

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

Кафедра вычислительных методов и программирования

***ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА***

Методические указания по типовому расчету
для студентов всех специальностей
заочной формы обучения

Минск БГУИР 2009

УДК 519.2(076.6)
ББК 22.17я7
Т33

С о с т а в и т е л и:
А. И. Волковец, А. Б. Гуринович, А. В. Аксенчик

Р е ц е н з е н т ы:
доцент кафедры высшей математики и физики
учреждения образования «Военная академия Республики Беларусь»,
кандидат физико-математических наук Т. А. Макарова;
доцент кафедры вычислительной техники Белорусского государственного
аграрного технического университета, кандидат технических наук А. И. Шакирин

Теория вероятностей и математическая статистика : метод. указания по типовому расчету для студ. всех спец. заоч. формы обуч. / сост. А. И. Волковец, А. Б. Гуринович, А. В. Аксенчик. – Минск : БГУИР, 2009. – 64 с. : ил.

ISBN 978-985-488-441-7

Методические указания содержат краткое изложение необходимого теоретического материала и руководство к решению всех 11 видов задач типового расчета по 17 темам, определенным Типовой рабочей программой изучения данной дисциплины. В учебный комплект, помимо указаний, входят «Теория вероятностей и математическая статистика: сборник задач по типовому расчету» (сост. А. В. Аксенчик и др.) и «Теория вероятностей и математическая статистика. Конспект лекций для студентов всех специальностей очной формы обучения БГУИР» А. И. Волковца, А. Б. Гуринович, изданные в БГУИР в 2007, 2004 гг. соответственно.

УДК 519.2(076.6)
ББК 22.17я7

ISBN 978-985-488-441-7

© Волковец А. И., Гуринович А. Б.,
Аксенчик А. В., составление, 2009
© УО «Белорусский государственный университет
информатики и радиоэлектроники», 2009

СОДЕРЖАНИЕ

1. СЛУЧАЙНЫЕ СОБЫТИЯ. ВЕРОЯТНОСТЬ СОБЫТИЯ	4
2. ТЕОРЕМЫ СЛОЖЕНИЯ И УМНОЖЕНИЯ ВЕРОЯТНОСТЕЙ.....	8
3. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ. ФОРМУЛА БАЙЕСА	11
4. ПОВТОРЕНИЕ НЕЗАВИСИМЫХ ОПЫТОВ	14
5. ДИСКРЕТНАЯ СЛУЧАЙНАЯ ВЕЛИЧИНА.....	16
6. НЕПРЕРЫВНАЯ СЛУЧАЙНАЯ ВЕЛИЧИНА.....	20
7. ЗАКОН РАСПРЕДЕЛЕНИЯ ФУНКЦИИ СЛУЧАЙНОГО АРГУМЕНТА	24
8. ДВУХМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ	26
9. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СУММЫ И ПРОИЗВЕДЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН.....	31
10. ОБРАБОТКА ОДНОМЕРНОЙ ВЫБОРКИ	34
11. ОБРАБОТКА ДВУХМЕРНОЙ ВЫБОРКИ.....	51
ЛИТЕРАТУРА	60
Приложение 1. Значения функции Лапласа.....	61
Приложение 2. Таблица распределения Стьюдента.....	62
Приложение 3. Таблица распределения Пирсона.....	63
Приложение 4. Таблица распределения Колмогорова	64

1. СЛУЧАЙНЫЕ СОБЫТИЯ. ВЕРОЯТНОСТЬ СОБЫТИЯ

Случайное событие – любой факт в опыте со случайным исходом, который может произойти или не произойти. Любое случайное событие A , возможное в данном опыте, есть некоторое подмножество универсального множества Ω исходов этого опыта.

Событие A называется **достоверным**, если $A = \Omega$, т. е. событие происходит в каждом опыте.

Событие A называется **невозможным**, если $A = \emptyset$, т. е. событие никогда не происходит в данном опыте.

Противоположным событием \bar{A} называется событие, которое происходит тогда, когда не происходит событие A .

Суммой (объединением) двух событий A и B (обозначается $A + B, A \cup B$) называется событие, состоящее в появлении **или** события A , **или** события B , **или** A и B одновременно.

Произведением (пересечением) двух событий A и B (обозначается $A \cdot B, A \cap B$) называется событие, состоящее в появлении **и** события A , **и** события B одновременно или совместно.

Несовместными событиями A и B называются такие, которые не могут произойти одновременно в одном опыте. Для несовместных событий $A \cap B = \emptyset$.

Аксиома 1. Вероятность $p(A)$ случайного события A есть функция множества элементарных исходов, благоприятных событию A ; вероятность любого события принимает значения

$$0 \leq p(A) \leq 1, \quad (1.1)$$

причем $p(\emptyset) = 0, p(\Omega) = 1$.

Аксиома 2. Вероятность суммы несовместных случайных событий равна сумме вероятностей этих событий:

$$p\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n p(A_i), \quad A_i \cdot A_j = \emptyset, \quad \forall i \neq j. \quad (1.2)$$

Следствие аксиом 1 и 2. Вероятность прямого события $p(A)$ и вероятность противоположного события $p(\bar{A})$ связаны соотношением

$$p(\bar{A}) = 1 - p(A). \quad (1.3)$$

Классическое определение вероятности: вероятность события A определяется по формуле

$$p(A) = \frac{m}{n}, \quad (1.4)$$

где m – число исходов, благоприятствующих появлению события;
 n – число всех возможных равновероятных исходов данного опыта.

Геометрическое определение вероятности. Пусть в некоторую область Ω случайным образом бросается точка T , причем все точки области Ω равноправны в отношении попадания точки T (рис. 1.1).

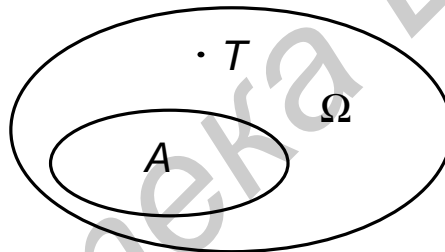


Рис. 1.1

Тогда за вероятность попадания точки T в область A принимается отношение

$$p(A) = \frac{S(A)}{S(\Omega)}, \quad (1.5)$$

где $S(A)$ и $S(\Omega)$ – геометрические меры (длина, площадь, объем и т. д.) областей A и Ω соответственно.

Основные комбинаторные формулы

При решении задач по классическому определению вероятности используются следующие комбинаторные формулы.

Размещения:

– без повторения элементов:

$$A_n^m = n(n-1) \cdot \dots \cdot (n-(m-1)) = \frac{n!}{(n-m)!}; \quad (1.6)$$

– с повторением элементов:

$$A_n^m = n^m. \quad (1.7)$$

Формулы для размещений позволяют подсчитать количество комбинаций из n элементов по m , где комбинации будут различаться как самими элементами, так и расположением относительно друг друга.

Сочетания:

– без повторений элементов:

$$C_n^m = \frac{n(n-1) \cdot \dots \cdot (n-(m-1))}{1 \cdot 2 \cdot 3 \cdot \dots \cdot m} = \frac{n!}{(n-m)!m!}; \quad (1.8)$$

– с повторением элементов:

$$C_n^m = C_{n+m-1}^m = \frac{(n+m-1)!}{m!(n-1)!}. \quad (1.9)$$

Формулы для сочетаний позволяют подсчитать количество комбинаций из n элементов по m , где комбинации будут различаться только самими элементами, т. е. каждая комбинация хотя бы одним элементом должна отличаться от другой.

Перестановки:

$$P_n = A_n^n = 1 \cdot 2 \cdot \dots \cdot n = n!, \quad P_0 = 0! = 1. \quad (1.10)$$

Формула для перестановок позволяет подсчитать количество комбинаций из n элементов, где комбинации будут различаться только расположением относительно друг друга.

Примеры

Пример 1.1. Какова вероятность того, что в наудачу взятом телефонном номере из 7 цифр все цифры различные?

Решение. Определим событие A . Событие A состоит в том, что в семизначном номере все цифры различны. Так как номер семизначный, а цифр всего 10, то общее число исходов n опыта равно числу размещений с повторением элементов из 10 по 7: $n = A_{10}^7 = 10^7$. Для подсчета

благоприятствующих исходов подходит формула для размещений без повторения элементов : $m = A_{10}^7$. Тогда

$$p(A) = \frac{m}{n} = \frac{A_{10}^7}{10^7} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4}{10^7} \approx 0,06.$$

Пример 1.2. Наудачу взяты два положительных числа x и y , причем $x \leq 5$, $y \leq 4$. Найти вероятность того, что $y + ax - b \leq 0$ и $y - cx \leq 0$, если $a = 2$, $b = 10$, $c = 2$.

Решение. Подставляя значения коэффициентов в неравенства, получаем

$$\begin{cases} y + 2x \leq 10, \\ y \leq 2x. \end{cases} \quad (1.11)$$

Строим оси координат и область, которая определяет пространство элементарных событий Ω (рис. 1.2). Она задается неравенствами $x \leq 5$, $y \leq 4$ и на рис. 1.2 отображается в виде прямоугольника.

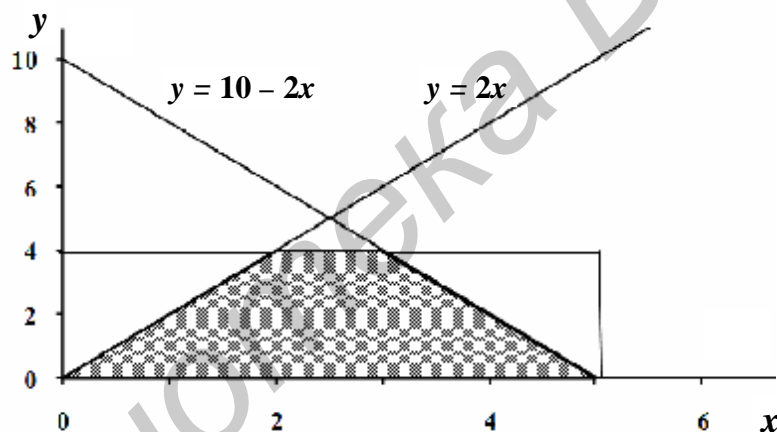


Рис. 1.2

Площадь прямоугольника $S_{\Omega} = 4 \cdot 5 = 20$ усл. ед. Область благоприятствующих исходов определяется неравенствами (1.11), поэтому на рис. 1.2 построены прямые, которые задаются этими неравенствами. Заштрихованная на рис. 1.2 область описывает благоприятствующие исходы (с учетом всех возможных значений), площадь этой заштрихованной трапеции равна

$$S_A = \frac{5+1}{2} \cdot 4 = 12 \text{ усл. ед. Тогда вероятность события } A$$

$$p(A) = \frac{S_A}{S_{\Omega}} = \frac{12}{20} = 0,6.$$

2. ТЕОРЕМЫ СЛОЖЕНИЯ И УМНОЖЕНИЯ ВЕРОЯТНОСТЕЙ

Вероятность **суммы** (объединения) двух произвольных случайных событий (т. е. тех, которые могут происходить совместно) равна сумме вероятностей каждого из событий минус вероятность их совместного появления:

$$p(A+B) = p(A) + p(B) - p(AB). \quad (2.1)$$

Для трех произвольных событий

$$p(A+B+C) = p(A) + p(B) + p(C) - p(AB) - p(BC) - p(AC) + p(ABC). \quad (2.2)$$

Для n произвольных событий

$$p(A_1 + A_2 + \dots + A_n) = 1 - p(\overline{A_1} \cdot \overline{A_2} \cdot \dots \cdot \overline{A_n}). \quad (2.3)$$

Событие A называется **независимым** от события B , если возможность наступления события A не зависит от того, произошло событие B или нет. В противном случае события являются **зависимыми**.

Условной вероятностью $p(B/A)$ называется вероятность события B , вычисленная при условии (в предположении), что событие A произошло. Для **независимых** событий $p(B/A) = p(B)$.

Вероятность **произведения** (пересечения) двух случайных событий равна вероятности одного из них, умноженной на условную вероятность второго при наличии первого:

$$p(A \cdot B) = p(A) \cdot p(B/A) = p(B) \cdot p(A/B). \quad (2.4)$$

Для независимых событий

$$p(A \cdot B) = p(A) \cdot p(B). \quad (2.5)$$

Вероятность произведения n произвольных событий A_i ($i = 1, 2, \dots, n$) равна

$$p(A_1 \cdot A_2 \cdot \dots \cdot A_n) = p(A_1) \cdot p(A_2/A_1) \cdot p(A_3/A_1 \cdot A_2) \cdot \dots \cdot p(A_n/A_1 \cdot A_2 \cdot \dots \cdot A_{n-1}), \quad (2.6)$$

где $p(A_k/A_1 \cdot \dots \cdot A_{k-1})$ – условная вероятность появления события A_k , при условии, что события A_1, A_2, \dots, A_{k-1} в данном опыте произошли, $k = 2, \dots, n$.

В случае независимых событий данная формула упрощается:

$$p(A_1 \cdot A_2 \cdot \dots \cdot A_n) = p(A_1) \cdot p(A_2) \cdot \dots \cdot p(A_n). \quad (2.7)$$

Примеры

Пример 2.1. Вычислительная машина (ВМ) состоит из n блоков. Вероятность безотказной работы в течении времени T (надежность) первого блока равна p_1 , второго – p_2 и т. д. Блоки отказывают независимо друг от друга. При отказе любого блока отказывает ВМ. Найти вероятность того, что ВМ откажет за время T .

Решение. Рассмотрим события A_1 – отказывает 1-й блок, A_2 – отказывает 2-й блок и т. д. Пусть событие B – отказ вычислительной машины. Это событие произойдет тогда, когда выполнится *или* событие A_1 , *или* событие A_2 и т. д. Следовательно, необходимо применить теорему о сумме, или объединении, n произвольных событий согласно формуле (2.3):

$$p(B) = p(A_1 + A_2 + \dots + A_n) = 1 - p(\bar{A}_1 \cdot \bar{A}_2 \cdot \dots \cdot \bar{A}_n).$$

События A_i являются независимыми, поэтому правая часть выражения запишется в виде

$$p(B) = 1 - p(\bar{A}_1 \cdot \bar{A}_2 \cdot \dots \cdot \bar{A}_n) = 1 - p(\bar{A}_1) \cdot p(\bar{A}_2) \cdot \dots \cdot p(\bar{A}_n).$$

Вероятности противоположных событий \bar{A}_i (здесь событие \bar{A}_i – i -й блок работает) даны в условии, т. е. $p(\bar{A}_i) = p_i$. Окончательно получаем

$$p(B) = 1 - p_1 \cdot p_2 \cdot \dots \cdot p_n = 1 - \prod_{i=1}^n p_i.$$

Пример 2.2. Дана схема электрической цепи (рис. 2.1).

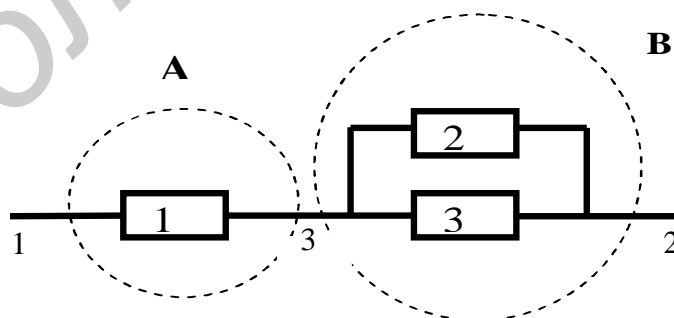


Рис. 2.1

Вероятности работы элементов цепи 1, 2, 3 соответственно равны $p_1 = 0,7$; $p_2 = 0,8$; $p_3 = 0,9$. Элементы цепи отказывают независимо друг от друга. Найти вероятность того, что ток пройдет из точки 1 в точку 2.

