле (2.1.5) вычисляем доверительный интервал для математического ожидания m_X.

П р и м е р 2.1.3. Для выборки примера 2.1.2 получить интервальную оценку математического ожидания m_X с доверительной вероятностью $\beta = 0{,}95$.

Р е ш е н и е. При решении примера 2.1.2 для выборки объема $n = 500$ были получены следующие точечные оценки математического ожидания и стандартного отклонения: $\bar{X} \approx 2{,}3764$, $s_X \approx 1{,}2509$.

1. Определяем значение стандартной функции нормального распределения по заданной доверительной вероятности $\beta = 0{,}95$:

$$N(u) = \frac{1+\beta}{2} = \frac{1+0{,}95}{2} = 0{,}975.$$

2. Из таблицы приложения 1 для $N(u) = 0{,}975$ находим аргумент $u = 1{,}96$.

3. Вычисляем предельную ошибку выборки:

$$\Delta = u\frac{s_X}{\sqrt{n}} = 1{,}96\Delta\frac{1{,}2509}{\sqrt{500}} \approx 0{,}1096.$$

4. По формуле (2.1.4) находим интервальную оценку для математического ожидания: $2{,}2668 \le m_X \le 2{,}4860$ с доверительной вероятностью $\beta = 0{,}95$.

Второй метод позволяет найти интервальную оценку математического ожидания при неизвестном законе распределения случайной величины X. При этом задача успешно решается как при выборках малого объема ($n < 30$), так и при выборках большого объема.

Метод интервальной оценки математического ожидания в этом случае основан на использовании t-распределения. Другое название этого распределения — распределение Стьюдента. Это распределение нормированного значения t выборочного среднего $\bar{X}$. Случайная величина t определяется формулой

$$t = \frac{\bar{X} - m_x}{s_{\bar{X}}} = \frac{\bar{X} - m_x}{s_x / \sqrt{n}}, \qquad (2.1.7)$$

где m_x — математическое ожидание случайной величины X; $s_{\bar{X}}$ — стандартное отклонение выборочного среднего $\bar{X}$ от математического ожидания m_X; s_x — стандартное отклонение случайной величины X от математического ожидания m_X; n — объем выборки.

Смысл перехода к этой переменной состоит в том, что ее закон распределения, в отличие от закона распределения $\bar{X}$, не зависит от неизвестных параметров закона распределения случайной величины X, а зависит только от одного параметра — числа степеней свободы k, связанного с объемом выборки n формулой $k = n - 1$.

Вероятностное распределение переменной t приближается к нормальному распределению, а при больших объемах выборки практически совпадает с ним.

Помимо нахождения доверительных интервалов для математического ожидания случайной величины X, t-распределение (или распределение Стьюдента) используется также при проверке гипотез. Речь об этом пойдет в следующих параграфах.

С доверительной вероятностью β связана величина α, называемая уровнем значимости:

$$\alpha = 1 - \beta. \tag{2.1.8}$$

Смысл величины α при интервальном оценивании параметра t-распределения можно понять из рисунка 2.1.7.

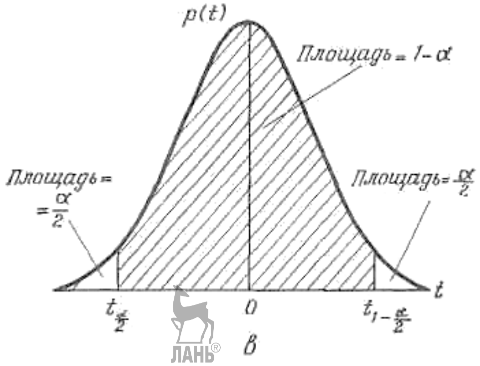

Рис. 2.1.7

Интервальная оценка по заданному уровню значимости α

На рисунке 2.1.7 площадь заштрихованной области равна доверительной вероятности $\beta = 1 - \alpha$.

Пользуясь терминологией статистического контроля качества продукции, можно сказать, что вероятность α представляет «риск поставщика», связанный с забраковкой по результатам выборочного контроля всей партии, удовлетворяющей стандарту.

Вероятность β — «риск потребителя», связанный с принятием по анализу выборки партии, не удовлетворяющей стандарту.

Зная t-распределение, можно найти такое критическое значение $t_{1-\alpha/2;\nu}$, что вероятность того, что статистика t (2.1.7) не превзойдет по абсолютной величине значения $t_{1-\alpha/2;\nu}$, будет равна β:

$$P\left(\left|\frac{\bar{X}-m_x}{s_X}\cdot\sqrt{n}\right|\le t_{1-\alpha/2;\nu}\right)=\beta, \qquad (2.1.9)$$

где β — доверительная вероятность.

Выражение (2.1.9) равносильно формуле

$$P\left(|\bar{X}-m_X|\le t_{1-\alpha/2;\nu}\cdot\frac{s_X}{\sqrt{n}}\right)=\beta. \qquad (2.1.10)$$

Доверительный интервал находится по формуле

$$\bar{X}-\Delta\le m_X\le\bar{X}+\Delta, \qquad (2.1.11)$$

где величина

$$\Delta=t_{1-\alpha/2;\nu}\cdot\frac{s_X}{\sqrt{n}} \qquad (2.1.12)$$

есть предельная ошибка выборки.

Величина доверительного интервала существенно зависит от объема выборки n (уменьшается с ростом n) и от значения доверительной вероятности β (увеличивается с приближением β к единице).

Критические значения нормированной переменной $t_{1-\alpha/2;\nu}$ для различных значений доверительной вероятности β и степеней свободы $k = n - 1$ приводятся в приложении 2.

П р и м е р 2.1.4. Дана выборка объема $n = 8$ случайной величины X: 76,48; 76,43; 77,20; 76,45; 76,25; 76,48; 76,48; 76,60.

Требуется получить интервальную оценку математического ожидания m_X с доверительной вероятностью β = 0,95.

Р е ш е н и е. Число степеней свободы: $k = n - 1 = 7$.

Выборочное среднее: $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}x_i=76,546.$

Выборочная дисперсия: $s_X^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=0,0790.$

Выборочное стандартное отклонение: $s_X \approx 0,2811$.

Уровень значимости: $\alpha = 1 - \beta = 0,05$.

Из таблицы приложения 2 для $\beta = 0,95$ и $k = 7$ находим $t_{0,975;\,7} = 2,365$.

Согласно формуле (2.1.12), предельная ошибка выборки:

$$\Delta = 2,365 \cdot \frac{0,2811}{\sqrt{8}} \approx 0,235.$$

Согласно формуле (2.1.11), доверительный интервал:

$$76,311 \le m_x \le 76,781.$$

П р и м е р 2.1.5. Для условий примера 2.1.5 найдем доверительный интервал математического ожидания с помощью MATLAB. Для нахождения доверительного интервала используется команда *ttest*, реализующая вычисление на основе распределения Стьюдента. Один из возможных вариантов синтаксиса этой команды:

• $[h, p, ci] = ttest(x, Mx, alpha)$, где x — вектор, содержащий выборку случайной величины X;

• Mx — выборочное среднее значение случайной величины X;

• *alpha* — уровень значимости;

• *ci* — вектор, содержащий вычисленные значения левой и правой границ доверительного интервала.

Параметры h и p для решения данной задачи не используются. Их смысл будет ясен из дальнейшего, при рассмотрении темы статистической проверки гипотез.

Функция *ttest* находит доверительный интервал по заданной выборке и известной точечной оценке математического ожидания случайной величины при заданном уровне значимости.

Найденные границы доверительного интервала *ci* отображаются в окне Workspace рабочей среды MATLAB (рис. 2.1.8). Простота решения задачи с помощью MATLAB очевидна.

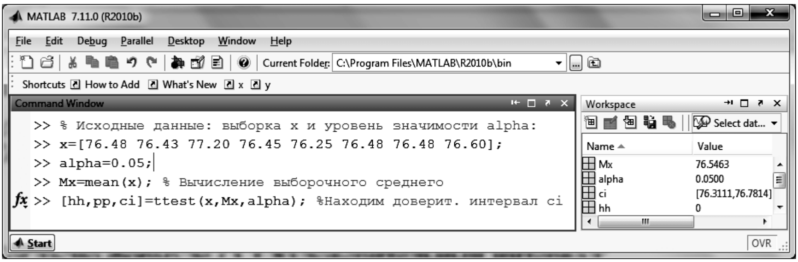

Рис. 2.1.8

Интервальная оценка математического ожидания в рабочей среде MATLAB

Как видим, результаты расчета доверительного интервала по выборке малого объема, полученные в примерах 2.1.4 и 2.1.5, совпадают.

Следует заметить, что решение этой же задачи первым методом, в основу которого положено предположение о нормальном законе распределения выборочного среднего, даст менее точный результат: [76,3514; 76,7412], который существенно отличается от результатов в примерах 2.1.4 и 2.1.5. Это объясняется тем, что при малых объемах выборки предположение о нормальном распределении выборочного среднего оказалсь несостоятельным.

При больших объемах выборки все три способа решения задачи дают одинаковый результат.

2.1.5.
Интервальная оценка дисперсии

Для интервального оценивания дисперсии нужно знать закон распределения ее точечной оценки s_X^2, который связан с известным распределением χ^2.

Установлено, что случайная величина $\sum_{i=1}^{n}\frac{(x_i-\bar{X})^2}{\sigma_X^2}$ имеет χ^2-распределение с $(n-1)$ степенями свободы (распределение Пирсона), где n — число наблюдений. То есть если число степеней свободы равно $(n-1)$, то $\chi^2=\sum_{i=1}^{n}\frac{(x_i-\bar{X})^2}{\sigma_X^2}$. Умножим и разделим правую часть этого равенства на $n-1$ и получим $\chi^2=\frac{(n-1)s_X^2}{\sigma_X^2}$, т. е. величина $\frac{(n-1)s_X^2}{\sigma_X^2}$ имеет распределение χ^2 с $n-1=k$ степенями свободы. Тогда можно найти распределение оценки s_X^2 дисперсии σ_X^2 по распределению χ^2:

$$s_X^2=\frac{\sigma_X^2}{n-1}\chi^2. \qquad (2.1.13)$$

Зная закон распределения случайной величины χ^2, можно найти интервал, в который оценка дисперсии s_X^2 попадает с заданной доверительной вероятностью β. Закон распределения χ^2 имеет вид, изображенный на рисунке 2.1.9.

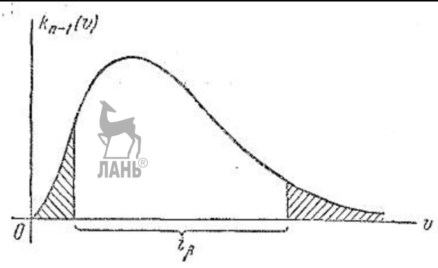

Рис. 2.1.9

Вид закона распределения χ^2 Пирсона

Интервал l_β выбирают так, чтобы вероятности выхода случайной величины χ^2 за его пределы вправо и влево (заштрихованные площади) были одинаковы и равны $\frac{\alpha}{2} = \frac{1-\beta}{2}$.

Чтобы построить интервал с таким свойством пользуются таблицей приложения 3. В ней приведены такие числа $\chi^2_{\text{кр}}$, что

$$P(\chi^2 > \chi^2_{\text{кр}}) = p,$$

где p — некоторое заданное значение вероятности.

Для числа степеней свободы $k = n - 1$ находим в соответствующей строке таблицы приложения 3 два критических значения $\chi^2_{\text{кр}}$: одно, отвечающее вероятности $p_1 = \frac{\alpha}{2}$; другое — вероятности $p_2 = 1 - \frac{\alpha}{2}$. Обозначим эти значения χ^2_1 и χ^2_2. Они называются квантилями распределения χ^2 Пирсона и являются границами интервала l_β.

Тогда искомая интервальная оценка дисперсии может быть записана в виде неравенства:

$$\frac{s_X^2 \cdot \nu}{\chi_1^2} < D_x < \frac{s_X^2 \cdot \nu}{\chi_2^2}. \tag{2.1.14}$$

П р и м е р 2.1.6. Дана выборка объема $n = 8$ случайной величины X: 76,48; 76,43; 77,20; 76,45; 76,25; 76,48; 76,48; 76,60.

Требуется получить интервальную оценку дисперсии D_X с доверительной вероятностью $\beta = 0{,}9$.

Р е ш е н и е. Решим эту задачу двумя способами: только что описанным способом и с помощью MATLAB.

1. Имеем $\alpha = 1 - \beta = 0{,}1$; $\dfrac{\alpha}{2} = 0{,}05$, $k = n - 1 = 7$. По таблице приложения 3 находим квантили χ_1^2 и χ_2^2 распределения Пирсона при $k = 7$: для $p_1 = \dfrac{\alpha}{2} = 0{,}05 \to \chi_1^2 = 14{,}1$; для $p_2 = 1 - \dfrac{\alpha}{2} = 0{,}95 \to \chi_2^2 = 2{,}17$. Оценка дисперсии $s_X^2 = 0{,}079$ (см. пример 2.1.4).

По формуле (2.1.14) находим доверительный интервал для дисперсии:

$$0{,}0392 < D_x < 0{,}2548.$$

2. Решим ту же задачу с помощью MATLAB. В пакете расширения MATLAB Statistics Toolbox для нахождения интервальной оценки дисперсии случайной величины можно воспользоваться функцией *vartest*, реализующую вычисление границ доверительного интервала на основе распределения χ^2 (распределения Пирсона). Один из возможных вариантов синтаксиса этой функции:

- $[h,\ p,\ ci] = vartest(x,\ Dx,\ alpha)$, где x — вектор, содержащий выборку случайной величины X;
- Dx — выборочная дисперсия случайной величины X;
- $alpha$ — уровень значимости;
- ci — вектор, содержащий вычисленные значения левой и правой границ доверительного интервала.

Параметры h и p для решения данной задачи не используются. Их смысл будет ясен из дальнейшего, при рассмотрении темы статистической проверки гипотез.

Результаты представлены на рисунке 2.1.10.