Решение. Опишем через события работу элементов цепи. Пусть событие A_1 состоит в том, что работает элемент 1, событие A_2 – элемент 2, событие A_3 – элемент 3. Тогда вероятности этих событий запишутся так: $p(A_1) = p_1, p(A_2) = p_2, p(A_3) = p_3$. Найдем вероятности противоположных событий (т.е. того, что элементы 1, 2, 3 не работают и ток через них не идет), используя (1.3):

$$p(\bar{A}_1) = 1 - p_1, p(\bar{A}_2) = 1 - p_2, p(\bar{A}_3) = 1 - p_3.$$

Анализируем заданную цепь и определяем участки цепи с последовательным и параллельным соединениями. На рис. 2.1 элементы 2, 3 соединены параллельно. А элемент 1 соединен последовательно с элементами 2, 3. Поэтому введем событие A , состоящее в том, что ток пройдет из точки 1 в точку 3, оно выполнится тогда, когда будет работать элемент 1. Можно записать $A = A_1$. Введем событие B , состоящее в том, что ток пройдет из точки 2 в точку 3; оно произойдет тогда, когда будут работать *или* элемент 2, *или* элемент 3. Тогда событие B можно описать так: $B = A_2 + A_3$. Рассмотрим событие C , состоящее в том, что ток пройдет из точки 1 в точку 2, оно выполнится тогда, когда выполнится *и* событие A , *и* событие B . Событие C запишется так: $C = A \cdot B$. По условию задачи необходимо найти вероятность события C (учтем, что события A и B независимы), используем (2.5):

$$p(C) = p(A \cdot B) = p(A) \cdot p(B). \quad (2.8)$$

Найдем вероятности событий, входящих в правую часть (2.8):

$$p(A) = p(A_1) = p_1,$$

а

$$\begin{aligned} p(B) &= p(A_2 + A_3) = \left| \text{см. (2.3)} \right| = \\ &= 1 - p(\bar{A}_2 \cdot \bar{A}_3) = 1 - p(\bar{A}_2) \cdot p(\bar{A}_3) = 1 - (1 - p_2)(1 - p_3). \end{aligned}$$

Подставив полученные значения в (2.8), получим

$$p(C) = p(A) \cdot p(B) = p_1 \cdot [1 - (1 - p_2)(1 - p_3)] = 0,7 \cdot [1 - 0,2 \cdot 0,1] = 0,686.$$

3. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ. ФОРМУЛА БАЙЕСА

Пусть проводится опыт, об условиях которого можно сделать n исключаящих друг друга предположений (гипотез): $H_1, H_2, \dots, H_n, H_i \cdot H_j = \emptyset, i \neq j,$

образующих полную группу $\sum_{i=1}^n H_i = \Omega$. Каждая из гипотез осуществляется

случайным образом и представляет собой случайное событие. Вероятности гипотез известны и равны

$$p(H_1), p(H_2), \dots, p(H_n), \sum_{i=1}^n p(H_i) = 1.$$

Рассмотрим некоторое событие A , которое может появиться только вместе с одной из гипотез. Известны условные вероятности события A для каждой из гипотез:

$$p(A/H_1), p(A/H_2), \dots, p(A/H_n).$$

Тогда **полная вероятность** события A определяется по формуле

$$p(A) = \sum_{i=1}^n p(H_i) \cdot p(A/H_i). \quad (3.1)$$

Пусть до проведения некоторого опыта об его условиях n можно сделать n исключаящих друг друга предположений (гипотез): $H_1, H_2, \dots, H_n, H_i \cdot H_j = \emptyset, i \neq j,$ образующих полную группу $\sum_{i=1}^n H_i = \Omega$. Вероятности гипотез $p(H_1), p(H_2), \dots, p(H_n)$

до опыта (априорные вероятности) известны, причем $\sum_{i=1}^n p(H_i) = 1$.

Опыт произведен, и произошло некоторое событие A . Тогда определить апостериорные (послеопытные) вероятности гипотез с учетом того, что произошло именно событие A $p(H_1/A), p(H_2/A), \dots, p(H_n/A)$, можно по **формуле Байеса**:

$$p(H_i/A) = \frac{p(H_i \cdot A)}{p(A)} = \frac{p(H_i)p(A/H_i)}{p(A)} = \frac{p(H_i)p(A/H_i)}{\sum_{j=1}^n p(H_j)p(A/H_j)}. \quad (3.2)$$

Примеры

Пример 3.1. Радиоприемное устройство имеет блок обработки сигналов, который позволяет отделить полезный сигнал от помехи без искажений. Если отношение уровня сигнала к уровню помехи менее 1,2, то вероятность выделить полезный сигнал без искажений равна 0,1; если отношение уровня сигнала к уровню помехи от 1,2 до 2, то вероятность – 0,8, а если превышает 2, то вероятность равна 1. Приемник принял сигнал, причем поступление сигнала с помехой любого уровня равновероятно. Найти вероятность того, что сигнал будет обработан без искажений.

Решение. Определим событие A – приемник обработал сигнал без искажений. Выдвигаем гипотезы: H_1 – приемник принял сигнал с отношением уровня сигнала к уровню помехи менее 1,2; H_2 – приемник принял сигнал с отношением уровня сигнала к уровню помехи от 1,2 до 2; H_3 – приемник принял сигнал с отношением уровня сигнала к уровню помехи более двух. Вероятности гипотез $p(H_1) = p(H_2) = p(H_3) = 1/3$ (так как по условию они равновероятны). Определим условные вероятности события A при каждой гипотезе: $p(A|H_1) = 0,1$, $p(A|H_2) = 0,8$, $p(A|H_3) = 1$. По формуле полной вероятности (3.1) найдем вероятность события A :

$$p(A) = \sum_{i=1}^3 p(H_i) \cdot p(A|H_i) = 1/3 \cdot (0,1 + 0,8 + 1) = 0,633.$$

Пример 3.2. Прибор состоит из двух блоков, обеспечивающих работу всего прибора. Вероятность безотказной работы в течение времени T (надежность) первого блока равна p_1 , второго – p_2 . Прибор испытывался в течение времени T и отказал. Найти вероятность того, что отказал только первый блок, а второй исправен ($p_1 = 0,5$; $p_2 = 0,7$).

Решение. Сформулируем событие A – прибор отказал. Это событие может произойти при следующих гипотезах: H_1 – отказал только первый блок, а второй исправен; H_2 – отказал второй блок, а первый исправен; H_3 – отказал первый блок, отказал второй блок; H_4 – работает первый блок, работает второй

блок. Как видим, гипотезы описывают сложные события. Для упрощения расчета вероятностей этих событий введем такие события: событие B_1 – работает первый блок; событие B_2 – работает второй блок. Тогда гипотезу H_1 через эти события можно описать так: $H_1 = \bar{B}_1 \cap B_2$, где \bar{B}_1 – противоположное событие, т. е. блок не работает. Аналогично распишем и другие гипотезы: $H_2 = B_1 \cap \bar{B}_2$, $H_3 = \bar{B}_1 \cap \bar{B}_2$, $H_4 = B_1 \cap B_2$. Так как события B_1 и B_2 независимы, то вероятности гипотез рассчитаем, используя теорему умножения вероятностей для независимых событий (2.5):

$$p(H_1) = p(\bar{B}_1 \cap B_2) = p(\bar{B}_1) \cdot p(B_2) = \begin{cases} p(B_1) = p_1, & p(\bar{B}_1) = 1 - p_1 \\ p(B_2) = p_2, & p(\bar{B}_2) = 1 - p_2 \end{cases} = (1 - p_1)p_2,$$

$$p(H_2) = p(B_1 \cap \bar{B}_2) = p(B_1) \cdot p(\bar{B}_2) = p_1(1 - p_2),$$

$$p(H_3) = p(\bar{B}_1 \cap \bar{B}_2) = p(\bar{B}_1) \cdot p(\bar{B}_2) = (1 - p_1)(1 - p_2),$$

$$p(H_4) = p(B_1 \cap B_2) = p(B_1) \cdot p(B_2) = p_1p_2.$$

Необходимо найти условные вероятности события A при каждой гипотезе. Тогда $p(A|H_1)$ – это условная вероятность того, что прибор вышел из строя при условии, что первый блок отказал, а второй исправен. Видим, что событие A всегда произойдет, если блок отказал, т. е. A – достоверное событие. Поэтому $p(A|H_1) = 1$. Аналогично $p(A|H_2) = 1$, $p(A|H_3) = 1$. А вот при четвертой гипотезе событие A никогда не произойдет, здесь A – невозможное событие, и вероятность его $p(A|H_4) = 0$.

Из условия задачи следует, что необходимо пересмотреть вероятность первой гипотезы, поэтому запишем формулу Байеса для H_1 :

$$p(H_1|A) = \frac{p(H_1) \cdot p(A|H_1)}{\sum_{i=1}^4 p(H_i) \cdot p(A|H_i)} = \frac{(1 - p_1)p_2 \cdot 1}{(1 - p_1)p_2 \cdot 1 + p_1(1 - p_2) \cdot 1 + (1 - p_1)(1 - p_2) \cdot 1 + p_1p_2 \cdot 0} = 0,538.$$

Из полученного решения следует, что до появления события A вероятность гипотезы H_1 была равна $p(H_1) = (1 - p_1)p_2 = (1 - 0,5) \cdot 0,7 = 0,35$, а с учетом появления события A изменилась значительно и стала равной 0,538.

4. ПОВТОРЕНИЕ НЕЗАВИСИМЫХ ОПЫТОВ

Пусть производится n независимых опытов. В результате событие A появляется в каждом опыте с вероятностью p и не появляется с вероятностью $q = (1 - p)$. Вероятность того, что в последовательности из n опытов событие A произойдет ровно k раз, вычисляется по **формуле Бернулли**:

$$P(n, k) = C_n^k \cdot p^k \cdot q^{n-k} = \frac{n!}{k!(n-k)!} p^k \cdot (1-p)^{n-k}. \quad (4.1)$$

Рекуррентная формула $P(n, k)$ имеет вид

$$P(n, k+1) = \frac{n-k}{k+1} \cdot \frac{p}{q} P(n, k). \quad (4.2)$$

Вероятность $P_n(k_1 \leq k \leq k_2)$ того, что в n опытах схемы Бернулли событие A появится от k_1 до k_2 раз ($0 \leq k_1 \leq k_2 \leq n$), равна

$$P(n, k_1 \leq k \leq k_2) = \sum_{k=k_1}^{k_2} P(n, k) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}. \quad (4.3)$$

Вероятность того, что при n независимых испытаниях событие A появится не менее m раз, вычисляется так:

$$P(n, m \leq k \leq n) = \sum_{k=m}^n P(n, k) = 1 - \sum_{k=0}^{m-1} P(n, k). \quad (4.4)$$

Здесь надо выбирать, какой ряд короче, и использовать его для расчета. Например, вероятность того, что в n опытах событие A появится хотя бы один раз, равна

$$P(n, 1 \leq k \leq n) = 1 - P(n, 0) = 1 - q^n. \quad (4.5)$$

Наивероятнейшее число m_0 наступлений события A при n опытах определяется из неравенства

$$np - q \leq m_0 \leq np + p. \quad (4.6)$$

Примеры

Пример 4.1. По каналу связи передается $n = 6$ сообщений, каждое из которых независимо от других оказывается искаженным с вероятностью $p = 0,2$. Найти вероятности следующих событий:

$A = \{\text{ровно два сообщения из шести искажены}\},$

$B = \{\text{все сообщения будут искажены}\},$

$C = \{\text{все сообщения будут переданы без искажений}\},$

$D = \{\text{не менее двух сообщений из шести искажены}\}.$

Решение. По формуле Бернулли (4.1) рассчитаем вероятности первых трех событий:

$$p(A) = P(6, 2) = C_6^2 \cdot p^2 \cdot (1-p)^4 = \frac{6!}{4! \cdot 2!} 0,2^2 \cdot 0,8^4 = 0,197,$$

$$p(B) = P(6, 6) = p^6 = 0,2^6 = 0,000064,$$

$$p(C) = P(6, 0) = (1-p)^6 = 0,262.$$

Вероятность события D определим по формуле (4.4), первый ряд требует для вычисления пять слагаемых, второй – только два:

$$\begin{aligned} p(B) = P(6, 2 \leq k \leq 6) &= P(6, 2) + P(6, 3) + P(6, 4) + P(6, 5) + P(6, 6) = 1 - P(6, 0) - P(6, 1) = \\ &= 1 - C_6^0 p^0 (1-p)^6 - C_6^1 p^1 (1-p)^5 = 1 - 0,8^6 - 6 \cdot 0,2^1 \cdot 0,8^5 = 0,345. \end{aligned}$$

Пример 4.2. Монету подбрасывают 10 раз. Чему будет равно наивероятнейшее число выпадений герба?

Решение. Пусть количество опытов $n = 10$, вероятность появления герба (событие A) симметричной монеты равна $p = 0,5$. Вероятность противоположного события \bar{A} $p(\bar{A}) = q = 1 - 0,5 = 0,5$. Тогда наивероятнейшее число m_0 определим, используя (4.6):

$$10 \cdot 0,5 - 0,5 \leq m_0 \leq 10 \cdot 0,5 + 0,5;$$

$$4,5 \leq m_0 \leq 5,5.$$

Отсюда следует, что $m_0 = 5$.

5. ДИСКРЕТНАЯ СЛУЧАЙНАЯ ВЕЛИЧИНА

Случайная величина (СВ) – это величина, которая в результате опыта со случайным исходом принимает то или иное значение, причем заранее до опыта неизвестно, какое именно. Обозначения случайной величины: X, Y ; а ее значения: x, y .

Случайная величина X называется **дискретной**, если ее множество возможных значений Ω_X – счетное, т. е. элементы множества можно расположить в определенном порядке и пронумеровать.

Закон распределения случайной величины – любое правило, устанавливающее соответствие между значениями случайной величины и вероятностями ее наступления.

Рядом распределения дискретной СВ X называется таблица, в верхней строке которой перечислены все возможные значения СВ $x_1, x_2, \dots, x_n (x_i < x_{i+1})$, а в нижней – вероятности их появления p_1, p_2, \dots, p_n , где $p_i = p(X = x_i)$:

x_i	x_1	x_2	...	x_n
p_i	p_1	p_2	...	p_n

Так как события $\{X = x_1\}, \{X = x_2\}, \dots, \{X = x_n\}$ несовместны и образуют полную группу, то справедливо контрольное соотношение

$$p_1 + p_2 + \dots + p_n = 1. \quad (5.1)$$

Функцией распределения случайной величины X называется вероятность того, что она примет значение меньше, чем аргумент x функции $F(x)$:
 $F(x) = p(X < x)$.

Свойства функции распределения:

1. $F(-\infty) = 0$ и $F(+\infty) = 1$.
2. Неубывающая функция: $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$.
3. Вероятность попадания значения СВ X в интервал $[a; b)$:

$$p\{a \leq X < b\} = F(b) - F(a). \quad (5.2)$$

Функция распределения дискретной СВ определяется так:

$$F(x) = \sum_{x_i < x} p(X = x_i) = \sum_{x_i < x} p_i, \quad (5.3)$$

где p_i – ряд распределения вероятности этой СВ.

В этом ряду суммируются вероятности всех тех значений x_i , которые по своей величине меньше, чем x – аргумент функции $F(x)$.

x_i	x_1	x_2	x_3	...	x_n	$> x_n$
p_i	p_1	p_2	p_3	...	p_n	0
$F(x_i)$	0	p_1	$p_1 + p_2$...	$p_1 + p_2 + \dots + p_{n-1}$	1

Функция распределения любой дискретной СВ есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям этих значений.

Математическое ожидание характеризует среднее значение случайной величины и для дискретной СВ определяется по формуле

$$m_X = M[X] = \sum_{i=1}^n x_i \cdot p_i. \quad (5.4)$$

Как видно из (5.4), в качестве математического ожидания СВ используется «среднее взвешенное значение», причем каждое из значений случайной величины учитывается с «весом», пропорциональным вероятности этого значения.

Дисперсия случайной величины характеризует степень рассеивания (разброса) значений случайной величины относительно ее математического ожидания и для дискретной СВ определяется по формуле

$$D_x = D[X] = \sum_{i=1}^n (x_i - m_X)^2 p_i = \sum_{i=1}^n x_i^2 p_i - m_X^2. \quad (5.5)$$

Примеры

Пример 5.1. По командному пункту противника производится запуск трех ракет, причем вероятность попадания в цель при пуске одной ракеты равна 0,8. Построить ряд распределения числа попаданий.

Решение. Определим случайную величину X как число попаданий в цель при трех пусках ракет. Эта случайная величина может принимать следующие значения: 0, 1, 2, 3. Найдем вероятность принятия величиной X этих значений, используя формулу Бернулли:

$$p\{X = 0\} = (1 - p)^3 = 0,2^3 = 0,008,$$

$$p\{X = 1\} = C_3^1 p(1 - p)^2 = 3 \cdot 0,8 \cdot 0,2^2 = 0,096,$$

$$p\{X = 2\} = C_3^2 p^2(1 - p) = 3 \cdot 0,8^2 \cdot 0,2 = 0,384,$$

$$p\{X = 3\} = p^3 = 0,8^3 = 0,512.$$

Ряд распределения имеет следующий вид:

x_i	0	1	2	3
p_i	0,008	0,096	0,384	0,512

Как видим, условие (5.1) выполняется.

Пример 5.2. Зная ряд распределения для случайной величина X , описанной в примере 5.1, построить график функции распределения. Найти математическое ожидание и дисперсию дискретной случайной величины X .

Решение. Рассчитаем значения функции распределения для фиксированных значений $X = x_i$, взятых из ряда распределения (см. пример 5.1).

$$1. \ x_1 = 0, \ F(0) = \sum_{x_i < 0} p(X = x_i) = 0.$$

$$2. \ x_1 = 1, \ F(1) = \sum_{x_i < 1} p(X = x_i) = p(X = 0) = 0,008.$$

$$3. \ x_2 = 2, \ F(2) = \sum_{x_i < 2} p(X = x_i) = p(X = 0) + p(X = 1) = 0,008 + 0,096 = 0,104.$$

4. $x_4 = 3, F(3) = \sum_{x_i < 3} p_i = p(X=0) + p(X=1) + p(X=2) = 0,008 + 0,096 + 0,384 = 0,488.$

5. При $x_5 = +\infty$ согласно свойствам функции распределения $F(+\infty) = 1.$

Опишем построение графика функции распределения $F(x)$ (рис. 5.1).

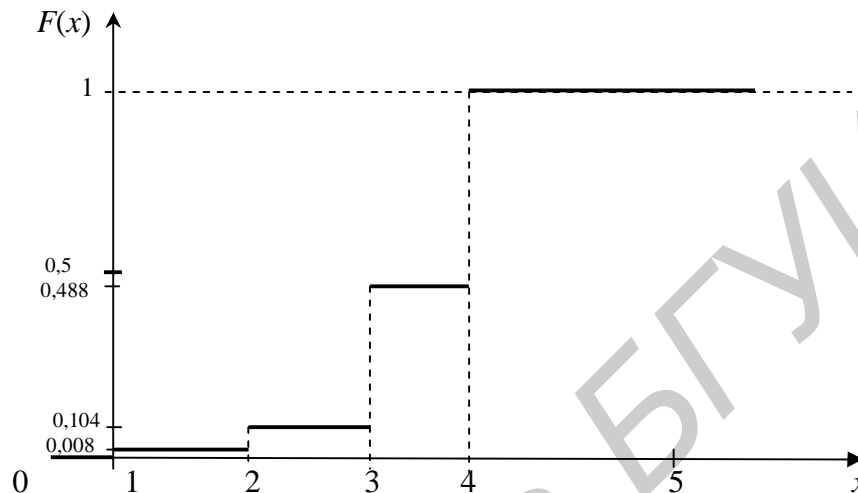


Рис. 5.1

Рассмотрим первый промежуток по оси X от $-\infty$ до 0; согласно п. 1 значение $F(x) = 0$ и линия идет по оси X до нуля включительно. Второй промежуток по оси X от 0 до 1; согласно п. 2 значение $F(x) = 0,008$, значит, проводим ступеньку высотой 0,008. Третий промежуток – от 1 до 2; согласно п. 3 значение $F(x) = 0,104$, следовательно, проводим ступеньку высотой 0,104. Четвертый промежуток – от 2 до 3; согласно п. 4 значение $F(x) = 0,488$, значит, проводим ступеньку высотой 0,488. Пятый промежуток – от 3 до $+\infty$; согласно п. 5 значение $F(x) = 1$, следовательно, проводим ступеньку высотой 1.

Математическое ожидание дискретной СВ X определим по формуле (5.4):

$$M[X] = \sum_{i=1}^n x_i \cdot p_i = 0 \cdot 0,008 + 1 \cdot 0,096 + 2 \cdot 0,384 + 3 \cdot 0,512 = 2,4.$$

Дисперсию дискретной СВ X определим по формуле (5.5):

$$D[X] = \sum_{i=1}^n x_i^2 p_i - m_X^2 = 0^2 \cdot 0,008 + 1^2 \cdot 0,096 + 2^2 \cdot 0,384 + 3^2 \cdot 0,512 - 2,4^2 = 0,48.$$

6. НЕПРЕРЫВНАЯ СЛУЧАЙНАЯ ВЕЛИЧИНА

Случайная величина X называется **непрерывной**, если ее функция распределения $F(x)$ – непрерывная и дифференцируемая функция для всех значений аргумента.

Плотность распределения (или плотность вероятности) $f(x)$ непрерывной случайной величины X в точке x характеризует плотность вероятности в окрестностях точки x и равна производной функции распределения этой СВ:

$$f(x) = \frac{dF(x)}{dx} = F'(x). \quad (6.1)$$

График плотности распределения называется **кривой распределения**.

Вероятность попадания случайной величины X на произвольный участок $[a; b)$ равна сумме элементарных вероятностей на этом участке:

$$p\{a \leq X < b\} = \int_a^b f(x)dx = F(b) - F(a). \quad (6.2)$$

В геометрической интерпретации вероятность $p\{a \leq X < b\}$ равна площади, ограниченной сверху кривой распределения $f(x)$ и отрезком $[a; b)$.

Соотношение (6.2) позволяет выразить функцию распределения $F(x)$ случайной величины X через ее плотность:

$$F(x) = p\{X < x\} = p\{-\infty < X < x\} = \int_{-\infty}^x f(t)dt. \quad (6.3)$$

Основные свойства плотности распределения

1. Плотность распределения неотрицательна: $f(x) \geq 0$. Причем $f(x) = 0$ для тех значений x , которые СВ никогда не принимает в опыте.

2. Условие **нормировки**:

$$\int_{-\infty}^{\infty} f(x)dx = p(-\infty \leq X < +\infty) = 1. \quad (6.4)$$

Математическое ожидание характеризует среднее значение случайной величины и для непрерывной СВ определяется по формуле

$$m_X = M[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx. \quad (6.5)$$

Дисперсия случайной величины характеризует степень рассеивания (разброса) значений случайной величины относительно ее математического ожидания и для непрерывной СВ определяется по формуле

$$D_X = D[X] = \int_{-\infty}^{\infty} (x - m_X)^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - m_X^2. \quad (6.6)$$

Дисперсия случайной величины имеет размерность квадрата случайной величины, поэтому для анализа диапазона значений величины X дисперсия не совсем удобна. Этому недостатка лишено среднее квадратическое отклонение (СКО), размерность которого совпадает с размерностью случайной величины.

Среднее квадратическое отклонение случайной величины X характеризует ширину диапазона значений X и равно

$$s_X = s[X] = +\sqrt{D[X]}. \quad (6.7)$$

Правило 3 σ . Практически все значения случайной величины находятся в интервале

$$[m_X - 3s_X; m_X + 3s_X]. \quad (6.8)$$

Примеры

Пример 6.1. Случайная величина X распределена по закону, определяемому плотностью вероятности вида

$$f(x) = \begin{cases} c \cos x, & -p/2 \leq x \leq p/2, \\ 0, & |x| > p/2. \end{cases}$$

Определить константу c , функцию распределения $F(x)$, математическое ожидание, дисперсию величины X , а также вероятность ее попадания в интервал $[0; 2)$.

Решение. Вначале вычислим значение константы c из условия нормировки (6.4). Условие нормировки представляет собой интегральное уравнение, из которого

можно определить неизвестный параметр плотности вероятности. Для этого определим значение интеграла в левой части условия нормировки:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-p/2}^{p/2} c \cos x dx = c \sin x \Big|_{-p/2}^{p/2} = c + c = 2c.$$

Из условия нормировки следует:

$$2c = 1 \Rightarrow c = \frac{1}{2}.$$

Плотность вероятности примет вид

$$f(x) = \begin{cases} 0, & x < -p/2, \\ \frac{1}{2} \cos x, & -p/2 \leq x \leq p/2, \\ 0, & x > p/2. \end{cases}$$

Определим функцию распределения $F(x)$. Так как плотность вероятности задана различными формулами на разных интервалах, то и ее первообразную – функцию распределения – будем искать по формуле (6.3) для каждого интервала в отдельности.

Для $x < -p/2$:

$$F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^x 0dt = 0,$$

для $-p/2 \leq x \leq p/2$:

$$F(x) = \int_{-\infty}^{-p/2} 0dt + \int_{-p/2}^x \frac{\cos t}{2} dt = \frac{\sin t}{2} \Big|_{-p/2}^x = \frac{1 + \sin x}{2},$$

для $x > p/2$:

$$F(x) = \int_{-\infty}^{-p/2} 0dt + \int_{-p/2}^{p/2} \frac{\cos t}{2} dt + \int_{p/2}^x 0dt = 0 + 1 + 0 = 1.$$

Окончательно имеем

$$F(x) = \begin{cases} 0, & x < -p/2, \\ \frac{1 + \sin x}{2}, & -p/2 \leq x \leq p/2, \\ 1, & x > p/2. \end{cases}$$

Вычислим вероятность $p(0 \leq X < 2)$ по формуле (6.2):

$$p\{0 \leq X < 2\} = F(3) - F(0) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Так как правый край интервала $[0; 2)$ больше, чем $p/2$, то $F(2) = 1$.

Вычислим математическое ожидание СВ по формуле (6.5):

$$\begin{aligned} m_X &= \int_{-\infty}^{\infty} x \cdot f(x) dx = \frac{1}{2} \int_{-p/2}^{p/2} x \cdot \cos x dx = \left| \begin{array}{l} u = x \quad du = dx \\ dv = \cos x \quad v = \sin x \end{array} \right| = \\ &= \frac{1}{2} \left(x \sin x \Big|_{-p/2}^{p/2} - \int_{-p/2}^{p/2} \sin x dx \right) = \frac{1}{2} \left(\frac{p}{2} - \frac{p}{2} + x \cos x \Big|_{-p/2}^{p/2} \right) = 0. \end{aligned}$$

Дисперсию случайной величины СВ вычислим по формуле (6.6):

$$\begin{aligned} D_x &= \int_{-\infty}^{\infty} x^2 f(x) dx - m_X^2 = \frac{1}{2} \int_{-p/2}^{p/2} x^2 \cdot \cos x dx = \left| \begin{array}{l} u = x^2 \quad du = 2x dx \\ dv = \cos x \quad v = \sin x \end{array} \right| = \\ &= \frac{1}{2} \left(x^2 \sin x \Big|_{-p/2}^{p/2} - 2 \int_{-p/2}^{p/2} x \sin x dx \right) = \left| \begin{array}{l} u = x \quad du = dx \\ dv = \sin x \quad v = -\cos x \end{array} \right| = \\ &= \frac{1}{2} \left(\frac{p^2}{4} + \frac{p^2}{4} + 2x \cos x \Big|_{-p/2}^{p/2} - 2 \int_{-p/2}^{p/2} \cos x dx \right) = \frac{p^2}{4} + 0 - \sin x \Big|_{-p/2}^{p/2} = \frac{p^2}{4} - 2. \end{aligned}$$

Пример 6.2. Определить по правилу 3σ диапазон возможных значений СВ X из примера 6.1.

Решение. Вычислим среднее квадратическое отклонение СВ по формуле (6.7):

$$s_X = s[X] = +\sqrt{D[X]} = \sqrt{\frac{p^2}{4} - 2} = \sqrt{0,467} = 0,684.$$

Оценим диапазона значений X по формуле (6.8):

$$[0 - 3 \cdot 0,684; 0 + 3 \cdot 0,684] = [-2,05; +2,05].$$

Как видим, получился интервал, полностью охватывающий точный диапазон значений СВ $[-\frac{p}{2}; +\frac{p}{2}]$, который можно определить по свойству 1 плотности вероятности.

7. ЗАКОН РАСПРЕДЕЛЕНИЯ ФУНКЦИИ СЛУЧАЙНОГО АРГУМЕНТА

Рассмотрим функцию одного случайного аргумента $Y = j(X)$. Если X – непрерывная случайная величина с известной плотностью вероятности $f(x)$, то алгоритм получения плотности вероятности $g(y)$ величины Y следующий:

1. Построить график $Y = j(X)$ и определить диапазон значений $Y \in [y_{\min}, y_{\max}]$.

2. Диапазон Y разбить на M интервалов, в каждом из которых одинаковая степень неоднозначности $k_i, i = 1, 2, \dots, M$:

$$[y_{\min}, y_1), [y_1, y_2), \dots, [y_{M-1}, y_{\max}].$$

Степень неоднозначности k_i – число значений X , соответствующих одному значению Y , или число обратных функций $y_j(y)$ для данного интервала, $j = 1, 2, \dots, k_i$.

3. Определить обратные функции $y_j(y) = j^{-1}(x)$ и вычислить модули производных обратных функций $|y'_j(y)|$. В общем случае число обратных функций $y_j(y)$ в i -м интервале равно k_i .

4. Определить плотность вероятностей $g(y)$ по следующей формуле:

$$g(y) = \begin{cases} 0, & y < y_{\min}, \\ \mathbf{M} \\ \sum_{j=1}^{k_i} f_X(y_j(y)) \cdot |y'_j(y)|, & y_{i-1} \leq y < y_i, \\ \mathbf{M} \\ 0, & y > y_{\max}. \end{cases} \quad (7.1)$$

Пример

Пример 7.1. Определить плотность вероятности величины $Y = X^2$, если X – случайная величина, равномерно распределенная на интервале $[-2; 1]$.

Решение

1. Построим график величины $Y = X^2$ для x в интервале $[-2; 1]$ и определим диапазон значений Y : $Y \in [0; 4]$ (рис. 7.1).

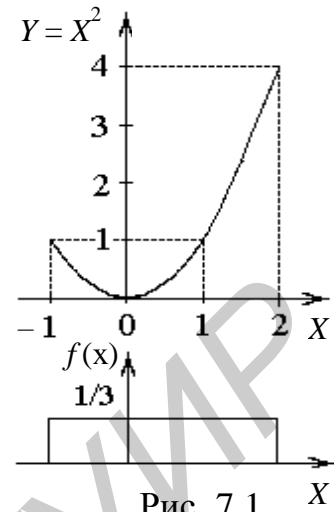


Рис. 7.1

2. В зависимости от числа k обратных функций выделим следующие интервалы для Y :

$$\begin{aligned} \text{при } k_1 &= 0 \quad [-\infty; 0), \\ k_2 &= 2 \quad [0; 1], \\ k_3 &= 1 \quad (1; 4], \\ k_4 &= 2 \quad (4; +\infty]. \end{aligned}$$

3. На интервалах $[-\infty; 0)$ и $(4; +\infty]$ обратных функций не существует.

В интервале $[0; 1]$ две обратные функции:

$$y_1(y) = +\sqrt{y} \text{ и } y_2(y) = -\sqrt{y}.$$

Вычислим модули производных обратных функций $|y'_j(y)|$:

$$|y'_1(y)| = \left| \frac{1}{2\sqrt{y}} \right| = \frac{1}{2\sqrt{y}}, \quad |y'_2(y)| = \left| -\frac{1}{2\sqrt{y}} \right| = \frac{1}{2\sqrt{y}}.$$

В интервале $(1; 4]$ одна обратная функция $y_1(y) = +\sqrt{y}$, следовательно,

$$|y'_1(y)| = \left| \frac{1}{2\sqrt{y}} \right| = \frac{1}{2\sqrt{y}}.$$

4. Так как X равномерно распределена в интервале $[-1; 2]$, то ее плотность вероятности равна

$$f(x) = \begin{cases} \frac{1}{3}, & -1 \leq x \leq 2, \\ 0, & x < -1, \quad x > 2. \end{cases}$$

По формуле (7.1) получим плотность вероятности величины Y :

$$g(y) = \begin{cases} 0, & y < 0, \\ f_x(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + f_x(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{3} \cdot \frac{1}{2\sqrt{y}} + \frac{1}{3} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{3\sqrt{y}}, & 0 \leq y \leq 1, \\ f_x(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{3} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{6\sqrt{y}}, & 1 < y \leq 4, \\ 0, & y > 4. \end{cases}$$

8. ДВУХМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Двухмерная случайная величина (X, Y) – совокупность двух одномерных случайных величин, которые принимают значения в результате проведения одного и того же опыта. Двухмерную случайную величину (X, Y) геометрически можно представить как случайную точку (X, Y) на плоскости xOy .