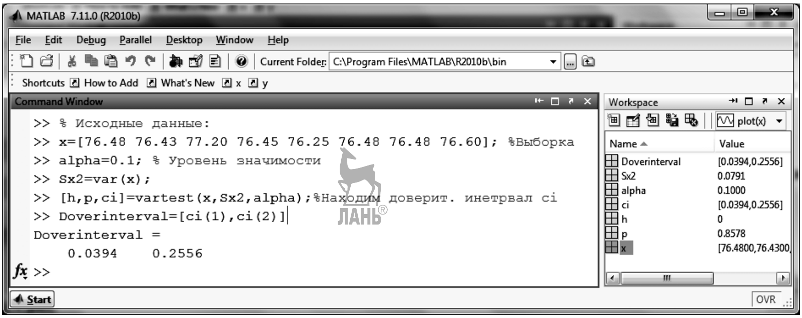

Рис. 2.1.10

Интервальная оценка дисперсии в рабочей среде MATLAB

Преимущество MATLAB при нахождении интервальной оценки дисперсии демонстрирует следующий пример.

П р и м е р 2.1.7. Для выборки объемом $n = 500$ примера 2.1.2 получить интервальную оценку дисперсии D_X с доверительной вероятностью $\beta = 0{,}95$.

Р е ш е н и е. Имеем $\alpha = 1 - \beta = 0{,}05;\ \dfrac{\alpha}{2} = 0{,}025,\ k = n - 1 = 499.$ Здесь мы сталкиваемся с проблемой нахождения квантилей χ_1^2 и χ_2^2 распределения Пирсона по таблице приложения 3 при $\nu = 499$: максимальное число степеней свободы в таблице $\nu = 30$. MATLAB позволяет решить эту проблему (рис. 2.1.11).

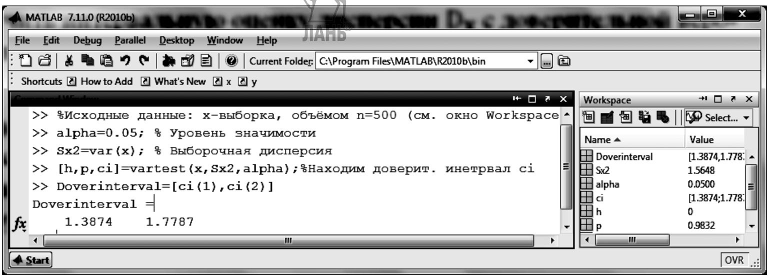

Рис. 2.1.11

Интервальная оценка дисперсии для выборки большого объема в рабочей среде MATLAB

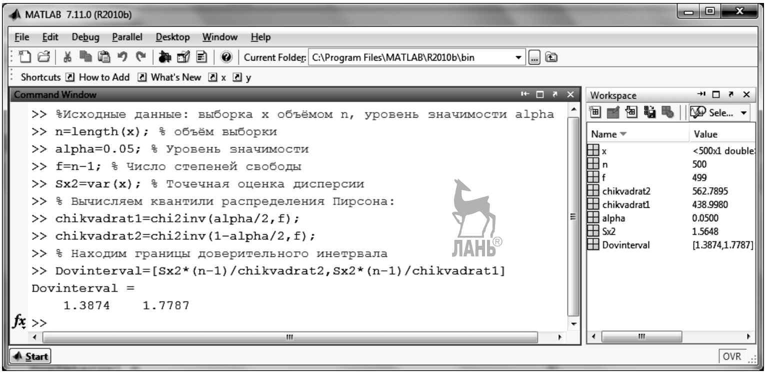

Рис. 2.1.12

Альтернативный способ интервальной оценки дисперсии в рабочей среде MATLAB

В рабочей среде MATLAB ту же задачу можно решить с помощью функции *chi2inv*(p, k), которая вычисляет обратное значение функции распределения χ^2 по заданным значениям вероятности $p_1 = \frac{\alpha}{2}$ или $p_2 = 1 - \frac{\alpha}{2}$ для любого числа степеней свободы. Пример такого решения представлен на рисунке 2.1.12.

Результаты интервальной оценки дисперсии тем и другим способом совпадают.

2.1.6. Критерий максимального правдоподобия

Следует подчеркнуть, что рассмотренные выше функции *ttest* и *vartest* можно применять для точечных и интервальных оценок параметров генеральной совокупности, закон распределения которой зависит от двух параметров: математического ожидания и стандартного отклонения. Это, прежде всего, нормальный закон распределения.

На практике часто приходится иметь дело со случайными величинами, распределенными по так называемым однопараметрическим законам распределения, т. е. зависящим от одного параметра. И далеко не всегда этим параметром является математическое ожидание или дисперсия случайной величины. В качестве примеров можно привести распределение χ^2, распределение Пуассона.

В распределении Пуассона параметром является математическое ожиданию случайной величины, которое равно дисперсии.

Рассмотрим метод точечной и интервальной оценки параметров распределения, основанный на критерии максимального правдоподобия. Метод максимального правдоподобия (ММП) позволяет находить эффективные и состоятельные оценки параметров законов распределения, зависящих от одного, двух и более параметров. Далее для общности будем обозначать исследуемый параметр распределения θ, а его выборочную оценку $\hat{\theta}$.

Если $\hat{\theta}$ является макимально правдоподобной оценкой некоторого параметра θ закона распределения $f(\theta)$, то $f(\hat{\theta})$ будет максимально правдоподобной оценкой закона распределения $f(\theta)$.

Рассмотрим коротко суть метода максимального правдоподобия на примере однопараметрического закона распределения случайной величины X. Согласно ММП, в качестве оценок выбираются те значения параметров, при которых данные результаты наблюдения «наиболее вероятны».

Будем предполагать, что результаты $x_1, x_2, ..., x_n$ наблюдений являются взаимно независимыми случайными величинами с одним и тем же распределением вероятностей, зависящим от одного неизвестного параметра θ. Для придания точного смысла принципу «наибольшей вероятности» поступают следующим образом. Вводят функцию от переменных $x_1, x_2, ..., x_n$ и параметра θ:

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta), \tag{2.1.15}$$

где $f(x_i; \theta)$ — значение плотности распределения случайной величины X при фиксированных величинах x_i и θ.

Формула (2.1.15) имеет силу для непрерывной случайной величины X.

Для дискретной случйной величины функция $L(x_1, x_2, ..., x_n; \theta)$ определяется формулой

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} P(X = x_i; \theta), \tag{2.1.16}$$

где $P(X = x_i; \theta)$ — вероятность того, что случайная величина X в i-м наблюдении примет значение x_i при фиксированном параметре распределения θ.

Случайную функцию $L(x_1, x_2, ..., x_n; \theta)$, рассматриваемую как функцию параметра θ при фиксированных значениях $x_1, x_2, ..., x_n$, называют функцией правдоподобия.

Поэтому функцию правдоподобия обычно записывают в виде $L(\theta|x_1, x_2, ..., x_n)$. Такое обозначение лучше отражает ее смысл. А именно, если вероятность позволяет предсказывать неизвестные результаты, основанные на знании параметров распределения, то правдоподобие позволяет оценивать неизвестные параметры по известным результатам наблюдений.

Оценкой максимального правдоподобия параметра θ называют такое случайное значение $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$, при котором функция правдоподобия достигает наибольшего возможного значения. Часто удобно пользоваться натуральным логарифмом функции правдоподобия. Поскольку логарифмическая функция является монотонной, то точка максимума для lnL та же, что и для L, и для нахождения оценок максимального правдоподобия следует решить так называемое уравнение правдоподобия:

$$\frac{d \ln L(\theta|x_1, x_2, \ldots, x_n)}{d\theta} = 0. \tag{2.1.17}$$

Данные выше определения непосредственно обобщаются и на случай нескольких неизвестных. В этом случае в формуле (2.1.17) находятся частные призводные по каждому параметру распределения.

П р и м е р 2.1.8. Случайная величина X — время безотказной работы изделия, имеет распределение, заданное статистическим рядом:

x_i	5	15	25	35	45	55	65
n_i	365	245	150	100	70	45	25

Найти методом максимального правдоподобия точечную оценку параметра (параметров) распределения случайной величины X.

Р е ш е н и е. Случайная величина X (время) по своей природе является непрерывной случайной величиной. Поэтому заданный статистический ряд будем считать интервальным статистическим рядом, в котором значения x_i являются центрами соответствующих интервалов.

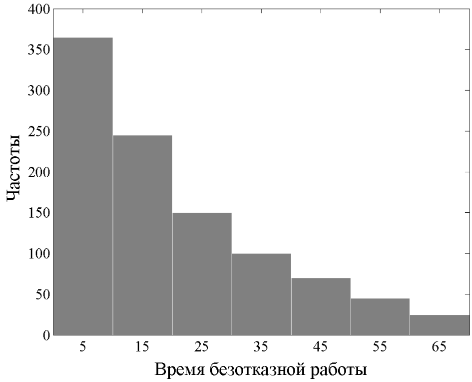

Рис. 2.1.13

Гистограмма распределения времени безотказной работы изделия

По виду гистограммы есть основания предполагать, что время безотказной работы имеет экспоненциальное распределение, плотность вероятности которого, как известно [4], определяется выражением

$$f(x;\, \theta) = \theta \cdot e^{-\theta x},$$

где θ — параметр распределения, который требуется найти.

Функция правдоподобия в нашем случае:

$$L(\theta \mid x) = \prod_{i=1}^{7} f(x_i; \theta) = \prod_{i=1}^{7} \theta \cdot e^{-\theta x_i} = \theta^7 \prod_{i=1}^{7} e^{-\theta x_i},$$

где $x = (x_1, x_2, ..., x_7)$ — массив вариантов x_i заданного статистического ряда.

Тогда логарифмическая функция правдоподобия

$$\ln L(\theta \mid x) = \ln\left(\theta^7 \prod_{i=1}^{7} e^{-\theta x_i}\right) = 7\ln\theta + \ln\prod_{i=1}^{7} e^{-\theta x_i} =$$

$$= 7\ln\theta + \sum_{i=1}^{7} \ln(e^{-\theta x_i}) = 7\ln\theta - \theta\sum_{i=1}^{7} x_i.$$

Условие экстремума (2.1.17):

$$\frac{d\ln L(\theta \mid \mathbf{x})}{d\theta} = \frac{7}{\theta} - \sum_{i=1}^{7} x_i = 0$$

или

$$\frac{1}{\theta} - \frac{1}{7}\sum_{i=1}^{7} x_i = \frac{1}{\theta} - \overline{X} = 0.$$

Отсюда получаем оценку $\hat{\theta}$ параметра θ:

$$\hat{\theta} = \frac{1}{\overline{X}},$$

где

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{m} x_i n_i = \frac{1}{1000}(5 \cdot 365 + 15 \cdot 24 + \ldots + 65 \cdot 25) = 20.$$

Окончательно

$$\hat{\theta} = \frac{1}{\overline{X}} = \frac{1}{20} = 0{,}05.$$

Решим эту же задачу в рабочей среде MATLAB, используя функцию *mle* (maximum likelihood estimates), с расширенным синтаксисом:

[phat, pci] = mle(data, ′Distribution′, name, ′alpha′, alpha),

где *phat* — в общем случае вектор, содержащий искомые точечные оценки параметров распределения (в нашем случае этот вектор будет состоять из одного элемента: $\text{phat} = \hat{\theta}$);

pci — в общем случае матрица с интервальными оценками параметров (в нашем случае эта матрица будет состоять из одного вектор-столбца, содержащего левую и правую границы доверительного интервала для параметра θ);

data — массив исходных данных в виде первичной выборки, где значения случайной величины могут повторяться (если данные в условии задачи представлены в виде вариационного ряда, то его надо преобразовать соответствующим образом);

name — вид распределения, который может принимать значения: 'normal' (нормальное распределение), 'exponential' (экспоненциальное распределение), 'binomial' (биномиальное распределение), 'poisson' (распределение Пуассона), 'rayleigh' (распределение Рэлея) и т. д.;

alpha — уровень значимости.

Более подробную информацию можно получить через меню «Help» в командном окне MATLAB.

Параметры *phat* и *pci* являются формальными параметрами, поэтому их обозначения могут быть выбраны пользователем произвольно в рамках правил системы MATLAB.

Решение задачи представлено на рисунке 2.1.14.

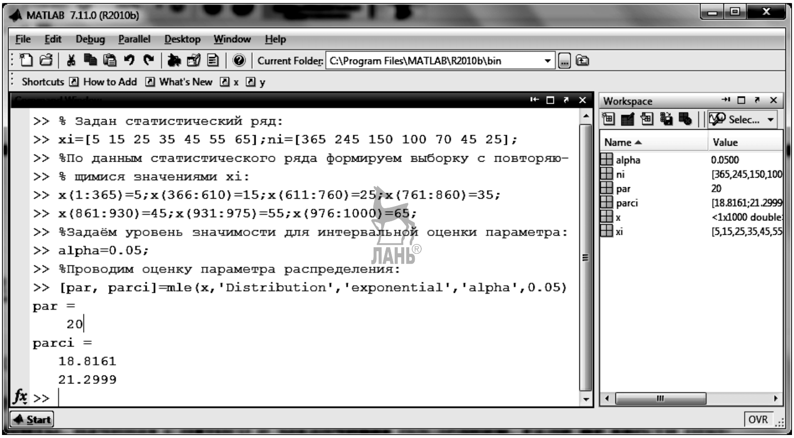

Рис. 2.1.14

Нахождение точечных и интервальных оценок параметра распределения методом максимального правдоподобия

Видим полное совпадение точечных оценок параметра распределения при решении задачи с помощью MATLAB и традиционным способом. Кроме того, решая эту задачу в среде MATLAB, мы получили еще и интервальную оценку параметра θ: [18,8161; 21,2999].

Добавим к сказанному, что гистограмма (рис. 2.1.13), которая позволила нам сделать предположение о виде закона распределения, была получена также с помощью MATLAB. Для этого, удерживая клавишу Ctrl, кликаем левой кнопкой в окне Workspace последовательно иконки *xi* и *ni* (значения вариантов и частот заданного статистического ряда). Затем в меню в окна Workspace выбираем вид графика и редактируем полученный график нужным образом.

2.1.7. Задания для самостоятельной работы

По заданной доверительной вероятности β и заданной выборке случайной величины X найти точечные и интервальные оценки ее математического ожидания и дисперсии. Предполагаем, что случайная величина X подчиняется нормальному закону распределения.

Для получения выборки следует использовать команду MATLAB: *x* = *normrnd*(*Mx*, *sigma*, 1, *n*), где *Mx* и *sigma* — математическое ожидание m_X и стандартное отклонение σ_X генеральной совокупности, *n* — объем выборки. Команда *normrnd* генерирует вектор-строку *x* (выборку) случайных значений генеральной совокупности, имеющей нормальный закон распределения с заданными параметрами.

Таблица 2.1.2

Исходные данные для выполнения самостоятельной работы

№	Параметры генеральной совокупности m_X	Параметры генеральной совокупности σ_X	Объем выборки *n*	β
1	**2**	**3**	**4**	**5**
1	23,4	3,5	14	0,9
2	123,8	15,9	21	0,9
3	5,2	0,8	10	0,9
4	12,2	1,7	15	0,9
5	40,1	2,6	14	0,9
6	23,4	2,3	11	0,9
7	123,8	13,2	20	0,9
8	34,5	3,5	10	0,95
9	143,8	15,9	12	0,95
10	5,2	0,4	21	0,95

Продолжение табл. 2.1.2

1	2	3	4	5
11	32,9	0,8	14	0,95
12	23,4	2,5	13	0,95
13	123,8	17,1	10	0,95
14	56,7	3,5	21	0,95
15	234,1	15,9	19	0,95
16	24,9	1,9	21	0,95
17	5,2	0,4	14	0,99
18	12,9	0,8	12	0,99
19	23,4	2,6	10	0,99
20	123,8	9,3	15	0,99
21	48.5	3,5	26	0,99
22	198,8	15,9	17	0,99
23	47,2	3,1	21	0,99
24	12,9	0,8	14	0,99
25	5,2	0,7	16	0,9
26	189,8	15,9	10	0,9
27	23,4	2,7	18	0,9
28	123,8	11,2	14	0,9
29	19,9	3,5	21	0,9

2.2.
СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

2.2.1.
Основные понятия

Проверка гипотез относительно параметров распределения случайной величины тесно связана с интервальным оцениванием параметров распределения, но имеет несколько другой аспект.

Определение 2.2.1. Статистической гипотезой называется любое предположение о виде или параметрах неизвестного закона распределения.

Проверяемую гипотезу обычно называют нулевой, или основной, и обозначают H_0. Альтернативную, или конкурирующую, гипотезу обозначают H_1.

Далее для общности будем обозначать исследуемый параметр случайной величины θ, а его выборочную оценку $\hat{\theta}$. Например, в случае математического ожидания, $m_x = \theta$, а его выборочная оценка $\bar{X} = \hat{\theta}$.

Чтобы принять решение относительно выдвигаемой гипотезы, зададим вопрос: если в качестве основной гипотезы принять, что параметр θ имеет какое-то конкретное значение $\theta = \theta_0$, то как сильно должна отличаться оценка $\hat{\theta}$ от θ_0, чтобы основная гипотеза была отвергнута как ложная?

Ответ на этот вопрос помогает понять рисунок 2.2.1.

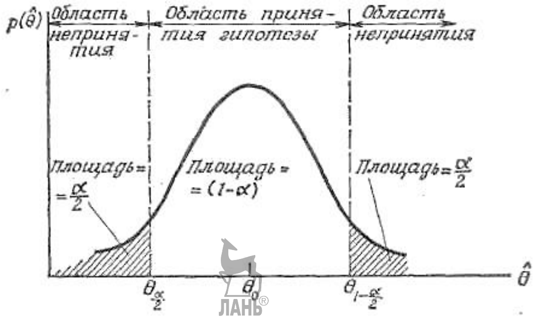

Рис. 2.2.1

Области принятия и непринятия гипотезы

Чтобы принять решение относительно выдвинутой гипотезы, нужно еще до получения выборки задать уровень значимости α для проверяемого критерия. Обычно α выбирают достаточно малым, чтобы можно было считать практически невозможным, что оценка $\hat{\theta}$ параметра θ выйдет за пределы интервала $[\theta_{\alpha/2}; \theta_{1-\alpha/2}]$. Затем берется выборка и вычисляется $\hat{\theta}$. Если $\hat{\theta} \le \theta_{\alpha/2}$ или $\hat{\theta} \ge \theta_{1-\alpha/2}$, то гипотеза отвергается. В противном случае она принимается.

Интервал значений $\hat{\theta}$, при которых основная гипотеза принимается, называется областью принятия гипотезы. Интервал значений $\hat{\theta}$, при которых основная гипотеза отвергается, называется критической областью.

Описанный выше критерий является двусторонним.

Односторонний критерий основывается на том, что гипотеза отвергается, если допускается выполнение только одного из неравенств: $\hat{\theta} \le \theta_{\alpha/2}$ — левосторонний критерий, $\hat{\theta} \ge \theta_{1-\alpha/2}$ — правосторонний критерий.

При проверке гипотез различают ошибки двух типов:

— ошибка первого рода, когда гипотеза верна, но отвергается;

— ошибка второго рода, когда гипотеза ошибочна, но принимается.

Ясно, что появление ошибки первого рода связано с тем, что уровень значимости α выбран отличным от нуля. Если гипотеза верна и, например, $\alpha = 0,05$, то эта гипотеза будет отвергнута в 5% испытаний. Если гипотеза оказалась верна, как предполагалось, то двусторонний критерий приведет к правильному решению в $100(1 - \alpha)$% испытаний.

Пример 2.2.1. По выборке объемом $n = 9$ случайной величины X получены точечные оценки ее математического ожидания и стандартного отклонения: $\bar{X} = 6,5$, $s_X = 0,50$. Проверить гипотезу H_0, состоящую в том, что с вероятностью $\beta = 0,95$ математическое ожидание случайной величины X равно $m_X = 6,8$ при условии, что дисперсия D_X генеральной совокупности неизвестна. Альтернативная гипотеза H_1 состоит в том, что $m_X \ne 6,8$.

Решение. Уровень значимости $\alpha = 1 - \beta = 0,05$. Предельную ошибку выборки найдем по формуле (2.1.11): $\Delta_{\text{в}} = t_{1-\alpha/2;\nu} \cdot \frac{s_X}{\sqrt{n}}$. Для $k = n - 1 = 8$ и $\alpha = 0,05$ из таблицы приложения 2 находим $t_{1-\alpha/2;\,\nu} = t_{0,975;\,8} = 2,306$. Тогда $\Delta_{\text{в}} = 2,306 \cdot \frac{0,5}{\sqrt{8}} \approx 0,395.$

По формуле (2.1.11) получаем: $6,5 - 0,395 < 6,8 < 6,5 + 0,395$.

Следовательно, гипотеза H_0 принимается.

2.2.2. Проверка гипотез относительно математического ожидания и дисперсии в среде MATLAB

В предыдущих параграфах мы рассмотрели теоретические основы вычисления интервальных оценок параметров случайной величины и проверки гипотез относительно параметров распределения.

Перейдем к решению тех же задач в среде MATLAB.

Если дисперсия генеральной совокупности неизвестна, проверка гипотез относительно математического ожидания и его интервальная оценка в среде MATLAB осуществляется с помощью уже рассмотренной нами ранее функции *ttest*, синтаксис которой имеет вид

$$[h, p, ci] = ttest(x, m_x, \alpha), \tag{2.2.1}$$

где x — выборка значений случайной величины X из генеральной совокупности; m_x — предполагаемое математическое ожидание случайной величины X; α — уровень значимости; h — параметр, который равен логическому нулю, если гипотеза H_0 о том, что математическое ожидание равно заданному значению m_x принимается, и равен логической единице, если гипотеза H_0 отвергается; p — вероятность совместимости данной выборки с гипотезой H_0; ci — вектор, содержащий значения левой и правой границ доверительного интервала.

Если дисперсия или стандартное отклонение генеральной совокупности известна, то проверку гипотезы относительно математического ожидания можно провести с помощью функции

$$[h, p, ci] = ztest(x, m_x, \sigma_X, \alpha),$$

где σ_X — заданное стандартное отклонение генеральной совокупности.

Рассмотрим применение функции (2.2.1) MATLAB для проверки гипотезы относительно математического ожидания. Для этого используем выборку примера 2.1.2.

Пример 2.2.2. По заданной выборке для уровня значимости $\alpha = 0{,}05$ проверить основную гипотезу H_0: математическое ожидание генеральной совокупности $m_X = 2{,}5$ при условии, что дисперсия генеральной совокупности неизвестна.

Решение задачи в командном окне MATLAB представлено на рисунке 2.2.2.

Таким образом, гипотеза H_0 о том, что данная выборка взята из генеральной совокупности с математическим ожиданием, равным 2,5, не подтвердилась ($h = 1$). Действительно, полученная вероятность $p = 0{,}0276$ совместимости предполагаемого математического ожидания $m_X = 2{,}5$ с заданной выборкой меньше уровня значимости $\alpha = 0{,}05$. Несостоятельность основной гипотезы H_0 подтверждается также и тем, что предполагаемое значение $m_X = 2{,}5$ для генеральной совокупности не попадает в найденный доверительный интервал [2,2665; 2,4863].

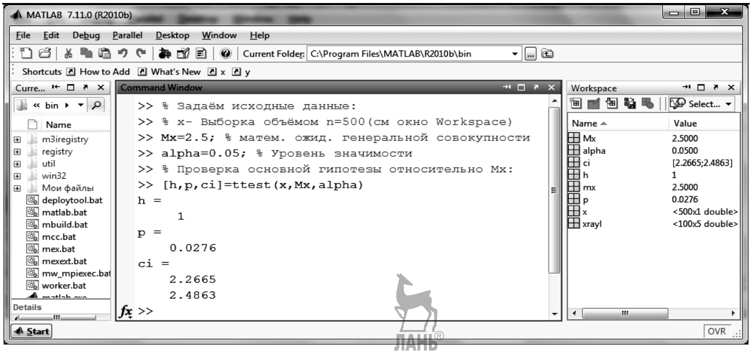

Рис. 2.2.2

Проверка гипотезы о совместимости заданного математического ожидания $m_x = 2{,}5$ *с выборкой случайной величины*

Проверка гипотез относительно дисперсии в среде MATLAB осуществляется с помощью уже рассмотренной нами ранее функции *vartest*, синтаксис которой имеет вид

$$[h, p, ci] = vartest(x, Dx, \alpha), \qquad (2.2.2)$$

где x — выборка значений случайной величины X; Dx — предполагаемая дисперсия случайной величины X; α — уровень значимости; h — параметр, который равен логическому нулю, если гипотеза H_0 о том, что дисперсия генеральной совокупности равна заданному значению Dx, принимается, и равен логической единице, если гипотеза H_0 отвергается; p — вероятность совместимости данной выборки с гипотезой H_0; ci — вектор, содержащий значения левой и правой границ доверительного интервала.

Рассмотрим применение функции (2.2.2) MATLAB для проверки гипотезы относительно дисперсии. Для этого используем выборку примера 2.1.2.

П р и м е р 2.2.3. По заданной выборке для уровня значимости $\alpha = 0{,}05$ проверить основную гипотезу H_0: дисперсия генеральной совокупности $Dx = 1{,}7$.

Решение задачи в командном окне MATLAB представлено на рисунке 2.2.3.

Так, параметр $h = 0$, то гипотеза H_0 принимается. При этом, хотя вероятность $p = 0{,}2041$ совместимости выборки с гипотезой H_0 невысока, но все же p больше заданного уровня значимости $\alpha = 0{,}05$. Кроме того, предполагаемая дисперсия $Dx = 1{,}7$ находится внутри доверительного интервала [1,3874; 1,7787].

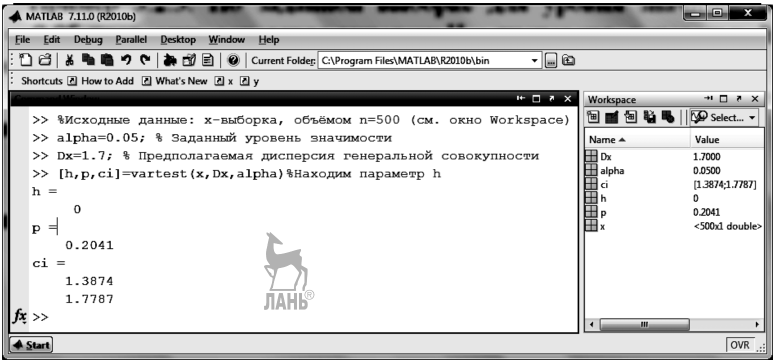

Рис. 2.2.3

Проверка гипотезы о совместимости заданной дисперсии $D_x = 1{,}7$ *с выборкой случайной величины*

Исходные данные и результаты всех промежуточных вычислений отображены в окне Workspace рабочей среды MATLAB.

2.2.3. Проверка гипотезы о равенстве математических ожиданий

Постановка задачи. Пусть имеются две случайные величины X и Y с неизвестными математическими ожиданиями m_X и m_Y и дисперсиями D_X и D_Y. Требуется по заданным выборкам случайных величин X и Y проверить основную гипотезу о принадлежности этих случайных величин к одной генеральной совокупности, т. е. проверить предположение о том, что $m_X = m_Y$, предполагая, что $D_X = D_Y$.

Исходными данными при решении этой задачи являются две выборки: одна — выборка значений случайной величины X объемом n_X, другая — выборка значений случайной величины Y объемом n_Y. Кроме того, задан уровень значимости α.

Проверку гипотезы о равенстве $m_X = m_Y$ будем проводить на основе t-распределения. Обоснованность применения этого критерия обсуждалась в п. 2.1.4.

Общая схема решения задачи следующая.

1. По заданным выборкам вычисляем: выборочные средние $\overline{X}$ и $\overline{Y}$ и выборочные дисперсии s_X^2 и s_Y^2.

2. Выдвигаем основную гипотезу H_0: $m_X = m_Y$ и одну из возможных альтернативных гипотез:

а) H_1: $m_X > m_Y$; или

б) H_1: $m_X < m_Y$.