Двухмерная случайная величина (X, Y) является непрерывной, если ее функция распределения $F(x, y)$ представляет собой непрерывную дифференцируемую функцию по каждому из аргументов и существует вторая смешанная производная $\frac{\partial^2 F(x, y)}{\partial x \partial y}$.

Двухмерная плотность распределения $f(x, y)$ характеризует плотность вероятности в окрестности точки с координатами (x, y) и равна второй смешанной производной функция распределения:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}. \quad (8.1)$$

Свойства двухмерной плотности распределения:

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

2. $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy.$ (8.2)

3. $p\{(X, Y) \in D\} = \iint_{(D)} f(x, y) dx dy.$ (8.3)

4. **Условие нормировки:** $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1.$ (8.4)

Геометрически интеграл условия нормировки вычисляет объем тела, ограниченный поверхностью распределения и плоскостью xOy .

5. $f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$ (8.5)

Математические ожидания компонент двумерной непрерывной случайной величины (X, Y) вычисляются по формулам

$$m_X = a_{1,0}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^1 y^0 f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy, \quad (8.6)$$

$$m_Y = a_{0,1}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^0 y^1 f(x, y) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy. \quad (8.7)$$

Дисперсии компонент двумерной непрерывной случайной величины (X, Y) вычисляются по формулам

$$D_X = a_{2,0}(x, y) - m_X^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 f(x, y) dx dy - m_X^2, \quad (8.8)$$

$$D_Y = a_{0,2}(x, y) - m_Y^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^2 f(x, y) dx dy - m_Y^2. \quad (8.9)$$

Корреляционный момент K_{XY} характеризует степень тесноты линейной зависимости величин X и Y , а также рассеивание их значений относительно точки (m_X, m_Y) :

$$K_{XY} = M[XY] - m_X m_Y = a_{1,1}(x, y) - m_X m_Y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f(x, y) dx dy - m_X m_Y. \quad (8.10)$$

Коэффициент корреляции R_{XY} характеризует только степень линейной зависимости величин и равен нормированному корреляционному моменту:

$$R_{XY} = \frac{K_{XY}}{\sqrt{D_X D_Y}} = \frac{K_{XY}}{S_X S_Y}. \quad (8.11)$$

Для любых случайных величин $|R_{XY}| \leq 1$. Если величины X и Y независимы, то $R_{XY} = 0$.

Пример

Пример 8.1. Двухмерный случайный вектор (X, Y) равномерно распределен внутри области **B**, выделенной жирными прямыми линиями на рис. 8.1.

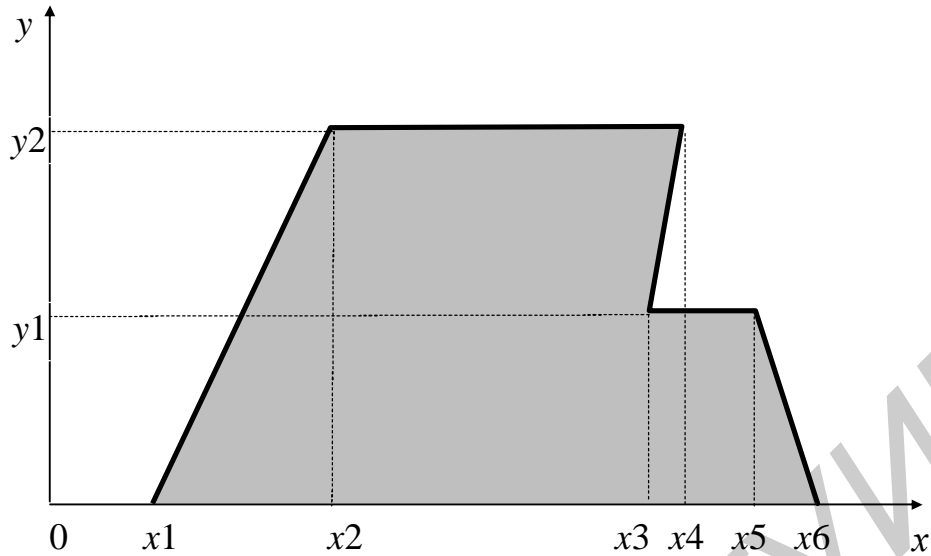


Рис. 8.1

Двухмерная плотность вероятности $f(x, y)$ одинакова для любой точки области **В**:

$$f(x, y) = \begin{cases} c, & (x, y) \in \mathbf{B}, \\ 0, & (x, y) \notin \mathbf{B}. \end{cases}$$

Вычислить коэффициент корреляции между величинами X и Y , если координаты вершин области **В** приведены в табл. 8.1.

Таблица 8.1

x_1	x_2	x_3	x_4	x_5	x_6	y_1	y_2
0	1	2	1	2	3	0,5	1

Решение. Построим область **В**. Соединим последовательно точки с координатами из табл. 8.1 согласно рис. 8.1:

- точку $(x_1; 0) = (0; 0)$ с точкой $(x_2; y_2) = (1; 1)$;
- точку $(x_2; y_2) = (1; 1)$ с точкой $(x_4; y_2) = (1; 1)$, т.е. остаемся на месте;
- точку $(x_4; y_2) = (1; 1)$ с точкой $(x_3; y_1) = (2; 0,5)$;
- точку $(x_3; y_1) = (2; 0,5)$ с точкой $(x_5; y_1) = (2; 0,5)$, т.е. остаемся на месте;
- точку $(x_5; y_1) = (2; 0,5)$ с точкой $(x_6; 0) = (3; 0)$.

В результате получим фигуру, представленную на рис. 8.2.

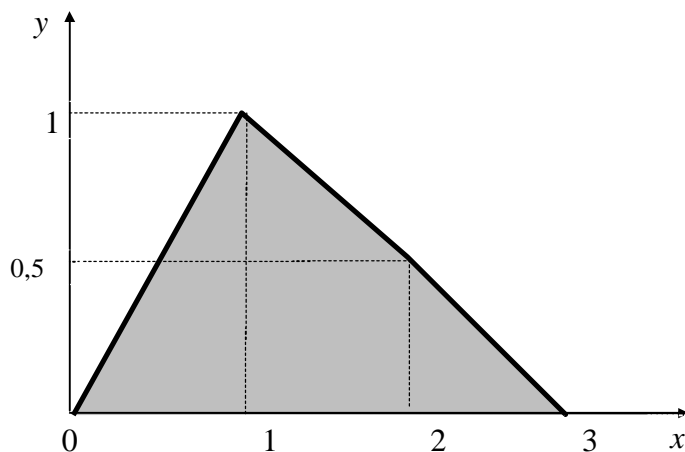


Рис. 8.2

Совместная плотность вероятности примет вид

$$f(x, y) = \begin{cases} c, & 0 \leq y \leq 1, \quad y \leq x \leq (3-2y), \\ 0, & \text{иначе.} \end{cases}$$

Неизвестную константу c определим, используя условие нормировки плотности вероятности (8.4):

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy &= \int_0^1 \left(\int_y^{3-2y} c dx \right) dy = \int_0^1 c(3-2y-y) dy = c \int_0^1 (3-3y) dy = \\ &= 3c \left(y \Big|_0^1 - \frac{y^2}{2} \Big|_0^1 \right) = \frac{3}{2} c = 1 \Rightarrow c = \frac{2}{3}. \end{aligned}$$

Таким образом,

$$f(x, y) = \begin{cases} \frac{2}{3}, & 0 \leq y \leq 1, \quad y \leq x \leq (3-2y), \\ 0, & \text{иначе.} \end{cases}$$

Проверим геометрически полученный результат. Объем тела, ограниченный поверхностью распределения и плоскостью xOy , должен быть равен единице, т. е. объем прямой треугольной призмы равен

$$V = h \cdot S_B = c \cdot S_B = \frac{2}{3} \cdot \frac{3}{2} = 1.$$

Вычислим математические ожидания по формулам (8.6) и (8.7):

$$m_X = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y) dx dy = \int_0^1 \left(\int_y^{3-2y} \frac{2}{3} x dx \right) dy = \int_0^1 \left(\frac{x^2}{3} \Big|_y^{3-2y} \right) dy = \int_0^1 (3 - 4y + y^2) dy = \frac{4}{3},$$

$$m_Y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x, y) dx dy = \int_0^1 y \left(\int_y^{3-2y} \frac{2}{3} dx \right) dy = \int_0^1 y \left(\frac{2x}{3} \Big|_y^{3-2y} \right) dy = \int_0^1 (2y + 2y^2) dy = \frac{1}{3}.$$

Вычислим дисперсии по формулам (8.8) и (8.9):

$$D_X = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 f(x, y) dx dy - m_X^2 = \int_0^1 \left(\int_y^{3-2y} \frac{2}{3} x^2 dx \right) dy - \frac{16}{9} = \int_0^1 \left(\frac{2x^3}{9} \Big|_y^{3-2y} \right) dy - \frac{16}{9} =$$

$$= \int_0^1 \left(6 - 12y + 8y^2 - \frac{14}{9} y^3 \right) dy - \frac{16}{9} = \frac{7}{18},$$

$$D_Y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y^2 f(x, y) dx dy - m_Y^2 = \int_0^1 y^2 \left(\int_y^{3-2y} \frac{2}{3} dx \right) dy - \frac{1}{9} = \int_0^1 y^2 \left(\frac{2x}{3} \Big|_y^{3-2y} \right) dy - \frac{1}{9} =$$

$$= \int_0^1 (2y^2 - 2y^3) dy - \frac{1}{9} = \frac{1}{18}.$$

Корреляционный момент вычислим по формуле (8.10):

$$K_{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) dx dy - m_X m_Y = \int_0^1 y \left(\int_y^{3-2y} \frac{2}{3} x dx \right) dy - \frac{4}{9} =$$

$$= \int_0^1 y \left(\frac{x^2}{3} \Big|_y^{3-2y} \right) dy - \frac{4}{9} = \int_0^1 (3y - 4y^2 + y^3) dy - \frac{4}{9} = -\frac{1}{36}.$$

После нормировки по формуле (8.11) получаем коэффициент корреляции

$$R_{XY} = \frac{K_{XY}}{\sqrt{D_X D_Y}} = \frac{-\frac{1}{36}}{\sqrt{\frac{7}{18} \cdot \frac{1}{18}}} = -\frac{1}{\sqrt{7}} = -0,189.$$

9. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СУММЫ И ПРОИЗВЕДЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Числовые характеристики суммы

Пусть $Y = a_0 + a_1X_1 + a_2X_2$, где a_i – не случайные коэффициенты, тогда:

– математическое ожидание Y равно

$$m_Y = a_0 + a_1m_1 + a_2m_2, \quad (9.1)$$

где m_i – математическое ожидание СВ X_i ;

– дисперсия Y равна

$$D_Y = a_1^2D_1 + a_2^2D_2 + 2a_1a_2K_{12}, \quad (9.2)$$

где D_i – дисперсия СВ X_i ;

K_{12} – корреляционный момент величин X_1 и X_2 .

Если $Y = a_0 + \sum_{i=1}^n a_iX_i$, a_i – не случайные коэффициенты, то математическое

ожидание и дисперсия величины Y равны

$$m_Y = M \left[a_0 + \sum_{i=1}^n a_iX_i \right] = a_0 + \sum_{i=1}^n a_i m_i; \quad (9.3)$$

$$D_Y = D \left[a_0 + \sum_{i=1}^n a_iX_i \right] = \sum_{i=1}^n a_i^2 D_i + 2 \sum_{i=1}^n \sum_{j=i+1}^n a_i a_j K_{ij}. \quad (9.4)$$

Числовые характеристики произведения

Пусть $Y = aX_1X_2$, где a – не случайный коэффициент, то математическое ожидание Y равно:

$$m_Y = a \cdot (m_1m_2 + K_{12}); \quad (9.5)$$

где m_i – математическое ожидание СВ X_i .

Если $Y = X \cdot X$, то математическое ожидание Y равно

$$m_Y = m_X m_X + K_{XX} = m_X^2 + D_X. \quad (9.6)$$

В случае независимых сомножителей X_1 и X_2 дисперсия $Y = a \cdot X_1 X_2$ может быть определена по формуле

$$D_Y = a^2 (D_1 D_2 + m_1^2 D_2 + m_2^2 D_1). \quad (9.7)$$

Если $Y = \prod_{i=1}^n X_i$, где X_i – независимые случайные величины, то математическое ожидание и дисперсия Y равны

$$m_Y = M \left[\prod_{i=1}^n X_i \right] = \prod_{i=1}^n m_i; \quad (9.8)$$

$$D_Y = D \left[\prod_{i=1}^n X_i \right] = \prod_{i=1}^n (D_i + m_i^2) - \prod_{i=1}^n m_i^2. \quad (9.9)$$

Пример

Пример 9.1. Вычислить математическое ожидание и дисперсию величин U и V , а также определить их коэффициент корреляции R_{UV} :

$$U = X_1 - 3X_2;$$

$$V = 4 - 2X_2 + X_3.$$

Величины X_1, X_2, X_3 имеют следующие числовые характеристики:

$$m_1 = 2; \quad m_2 = -3; \quad m_3 = 0;$$

$$D_1 = 4; \quad D_2 = 6; \quad D_3 = 12;$$

$$K_{12} = 0; \quad K_{13} = -4; \quad K_{23} = 9.$$

Решение. Вычислим математические ожидания U и V по формуле (9.1):

$$m_U = m_1 - 3m_2 = 2 - 3 \cdot (-3) = 11,$$

$$m_V = 4 - 2m_2 + m_3 = 4 - 2 \cdot (-3) + 0 = 10.$$

Вычислим дисперсии D_U и D_V по формуле (9.2):

$$D_U = D_1 + (-3)^2 D_2 + 2 \cdot (-3) K_{12} = 4 + 9 \cdot 6 = 58,$$

$$D_V = (-2)^2 D_2 + D_3 + 2 \cdot (-2) K_{23} = 4 \cdot 6 + 12 + 36 = 64.$$

Рассчитаем корреляционный момент K_{UV} по формуле (8.10):

$$K_{UV} = M[UV] - m_U m_V.$$

Для этого определим математическое ожидание произведения величин U и V :

$$\begin{aligned} M[UV] &= M[(X_1 - 3X_2)(4 - 2X_2 + X_3)] = \\ &= M[4X_1 - 2X_1X_2 + X_1X_3 - 12X_2 + 6X_2X_2 - 3X_2X_3] = \text{применяем (9.3)} = \\ &= 4m_1 - 2 \cdot M[X_1X_2] + M[X_1X_3] - 12m_2 + 6 \cdot M[X_2X_2] - 3 \cdot M[X_2X_3] = \\ &= \text{применяем (9.5) и (9.6)} = \\ &= 4m_1 - 2(m_1m_2 + K_{12}) + (m_1m_3 + K_{13}) - 12m_2 + 6(m_2^2 + D_2) - 3(m_2m_3 + K_{23}) = \\ &= 8 + 12 - 4 + 36 + 90 - 27 = 115. \end{aligned}$$

Таким образом,

$$K_{UV} = M[UV] - m_U m_V = 115 - 11 \cdot 10 = 5.$$

Величину R_{UV} определим по формуле (8.11):

$$R_{UV} = \frac{K_{UV}}{\sqrt{D_U D_V}} = \frac{5}{\sqrt{58 \cdot 64}} = 0,082.$$

10. ОБРАБОТКА ОДНОМЕРНОЙ ВЫБОРКИ

Генеральной совокупностью опыта называется множество объектов, из которых производится выборка. **Выборка** – множество $\{x_1, x_2, \dots, x_n\}$ случайно отобранных объектов (значений) из генеральной совокупности. **Объемом** выборки n называется число входящих в нее объектов.

Вариационным рядом называется выборка $\{x'_1, x'_2, \dots, x'_n\}$, полученная в результате расположения значений исходной выборки в порядке возрастания. Значения x'_i называются **вариантами**.

Оценка закона распределения

Эмпирическая функция распределения случайной величины X равна частоте того, что X примет значение меньше, чем аргумент функции x , и определяется формулой

$$F^*(x) = p^*(X < x) = \begin{cases} 0, & x \leq x'_1, \\ \frac{i}{n}, & x_i < x \leq x_{i+1}, \\ 1, & x > x'_n. \end{cases} \quad (10.1)$$

При $n \rightarrow \infty$ эмпирическая функция распределения $F^*(x)$ сходится по вероятности к теоретической функции распределения $F(x)$.

Интервальный статистический ряд вероятностей строится по исходной выборке, если анализируемая случайная величина X является непрерывной, и представляет собой следующую таблицу:

Таблица 10.1

j	A_j	B_j	h_j	n_j	p_j^*	f_j^*
1	A_1	B_1	h_1	n_1	p_1^*	f_1^*
\mathbf{M}	\mathbf{M}	\mathbf{M}	\mathbf{M}	\mathbf{M}	\mathbf{M}	\mathbf{M}
M	A_M	B_M	h_M	n_M	p_M^*	f_M^*

В табл. 10.1 j – номер интервала;

M – число непересекающихся и примыкающих друг к другу интервалов, на которые разбивается диапазон значений $[\underline{x}_1, \underline{x}_n]$:

$$M \approx \begin{cases} \text{int}(\sqrt{n}), & n \leq 100, \\ \text{int}((2 \dots 4) \cdot \lg(n)), & n > 100, \end{cases} \quad (10.2)$$

где $\text{int}(x)$ – целая часть числа x . Желательно, чтобы n без остатка делилось на M ;

A_j, B_j – левая и правая границы j -го интервала ($B_j = A_{j+1}$ – интервалы

примыкают друг к другу), причем $A_1 = \underline{x}_1, B_M = \underline{x}_n$;

$h_j = B_j - A_j$ – длина j -го интервала;

n_j – количество чисел в выборке, попадающих в j -й интервал, $\sum_{j=1}^M n_j = n$;

$p_j^* = \frac{n_j}{n}$ – частота попадания в j -й интервал, $\sum_{j=1}^m p_j^* = 1$;

$f_j^* = \frac{p_j^*}{h_j} = \frac{n_j}{nh_j}$ – статистическая плотность вероятности в j -м интервале.

При построения интервального статистического ряда вероятностей используют следующие методы разбиения диапазона значений на интервалы:

1) **равноинтервальный**, т. е. все интервалы одинаковой длины:

$$h_j = h = \frac{\underline{x}_n - \underline{x}_1}{M}, \quad \forall j \Rightarrow A_j = \underline{x}_1 + (j-1)h, \quad j = \overline{2, M}; \quad (10.3)$$

2) **равновероятностный**, т. е. границы интервалов выбирают так, чтобы в каждом интервале было одинаковое число выборочных значений (необходимо, чтобы n без остатка делилось на M):

$$n_j = n = \frac{n}{M}, \quad p_j^* = \frac{1}{M} \quad \forall j \Rightarrow A_j = \frac{\underline{x}_{(j-1)n} + \underline{x}_{(j-1)n+1}}{2}, \quad j = \overline{2, M}. \quad (10.4)$$

Гистограмма строится по интервальному статистическому ряду и представляет собой статистический аналог графика плотности вероятности $f^*(x)$ случайной величины. **Гистограмма** – совокупность прямоугольников, построенных, как на основаниях, на интервалах h_j статистического ряда, с высотой, равной статистической плотности вероятности f_j^* в соответствующем интервале. Для равноинтервального метода все прямоугольники гистограммы

имеют одинаковую ширину, а для равновероятностного метода – одинаковую площадь. Сумма площадей всех прямоугольников гистограммы равна 1.

Точечные оценки числовых характеристик

Статистической оценкой Q^* параметра Q распределения называется приближенное значение параметра, вычисленное по результатам эксперимента (по выборке). Статистические оценки делятся на точечные и интервальные.

Точечной называется оценка, определяемая одним числом. Точечная оценка Q^* параметра Q случайной величины X в общем случае равна $Q^* = \varphi(x_1, x_2, \dots, x_n)$, где x_i – значения выборки. Очевидно, что оценка Q^* – это случайная величина, и значения Q^* будут изменяться от выборки к выборке случайным образом. К оценкам предъявляется ряд требований.

1. Оценка Q^* называется **состоятельной**, если при увеличении объема выборки n она сходится по вероятности к значению параметра Q :

$$Q^* \xrightarrow[n \rightarrow \infty]{P} Q \Rightarrow \lim_{n \rightarrow \infty} (P(|Q^* - Q| < \epsilon)) = 1, \forall \epsilon > 0.$$

Состоятельность – это минимальное требование к оценкам.

2. Состоятельная оценка Q^* называется **несмещенной**, если ее математическое ожидание точно равно параметру Q для любого объема выборки:

$$M[Q^*] = Q, \forall n.$$

3. Состоятельная несмещенная оценка Q^* является **эффективной**, если ее дисперсия минимальна по отношению к дисперсии любой другой оценки этого параметра:

$$D[Q^*] = \min.$$

Состоятельная, несмещенная и эффективная точечные оценки **математического ожидания** вычисляется как среднее арифметическое значений выборки \bar{x} , называемое **выборочным средним**:

$$m_X^* = \frac{1}{n} \sum_{i=1}^n x_i. \quad (10.5)$$

Состоятельная несмещенная точечная оценка **дисперсии** равна

$$D_X^* = S_0^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}^2. \quad (10.6)$$

Состоятельная несмещенная точечная оценка **среднеквадратического отклонения** равна

$$S_X^* = S_0 = \sqrt{S_0^2}. \quad (10.7)$$

Интервальные оценки числовых характеристик

Доверительным называется интервал

$$I_g(Q) = (Q^* - e; Q^* + e),$$

в который с заданной вероятностью (*надежностью*) γ попадает истинное значения параметра Q , где Q^* – несмещенная точечная оценка параметра Q . Вероятность γ выбирается близкой к 1: 0,9; 0,95; 0,975; 0,99.

Согласно центральной предельной теореме при достаточно большом объеме выборки n ($n > 20 \dots 50$) закон распределения несмещенных точечных оценок m_X^* и D_X^* можно считать нормальным при **любом** законе распределения случайной величины, и доверительные интервалы для математического ожидания и дисперсии могут быть определены по следующим формулам.

Доверительный интервал для математического ожидания имеет вид

$$I_g(m_X) = \left[\bar{x} - z_g \frac{S_0}{\sqrt{n}}; \bar{x} + z_g \frac{S_0}{\sqrt{n}} \right], \quad (10.8)$$

где $z_g = \arg \Phi\left(\frac{g}{2}\right)$ – значение аргумента функции Лапласа, т. е. $\Phi(z\gamma) = \frac{g}{2}$.

Доверительный интервал для дисперсии имеет вид

$$I_g(D_X) = \left[S_0^2 - z_g \sqrt{\frac{2}{n-1}} S_0^2; S_0^2 + z_g \sqrt{\frac{2}{n-1}} S_0^2 \right].$$

(10.9)

Проверка статистических гипотез

Статистической гипотезой называется всякое непротиворечивое множество утверждений $\{H_0, H_1, \dots, H_{k-1}\}$ относительно свойств распределения случайной величины. Простейшей гипотезой является двухальтернативная: $\{H_0, H_1\}$. В этом случае альтернативу H_0 называют нулевой гипотезой, а H_1 – конкурирующей гипотезой.

Критерием называется случайная величина $U = j(x_1, \mathbf{K}, x_n)$, где x_i – значения выборки, которая позволяет принять или отклонить нулевую гипотезу H_0 . *Ошибка первого рода* состоит в том, что будет отклонена гипотеза H_0 , если она верна («пропуск цели»). Вероятность совершить ошибку первого рода обозначается α и называется **уровнем значимости**. Наиболее часто на практике принимают, что $\alpha = 0,05$ или $\alpha = 0,01$.

Критериями согласия называются критерии, используемые для проверки гипотез о предполагаемом законе распределения.

Гипотеза о законе распределения выдвигается следующим образом.

1. Построить по вариационному ряду график эмпирической функции распределения $F^*(x)$ и гистограммы по интервальным статистическим рядам (равноинтервальному и равновероятностному).

2. По виду графиков выдвинуть двухальтернативную гипотезу о предполагаемом (гипотетическом) законе распределения:

H_0 – величина X распределена по *такому-то* закону:

$$f(x) = f_0(x), \quad F(x) = F_0(x);$$

H_1 – величина X не распределена по *такому-то* закону:

$$f(x) \neq f_0(x), \quad F(x) \neq F_0(x),$$

где $f_0(x), F_0(x)$ – соответственно плотность и функция распределения гипотетического закона распределения.

График эмпирической функции распределения $F^*(x)$ должен быть похож на график функции распределения $F_0(x)$ гипотетического закона, а гистограммы – на график плотности гипотетического распределения $f_0(x)$. Ниже приведены графики и аналитические выражения плотности и функции распределения для часто встречающихся на практике законов.

Равномерное распределение имеет непрерывная случайная величина X , если ее плотность вероятности в некотором интервале $[a; b]$ постоянна:

$$f(x) = \begin{cases} 0, & x < a, \\ \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & x > b, \end{cases} \quad F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b, \end{cases} \quad (10.10)$$

где a, b – параметры распределения, $b > a$.

Графики плотности и функции равномерного распределения при $a = 1$ и $b = 3$ показаны на рис. 10.1.

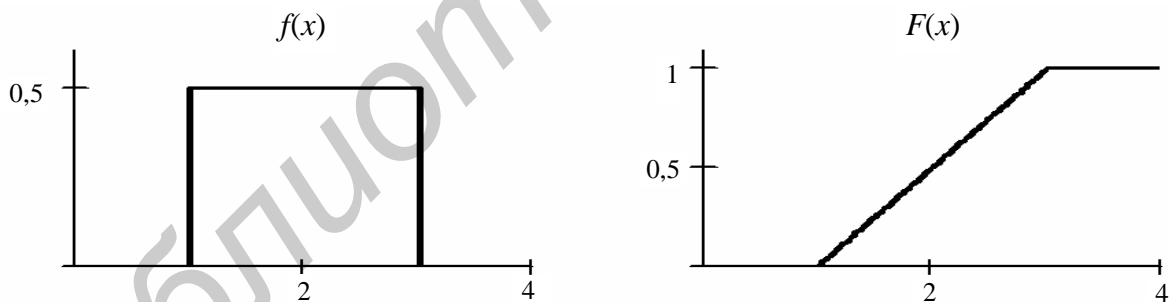


Рис. 10.1

Экспоненциальное распределение имеет непрерывная случайная величина T , принимающая только положительные значения, если ее плотность вероятности и функция распределения равны:

$$f(t) = \begin{cases} 0, & t < 0, \\ \lambda e^{-\lambda t}, & t > 0, \end{cases} \quad F(t) = \begin{cases} 0, & t < 0, \\ 1 - e^{-\lambda t}, & t > 0, \end{cases} \quad (10.11)$$

где λ – параметр распределения, $\lambda > 0$.

Графики плотности и функции экспоненциального распределения при $\lambda = 1$ показаны на рис. 10.2.

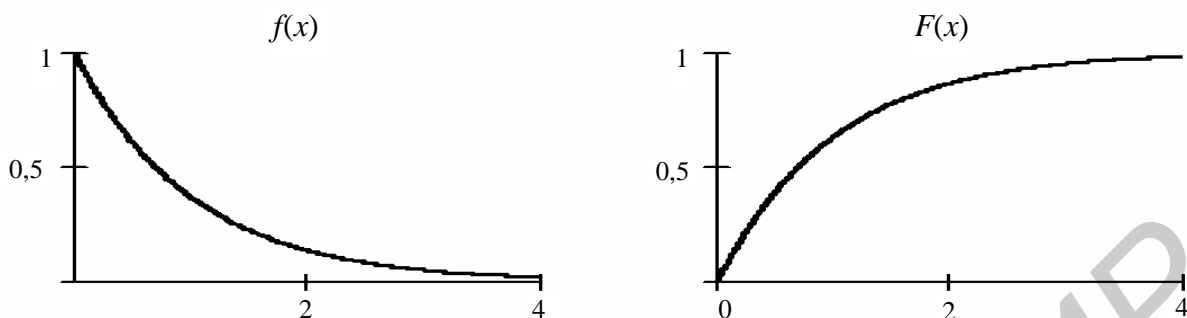


Рис. 10.2

Нормальное распределение (распределение Гаусса) имеет непрерывная случайная величина X , если ее плотность вероятности и функция распределения равны:

$$f(x) = \frac{1}{s\sqrt{2p}} \exp\left\{-\frac{(x-a)^2}{2s^2}\right\}, \quad F(x) = 0,5 + \Phi\left(\frac{x-m}{s}\right), \quad (10.12)$$

где m, σ – параметры распределения, $\sigma > 0$;

$$\Phi(x) = \frac{1}{\sqrt{2p}} \int_0^x e^{-\frac{t^2}{2}} dt - \text{функция Лапласа.}$$

Графики плотности и функции нормального распределения при $m = 0$, $\sigma = 1$ показаны на рис. 10.3.

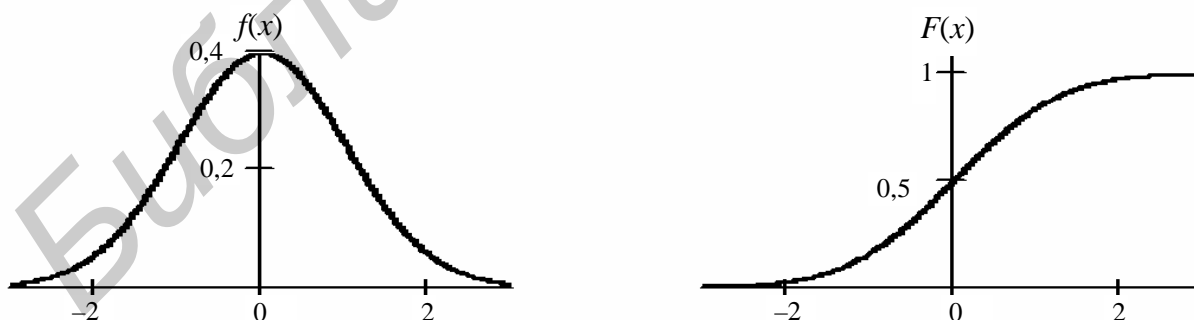


Рис. 10.3

3. Вычислить точечные оценки математического ожидания \bar{x} и дисперсии S_0^2 и, используя метод моментов, определить оценки неизвестных параметров

$\hat{Q}_1, \dots, \hat{Q}_s$ гипотетического закона распределения, где $s \leq 2$ – число неизвестных параметров гипотетического закона распределения.

Оценки неизвестных параметров a, b **равномерного** распределения можно определить по формулам

$$a^* = \bar{x} - \sqrt{3} \cdot S_0, \quad b^* = \bar{x} + \sqrt{3} \cdot S_0 \quad (10.13)$$

или

$$a^* = x_1, \quad b^* = x_n, \quad (10.14)$$

где x_1, x_n – первое и последнее значение вариационного ряда соответственно.

Оценку неизвестного параметра λ **экспоненциального** распределения можно определить по формуле

$$I^* = \frac{1}{\bar{x}}. \quad (10.15)$$

Оценки неизвестных параметров m, σ **нормального** распределения можно определить по формулам

$$m^* = \bar{x}, \quad \sigma^* = S_0. \quad (10.16)$$

4. Проверить гипотезу о предполагаемом законе распределения при помощи критерия согласия.

Критерий согласия Пирсона (χ^2) – один из наиболее часто применяемых критериев. Алгоритм проверки гипотезы о предполагаемом законе распределения следующий.