3. В качестве критерия выбираем случайную величину

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}} \sqrt{\frac{n_X n_Y (n_X + n_Y - 2)}{n_X + n_Y}}, \qquad (2.2.3)$$

которая имеет распределение Стьюдента с числом степеней свободы $k = n_X + n_Y - 2$.

4. Вычисляем $P(t > t_{кр}) = 1 - \alpha$ и по таблице приложения 2 определяем критическую точку $t_{кр}$.

5. Принимаем решение: если $t > t_{кр}$, то критерий t оказался в критической области и, следовательно, основная гипотеза H_0 отклоняется, а принимается одна из выдвинутых альтернативных гипотез.

П р и м е р 2.2.4. Пусть случайные величины X и Y — ежедневные объемы выручки от реализации одного и того же вида продукции в двух магазинах, расположенных в различных районах города. В первом магазине наблюдение за выручкой проводилось в течение 6 дней, во втором — в течение 5 дней. Получили выборки ежедневных наблюдений X (в тыс. руб.): 282,4998; 237,2937; 156,0610; 217,7281; 424,5705 и 279,5696; Y: 262,9059; 225,4591; 140,8329; 258,1359 и 65,2622.

Объемы выборок для случайных величин X и Y: $n_X = 6$, $n_Y = 5$.

Требуется определить, принадлежат ли эти выборки одной генеральной совокупности или ожидаемый объем ежедневной выручки зависит от расположения магазина. Принимаем уровень значимости $\alpha = 0{,}05$.

Р е ш е н и е. 1. Получим точечные оценки (табл. 2.1.1) математического ожидания для X и Y: $\bar{X} = 266{,}2871$, $\bar{Y} = 190{,}5192$ и дисперсии $s_X^2 = 8{,}1683$, $s_X^2 = 7{,}2977$.

2. Основной является гипотеза о равенстве математических ожиданий величин X и Y, т. е. H_0: $m_X = m_Y$. Альтернативной — гипотеза H_1: $m_X > m_Y$. Основанием для такой альтернативы является то, что $\bar{X} > \bar{Y}$.

3. По формуле (2.2.3) вычисляем значение критерия $t = 1{,}4185$.

4. Вычисляем $P(t < t_{кр}) = 1 - \alpha = 0{,}95$ и число степеней свободы $k = n_X + n_Y - 2 = 9$. Затем по таблице приложения 2 определяем критическую точку $t_{кр} = 1{,}833$.

5. Так как $t < t_{кр}$, то критерий t оказался в области принятия основной гипотезы: $m_X = m_Y$, т. е. обе выборки принадлежат одной генеральной совокупности. Иными словами, объективно ожидаемые зна-

чения ежедневной выручки в обоих магазинах одинаковы. Таким образом, с вероятностью 0,95 смена местоположения второго магазина кардинально не изменит эффективность его работы.

Решим эту задачу в среде MATLAB. Используем функцию *ttest*2 из пакета Statistics Toolbox. Эта функция имеет много возможностей, с которыми, как и с другими функциями, при желании можно ознакомиться, выбрав в меню Help рабочей среды пункт Product Help. В левой части открывшегося окна на вкладке Contents появится оглавление справочной системы, в котором выбираем раздел Statistics Toolbox, а затем в подразделе Functions — описание требуемой функции. Будем использовать вариант синтаксиса функции *ttest*2, который наиболее подходит для нашего случая:

$$[h,p,ci,stats]=ttest2(x,y,alpha,'tail'). \qquad (2.2.4)$$

где *stats* — параметр, содержащий информацию о критерии $t = tstat$, числе степеней свободы $df = n_X + n_Y - 2$ и объединенной дисперсии sd; *tail* — параметр, задающий одну из альтернативных гипотез: 1) если $tail =$ '*right*', то альтернативной является гипотеза H_1: $m_X > m_Y$; 2) если $tail =$ '*left*', то альтернативной является гипотеза H_1: $m_X < m_Y$. Если параметр в командной строке отсутствует, то по умолчанию за альтернативную гипотезу принимается гипотеза H_1: $m_X \neq m_Y$.

Смысл остальных параметров *h*, *p*, *ci*, *alpha* пояснялся ранее в примере 2.2.2.

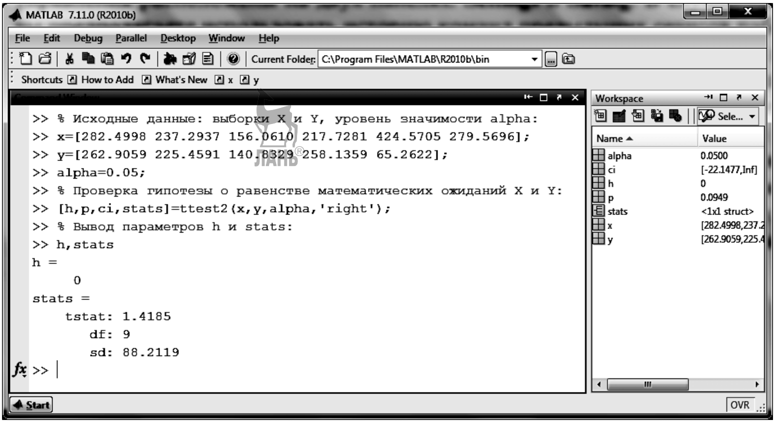

Рис. 2.2.4

Проверка гипотезы о равенстве математических ожиданий двух случйных величин с помощью MATLAB

Результат решения задачи примера 2.2.4 с помощью MATLAB приведен на рисунке 2.2.4.

Так как параметр $h = 0$, то принимается основная гипотеза о равенстве математических ожиданий.

Значение параметра *tstat* (см. командное окно на рисунке 2.2.4) совпадает со значением критерия $t = 1{,}4185$, вычисленного по формуле (2.2.3) (см п. 3 решения без применения MATLAB).

2.2.4.
Проверка гипотезы о равенстве дисперсий

Постановка задачи. Пусть имеются две случайные величины X и Y с неизвестными математическими ожиданиями m_X и m_Y и дисперсиями σ_X^2 и σ_Y^2. Требуется проверить основную гипотезу о том, что $\sigma_X^2 = \sigma_Y^2$. Предполагаем, что $m_X = m_Y$.

Исходными данными при решении этой задачи являются две выборки: одна из них — это выборка значений случайной величины X объемом n_X, другая — выборка значений случайной величины Y объемом n_Y. По этим выборкам могут быть получены точечные оценки s_X^2 и s_Y^2 для дисперсий случайных величин X и Y, которые потребуются для решения поставленной задачи. Кроме того, задан уровень значимости α.

Проверка основной гипотезы о равенстве $\sigma_X^2 = \sigma_Y^2$ основана на использовании критерия F:

$$F = \frac{s_X^2}{s_Y^2}, \text{ если } s_X^2 > s_Y^2, \text{ или } F = \frac{s_Y^2}{s_X^2}, \text{ если } s_Y^2 > s_X^2. \qquad (2.2.5)$$

Случайная величина F имеет распределение Фишера — Снедекора или F-распределение с числами степеней свободы $k_1 = n_X - 1$, $k_2 = n_Y - 1$. Критическая точка $F_{\text{кр}}$ статистики F определяется из условия

$$P(F > F_{\text{кр}}) = \int_{F_{\text{кр}}}^{\infty} \varphi(F, k_1, k_2)df = \alpha, \qquad (2.2.6)$$

где $\varphi(F, k_1, k_2)$ — плотность распределения критерия F; α — заданный уровень значимости.

Значения интеграла (2.2.6) сведены в таблицу, которая позволяет по заданным значениям k_1, k_2 и α определить критическую точку $F_{\text{кр}}$ статистики F (см. таблицу приложения 4).

Общая схема решения задачи следующая.

1. По заданным выборкам вычисляем: выборочные средние $\overline{X}$ и $\overline{Y}$ и выборочные дисперсии s_X^2 и s_Y^2.

2. Выдвигаем основную гипотезу H_0: $\sigma_X^2 = \sigma_Y^2$ и одну из возможных альтернативных гипотез:

а) H_1: $\sigma_X^2 > \sigma_Y^2$; или

б) H_1: $\sigma_X^2 < \sigma_Y^2$.

3. В качестве критерия выбираем случайную величину F, вычисляемую по формуле (2.2.5).

4. Вычисляем $P(F > F_{кр}) = \alpha$ и по таблице приложения 4 определяем критическую точку $F_{кр}$.

5. Принимаем решение: если $F > F_{кр}$, то критерий F оказался в критической области и, следовательно, основная гипотеза H_0 отклоняется, а принимается одна из выдвинутых альтернативных гипотез. Если $F \leq F_{кр}$, то критерий F оказался в допустимой области и основная гипотеза H_0 принимается.

Пример 2.2.5. При уровне значимости $\alpha = 0{,}05$ требуется принять решение о регулировке весов путем сопоставления разброса их показаний x_i с разбросом показаний y_i этих же весов непосредственно после их предыдущей регулировки. Было проведено 9 взвешиваний одного и того же груза на обоих весах и получены следующие результаты в граммах: x_i: 99,77; 100,02; 100,14; 100,52; 99,87; 100,04; 99,98; 99,61; 99,91; y_i: 99,82; 100,08; 99,91; 100,01; 99,95; 100,03; 99,94; 100,05; 100,07.

Сначала решим эту задачу, пользуясь общей схемой решения, изложенной выше.

1. По формулам таблицы 2.1.1 вычисляем: выборочные средние $\overline{X}$ и $\overline{Y}$ и выборочные дисперсии s_X^2 и s_Y^2 случайных величин X и Y: $\overline{X} = 99{,}9851$, $\overline{Y} = 99{,}9850$, $s_X^2 = 0{,}0844$, $s_Y^2 = 0{,}0074$.

2. Основная гипотеза H_0: $\sigma_X^2 = \sigma_Y^2$. Альтернативная гипотеза: H_1: $\sigma_X^2 > \sigma_Y^2$, так как $s_X^2 > s_Y^2$.

3. По формуле (2.2.5) вычисляем критерий нашего теста: $F = \dfrac{s_X^2}{s_Y^2} = 11{,}4049.$

4. По таблице приложения 4 находим критическую точку $F_{кр} = 3{,}44$ для уровня значимости $\alpha = 0{,}05$ и чисел степеней свободы $k_1 = n_X - 1 = 8$, $k_2 = n_Y - 1 = 8$.

5. Так как выполняется неравенство $F > F_{кр}$, то критерий F оказался в критической области и, следовательно, основная гипотеза H_0

отклоняется, а принимается альтернативная гипотеза: $\sigma_X^2 > \sigma_Y^2$, т. е. весы нуждаются в регулировке.

Теперь решим эту задачу в рабочей среде MATLAB.

Используем функцию *vartest*2 из пакета Statistics Toolbox. Будем использовать вариант синтаксиса функции *vartest*2, который наиболее подходит для нашего случая:

$$[h, p, ci, stats] = vartest2(x, y, alpha, \text{'}tail\text{'}), \qquad (2.2.7)$$

где *stats* — параметр, содержащий информацию о критерии $F = fstat$ и числах степеней свободы: с большей выборочной дисперсией — $df1$ (в примере 2.2.5 $df1 = n_X - 1 = 8$) и меньшей выборочной дисперсией — $df2$ (в примере 2.2.5 $df2 = n_Y - 1 = 8$); *tail* — параметр, задающий одну из альтернативных гипотез: если *tail* = '*right*', то альтернативной является гипотеза H_1: $\sigma_X^2 > \sigma_Y^2$, если *tail* = '*left*', то альтернативной является гипотеза H_1: $\sigma_X^2 < \sigma_Y^2$. Если параметр в командной строке отсутствует, то по умолчанию за альтернативную гипотезу принимается гипотеза H_1: $\sigma_X^2 \neq \sigma_Y^2$; *p* — параметр, равный вероятности, что нулевая гипотеза окажется верной.

Результат решения задачи примера 2.2.5 с помощью MATLAB приведен на рисунке 2.2.5. Все необходимые пояснения даются в командном окне MATLAB.

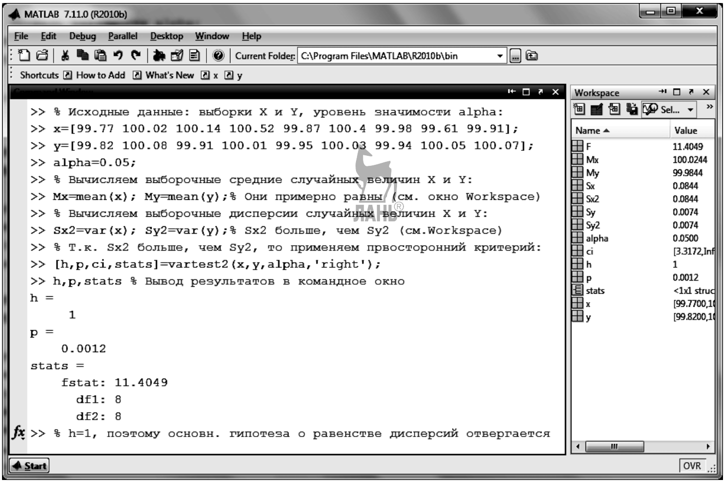

Рис. 2.2.5

Проверка гипотезы о равенстве дисперсий двух случйных величин с помощью MATLAB

2.2.5.
Проверка гипотезы о виде закона распределения по критерию χ^2

При выдвижении гипотезы о виде закона распределения могут быть использованы такие предпосылки, как использование свойств распределений. Например, равенство математического ожидания и дисперсии (характерно для распределения Пуассона), или равенство нулю коэффициентов ассиметрии и эксцесса (для нормального распределения), или равенство математического ожидания и стандартного отклонения (экспоненциальное распределение). Предположение о виде закона распределения может быть сделано также по гистограмме заданного статистического ряда, как это было сделано в примере 2.1.8.

Одним из критериев, по которому проводится проверка гипотезы о виде закона распределения, является критерий согласия χ^2 (критерий Пирсона).

Критерий χ^2 применим при больших объемах выборки ($n \geq 200$) и требует, чтобы при представлении ее в виде интервального статистического ряда число m интервалов было не меньше 8. Кроме того, число попаданий случайного значения в каждый интервал должно быть не менее 5.

Рассмотрим, как применяется критерий χ^2 при решении поставленной задачи.

Доказано, что при объеме выборки $n > 50$ случайная величина χ^2, определяемая как

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i}, \tag{2.2.8}$$

имеет распределение Пирсона с числом степеней свободы $k = m - 1 - s$, где m — число разрядов заданного статистического ряда; n_i — частоты заданного статистического ряда (эмпирические частоты); np_i — теоретические частоты, соответствующие предполагаемому закону распределения; s — число неизвестных параметров предполагаемого закона распределения.

Основной проблемой при вычислении критерия χ^2 по формуле (2.2.8) является нахождение теоретической вероятности p_i.

Для дискретной случайной величины: $p_i = P(X = x_i)$, т. е. это вероятность того, что случайная величина X примет конкретное значение x_I ряда распределения.

Для непрерывной случайной величины p_i есть вероятность попадания случайной величины X в i-й разряд интервального статистического ряда. В этом случае вероятности p_i ($i = 1, 2, ..., m$) определяются по-разному, в зависимости от предполагаемого вида закона распределения.

В частности, при выдвижении в качестве основной гипотезы предположения о нормальном законе распределения имеем

$$p_i = h_{\text{нормир}} \cdot \varphi(y_i), \tag{2.2.9}$$

где $h_{\text{нормир}} = \dfrac{x_{i+1} - x_i}{s_X}$ — нормированная ширина разряда заданного статистического ряда; $\varphi(y_i) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{y_i^2}{2}}$ — функция плотности вероятности нормированного нормального распределения (см. таблицу приложения 5); $y_i = \dfrac{(x_i - \overline{\mathrm{X}})}{s_X}$ — нормированные значения центров группирования x_i заданного интервального статистического ряда.

В остальном схема применения критерия χ^2 такая же, как при нахождении интервальной оценки дисперсии (п. 2.1.5).

Пример 2.2.6. Применив критерий согласия χ^2 при уровне значимости $\alpha = 0{,}05$ определить закон распределения генеральной совокупности по выборке, заданной в форме интервального статистического ряда с числом разрядов $m = 11$:

Интервалы	(10,5; 13,5)	(13,5; 16,5)	(16,5; 19,5)	(19,5; 22,5)	(22,5; 25,5)	(25,5; 28,5)
Частоты, n_i	6	11	21	20	30	31
Интервалы	(28,5; 31,5)	(31,5; 34,5)	(34,5; 37,5)	(37,5; 40,5)	(40,5; 43,5)	
Частоты, n_i	26	29	13	9	4	

Решение. Строим гистограмму заданного статистического ряда (рис. 2.2.6) с шагом $h = 3$, равным ширине частичных интервалов. По горизонтальной оси графика отложены центры интервалов x_i статистического ряда, по вертикальной оси — соответствующие им частоты n_i.

Анализируя гистограмму, приходим к выводу, что исследуемая случайная величина X, скорее всего, имеет нормальное распределение.

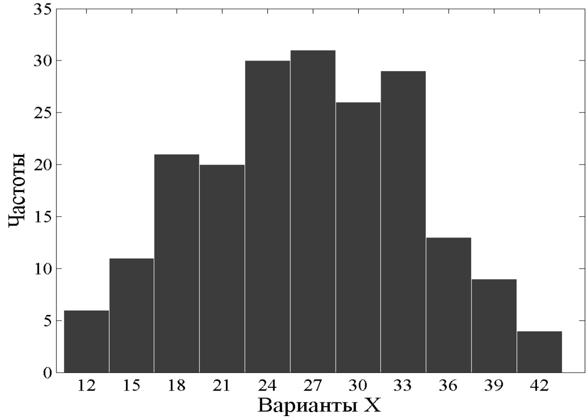

Рис. 2.2.6

Гистограмма распределения случайной величины X

Проверим эту гипотезу, выполнив следующую последовательность действий.

1. Выдвигаем нулевую и альтернативную гипотезы:

• H_0: закон распределения случайной величины X нормальный;

• H_1: закон распределения случайной величины X отличен от нормального.

2. Рассчитываем выборочное среднее $\overline{X} = 26,58$, выборочную дисперсию $s_X^2 = 50,58$ и выборочное стандартное отклонение $s_X = 7,09$ случайной величины X, принимая за x_i центры интервалов статистического ряда.

Нормируем значения x_i по формуле $y_i = \dfrac{x_i - \overline{X}}{s_X}$ и определяем соответствующие значения $\varphi(y_i)$ теоретической плотности вероятности по таблице приложения 5.

Определяем теоретические вероятности $p_i = \dfrac{h}{s_X}\varphi(y_i)$ попадания случайной величины X в i-й частичный интервал и теоретические частоты np_i попадания случайной величины в i-й частичный интервал, соответствующие предполагаемому закону распределения $\varphi(y_i)$.

Рассчитываем критерий $\chi^2 \approx 10,9$ по формуле (2.2.8).

Все расчеты, связанные с решением этой задачи, проводились в Excel и сведены в таблицу 2.2.1.

Таблица 2.2.1

Расчет критерия согласия χ^2

$x_i x_i$	$x_i x_i$	$\frac{x_i n_i}{n}$	$\frac{(x_i - \bar{X})^2 n_i}{n-1}$	$y_i = \frac{x_i - \bar{X}}{s_X}$	$\varphi(y_i)$	$p_i = \frac{h}{s_X}\varphi(y_i)$	$\frac{(n_i - np_i)^2}{np_i}$
12	6	0,36	6,3773	2,0500	0,053	0,0224	0,5227
15	11	0,825	7,3753	1,6282	0,1145	0,0483	0,1860
18	21	1,89	7,7297	1,2064	0,2059	0,0869	0,7585
21	20	2,1	3,1136	0,7846	0,3056	0,1289	1,2963
24	30	3,6	0,9985	0,3628	0,3808	0,1606	0,1406
27	31	4,185	0,0273	0,0591	0,3961	0,1671	0,1746
30	26	3,9	1,5205	0,4809	0,3983	0,1680	1,7196
33	29	4,785	5,9764	0,9027	0,2504	0,1056	2,9363
36	13	2,34	5,7679	1,3245	0,2323	0,0980	2,2209
39	9	1,755	6,9415	1,7463	0,0775	0,0327	0,9270
42	4	0,84	4,7555	2,1681	0,0514	0,0216	0,0261
Σ	$n =$ $= 200$	$\bar{X} =$ $= 26{,}58$	$s_X^2 =$ $= 50{,}33$				$\chi^2 \approx$ $\approx 10{,}9$

3. По таблице приложения 3 при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $k = m - 1 - s = 11 - 1 - 2 = 8$ находим критическую точку $\chi^2_{\text{кр}} = 19{,}7$.

4. Так как $\chi^2 < \chi^2_{\text{кр}}$, т. е. критерий оказался в области допустимых значений, то основная гипотеза H_0 о нормальном законе распределения случайной величины X принимается.

Пример 2.2.7. Решим задачу примера 2.2.6, используя инструменты MATLAB.

Решение. Для проверки гипотезы о виде закона распределения на основе критерия согласия χ^2 в пакете расширения Statistics Toolbox предусмотрена функция *chi2gof* (chi-square goodness-of-fit test). С ее подробным описанием можно познакомиться в справочной системе MATLAB.

Функция *chi2gof* предоставляет возможность тестирования различных основных гипотез относительно законов распределения с различным числом параметров. Исходная выборка может быть задана как в форме статистического ряда, так и в форме простой статистической совокупности.

Для решения поставленной задачи подходит следующий вариант применения функции *chi2gof*:

[*H*, *P*, *stats*] = *chi2gof*(*xi*, ′*nbins*′, *m*, ′*freqency*′, *ni*, ′*expected*′, *npi*, '*nparam*′, *nparam*, ′*emin*′, *emin*, ′*alpha*′, *alpha*),

где *H* — параметр, который равен логическому нулю, если основная гипотеза H_0 о теоретическом законе распределения принимается, и равен логической единице, если гипотеза H_0 отвергается;

P — вероятность совместимости заданной выборки с гипотезой H_0 о законе распределения генеральной совокупности;

stats — формальный параметр, содержащий следующие поля:

- *chi2stat* — вычисленная статистика χ^2;
- *df* — число степеней свободы распределения χ^2;
- *edges* — вектор границ интервалов статистического ряда;
- *O* — эмпирические частоты n_i заданного статистического ряда;
- *E* — теоретические частоты np_i;

xi — вектор-строка центров частичных интервалов заданного интервального статистического ряда;

m — число разрядов интервального статистического ряда;

ni — эмпирические частоты n_i статистического ряда;

npi — теоретические частоты np_i попадания случайной величины *X* в *i*-й частичный интервал;

nparam — число параметров теоретического закона распределения;

emin — минимально допустимая частота *ni* интервального статистического ряда;

alpha — уровень значимости.

На рисунке 2.2.7 показано решение поставленной задачи в рабочей среде MATLAB.

Так как $H = 0$, то основная гипотеза о нормальном распределении генеральной совокупности принимается.

Полученное значение *chi2stat* = 6,8954 (рис. 2.2.7) — это и есть значение критерия χ^2. Отличие величин χ^2, полученных в примерах 3.2.6 и 3.2.7, можно объяснить тем, что таблица приложения 5 дает округленные значения плотности распределения $\varphi(y_i)$. Для получения точных значений плотности по таблице необходимо прибегать к интерполированию.

Вероятность совместимости заданной выборки с основной гипотезой равна $P = 0{,}5480$ (рис. 2.2.7).

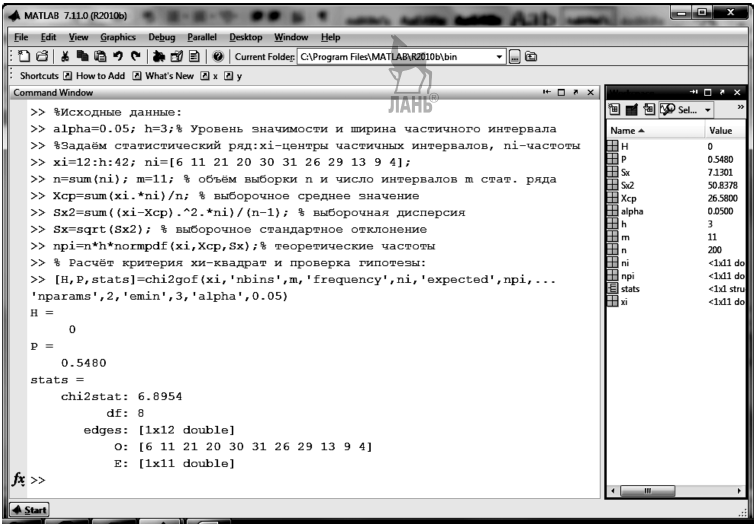

Рис. 2.2.7

Проверка гипотезы о законе распределения
по заданному статистическому ряду на основе критерия согласия χ^2

Для просмотра найденных теоретических частот np_i следует дважды кликнуть левой кнопкой мыши по иконке с обозначением *npi* в окне Workspase рабочей среды MATLAB.

В примере 2.2.6 выдвижение гипотезы о законе распределения делалось по виду гистограммы, построенной на основе заданного интервального статистического ряда.

Инструменты MATLAB упрощают процедуру выдвижения гипотезы и делают ее более надежной. Гистограмму распределения можно построить непосредственно по простой статистической совокупности, не прибегая к составлению интервального статистического ряда. Более того, средствами MATLAB гистограмму можно совместить с графиками предполагаемых теоретических законов распределения и затем делать вывод о том, какая из возможных гипотез о виде закона распределения наиболее правдоподобна.

Чтобы полнее проиллюстрировать возможности MATLAB, проведем следующий численный эксперимент. Решим задачу о согласии выборки случайной величины X с предполагаемым законом распределения, если выборка представлена в виде простой статистической совокупности, извлеченной из некоторой генеральной совокупности.

Задачу будем решать в следующей последовательности.

1. Средствами MATLAB моделируем генеральную статистическую совокупность, подчиненную какому-либо закону распределения с заданными параметрами. Будем далее предполагать, что закон распределения и его параметры нам неизвестны.

2. Из полученной генеральной совокупности сделаем случайную выборку значений объемом $n = 200$ значений.

3. Совместим гистограмму выборки и графики предполагаемых законов распределения на одном рисунке.

4. По полученным рисункам выдвинем основную гипотезу о виде закона распределения.

5. Вычислим выборочные параметры генеральной совокупности по простой статистической совокупности, полученной в п. 2.

6. Применяя критерий согласия χ^2, проверим справедливость основной гипотезы.

В пакете расширения Statistics Toolbox имеются функции, позволяющие генерировать последовательности независимых случайных чисел, подчиненных различным законам распределения с заданными параметрами. К их числу относятся, в частности, функции: *normrnd* (нормальный закон распределения), *unifrnd* (закон равномерной плотности распределения), *poissrnd* (закон Пуассона), *exprnd* (экспоненциальный закон распределения) и т. д. Подробнее узнать о правилах их применения можно в справочной системе MATLAB.

На рисунке 2.2.8 показано, как реализована вышеописанная последовательность действий в командном окне MATLAB.

На рисунке 2.2.9 изображена гистограмма выборки и графики теоретических функций нормального распределения и распределения Рэлея. Эти результаты получены после выполнения команд *histfit*(*x*, 15, ′*normal*′) и *histfit*(*x*, 15, ′*rayleigh*′). Число 15 означает число интервалов группирования выборки. Команда *subplot* позволяет последовательно рисовать несколько графиков в одном окне.

Анализируя рисунок 2.2.9, можно сделать вывод, что распределение Рэлея лучше согласуется с данной выборкой, чем нормальное распределение.

Дальнейший расчет по критерию χ^2 (рис. 2.2.8) подтвердил правильность выбора основной гипотезы, так как возвращаемый параметр *h* функции *chi2gof* равен нулю. При этом правдоподобность выбора распределения Рэлея в качестве основной гипотезы равна $p = 0{,}7352$. Численные значения параметров *h* и *p* отображены на рисунке 2.2.8 в окне Workspace рабочей среды MATLAB.