1. По интервальному статистическому ряду (равноинтервальному или равновероятностному) вычислить значение критерия χ^2 по формуле

$$\chi^2 = n \sum_{j=1}^M \frac{(p_j - p_j^*)^2}{p_j} = \sum_{j=1}^M \frac{(n_j - np_j)^2}{np_j}, \quad (10.17)$$

где n – объем выборки;

M – число интервалов интервального статистического ряда;

p_j – теоретическая вероятность попадания случайной величины в

j -й интервал при условии, что гипотеза H_0 верна:

$$p_j = p(A_j \leq X < B_j) = \int_{A_j}^{B_j} f_0(x) dx = F_0(B_j) - F_0(A_j); \quad (10.18)$$

$f_0(x)$, $F_0(x)$ – плотность и функция распределения гипотетического закона распределения;

p_j^* – частота попадания в j -й интервал;

n_j – количество чисел в выборке, попадающих в j -й интервал.

При расчете p_1 и p_M в качестве крайних границ первого и последнего интервалов A_1 , B_M следует использовать теоретические границы гипотетического закона распределения.

Если проверяется гипотеза о **равномерном** законе распределения, то $A_1 = x_1$, $B_M = x_n$, а гипотетическая функция распределения будет иметь следующий вид (см. (10.10) и (10.14)):

$$F_0(x) = \begin{cases} 0, & x \leq x_1, \\ \frac{x - x_1}{x_n - x_1}, & x_1 \leq x \leq x_n, \\ 1, & x \geq x_n, \end{cases} \quad (10.19)$$

и теоретические вероятности попадания в интервалы будет вычисляться по формуле

$$p_j = F_0(B_j) - F_0(A_j) = \frac{B_j - A_j}{x_n - x_1}. \quad (10.20)$$

Если проверяется гипотеза об **экспоненциальном** законе распределения, то $A_1 = 0$, $B_M = +\infty$, и гипотетическая функция распределения будет иметь вид (см. (10.11) и (10.15)):

$$F_0(x) = \begin{cases} 1 - e^{-\frac{x}{\bar{x}}}, & x \geq 0, \\ 0, & x \leq 0, \end{cases} \quad (10.21)$$

а теоретические вероятности попадания в интервалы будет вычисляться по формуле

$$p_j = F_0(B_j) - F_0(A_j) = e^{-\frac{A_j}{\bar{x}}} - e^{-\frac{B_j}{\bar{x}}}. \quad (10.22)$$

Если проверяется гипотеза о **нормальном** законе распределения, то $A_1 = -\infty$, $B_M = +\infty$, и гипотетическая функция распределения будет иметь вид (см. (10.12) и (10.16)):

$$F_0(x) = 0,5 + \Phi\left(\frac{x - \bar{x}}{S_0}\right), \quad (10.23)$$

а теоретические вероятности попадания в интервалы будет вычисляться по формулам

$$p_j = F_0(B_j) - F_0(A_j) = \Phi\left(\frac{B_j - \bar{x}}{S_0}\right) - \Phi\left(\frac{A_j - \bar{x}}{S_0}\right). \quad (10.24)$$

При правильном вычислении вероятностей p_i должно выполняться

контрольное соотношение $\left|1 - \sum_{j=1}^M p_j\right| \leq 0,01$.

Величина c^2 распределена по закону, который называется **распределением c^2** . Данное распределение не зависит от закона распределения величины X , а зависит от параметра k , который называется **числом степеней свободы**.

2. Из таблицы распределения c^2 выбирается критическое значение $c_{\alpha, k}^2$, где α – заданный уровень значимости ($\alpha = 0,05$ или $\alpha = 0,01$), а k – число степеней свободы, которое определяется по формуле

$$k = M - 1 - s, \quad (10.25)$$

где M – число слагаемых в формуле (10.17), т.е. число интервалов интервального статистического ряда;

s – число неизвестных параметров гипотетического закона распределения, значения (для равномерного закона $s = 2$, экспоненциального $s = 1$, нормального $s = 2$).

3. Если значение c^2 , вычисленное по выборочным данным на шаге 1, больше, чем критическое значение, т.е. $c^2 > c_{\alpha, k}^2$, то гипотеза H_0 отклоняется, в противном случае нет оснований ее отклонить.

Критерий согласия Колмогорова. Алгоритм проверки гипотезы о предполагаемом законе распределения следующий.

1. На основании эмпирической функции распределения $F^*(x)$ вычислить значение критерия Колмогорова

$$= \sqrt{n} \cdot Z, \quad (10.26)$$

где n – объем выборки;

$Z = \max_{i=1}^n |F^*(x_i) - F_0(x_i)|$ – максимальный модуль отклонения эмпирической функции распределения $F^*(x)$ от гипотетической функции распределения $F_0(x)$, определенный по **всем** n значениям x_i исходной выборки.

Значение Z с достаточной точностью может быть определено по графикам функций $F^*(x)$ и $F_0(x)$, которые строят в одной системе координат на масштабной-координатной бумаге (миллиметровке). Для построения графика $F_0(x)$ достаточно рассчитать значения функции $F_0(x)$ в 10 – 20 равноотстоящих точках, которые затем соединить плавной кривой.

Величина λ распределена по закону Колмогорова, который не зависит от закона распределения величины X .

2. Из таблицы распределения Колмогорова (прил. 4) выбрать критическое значение I_g , $g = 1 - \alpha$, где α – заданный уровень значимости ($\alpha = 0,05$ или $\alpha = 0,01$).

3. Если значение λ , вычисленное на шаге 1, больше, чем критическое значение, т. е. $\lambda > \lambda_g$, то гипотеза H_0 отклоняется, в противном случае нет оснований ее отклонить.

Пример

Пример 10.1. По вариационному ряду случайной величины X ($n = 100$)
 –6,237 –6,229 –5,779 –5,139 –4,950 –4,919 –4,636 –4,560 –4,530 –4,526 –4,523
 –4,511 –4,409 –4,336 –4,259 –4,055 –4,044 –4,006 –3,972 –3,944 –3,829 –3,794
 –3,716 –3,542 –3,541 –3,431 –3,406 –3,384 –3,307 –3,181 –3,148 –3,124 –3,116

-2,892 -2,785 -2,734 -2,711 -2,637 -2,633 -2,428 -2,381 -2,339 -2,276 -2,222
-2,167 -2,111 -2,034 -1,958 -1,854 -1,803 -1,774 -1,755 -1,745 -1,713 -1,709
-1,566 -1,548 -1,480 -1,448 -1,353 -1,266 -1,229 -1,179 -1,130 -1,102 -1,060
-1,046 -1,035 -0,969 -0,960 -0,903 -0,885 -0,866 -0,865 -0,774 -0,721 -0,688
-0,673 -0,662 -0,626 -0,543 -0,445 -0,241 -0,174 -0,131 0,115 0,205 0,355
0,577 0,591 0,795 0,986 1,068 1,099 1,195 1,540 2,008 2,160 2,534 2,848:

- построить график эмпирической функции распределения $F^*(x)$;
- построить гистограмму равноинтервальным способом;
- построить гистограмму равновероятностным способом;
- вычислить точечные оценки математического ожидания и дисперсии;
- вычислить интервальные оценки математического ожидания и дисперсии ($\gamma = 0,95$);

– выдвинуть гипотезу о законе распределения случайной величины и проверить ее при помощи критерия согласия χ^2 и критерия Колмогорова ($\alpha = 0,05$). График гипотетической функции распределения $F_0(x)$ построить совместно с графиком $F^*(x)$ в той же системе координат и на том же листе.

Решение. По формуле (10.1) построим график эмпирической функции распределения $F^*(x)$ (рис. 10.4). Так как $F^*(x)$ является неубывающей функцией и все ступеньки графика $F^*(x)$ имеют одинаковую величину $1/n$ (или ей кратны – для одинаковых значений), то таблицу значений эмпирической функции распределения $F^*(x)$ можно не вычислять, а построить ее график непосредственно по вариационному ряду начиная с его первого значения (см. *Пример 5.2*). На рисунке 10.4 изображены графики эмпирической $F^*(x)$ и гипотетической $F_0(x)$ функций распределения.

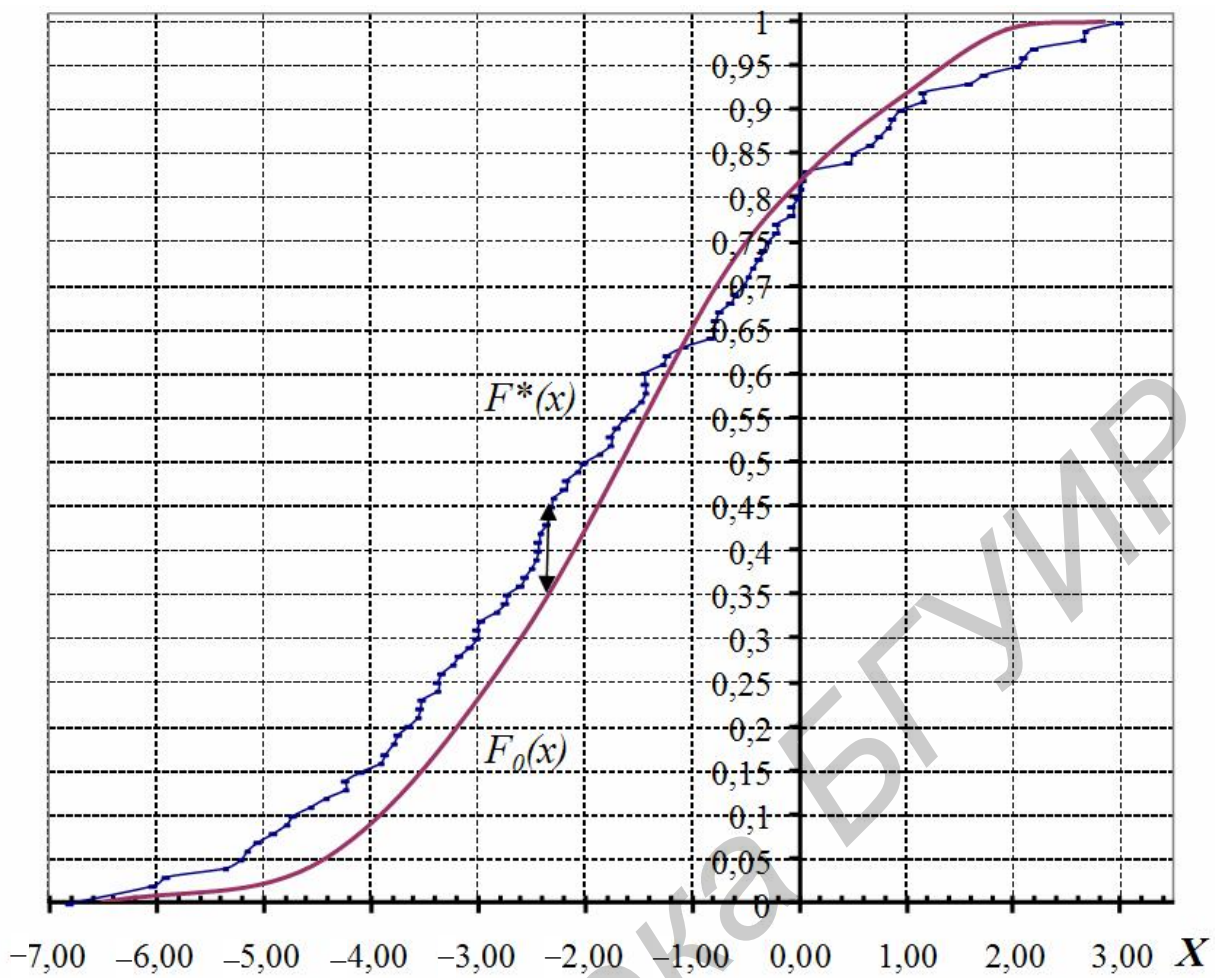


Рис. 10.4

Количество интервалов M , необходимое для построения гистограмм, определим по объему выборки (см. (10.2)):

$$M \approx \sqrt{n} = \sqrt{100} = 10.$$

Для равноинтервальной гистограммы величины h_j , A_j , B_j рассчитаем по формуле (10.3) и заполним все колонки интервального статистического ряда (табл. 10.2).

Таблица 10.2

j	A_j	B_j	h_j	v_j	p_j^*	f_j^*
1	-6,237	-5,3345	0,9085	3	0,03	0,033
2	-5,3345	-4,426	0,9085	9	0,09	0,099
3	-4,426	-3,5175	0,9085	13	0,13	0,143
4	-3,5175	-2,609	0,9085	14	0,14	0,154
5	-2,609	-1,7005	0,9085	16	0,16	0,176
6	-1,7005	-0,792	0,9085	19	0,19	0,209
7	-0,792	0,1165	0,9085	12	0,12	0,132
8	0,1165	1,025	0,9085	6	0,06	0,066
9	1,025	1,9335	0,9085	4	0,04	0,044
10	1,9335	2,848	0,9085	4	0,04	0,044

Равноинтервальная гистограмма представлена на рис. 10.5.

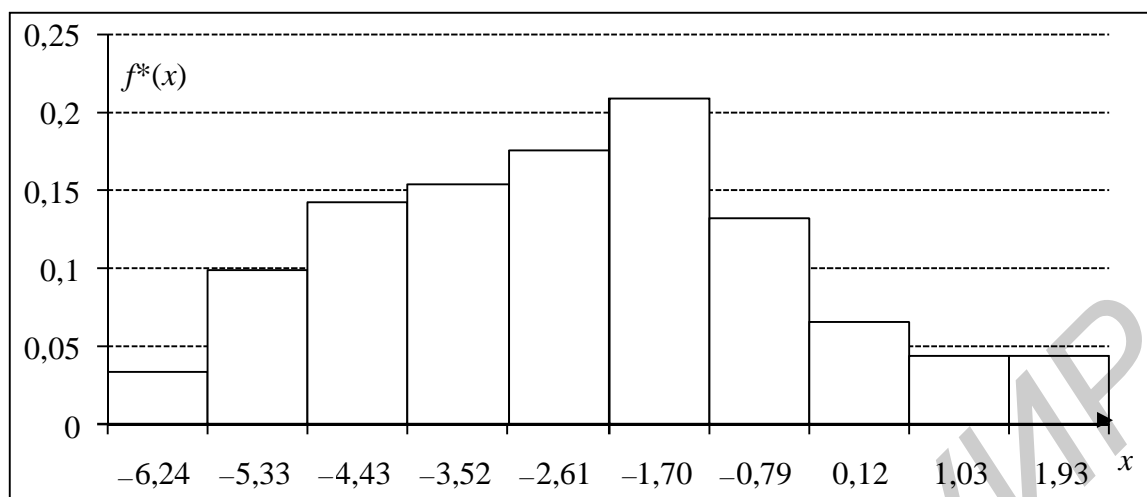


Рис. 10.5

Для равновероятностной гистограммы величины v_j, p_j^*, A_j, B_j рассчитаем по формуле (10.4) и заполним все колонки интервального статистического ряда (табл. 10.3).

Таблица 10.3

j	A_j	B_j	h_i	n_j	p_j^*	f_j^*
1	-6,2370	-4,5245	1,7125	10	0,1	0,0584
2	-4,5245	-3,8865	0,6380	10	0,1	0,1567
3	-3,8865	-3,1645	0,7220	10	0,1	0,1385
4	-3,1645	-2,4045	0,7600	10	0,1	0,1316
5	-2,4045	-1,7885	0,6160	10	0,1	0,1623
6	-1,7885	-1,3095	0,4790	10	0,1	0,2086
7	-1,3085	-0,9319	0,3766	10	0,1	0,2655
8	-0,9319	-0,5843	0,3476	10	0,1	0,2877
9	-0,5843	0,6932	1,2775	10	0,1	0,0783
10	0,6932	2,8480	2,1548	10	0,1	0,0464

Равновероятностная гистограмма представлена на рис. 10.6.

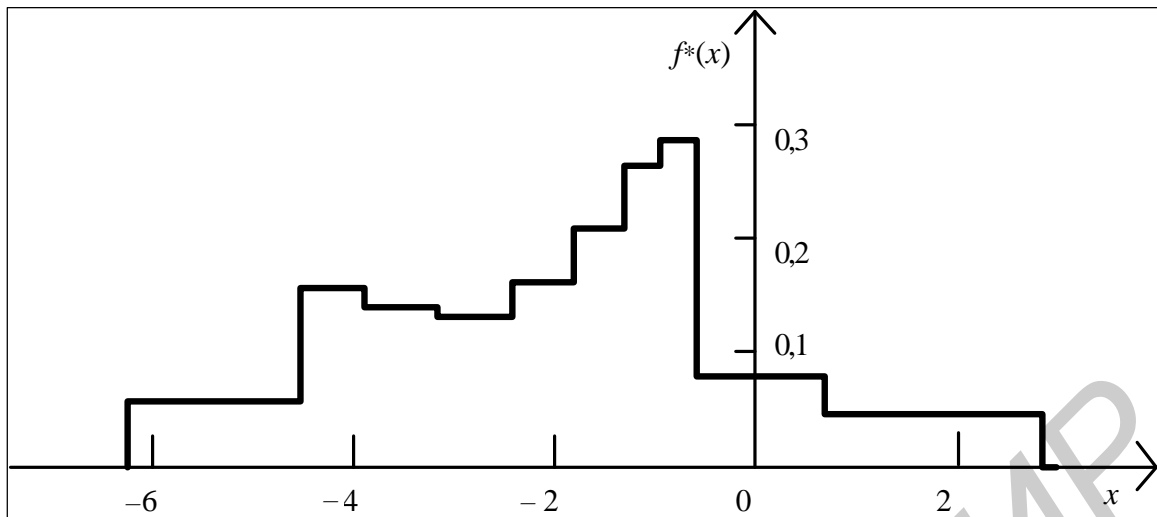


Рис. 10.6

Вычислим *точечную оценку математического ожидания* по формуле (10.5):

$$m_X^* = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = -1,7.$$

Вычислим *точечную оценку дисперсии* по формуле (10.6):

$$D_X^* = S_0^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}^2 = 3,92.$$

Построим *доверительный интервал для математического ожидания* с надежностью $\gamma = 0,95$ по формуле (10.8). Для этого в таблице функции Лапласа (прил. 1) найдем значение, равное $\frac{g}{2} = 0,475$, и определим значение аргумента, ему соответствующее: $z_{0,95} = \arg \Phi(0,475) = 1,96$ (строка 1,9; столбец 6). Затем вычислим $z_{0,95} \frac{S_0}{\sqrt{n}} = 0,388$ и получим *доверительный интервал для математического ожидания*:

$$I_{0,95}(m_X) = [-2,088; -1,312].$$

Построим *доверительный интервал для дисперсии* с надежностью $\gamma = 0,95$ по формуле (10.9). Вычислим $z_{0,95} \cdot \sqrt{\frac{2}{n-1}} \cdot S_0^2 = 1,092$ и получим *доверительный интервал для дисперсии*:

$$I_{0,95}(D_X) = [2,828; 5,012].$$

По виду графика эмпирической функции распределения $F^*(x)$ и гистограмм выдвигаем двухальтернативную гипотезу о законе распределения случайной величины:

H_0 – величина X распределена по нормальному закону:

$$f(x) = f_0(x) = \frac{1}{s\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2s^2}\right\}, \quad F(x) = F_0(x) = 0,5 + \Phi\left(\frac{x-m}{s}\right);$$

H_1 – величина X не распределена по нормальному закону:

$$f(x) \neq f_0(x), \quad F(x) \neq F_0(x).$$

Определим оценки неизвестных параметров m и σ гипотетического (нормального) закона распределения по формулам (10.16):

$$m^* = \bar{x} = -1,7; \quad s^* = \sqrt{S_0^2} = S_0 = 1,98.$$

Таким образом, получаем полностью определенную гипотетическую функцию распределения (см. (10.23)):

$$F_0(x) = 0,5 + 0,5 \cdot \Phi\left(\frac{x - \bar{x}}{S_0}\right) = 0,5 + 0,5 \cdot \Phi\left(\frac{x + 1,7}{1,98}\right).$$

Проверим гипотезу о нормальном законе с помощью критерия c^2 . Вычислим значение критерия c^2 на основе равноинтервального статистического ряда (см. табл. 10.2) по формуле (10.17):

$$c^2 = 100 \sum_{j=1}^{10} \frac{(p_j - p_j^*)^2}{p_j}.$$

Теоретические вероятности p_j попадания в интервалы равноинтервального статистического ряда нормальной случайной величины с параметрами $m^* = -1,7; s^* = 1,98$ вычислим по формуле (10.24):

$$p_j = F_0(B_j) - F_0(A_j) = \Phi\left(\frac{B_j + 1,7}{1,98}\right) - \Phi\left(\frac{A_j + 1,7}{1,98}\right).$$

Значения функции Лапласа определяем с помощью таблицы, приведенной в прил. 2. При использовании таблицы функции Лапласа следует учитывать, что $\Phi(-x) = -\Phi(x)$, $\Phi(0) = 0$, $\Phi(+\infty) = 0,5$. Результаты расчета можно свести в табл. 10.4.

Таблица 10.4

j	A_j	B_j	$F_0(A_j)$	$F_0(B_j)$	p_j	p_j^*	$\frac{(p_j^* - p_j)^2}{p_j}$
1	$-\infty$	-5,335	0	0,0336	0,0336	0,03	0
2	-5,335	-4,426	0,0336	0,0708	0,0372	0,09	0,0625
3	-4,426	-3,518	0,0708	0,1768	0,106	0,13	0,003636
4	-3,518	-2,609	0,1768	0,3228	0,146	0,14	0,000667
5	-2,609	-1,701	0,3228	0,5	0,1772	0,16	0,000588
6	1,7005	-0,792	0,5	0,6772	0,1772	0,19	0,000556
7	-0,792	0,1165	0,6772	0,8212	0,144	0,12	0,002857
8	0,1165	1,025	0,8212	0,9162	0,095	0,06	0,01
9	1,025	1,9335	0,9162	0,989	0,0728	0,04	0,012857
10	1,9335	$+\infty$	0,989	1	0,011	0,04	0,02
				Сумма:	0,999	1	0,113661

Проверяем выполнение контрольного соотношения для p_j :

$$\left| 1 - \sum_{j=1}^{10} p_j \right| = 0,001 < 0,01.$$

В результате получаем $c^2 = 100 \cdot 0,113661 = 11,37$.

Вычислим число степеней свободы по формуле (10.25):
 $k = M - 1 - s = 10 - 1 - 2 = 7$, и по заданному уровню значимости $\alpha = 0,05$ из таблицы распределения c^2 (прил. 3) выбираем критическое значение $c_{\alpha; 7}^2 = c_{0,05; 7}^2 = 14,07$.

Так как $c^2 = 11,37 < c_{0,05; 7}^2 = 14,07$, то гипотеза H_0 о нормальном законе распределения принимается (нет основания ее отклонить).

Проверим гипотезу о нормальном законе с помощью критерия Колмогорова. Построим график $F_0(x)$ в одной системе координат с графиком

эмпирической функции распределения $F^*(x)$ (см. рис. 10.1). В качестве опорных точек для графика $F_0(x)$ используем 10 значений $F_0(A_j)$ из табл. 10.4.

По графику определим максимальное по модулю отклонение между функциями $F^*(x)$ и $F_0(x)$ (см. рис. 10.1):

$$Z = \max_{i=1}^n |F^*(x_i) - F_0(x_i)| = 0,09.$$

Вычислим значение критерия Колмогорова по формуле (10.26):

$$I = \sqrt{n} \cdot Z = \sqrt{100} \cdot 0,09 = 0,9.$$

Из таблицы Колмогорова (прил. 4) по заданному уровню значимости $\alpha = 0,05$ выбираем критическое значение $I_g = I_{1-\alpha} = I_{0,95} = 1,36$.

Так как $I = 0,9 \leq I_{0,95} = 1,36$, то гипотезу H_0 о нормальном законе распределения отвергать нет основания.

11. ОБРАБОТКА ДВУХМЕРНОЙ ВЫБОРКИ

Пусть проводится n независимых опытов, в каждом из которых двумерная случайная величина (X, Y) принимает определенные значения, и результаты опытов представляют собой двумерную выборку вида

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}.$$

Статистическая обработка двумерных массивов данных включает в себя обработку и анализ составляющих X и Y как одномерных величин и вычисление оценок и анализ параметров, присущих только двумерным (многомерным) случайным величинам.

Как правило, определяются следующие оценки:

– математических ожиданий случайных величин X и Y :

$$m_X^* = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (11.1)$$

$$m_Y^* = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i;$$

– дисперсий случайных величин X и Y :

$$D_X^* = S_0^2(x) = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}^2, \quad (11.2)$$

$$D_Y^* = S_0^2(y) = \frac{1}{n-1} \cdot \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^n y_i^2 - \frac{n}{n-1} \bar{y}^2.$$

Состоятельная несмещенная **оценка корреляционного момента** равна

$$K_{XY}^* = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} \sum_{i=1}^n x_i y_i - \frac{n}{n-1} \bar{x} \cdot \bar{y}, \quad (11.3)$$

где x_i, y_i – значения, которые приняли случайные величины X и Y в i -м опыте;

\bar{x}, \bar{y} – средние значения случайных величин X и Y соответственно.

Состоятельная **оценка коэффициента корреляции** равна

$$R_{XY}^* = \frac{K_{XY}^*}{\sqrt{S_0^2(x) \cdot S_0^2(y)}} = \frac{K_{XY}^*}{S_0(x) \cdot S_0(y)}, \quad (11.4)$$

где $S_0(x), S_0(y)$ – оценки среднеквадратичного отклонения случайных величин X и Y соответственно.

Доверительный интервал для коэффициента корреляции с надежностью γ для случая двумерного нормального распределения имеет вид

$$I_g(R_{XY}) = \left[\frac{e^{2a} - 1}{e^{2a} + 1}; \frac{e^{2b} - 1}{e^{2b} + 1} \right], \quad (11.5)$$

где $a = 0,5 \cdot \ln \left(\frac{1 + R_{XY}^*}{1 - R_{XY}^*} \right) - \frac{z_g}{\sqrt{n-3}}$;

$b = 0,5 \cdot \ln \left(\frac{1 + R_{XY}^*}{1 - R_{XY}^*} \right) + \frac{z_g}{\sqrt{n-3}}$;

$z_g = \arg \Phi \left(\frac{g}{2} \right)$ – значение аргумента функции Лапласа, т. е. $\Phi(z_g) = \frac{g}{2}$.

Гипотеза об отсутствии корреляционной зависимости.

Предполагается, что двухмерная случайная величина (X, Y) распределена по нормальному закону. Алгоритм проверки следующий.

1. Формулируется гипотеза:

$$H_0: R_{XY} = 0;$$

$$H_1: R_{XY} \neq 0.$$

Здесь R_{XY} – теоретический коэффициент корреляции.

2. Вычисляется оценка коэффициента корреляции R_{XY}^* по формуле (11.4).

3. Если *объем выборки не велик* ($n < 50$), то определяется значение критерия

$$t = \frac{|R_{XY}^*| \sqrt{n-2}}{\sqrt{1-(R_{XY}^*)^2}}, \quad (11.6)$$

который распределен по закону Стьюдента с $(n-2)$ степенями свободы, если гипотеза H_0 верна.

4. По заданному уровню значимости α вычисляется доверительная вероятность $g = 1 - \alpha$ и из таблицы Стьюдента (прил. 2) выбирается критическое значение $t_{g, n-2}$.

5. Если $t > t_{g, n-2}$, то гипотеза H_0 отклоняется, т. е. величины X, Y коррелированы. В противном случае гипотеза H_0 принимается.

3*. Если *объем выборки велик* ($n \geq 50$), то определяется значение критерия

$$Z = \frac{|R_{XY}^*| \sqrt{n}}{1-(R_{XY}^*)^2}, \quad (11.7)$$

который распределен по нормальному закону, если гипотеза H_0 верна.

4*. По заданному уровню значимости α из таблицы функции Лапласа определяется критическое значение $Z_a = \arg \Phi\left(\frac{1-\alpha}{2}\right)$, т. е. $\Phi(Z_a) = \frac{1-\alpha}{2}$.

5*. Если $Z > Z_a$, то гипотеза H_0 отклоняется, а следовательно, величины X, Y коррелированы. В противном случае гипотеза H_0 принимается.

Оценка регрессионных характеристик

Регрессией случайной величины Y на x называется условное математическое ожидание $m_{Y/x} = M[Y/X = x]$ случайной величины Y при условии, что $X = x$. Регрессия Y на x устанавливает зависимость среднего значения величины Y от величины X . Если случайные величины X и Y независимы, то $m_{Y/x} = m_Y = \text{const}$.

Необходимо на основании имеющейся выборки выявить характер связи между величинами X и Y , т. е. получить оценку условного математического ожидания $m_{Y/x}^*$ – оценку регрессии Y на x . Данная оценка представляет собой некоторую функцию

$$m_{Y/x}^* = \bar{y}(x) = j(x, a_0, a_1, \dots, a_m),$$

где a_0, a_1, \dots, a_m – неизвестные параметры.

Для определения типа зависимости строится *диаграмма рассеивания* или *корреляционное поле*, которую можно получить, если результаты опытов изобразить в виде точек на плоскости в декартовой системе координат. На основании анализа корреляционного поля выбираем тип *линии регрессии* $\bar{y}(x) = j(x, a_0, a_1, \dots, a_m)$. Значения параметров a_0, a_1, \dots, a_m для выбранного типа определяются так, чтобы функция $\bar{y}(x) = j(x, a_0, a_1, \dots, a_m)$ наилучшим образом соответствовала бы неизвестной регрессии $m_{Y/x}$, т. е. ее значения должны быть приблизительно равны средним арифметическим значений Y для каждого значения $X = x$.

Если величины X и Y распределены по нормальному закону, то регрессия является линейной:

$$m_{Y/x} = a_0 + a_1x.$$

Оценки параметров для линейной регрессии $\bar{y}(x) = a_0^* + a_1^*x$ определяются по формулам

$$a_1^* = \frac{a_{1,1}^*(x, y) - \bar{x} \cdot \bar{y}}{a_2^*(x) - \bar{x}^2} = \frac{K_{XY}^*}{S_0^2(x)}, \quad (11.8)$$

$$a_0^* = \bar{y} - a_1^* \cdot \bar{x},$$

где \bar{x} , \bar{y} – оценки математического ожидания величин X и Y ;

K_{XY}^* – оценка корреляционного момента величин X и Y ;

$S_0^2(x)$ – оценка дисперсии величины X .

Для визуальной проверки правильности вычисления величин a_0^* , a_1^* необходимо построить диаграмму рассеивания и график $\bar{y}(x) = a_0^* + a_1^*x$. Если оценки параметров a_0 , a_1 рассчитаны без грубых ошибок, то сумма квадратов отклонений всех значений (точек) двумерной выборки $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ от прямой $\bar{y}(x) = a_0^* + a_1^*x$ должна быть минимально возможной.

Пример

Пример 11.1. По выборке двумерной случайной величины, которая содержит 50 пар значений (x, y) (первые два столбца табл. 11.1):

- вычислить точечную оценку коэффициента корреляции;
- вычислить интервальную оценку коэффициента корреляции ($\gamma = 0,95$);
- проверить гипотезу об отсутствии корреляционной зависимости ($\alpha = 0,05$);
- вычислить оценки параметров a_0^* и a_1^* линии регрессии $\bar{y}(x) = a_0^* + a_1^*x$;
- построить диаграмму рассеивания и линию регрессии.