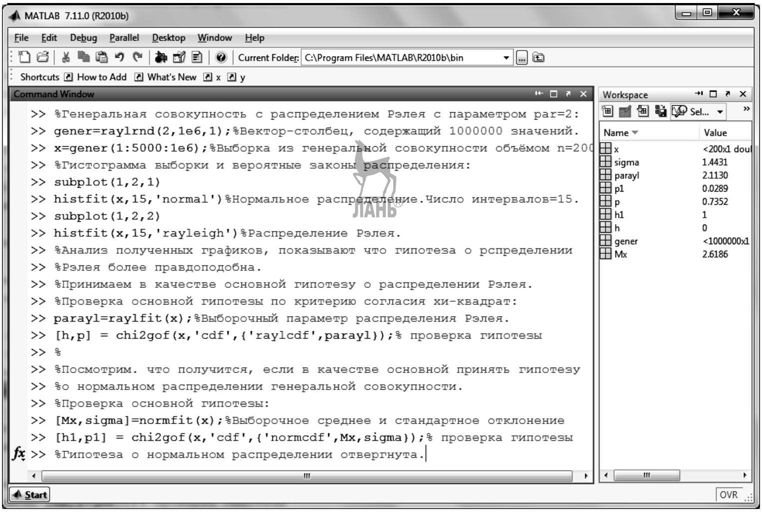

Рис.2.2.8

Проверка гипотезы о законе распределения
по простой статистической совокупности на основе критерия χ^2

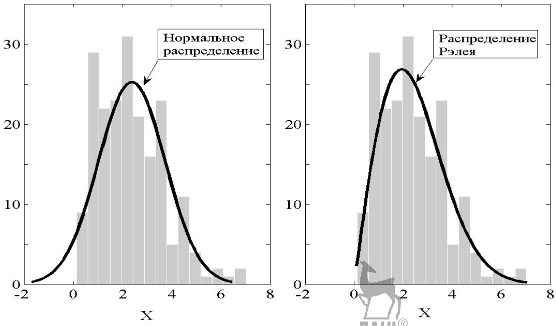

Рис. 2.2.9

Выбор основной гипотезы о законе распределения
генеральной совокупности

При выборе в качестве основной гипотезы нормального распределения, возвращаемые параметры $h1$ и $p1$ функции *chi2gof* равны 1 и 0,0289 соответственно (см. окно Workspace на рисунке 2.2.8). Поэтому гипотеза о нормальном распределении генеральной статистической совокупности отвергается.

2.2.6.
Проверка гипотезы о законе распределения по критерию Колмогорова — Смирнова

Критерий Колмогорова — Смирнова позволяет проверить гипотезу о согласии эмпирического и теоретического распределений даже при небольшом объеме выборки. При этом имеются следующие ограничения:

1) действие критерия распространияется только на непрерывные случайные величины;

2) вид предполагаемого закона распределения и его параметры должны быть известны заранее из каких-либо теоретических соображений.

В качестве критерия используется следующая мера отклонения эмрирического $F^*(x)$ и предполагаемого теоретического $F(x)$ распределений

$$\psi = \sqrt{n} \cdot \max |F(x) - F^*(x)|, \tag{2.2.10}$$

где n — объем выборки.

Колмогоров доказал, что если теоретическая функция распределения $F(x)$ непрерывна, то при достаточно большом числе наблюдений n можно считать, что

$$P(\psi < \psi_{кр}) = K(\psi_{кр}) = 1 - \alpha, \tag{2.2.11}$$

где $\psi_{кр}$ — некоторая критическая точка критерия ψ; α — уровень значимости; $K(\psi_{кр})$ — функция Колмогорова, определяемая формулой

$$K(\psi_{кр}) = \sum_{i=1}^{\infty}(-1)^i e^{-2i^2\psi_{кр}^2}. \tag{2.2.12}$$

Результаты расчетов по форуле (2.2.12) сведены в таблицу (см. таблицу приложения 6).

Функция Колмогорова (2.2.12) не зависит от теоретической функции распределения $F(x)$. Но если парметры функции $F(x)$ неизвестны и они оцениваютя по заданной выборке, то распределение статистики $\psi_{кр}$ уже зависит от $F(x)$. При этом статистика $\psi_{кр}$ будет зависеть только от вида распределения $F(x)$. Если в теоретической функции распределения есть только параметры сдвига и масштаба, то применимость критерия Колмогорова — Смирнова корректна. Применение критерия Колмогорова — Смирнова о согласии эмпирического и теоретического распределений покажем на следующем примере.

П р и м е р 2.2.8. По выборке объемом $n = 30$, представленной в виде простой статистической совокупности:

1,79; 5,47; 5,10; 2,07; 3,98; 1,89; 4,32; 1,79; 3,79; 2,28; 0,80; 1,00;
2,04; 2,54; 3,24; 2,07; 2,74; 2,97; 3,29; 2,66; 1,83; 1,42; 3,42; 1,69;
2,71; 1,99; 1,70; 5,42; 3,17; 1,22,

проверить по критерию Колмогорова — Смирнова гипотезу о согласии эмпирической функции распределения с функцией распределения Рэлея с параметром $\sigma = 1{,}9$ при уровне значимости $\alpha = 0{,}05$.

Р е ш е н и е. 1. Выдвигаем в качестве нулевой гипотезы, что функцией распределения генеральной совокупности является функция Рэлея:

$$F(x) = \int_0^x \frac{u^2}{\sigma^2} e^{-\frac{u^2}{2\sigma^2}}, \quad u > 0 \tag{2.2.13}$$

с параметром $\sigma = 1{,}9$.

Альтернативной гипотезой является какая-то другая функция распределения.

2. По исходной выборке строим интервальный статистический ряд. Оптимальную ширину интервала ряда определяем по формуле Стерджесса (2.1.1).

Интервалы	(0,76; 1,59)	(1,59; 2,42)	(2,42; 3,25)	(3,25; 4,08)	(4,08; 4,90)	(4,90; 5,73)
Частоты, n_i	4	11	5	5	2	3

3. Вычисляем значения эмпирической функции распределения на правых границах интервалов: $F^*(x) = \dfrac{n_i^{\Sigma}}{n}$, где n_i^{Σ} — накопленные частоты (п. 2.1.1).

4. Определяем значения $F(x_i)$ теоретической функции распределения по формуле (2.2.13) или по таблице.

5. Находим расчетное значение критерия $\psi \approx 0{,}89$ по формуле (2.2.10). Все вычисления, связанные с нахождением статистики ψ, сведены в таблицу 2.2.2.

6. Для $K(\psi_{кр}) = 1 - \alpha = 0{,}95$ определяем по таблице приложения 6 критическую точку $\psi_{кр} = 1{,}38$.

7. Поскольку $\psi < \psi_{кр}$, то основная гипотеза о рэлеевском распределении генеральной совокупности принимается.

Таблица 2.2.2

Расчет критерия согласия ψ. $n = 30$

x_i	n_i^{Σ}	$F^*(x_i)$	$F(x_i)$	$\lvert F(x_i) - F^*(x_i)\rvert$	
1,59	4	0,1333	0,2954	0,1621	
2,42	15	0,5000	0,5556	0,0556	
3,25	20	0,6667	0,7684	0,1017	
4,08	25	0,8333	0,9003	0,0669	
4,9	27	0,9000	0,9640	0,0640	
5,73	30	1,0000	0,9894	0,0106	
$\max\lvert F(x_i) - F^*(x_i)\rvert = 0{,}1621$					$\psi = \sqrt{n}\cdot\max = \sqrt{30}\cdot 0{,}1621 = 0{,}8877$

П р и м е р 2.2.9. Решим задачу, сформулированную в примере 2.2.8, с помощью MATLAB.

Для проверки гипотезы о виде закона распределения на основе критерия согласия Колмогорова — Смирнова в пакете расширения Statistics Toolbox предусмотрена функция *kstest* (Kolmogorov — Smirnov test). С ее подробным описанием можно познакомиться в справочной системе MATLAB.

Функция *kstest* тестирует согласие выборки и куммулятивной функции распределения генеральной совокупности, выдвинутой в качестве основной гипотезы. При этом выборка может быть задана в форме простой статистической совокупности.

Для решения задачи примера 2.2.8 выберем следующий вариант применения функции *kstest*:

[H, P, kstat, krit] = kstest(x, CDF, alpha),

где H — параметр, равный логическому нулю, если основная гипотеза H_0 принимается, и равен логической единице, если отвергается;

P — вероятность совместимости заданной выборки с гипотезой H_0 о законе распределения генеральной совокупности;

kstat — вычисленная статистика критерия Колмогорова — Смирнова, равная $\max\lvert F(x_i) - F^*(x_i)\rvert$;

krit — критическое значение критерия;

x — вектор-столбец, содержащий выборку значений случайной величины X;

CDF (Cumulative Distribution Functions) — матрица, состоящая из двух вектор-столбцов: первый содержит заданную выборку x, второй — соответствующие значения теоретической куммулятивной функции распределения $F(x)$;

alpha — уровень значимости.

Расчет теоретической куммулятивной функции распределения производится с помощью функции

$$cdf('name', x, A, B),$$

где *name* — имя, соответствующее функции распределения, выбранной в качестве основной гипотезы (в нашем случае в апострофах должно быть указано имя функции распределения Рэлея: *rayl*);

A и *B* — параметры функции распределения (для однопараметрических функций распределения указывается один параметр).

Более подробную информацию можно получить в справочной системе MATLAB.

На рисунке 2.2.10 показано решение поставленной задачи в рабочей среде MATLAB.

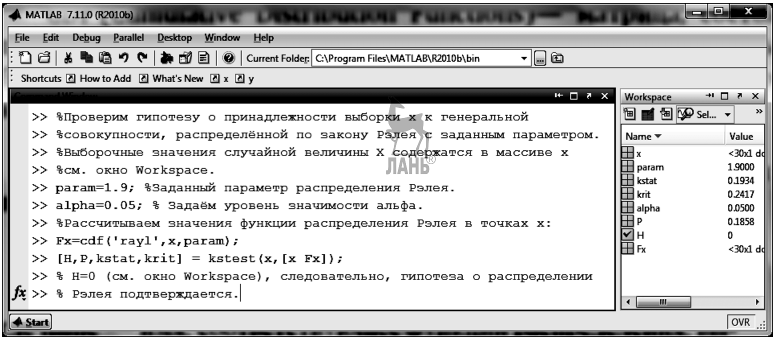

Рис. 2.2.10

Проверка гипотезы о законе распределения по критерию Колмогорова — Смирнова в среде MATLAB

Заметим, что полученные здесь значения $kstat = 0{,}1934$ и $krit = 0{,}2417$, примерно равны соответствующим значениям $\psi = 0{,}8877$ и $\psi_{кр} = 1{,}38$, полученным в примере 2.2.8, с точностью до множителя $\sqrt{n} = \sqrt{30}$, где n — объем выборки. То есть статистикой теста при решении задачи в MATLAB является величина $kstat = \max|F(x_i) - F^*(x_i)|$, а не критерий $\psi = \sqrt{n} \cdot \max|F(x_i) - F^*(x_i)|$, как в примере 2.2.8.

2.2.7.
Задания для самостоятельной работы

По данным таблицы 2.2.3 подобрать законы распределения, применяя критерии согласия χ^2 и Колмогорова — Смирнова при уровне значимости $\alpha = 0{,}05$.

Таблица 2.2.3

Таблица исходных данных

№	Критерий χ^2	Критерий Колмогорова — Смирнова	
	Имя файла с выборкой	**Закон распределения**	**Имя файла с выборкой**
1	**2**	**3**	**4**
1	sample01.txt	Нормальный, $m_X = 10$, $\sigma_X = 2{,}5$	ksn01.txt
2	sample02.txt	Нормальный, $m_X = 7$, $\sigma_X = 1{,}0$	ksn02.txt
3	sample03.txt	Равномерный, $a = 2$, $b = 9$	ksu03.txt
4	sample04.txt	Рэлея, $\sigma = 4{,}0$	ksr04.txt
5	sample05.txt	Рэлея, $\sigma = 5{,}0$	ksr05.txt
6	sample06.txt	Экспоненциальный,	kse06.txt
7	sample07.txt	Равномерный, $a = 2$, $b = 9$	ksu07.txt
8	sample08.txt	Нормальный, $m_X = 15$, $\sigma_X = 3{,}0$	ksn08.txt
9	sample09.txt	Рэлея, $\sigma = 6{,}0$	ksr09.txt
10	sample10.txt	Рэлея, $\sigma = 2{,}0$	ksr10.txt
11	sample11.txt	Экспоненциальный,	kse11.txt
12	sample12.txt	Нормальный, $m_X = 9$, $\sigma_X = 2{,}0$	ksn12.txt
13	sample13.txt	Экспоненциальный, $\lambda = 0{,}5$	kse13.txt
14	sample14.txt	Равномерный, $a = -4$, $b = 8$	ksu14.txt
15	sample15.txt	Нормальный, $m_X = 23$, $\sigma_X = 4{,}5$	ksn15.txt
16	sample16.txt	Рэлея, $\sigma = 1{,}0$	ksr16.txt
17	sample17.txt	Нормальный, $m_X = 45$, $\sigma_X = 7{,}0$	ksn17.txt
18	sample18.txt	Равномерный, $a = 11$, $b = 19$	ksu18.txt
19	sample19.txt	Нормальный, $m_X = 65$, $\sigma_X = 14{,}0$	ksn19.txt
20	sample20.txt	Экспоненциальный, $\lambda = 0{,}2$	kse20.txt
21	sample21.txt	Равномерный, $a = 20$, $b = 29$	ksu21.txt

Продолжение табл. 2.2.3

1	2	3	4
22	sample22.txt	Рэлея, $\sigma = 6,0$	ksr22.txt
23	sample23.txt	Нормальный, $m_X = 20$, $\sigma_X = 5,5$	ksn23.txt
24	sample24.txt	Рэлея, $\sigma = 7,0$	ksr24.txt
25	sample25.txt	Нормальный, $m_X = 97$, $\sigma_X = 15,0$	ksn25.txt
26	sample26.txt	Равномерный, $a = 12$, $b = 19$	ksu26.txt
27	sample27.txt	Экспоненциальный, $\lambda = 1,0$	kse27.txt
28	sample28.txt	Рэлея, $\sigma = 4,5$	ksr28.txt
29	sample29.txt	Равномерный, $a = 22$, $b = 39$	ksu29.txt
30	sample30.txt	Нормальный, $m_X = 19$, $\sigma_X = 4,0$	ksn30.txt

ЗАКЛЮЧЕНИЕ

На основе изложенного учебного материала можно утверждать, что применение системы MATLAB позволяет решать задачи теории вероятностей, математической статистики и эконометрики. Это доказывается сравнением результатов расчетов, полученных с использованием MATLAB и традиционными способами.

Важным является и то, что с помощью пакета символьной математики MuPAD можно с легкостью находить законы распределения функции системы случайных величин в аналитической форме, даже в том случае, когда эта функция нелинейна, что доказывается приведенным в п. 1.15.2 примером. Решение этой задачи традиционным способом часто бывает трудновыполнимым.

Достоинство MATLAB при решении задач математической статистики состоит в том, что он без труда позволяет проводить статистическую обработку больших объемов информации, в том числе и носящих многомерный характер. Это вытекает из самой концепции построения системы MATLAB (Matrix Laboratory), когда любой идентификатор является в общем случае многомерным массивом.

Применение рассмотренных в работе функций интегрирования позволяет с успехом обходиться без таблиц распределений Стьюдента, Пирсона, Фишера — Снедекора, Колмогорова — Смирнова, нормального распределения. Это также подтверждается многочисленными примерами.

Сдедует отметить возможности и удобство наглядного представления результатов, которые дает пакет расширения MuPAD. Удобство пользования этим пакетом для символьных преобразований и графического отображения информации демонстрируется при решении многих задач. В приводимой в библиогафическом списке литературе этому пакету не уделяется внимание.

Все возможности системы MATLAB, конечно, не могут быть изложены в рамках одного учебного пособия. В частности, не рассмотрены вопросы создания собственных программ на языке

MATLAB, применение логических операторов, за исключением тех, которые применяются при задании кусочных функций. Не рассмотрены задачи корреляционного анализа, дисперсионного анализа. Большие возможности MATLAB представляет в области имитационного моделирования сложных систем, в области финансового анализа и других областях прикладных наук. Эти темы достаточно обширны и требуют отдельного рассмотрения.

ПРИЛОЖЕНИЯ

Приложение 1

ПРИБЛИЖЕННЫЕ ЗНАЧЕНИЯ СТАНДАРТНОГО НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ $N(u)=\dfrac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{u} e^{-x^2/2}dx$, УМНОЖЕННЫЕ НА 10^5

u	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	50 000	50 399	50 798	51 197	51 595	51 994	52 392	52 790	53 188	53 586
0,1	53 983	54 380	54 776	55 172	55 567	55 962	56 356	56 749	57 142	57 535
0,2	57 926	58 317	58 706	59 095	59 483	59 871	60 257	60 642	61 026	61 409
0,3	61 791	62 172	62 552	62 930	63 307	63 683	64 058	64 431	64 803	65 173
0,4	65 542	65 910	66 276	66 640	67 003	67 364	67 724	68 082	68 439	68 793
0,5	69 146	69 497	69 847	70 194	70 540	70 884	71 226	71 566	71 904	72 240
0,6	72 575	72 907	73 237	73 565	73 891	74 215	74 537	74 857	75 175	75 490
0,7	75 804	76 115	76 424	76 730	77 035	77 337	77 637	77 935	78 230	78 524
0,8	78 814	79 103	79 389	79 673	79 955	80 234	80 511	80 785	81 057	81 327
0,9	81 594	81 859	82 121	82 381	82 639	82 894	83 147	83 398	83 646	83 891
1	84 134	84 375	84 614	84 849	85 083	85 314	85 543	85 769	85 993	86 214
1,1	86 433	86 650	86 864	87 076	87 286	87 493	87 698	87 900	88 100	88 298
1,2	88 493	88 686	88 877	89 065	89 251	89 435	89 617	89 796	89 973	90 147
1,3	90 320	90 490	90 658	90 824	90 988	91 149	91 308	91 456	91 621	91 774
1,4	91 924	92 073	92 220	92 364	92 507	92 647	92 785	92 922	93 056	93 189
1,5	93 319	93 448	93 574	93 699	93 822	93 943	94 062	94 179	94 295	94 408
1,6	94 520	94 630	94 738	94 845	94 950	95 053	95 154	95 254	95 352	95 449
1,7	95 543	95 637	95 728	95 818	95 907	95 994	96 080	96 164	96 246	96 327
1,8	96 407	96 485	96 562	96 638	96 712	96 784	96 855	96 926	96 995	97 062
1,9	97 128	97 193	97 257	97 320	97 381	97 441	97 500	97 558	97 615	97 670
2	97 725	97 778	97 831	97 882	97 932	97 982	98 030	98 077	98 124	98 169
2,1	98 214	98 257	98 300	98 341	98 382	98 422	98 461	98 500	98 537	98 574
2,2	98 610	98 645	98 679	98 713	98 745	98 778	98 809	98 840	98 870	98 899

2,3	98 928	98 956	98 983	99 010	99 036	99 061	99 086	99 111	99 134	99 158
2,4	99 180	99 202	99 224	99 245	99 266	99 286	99 305	99 324	99 343	99 361
2,5	99 379	99 396	99 413	99 430	99 446	99 461	99 477	99 492	99 506	99 520
2,6	99 534	99 547	99 560	99 573	99 585	99 598	99 609	99 621	99 632	99 643
2,7	99 653	99 664	99 674	99 683	99 693	99 702	99 711	99 720	99 728	99 736
2,8	99 744	99 752	99 760	99 767	99 774	99 781	99 788	99 795	99 801	99 807
2,9	99 813	99 819	99 825	99 831	99 836	99 841	99 846	99 851	99 856	99 861
3	99 865	99 869	99 874	99 878	99 882	99 886	99 889	99 893	99 896	99 900
3,1	99 903	99 906	99 910	99 913	99 916	99 918	99 921	99 924	99 926	99 929
3,2	99 931	99 934	99 936	99 938	99 940	99 942	99 944	99 946	99 948	99 950
3,3	99 952	99 953	99 955	99 957	99 958	99 960	99 961	99 962	99 964	99 965
3,4	99 966	99 968	99 969	99 970	99 971	99 972	99 973	99 974	99 975	99 976
3,5	99 977	99 978	99 978	99 979	99 980	99 981	99 981	99 982	99 983	99 983

Приложение 2

КРИТИЧЕСКИЕ ТОЧКИ *t*-РАСПРЕДЕЛЕНИЯ

$$\beta = P(t < t_0) = \int_{-\infty}^{t_0} p(t)dt$$

k	Доверительная вероятность β										
	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99
1	0,1584	0,3249	0,5095	0,7265	1,0000	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205
2	0,1421	0,2887	0,4447	0,6172	0,8165	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646
3	0,1366	0,2767	0,4242	0,5844	0,7649	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407
4	0,1338	0,2707	0,4142	0,5686	0,7407	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469
5	0,1322	0,2672	0,4082	0,5594	0,7267	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649
6	0,1311	0,2648	0,4043	0,5534	0,7176	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427
7	0,1303	0,2632	0,4015	0,5491	0,7111	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980
8	0,1297	0,2619	0,3995	0,5459	0,7064	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965
9	0,1293	0,2610	0,3979	0,5435	0,7027	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214
10	0,1289	0,2602	0,3966	0,5415	0,6998	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638
11	0,1286	0,2596	0,3956	0,5399	0,6974	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181
12	0,1283	0,2590	0,3947	0,5386	0,6955	0,8726	1,0832	1,3562	1,7823	2,1788	2,6810
13	0,1281	0,2586	0,3940	0,5375	0,6938	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503
14	0,1280	0,2582	0,3933	0,5366	0,6924	0,8681	1,0763	1,3450	1,7613	2,1448	2,6245
15	0,1278	0,2579	0,3928	0,5357	0,6912	0,8662	1,0735	1,3406	1,7531	2,1314	2,6025
16	0,1277	0,2576	0,3923	0,5350	0,6901	0,8647	1,0711	1,3368	1,7459	2,1199	2,5835
17	0,1276	0,2573	0,3919	0,5344	0,6892	0,8633	1,0690	1,3334	1,7396	2,1098	2,5669
18	0,1274	0,2571	0,3915	0,5338	0,6884	0,8620	1,0672	1,3304	1,7341	2,1009	2,5524
19	0,1274	0,2569	0,3912	0,5333	0,6876	0,8610	1,0655	1,3277	1,7291	2,0930	2,5395
20	0,1273	0,2567	0,3909	0,5329	0,6870	0,8600	1,0640	1,3253	1,7247	2,0860	2,5280
21	0,1272	0,2566	0,3906	0,5325	0,6864	0,8591	1,0627	1,3232	1,7207	2,0796	2,5176
22	0,1271	0,2564	0,3904	0,5321	0,6858	0,8583	1,0614	1,3212	1,7171	2,0739	2,5083
23	0,1271	0,2563	0,3902	0,5317	0,6853	0,8575	1,0603	1,3195	1,7139	2,0687	2,4999
24	0,1270	0,2562	0,3900	0,5314	0,6848	0,8569	1,0593	1,3178	1,7109	2,0639	2,4922
25	0,1269	0,2561	0,3898	0,5312	0,6844	0,8562	1,0584	1,3163	1,7081	2,0595	2,4851
26	0,1269	0,2560	0,3896	0,5309	0,6840	0,8557	1,0575	1,3150	1,7056	2,0555	2,4786

27	0,1268	0,2559	0,3894	0,5306	0,6837	0,8551	1,0567	1,3137	1,7033	2,0518	2,4727
28	0,1268	0,2558	0,3893	0,5304	0,6834	0,8546	1,0560	1,3125	1,7011	2,0484	2,4671
29	0,1268	0,2557	0,3892	0,5302	0,6830	0,8542	1,0553	1,3114	1,6991	2,0452	2,4620
30	0,1267	0,2556	0,3890	0,5300	0,6828	0,8538	1,0547	1,3104	1,6973	2,0423	2,4573
31	0,1267	0,2555	0,3889	0,5298	0,6825	0,8534	1,0541	1,3095	1,6955	2,0395	2,4528
32	0,1267	0,2555	0,3888	0,5297	0,6822	0,8530	1,0535	1,3086	1,6939	2,0369	2,4487
33	0,1266	0,2554	0,3887	0,5295	0,6820	0,8526	1,0530	1,3077	1,6924	2,0345	2,4448
34	0,1266	0,2553	0,3886	0,5294	0,6818	0,8523	1,0525	1,3070	1,6909	2,0322	2,4411
35	0,1266	0,2553	0,3885	0,5292	0,6816	0,8520	1,0520	1,3062	1,6896	2,0301	2,4377
36	0,1266	0,2552	0,3884	0,5291	0,6814	0,8517	1,0516	1,3055	1,6883	2,0281	2,4345
37	0,1265	0,2552	0,3883	0,5289	0,6812	0,8514	1,0512	1,3049	1,6871	2,0262	2,4314
38	0,1265	0,2551	0,3882	0,5288	0,6810	0,8512	1,0508	1,3042	1,6860	2,0244	2,4286
39	0,1265	0,2551	0,3882	0,5287	0,6808	0,8509	1,0504	1,3036	1,6849	2,0227	2,4258
40	0,1265	0,2550	0,3881	0,5286	0,6807	0,8507	1,0500	1,3031	1,6839	2,0211	2,4233
41	0,1264	0,2550	0,3880	0,5285	0,6805	0,8505	1,0497	1,3025	1,6829	2,0195	2,4208
42	0,1264	0,2550	0,3880	0,5284	0,6804	0,8503	1,0494	1,3020	1,6820	2,0181	2,4185
43	0,1264	0,2549	0,3879	0,5283	0,6802	0,8501	1,0491	1,3016	1,6811	2,0167	2,4163
44	0,1264	0,2549	0,3878	0,5282	0,6801	0,8499	1,0488	1,3011	1,6802	2,0154	2,4141
45	0,1264	0,2549	0,3878	0,5281	0,6800	0,8497	1,0485	1,3006	1,6794	2,0141	2,4121
46	0,1264	0,2548	0,3877	0,5281	0,6799	0,8495	1,0483	1,3002	1,6787	2,0129	2,4102
47	0,1263	0,2548	0,3877	0,5280	0,6797	0,8493	1,0480	1,2998	1,6779	2,0117	2,4083
48	0,1263	0,2548	0,3876	0,5279	0,6796	0,8492	1,0478	1,2994	1,6772	2,0106	2,4066
49	0,1263	0,2547	0,3876	0,5278	0,6795	0,8490	1,0475	1,2991	1,6766	2,0096	2,4049
50	0,1263	0,2547	0,3875	0,5278	0,6794	0,8489	1,0473	1,2987	1,6759	2,0086	2,4033
100	0,1260	0,2540	0,3864	0,5261	0,6770	0,8452	1,0418	1,2901	1,6602	1,9840	2,3642