Решение. Для решения задачи удобно воспользоваться табл. 11.1. Значения в 3, 4 и 5-м столбцах вычисляются по формулам, приведенными в головке таблицы. В последней строке таблицы приведены средние арифметические значений каждого из столбцов. Таким образом получены:

– оценки математических ожиданий по каждой переменной (см. табл. 11.1):

$$m_X^* = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 5,08 \text{ (см. столбец 2),}$$

$$m_Y^* = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 5,21 \text{ (см. столбец 3);}$$

– оценки начальных моментов второго порядка по каждой переменной:

$$a_2^*(x) = \frac{1}{n} \sum_{i=1}^n x_i^2 = 34,55755 \text{ (см. столбец 4),}$$

$$a_2^*(y) = \frac{1}{n} \sum_{i=1}^n y_i^2 = 36,09954 \text{ (см. столбец 5);}$$

– оценка смешанного начального момента второго порядка:

$$a_{1,1}^*(x, y) = \frac{1}{n} \sum_{i=1}^n x_i y_i = 36,09954 \text{ (см. столбец 6).}$$

Таблица 11.1

	x	y	x²	y²	x*y
1	2	3	4	5	6
1	8,974883	9,784539	80,54853	95,73721	87,8151
2	1,271096	5,058748	1,615685	25,59093	6,430154
3	3,967406	6,383251	15,74031	40,7459	25,32495
4	6,841945	1,953795	46,81221	3,817315	13,36776
5	3,341777	5,445723	11,16747	29,6559	18,19839
6	6,009095	1,657155	36,10922	2,746163	9,958001
7	3,806879	1,750542	14,49233	3,064396	6,6641
8	4,714805	0,509049	22,22938	0,259131	2,400065
9	8,8464	2,334056	78,2588	5,447816	20,64799
10	4,395581	1,568651	19,32113	2,460667	6,895134
11	2,179632	2,34901	4,750795	5,517846	5,119977
12	5,651112	9,857173	31,93507	97,16387	55,70399
13	3,278298	4,774926	10,74724	22,79992	15,65363
14	0,369579	2,23365	0,136589	4,989191	0,82551
15	8,991363	1,784112	80,84461	3,183056	16,0416
16	8,873562	2,211371	78,7401	4,890163	19,62274
17	0,347606	0,58504	0,12083	0,342272	0,203363
18	3,643605	5,025178	13,27586	25,25241	18,30976
19	8,600116	1,547594	73,96199	2,395046	13,30948

1	2	3	4	5	6
20	6,193731	3,268838	38,36231	10,6853	20,2463
21	9,565111	1,426435	91,49135	2,034717	13,64401
22	8,646809	8,410901	74,76731	70,74326	72,72746
23	0,328074	9,496139	0,107633	90,17666	3,115436
24	6,583453	8,498489	43,34185	72,22432	55,9494
25	7,376934	9,40611	54,41916	88,4749	69,38825
26	4,722129	7,369304	22,2985	54,30665	34,79881
27	0,216987	4,574725	0,047083	20,9281	0,992654
28	1,993774	5,678579	3,975136	32,24626	11,3218
29	9,5468	9,927671	91,14139	98,55865	94,77749
30	7,572253	9,053316	57,33901	81,96253	68,55399
31	4,035768	7,796869	16,28742	60,79116	31,46635
32	4,425794	3,689077	19,58765	13,60929	16,3271
33	4,788659	0,793786	22,93126	0,630097	3,801173
34	1,951964	4,702902	3,810163	22,11729	9,179895
35	1,539354	9,467757	2,36961	89,63843	14,57423
36	4,251534	7,547838	18,07554	56,96985	32,08989
37	9,650868	7,558214	93,13926	57,1266	72,94333
38	5,616932	7,811213	31,54992	61,01504	43,87505
39	1,975768	2,663045	3,90366	7,091809	5,26156
40	9,783319	9,700919	95,71332	94,10782	94,90718
41	4,645833	5,125278	21,58376	26,26848	23,81119
42	4,516434	8,537248	20,39818	72,8846	38,55792
43	0,844447	2,955412	0,713091	8,734463	2,49569
44	8,093509	7,561266	65,50488	57,17274	61,19717
45	1,636402	5,603198	2,677813	31,39583	9,169088
46	9,240089	4,370251	85,37925	19,09909	40,3815
47	7,904599	4,388867	62,48269	19,26215	34,69223
48	7,087313	7,297891	50,23001	53,25922	51,72244
49	2,466811	2,405164	6,085157	5,784813	5,933085
50	2,71218	7,043977	7,35592	49,61761	19,10453
Средние	5,080367	5,218885	34,55755	36,09954	27,98996

На основании этих данных легко вычислить оценки дисперсий (см. (11.2)):

$$D^*(x) = S_0^2(x) = \frac{1}{n-1} \sum_{i=1}^n x_i^2 - \frac{n}{n-1} \bar{x}^2 = \frac{n}{n-1} a_2^*(x) - \frac{n}{n-1} \bar{x}^2 = 8,74746;$$

$$D_Y^* = S_0^2(y) = \frac{1}{n-1} \sum_{i=1}^n y_i^2 - \frac{n}{n-1} \bar{y}^2 = \frac{n}{n-1} a_2^*(y) - \frac{n}{n-1} \bar{y}^2 = 8,86278$$

и оценку корреляционного момента (см. (11.3)):

$$K_{XY}^* = \frac{1}{n-1} \sum_{i=1}^n x_i y_i - \frac{n}{n-1} \bar{x} \cdot \bar{y} = \frac{n}{n-1} a_{1,1}^*(x, y) - \frac{n}{n-1} \bar{x} \cdot \bar{y} = 1,476106.$$

Вычислим *точечную оценку коэффициент корреляции* по формуле (11.4):

$$R_{XY}^* = \frac{K_{XY}^*}{\sqrt{S_0^2(x) \cdot S_0^2(y)}} = 0,168.$$

Вычислим *интервальную оценку коэффициента корреляции* с надежностью $\gamma = 0,95$ по формуле (11.5). Для этого в таблице функции Лапласа (см. прил. 1) найдем значение, равное $\frac{g}{2} = 0,475$ и определим значение аргумента, ему соответствующее: $z_{0,95} = \arg \Phi(0,475) = 1,96$ (строка 1,9; столбец 6).

Вычислим вспомогательные значения a, b :

$$a = 0,5 \cdot \ln \left(\frac{1 + R_{XY}^*}{1 - R_{XY}^*} \right) - \frac{z_g}{\sqrt{n-3}} = 0,5 \cdot \ln \left(\frac{1,1676}{0,8324} \right) - \frac{0,196}{\sqrt{47}} = 0,5 \cdot 0,3384 - 0,0286 = 0,1406;$$

$$b = 0,5 \cdot \ln \left(\frac{1 + R_{XY}^*}{1 - R_{XY}^*} \right) + \frac{z_g}{\sqrt{n-3}} = 0,5 \cdot \ln \left(\frac{1,1676}{0,8324} \right) + \frac{0,196}{\sqrt{47}} = 0,5 \cdot 0,3384 + 0,0286 = 0,1978.$$

Таким образом, доверительный интервал для коэффициента корреляции имеет вид

$$I_g(R_{XY}) = \left[\frac{e^{2a} - 1}{e^{2a} + 1}; \frac{e^{2b} - 1}{e^{2b} + 1} \right] = [0,1406; 0,1978].$$

Проверим *гипотезу об отсутствии корреляционной зависимости*:

$$H_0: R_{XY} = 0;$$

$$H_1: R_{XY} \neq 0.$$

Так как объем выборки велик ($n \geq 50$), то вычислим значение критерия по формуле (11.7):

$$Z = \frac{|R_{XY}^*| \cdot \sqrt{n}}{1 - (R_{XY}^*)^2} = \frac{0,168 \cdot \sqrt{50}}{1 - (0,168)^2} = \frac{0,168 \cdot 7,0711}{1 - 0,028224} = \frac{1,1851}{0,9712} = 1,2195.$$

Определим значение Z_α из таблицы функции Лапласа (см. прил. 1):

$$Z_\alpha = \arg \Phi \left(\frac{1 - \alpha}{2} \right) = 1,96.$$

Так как $Z < Z_{\alpha}$, то гипотеза H_0 принимается, т. е. величины X и Y некоррелированы.

Вычислим оценки параметров a_0^* и a_1^* линии регрессии $\bar{y}(x) = a_0^* + a_1^*x$ по формуле (11.8):

$$a_1^* = \frac{K_{XY}^*}{S_0^2(x)} = \frac{1,4761}{8,7475} = 0,1687;$$

$$a_0^* = \bar{y} - a_1^* \cdot \bar{x} = 5,218 - 0,1687 \cdot 5,08 = 4,361.$$

Уравнение линии регрессии имеет вид:

$$\bar{y}(x) = 4,361 - 0,169x.$$

Построим диаграмму рассеивания, изобразив значения исходной двумерной выборки $\{(x_1, y_1), (x_2, y_2), \dots, (x_{50}, y_{50})\}$ в виде точек с координатами (x_i, y_i) на плоскости в декартовой системе координат и линию регрессии (рис. 11.1).

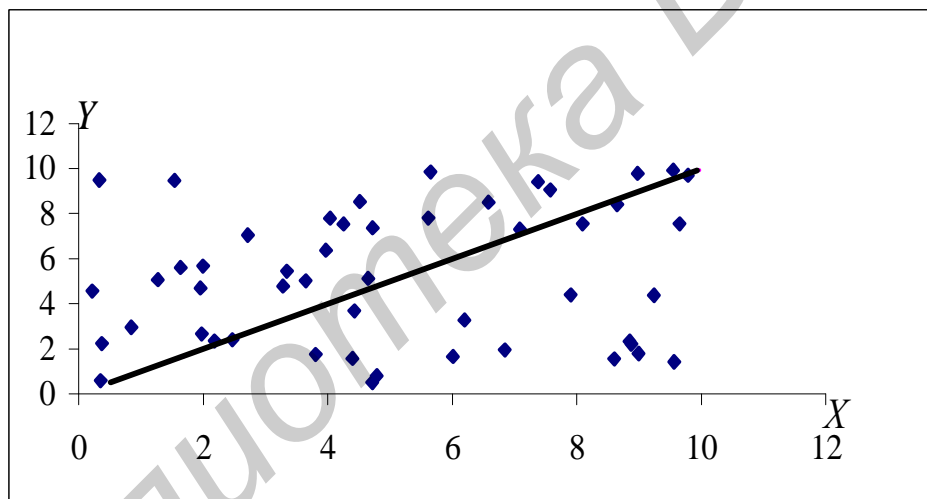


Рис. 11.1

ЛИТЕРАТУРА

1. Вентцель, Е. С. Теория вероятностей и ее инженерные приложения / Е. С. Вентцель, Л. А. Овчаров. – М. : Наука, 1988. – 416 с.
2. Вентцель, Е. С. Теория вероятностей и математическая статистика : учеб. пособие / Е. С. Вентцель. – 5-е изд., стереотип. – М. : Высш. шк., 1999. – 576 с.
3. Герасимович, А. И. Математическая статистика / А. И. Герасимович. – Минск : Высш. шк., 1983. – 279 с.
4. Жевняк, Р. М. Теория вероятностей и математическая статистика : учеб. пособие для студ. инж.-экон. спец. / Р. М. Жевняк, А. А. Карпук, В. Т. Унукович. – Минск : Харвест, 2000. – 384 с.
5. Волковец, А. И. Теория вероятностей и математическая статистика : практикум для студ. всех спец. очной формы обуч. БГУИР / А. И. Волковец, А. Б. Гуринович. – Минск : БГУИР, 2003. – 68 с. : ил.
6. Волковец, А. И. Теория вероятностей и математическая статистика : конспект лекций для студ. всех спец. очной формы обуч. БГУИР / А. И. Волковец, А. Б. Гуринович. – Минск : БГУИР, 2003. – 82 с. : ил.
7. Теория вероятностей и математическая статистика : сб. задач по типовому расчету / сост. А. В. Аксенчик [и др.]. – Минск : БГУИР, 2007. – 84 с.

Значения функции Лапласа $\Phi(x) = \frac{1}{\sqrt{2p}} \int_0^x \exp\left(-\frac{x^2}{2}\right) dx$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0004	0080	0120	0159	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0556	0596	0636	0675	0714	0753
0,2	0792	0832	0871	0909	0948	0987	1025	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1949	1985	2019	2054	2088	2126	2156	2190	2224
0,6	2257	2291	2324	2356	2389	2421	2453	2485	2517	2549
0,7	2580	2611	2642	2673	2704	2734	2764	2793	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3105	3123
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3364	3389
1,0	0,341	3437	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3686	3707	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3888	3906	3925	3943	3961	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4250	4265	4278	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4544
1,7	4554	4563	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4648	4656	4664	4671	4678	4685	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4755	4761	4767
2,0	0,477	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4825	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4874	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4924	4927	4929	4930	4932	4934	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
2,6	4953	4954	4956	4957	4958	4959	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986

3,0	0,49865	3,1	0,49903	3,2	0,49931	3,3	0,49952	3,4	0,49966	3,5	0,49977
3,6	0,49984	3,7	0,49989	3,8	0,49993	3,9	0,49995	4,0	0,499968	5,0	0,49999997

Таблица распределения Стьюдента $g = \int_{-t_{g,k}}^{t_{g,k}} f_t(x) dx$

k	γ			
	0,90	0,95	0,98	0,99
1	6,31	12,71	31,8	63,7
2	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,77	3,75	4,60
5	2,02	2,57	3,36	4,03
6	1,943	2,45	3,14	4,71
7	1,895	2,36	3,00	3,50
8	1,860	2,31	2,90	3,36
9	1,833	2,26	2,82	3,25
10	1,812	2,23	2,76	3,17
12	1,782	2,18	2,68	3,06
14	1,761	2,14	2,62	2,98
16	1,746	2,12	2,58	2,92
18	1,734	2,10	2,55	2,88
20	1,725	2,09	2,53	2,84
22	1,717	2,07	2,51	2,82
24	1,711	2,06	2,49	2,80
30	1,697	2,04	2,46	2,75
40	1,684	2,02	2,42	2,70
60	1,671	2,00	2,39	2,66
120	1,658	1,980	2,36	2,62
∞	1,645	1,960	2,33	2,58

Таблица распределения Пирсона c^2 $p(c^2 > c_{a,k}^2) = a$

k	α					
	0,01	0,02	0,05	0,95	0,98	0,99
1	6,64	5,41	3,84	0,004	0,001	0,000
2	9,21	7,82	5,99	0,103	0,040	0,020
3	11,34	9,84	7,82	0,352	0,185	0,115
4	13,28	11,67	9,49	0,711	0,429	0,297
5	15,09	13,39	11,07	1,145	0,752	0,554
6	16,81	15,03	12,59	1,635	1,134	0,872
7	18,48	16,62	14,07	2,17	1,564	1,239
8	20,10	18,17	15,51	2,73	2,03	1,646
9	21,07	19,68	16,92	3,32	2,53	2,09
10	23,20	21,2	18,31	3,94	3,06	2,56
12	26,2	24,1	21,0	5,23	4,18	3,57
14	29,1	26,9	23,7	6,57	5,37	4,66
16	32,0	29,6	26,3	7,96	6,61	5,81
18	34,8	32,3	28,9	9,39	7,91	7,02
20	37,6	35,0	31,4	10,85	9,24	8,26
22	40,3	37,7	33,9	12,34	10,60	9,54
24	43,0	40,3	36,4	13,85	11,99	10,86
26	45,6	42,9	38,9	15,38	13,41	12,20
28	48,3	45,4	41,3	16,93	14,85	13,56
30	50,9	48,0	43,8	18,49	16,31	14,95

Таблица распределения Колмогорова $p(0 \leq l < l_g) = g$

λ_γ	γ
0,50	0,0361
0,54	0,0675
0,58	0,1104
0,62	0,1632
0,66	0,2236
0,70	0,2888
0,74	0,3560
0,78	0,4230
0,82	0,4880
0,86	0,5497
0,90	0,6073
0,94	0,6601
0,98	0,7079
1,02	0,7500
1,06	0,7889
1,10	0,8223
1,14	0,8514
1,18	0,8765
1,22	0,8981
1,26	0,9164
1,30	0,9319
1,34	0,9449
1,38	0,9557
1,42	0,9646
1,46	0,9718
1,50	0,9778
1,54	0,9826
1,58	0,9864
1,62	0,9895
1,66	0,9918
1,70	0,9938
1,74	0,9953
1,78	0,9965
1,82	0,9973
1,86	0,9980
1,90	0,9985
1,94	0,9989
1,98	0,9992

Учебное издание

**ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА**

Методические указания по типовому расчету
для студентов всех специальностей
заочной формы обучения

С о с т а в и т е л и:

Волковец Александр Иванович
Гуринович Алевтина Борисовна
Аксенчик Анатолий Владимирович

Редактор Т. П. Андрейченко
Корректор Л. А. Шичко
Компьютерная верстка Е. Г. Бабичева

Подписано в печать 16.10.2009.	Формат 60x84 1/16.	Бумага офсетная.
Гарнитура «Таймс».	Печать ризографическая.	Усл. печ. л. 3,95.
Уч.-изд. л. 3,0.	Тираж 200 экз.	Заказ 151.

Издатель и полиграфическое исполнение: Учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники»
ЛИ №02330/0494371 от 16.03.2009. ЛП №02330/0494175 от 03.04.2009.
220013, Минск, П. Бровки, 6