Приложение 3

КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ ПИРСОНА

k	**Вероятности P($\chi^2 > \chi^2_{\text{кр}}$)**														
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,005	0,001
1	1,00026	0,0006	0,0039	0,016	0,064	0,148	0,455	1,07	1,64	2,7	3,8	5,4	6,6	7,9	10,83
2	0,020	0,040	0,103	0,211	0,446	0,713	1,386	2,41	3,22	4,6	6,0	7,8	9,2	11,6	13,8
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,66	4,64	6,3	7,8	9,8	11,3	12,8	16,3
4	0,30	0,43	0,71	1,06	1,65	2,19	3,36	4,9	6,0	7,8	9,5	11,7	13,3	14,9	18,5
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,1	7,3	9,2	11,1	13,4	15,1	16,3	20,5
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,2	8,6	10,6	12,6	15,0	16,8	18,6	22,5
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,4	9,8	12,0	14,1	16,6	18,5	20,3	24,3
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,5	11,0	13,4	15,5	18,2	20,1	21,9	26,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7	12,2	14,7	16,9	19,7	21,7	23,6	27,9
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2	25,2	29,6
11	3,1	3,6	4,6	5,6	7,0	8,1	10,3	12,9	14,6	17,3	19,7	22,6	24,7	26,8	31,3
12	3.6	4,2	5,2	6,3	7,8	9,0	11,3	14,0	15,8	18,5	21,0	24,1	26,2	28,3	32,9
13	4,1	4,8	5,9	7,0	8,6	9,9	12,3	15,1	17,0	19,8	22,4	25,5	27,7	29,8	34,5
14	4,7	5,4	6,6	7,8	9,5	10,8	13,3	16,2	18,2	21,1 .	23,7	26,9	29,1	31,0	36,1
15	5,2	6,0	7,3	8,5	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6	32,5	37,7
16	5,8	6,6	8,0	9,3	11,2	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0	34,0	39,2
17	6,4	7,3	8,7	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4	35,5	40,8
18	7,0	7,9	9,4	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8	37,0	42,3
19	7,6	8,6	10,1	11,7	13,7	15,4	18,3	21,7	23,9	27,2	30,1	33,7	36,2	38,5	43,8
20	8,3	9,2	10,9	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6	40,0	45,3
21	8,9	9,9	11,6	13.2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9	41,5	46,8
22	9,5	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	37,7	40,3	42,5	48,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6	44,0	49,7
24	l0,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0	45,5	51,2
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2	30,7	34,4	37,7	41,6	44,3	47,0	52,6
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,3	31,8	35,6	38,9	42,9	45,6	48,0	54,1
27	12,9	14,1	16,2	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0	49,5	55,5
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3	51,0	56,9
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6	52,5	58,3
30	15,0	16,3	18,5	20,6	23,4	25,5	29,3	33,5	36,3	40,3	43,8	48,0	50,9	54,0	59,7

Приложение 4

КРИТИЧЕСКИЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ ФИШЕРА

(k_1 — число степеней свободы большей дисперсии,
k_2 — число степеней свободы меньшей дисперсии)

Уровень значимости $\alpha = 0{,}01$

k_2	k_1											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	608	6106
2	98,49	99,01	99,17	99,25	99,30	99,33	99,34	99,36	99,38	99,40	99,4	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,1	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,4	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94.	3,86	3,80
15	8,68	6,36	5,42	4;89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости $\alpha = 0{,}05$

k_2	k_1											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,93	5,91

Продолжение табл.

k_2	k_1											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,62	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	*2,72*	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	*2,65*	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Приложение 5

НОРМИРОВАННАЯ ФУНКЦИЯ ГАУССА

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

y	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0604
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	02831	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	ОНО	0107
2,7	0104	010Г	0099	0096	0093	0091	0088	0086	0084	0081

2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3.3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0005	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Приложение 6
КРИТИЧЕСКИЕ ТОЧКИ КРИТЕРИЯ КОЛМОГОРОВА — СМИРНОВА И ФУНКЦИИ $K(\psi_{кр}) = \sum_{i=1}^{\infty}(-1)^i e^{-2i^2\psi_{кр}^2}$

$\psi_{кр}$	$K(\psi_{кр})$	$\psi_{кр}$	$K(\psi_{кр})$	$\psi_{кр}$	$K(\psi_{кр})$	$\psi_{кр}$	$K(\psi_{кр})$
0,28	0,0000	0,90	0,6073	1,52	0,9803	2,14	0,9998
0,32	0,0000	0,94	0,6601	1,56	0,9846	2,18	0,9999
0,36	0,0005	0,98	0,7079	1,60	0,9880	2,22	0,9999
0,40	0,0028	1,02	0,7500	1,64	0,9908	2,26	0,9999
0,44	0,0097	1,06	0,7889	1,68	0,9929	2,30	0,9999
0,48	0,0247	1,10	0,8223	1,72	0,9946	2,34	0,9999
0,52	0,0503	1,14	0,8514	1,76	0,9956	2,38	0,9999
0,56	0,0876	1,18	0,8765	1,80	0,9969	2,42	0,9999
0,60	0,1357	1,22	0,8981	1,84	0,9977	2,46	1,0
0,64	0,1927	1,26	0,9164	1,88	0,9983	2,50	1,0
0,68	0,2558	1,30	0,9319	1,92	0,9987	2,60	1,0
0,72	0,3223	1,34	0,9449	1,96	0,9991	2,70	1,0
0,76	0,3896	1,38	0,9557	2,00	0,9993	2,80	1,0
0,80	0.4559	1,42	0,9646	2,04	0,9995	2,90	1,0
0,84	0,5194	1,46	0,9718	2,08	0,9997	3,00	1,0
0,88	0,5791	1,5.0	0,9778	2,12	0,9998	-	-

Приложение 7

ФУНКЦИИ MATLAB

nchoosec(k,r) вычисляет число сочетаний C_k^r.

sym2poly(s) вычисляет коэффициенты произвольного многочлена *s* при степенях аргумента эквивалентного полинома.

double(x) переводит результат вычисления в формат числа с плавающей точкой двойной точности.

norminv(F, m, σ) вычисляет аргумент *u* по заданному значению интеграла $F(u;m,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{u} e^{-(x-m)^2/2\sigma^2}dx$ с параметрами *m*, σ.

binopdf(x, n, p) возвращает значения вероятностей биномиального распределения с параметрами *n*, *p* для заданных значений *x*.

binocdf(x, n, p) возвращает значения функции биномиального распределения с параметрами *n*, *p* для заданных значений *x*.

poisspdf(x, λ) вычисляет вероятность того, что случайная величина, распределенная по закону Пуассона с параметром λ, примет заданное значение *x*.

poisscdf(x, λ) вычисляет вероятность того, что случайная величина, распределенная по закону Пуассона с параметром λ, примет значение не больше заданного значения *x*.

unifpdf(x, a, b) возвращает численное значение плотности равномерного распределения с параметрами *a*, *b* в точке *x*.

unifcdf(x, a, b) возвращает численное значение функции распределения с параметрами *a*, *b* в точке *x*.

unifstat(a, b) возвращает математическое ожидание *mx* и дисперсию *Dx* случайной величины *X*, равномерно распределенной на интервале [*a*, *b*].

normpdf(x, mx, D) возвращает численное значение нормальной плотности распределения с параметрами *mx* и *D* в точке *x*.

normcdf(x, mx, D) возвращает численное значение нормальной функции распределения с параметрами *mx* и *D* в точке *x*.

normpdf(x,1,1) вычисляет значения нормированной функции Гаусса с заданными параметрами $mx = 1$, $D = 1$ в точке x.

norminv(P,mx,D) возвращает значение аргумента, соответствующее заданному значению P нормальной функции распределения, т. е. вычисляет значения функции, обратной по отношению к заданной нормальной функции рапределения.

int(f, x = a...b) вычисляет определенный интеграл функции f в пределах изменения аргумента x от a до b (функция в пакете MuPAD).

int(f, a, b) вычисляет определенный интеграл функции f в пределах изменения аргумента от a до b (функция MATLAB).

int(f,x) находит первообразную функцию для подынтегральной функции f (функция в пакете MuPAD).

int(f) находит первообразную функцию для подынтегральной функции f (функция MATLAB).

piecewise([условие1, объект1], [условие2, объект2], ...) создает кусочную функцию (функция в пакете MuPAD).

plot — функция рисования графиков.

stairs — функция рисования «ступенчатого» графика.

cumsum(x) возвращает накопленные (кумулятивные) суммы элементов вектора x.

limit(f,x,a) вычисляет двусторонний предел функции f при $x \to a$.

limit(f,x,a,'left') вычисляет левосторонний предел функции $f(x)$ при $x \to a$ $(x < a)$.

limit(f,x,a,'right') вычисляет правосторонний предел функции $f(x)$ при $x \to a$ $(x > a)$.

limit(f, x, infinity) вычисляет предел $\lim\limits_{x\to+\infty} f(x)$.

limit(f, x, –infinity) вычисляет предел $\lim\limits_{x\to-\infty} f(x)$.

diff(f,n,x) вычисляет производную n-го порядка функции $f(x)$.

diff(f,arg) вычисляет частную производную первого порядка от функции f нескольких переменных по переменной arg.

diff(f,arg,n) вычисляет частную производную n-го порядка от функции f нескольких переменных по переменной arg.

polival(A,arg) вычисляет полином для заданного значения аргумента arg, где A — вектор коэффициентов полинома.

roots(A) вычисляет корни полинома, заданного вектором коэффициентов.

solve('f$_1$','f$_2$', ...,'f$_N$') — символьное решение систем уравнений, где f_1, f_2, ..., f_N — функции уравнений системы, заданные в неявном виде.

symsum(an, n1, n2) вычисляет сумму ряда, заданного общим членом *an* для номеров членов ряда с $n1$ по $n2$.

taylor(f, n, a) — разложение функции $f(x)$ в ряд Тэйлора, содержащий n членов по степеням $(x - a)$;

sort(x(:)) преобразует двумерный массив x исходных данных в одномерный и упорядочивает его члены в порядке возрастания.

length(x) вычисляет число элементов массива данных x.

range(x) вычисляет размах выборки x.

mean(x) вычисляет выборочное среднее выборки x.

var(x) вычисляет выборочную дисперсию выборки x.

std(x) вычисляет выборочное стандартное отклонение выборки x.

skewness(x,0) вычисляет несмещенную оценку ассиметрии.

kurtosis(x,0) вычисляет несмещенную оценку эксцесса.

median(x) вычисляет несмещенную оценку медианы.

hist(x) строит гистограмму распределения по выборке x.

ttest(x, Mx, alpha) выполняет проверку основной гипотезы по критерию Стьюдента о том, что выборка x подчиняется нормальному распределению с математическим ожиданием Mx при уровне значимости alpha.

vartest(x, Dx, alpha) выполняет проверку основной гипотезы по критерию χ^2 о том, что выборка x подчиняется нормальному распределению с дисперсией Dx при уровне значимости alpha.

chi2inv(p,k) вычисляет обратное значение функции распределения χ^2 по заданным значением вероятности $p_1 = \frac{\alpha}{2}$ или $p_2 = 1 - \frac{\alpha}{2}$ для k степеней свободы.

mle(x, ′Distribution′, name, ′alpha′, alpha) вычисляет методом макимального правдоподобия (ММП) интервальные оценки параметров для заданного распределения по выборке x при уровне значимости alpha.

ttest2(x,y,alpha) — проверка основной гипотезы о равенстве математических ожиданий генеральных совокупностей, представленных выборками x и y при уровне значимости alpha.

vartest2(x,y,alpha) — проверка основной гипотезы о равенстве дисперсий генеральных совокупностей, представленных выборками x и y при уровне значимости alpha.

chi2gof(x,'cdf',{'Distribution',param}) — тест по критерию χ^2 о принадлежности выборки x с выборочными параметрами param заданному кумулятивному закону распределения.

kstest(x, CDF, alpha) — тест по критерию Колмогорова — Смирнова о принадлежности выборки *x* теоретическому закону распределения CDF с известнами параметрами. *CDF* (Cumulative Distribution Functions) — матрица, состоящая из двух вектор-столбцов: первый содержит заданную выборку *x*, второй — соответствующие значения теоретической куммулятивной функции распределения $F(x)$; *alpha* — уровень значимости.

cdf('name',x,A,B) расчитывает теоретическую функцию распределения, с именем 'name' с заданными параметрами *A*, *B* для массива значений *x*.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Вентцель*, *Е. С.* Теория вероятностей. — М. : Высш. шк., 1999. — 576 с.

2. *Алибеков*, *И. Ю.* Избранные главы высшей математики в экономических приложениях в среде MATLAB : учебн. пособие. Ч. 1 / И. Ю. Алибеков, Н. И. Царькова, О. Л. Казаков. — М. : МГИУ, 2015. — 10 с.

3. *Иглин*, *С. П.* Теория вероятностей и математическая статистика на базе MATLAB. — Харьков : НТУ ХПИ, 2006. — 612 с.

4. *Кремер*, *Н. Ш.* Теория вероятностей и математическая статистика. — М. : ЮНИТИ-ДАНА, 2007. — 551 с.

5. *Дьяконов*, *В. П.* Математические пакеты расширения MATLAB. Специальный справочник / В. П. Дьяконов, В. В. Круглов. — СПб. : Питер, 2001. — 480 с.

6. *Чен*, *К.* MATLAB в математических исследованиях / К. Чен, П. Джиблин, А. Ирвинг. — М. : Мир, 2001. — 346 с.

7. *Лазарев*, *Ю. Ф.* Начала программирования в среде MatLAB. — Киев : НГУУ «КПИ», 2003. — 424 с.

8. *Ануфриев*, *И. Е.* MATLAB 7 / И. Е. Ануфриев, А. Б. Смирнов, Е. Н. Смирнова. — СПб. : БХВ-Петербург, 2005. — 1104 с.

9. *Половко*, *А. М.* Matlab для студентов / А. М. Половко, П. Н. Бутусов. — СПб. : БХВ-Петербург, 2005. — 321 с.

Игорь Юсупович АЛИБЕКОВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА В СРЕДЕ MATLAB

Учебное пособие

Издание второе, стереотипное

Зав. редакцией литературы по информационным технологиям и системам связи *О. Е. Гайнутдинова*

ЛР № 065466 от 21.10.97
Гигиенический сертификат 78.01.10.953.П.1028
от 14.04.2016 г., выдан ЦГСЭН в СПб

Издательство «ЛАНЬ»
lan@lanbook.ru; www.lanbook.com
196105, Санкт-Петербург, пр. Ю. Гагарина, д. 1, лит. А.
Тел./факс: (812) 336-25-09, 412-92-72.
Бесплатный звонок по России: 8-800-700-40-71

Подписано в печать 20.10.20.
Бумага офсетная. Гарнитура Школьная. Формат 84×108 $^{1}/_{32}$.
Печать офсетная. Усл. п. л. 9,66. Тираж 50 экз.

Заказ № 1348-20.

Отпечатано в полном соответствии
с качеством предоставленного оригинал-макета
в АО «Т8 Издательские Технологии».
109316, г. Москва, Волгоградский пр., д. 42, к. 5.