МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ МОРСЬКОЙ УНІВЕРСИТЕТ

Ю.І. Буріменко, М.Я. Фомичев

ТЕОРІЯ ЙМОВІРНОСТІ, ЙМОВІРНІСТНІ ПРОЦЕСИ І МАТЕМАТИЧНА СТАТИСТИКА

Навчально-методичний посібник
для студентів денної та заочної форми навчання

Одеса 2005

Пособие подготовлено доктором технических наук, профессором **Бурименко Юрием Ивановичем** – заведующим кафедрой «Высшая и прикладная математика» Одесского национального морского университета и **Фомичевым Михаилом Яковлевичем** – старшим преподавателем этой же кафедры.

Пособие одобрено кафедрой «Высшая и прикладная математика» ОНМУ «_____» сентября 2004 г. (протокол №3.)

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Настоящее учебно-методическое пособие предназначено для студентов высших учебных заведений, в которых предусмотрено изучение элементов теории вероятностей и математической статистики в объеме до 108 часов.

Для усвоения материала пособия достаточно знания математического аппарата в объеме обычной программы по высшей математике (до 320 часов). Пособие знакомит с основными понятиями и некоторыми методами теории вероятностей и математической статистики, знание которых необходимо во многих областях техники, экономики, управления.

Теория вероятностей, зародившись в процессе анализа азартных игр, давно перешагнула этот рубеж развития; став наукой о случае, случайных событиях и процессах.

Задачи математической статистики связаны с разработкой методов сбора и обработки статистических данных различных явлений, опытов, экспериментов, процессов технического, производственного, социально-экономического и другого характера с целью получения научно достоверных знаний о них.

Пособие имеет следующую структуру.

Раздел теории вероятностей включает десять тем.

Раздел математической статистики – семь.

Подразделы включают:

- основные понятия и формулы;
- литературу;
- типовые примеры;
- методические указания;
- упражнения.

Примеры, приводимые в тексте пособия, играют важную роль в разъяснении и усвоении основных понятий. Их рекомендуется разбирать подробно.

РАЗДЕЛ 1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ

1.1. Предмет теории вероятностей. Основные понятия

Теория вероятностей занимается разработкой и изучением математических моделей случайных событий, носящих массовый характер и обладающих статистической устойчивостью параметров, характеризующих эти события.

Событием называется результат наблюдения, опыта, испытания. Чтобы событие произошло необходимо выполнение ряда условий – S.

События бывают ***достоверные*** (обозначается достоверное событие через U), ***невозможные*** (обозначается через V) и ***случайные.***

Достоверное событие обязательно происходит при заданных условиях S опыта. Невозможные не происходят.

Случайным событием называется событие, которое при выполнении условий S может произойти или не произойти. Например, выпадение герба (решки) при однократном бросании монеты.

Случайные события подразделяются на:

- зависимые;
- независимые;
- совместные;
- несовместные;
- равновозможные и т.п.

Обозначаются случайные события заглавными буквами латинского алфавита $A, B, C,...$.

Простейший (неразложимый) результат опыта называется ***элементарным событием*** или ***исходом*** и обозначается через ω. При опыте неизбежно наступает какой-то исход и только один. Исход – первичное (формально не определяемое) понятие. Множество всех возможных исходов опыта называется ***пространством элементарных событий*** и обозначается Ω.

Введение пространства элементарных событий позволяет формально определить случайное событие – это любое множество исходов, т.е. подмножество пространства Ω. Исходы, из которых состоит событие A, называются ***благоприятными*** для A $(\omega \in A \subset \Omega)$.

Событие A наступает при опыте тогда и только тогда, когда наступают благоприятные для него исходы.

Очевидно, множество Ω есть достоверное событие, пустое множество исходов (∅) – невозможное, несовместные события не содержат общих благоприятных исходов.

Система случайных событий. Множество событий $\{A_1, A_2, ..., A_n\}$ называется ***системой событий,*** если это множество не содержит ни одной пары равных событий. Если все события в системе попарно несовместны, то она называется ***системой несовместных событий.*** Последняя называется ***полной системой,*** если ее элементы образуют достоверное событие *U*, т.е.

$$\sum_{i=1}^{n} A_i = U .$$

Литература [5, С. 7-15; 13, С. 12-29; 10, С. 21-30].

Типовые примеры

1. Одноразовое подбрасывание игральной кости.

При этом испытании множество исходов Ω = $\{\omega_i\}$, где (ω_i = «при подбрасывании выпала цифра *i*» (*i* = 1, 2,..., 6).

Событиями могут быть, например, следующие:

A = «выпало четное число»;

B = «выпало нечетное число»;

C = «выпало число, делящееся на 3».

События *A*, *B* и C выражаются через ω_i и являются сложными (составными). Очевидно, что *A* происходит тогда и только тогда, когда имеют место исходы ω_2, или ω_4, или ω_6. В этом случае пишут

$$A = \{\omega_2, \omega_4, \omega_6\}.$$

Аналогично,

$$B = \{\omega_1, \omega_3, \omega_5\}, \quad C = \{\omega_3, \omega_6\}.$$

В данном примере пространство элементарных событий Ω состоит из шести элементов.

2. Физик-экспериментатор обстреливает некой элементарной частицей квадратный экран.

Исходами этого испытания будут попадания в точки квадрата. Введем на плоскости экрана прямоугольную систему координат *xOy*. Тогда множество всех исходов ω = (x, *y*) можно записать в виде

$$\Omega = \{\omega\} = \{(x, y):\ -a < x < a,\quad -a < y < a\}.$$

Здесь множество исходов испытания несчетно. Сложные события могут определяться как любые подмножества квадрата. Например, *A* = «попадание частицы в замкнутый круг с центром в начале координат радиуса *r* »(*r* < *a*).

Методические указания

Следует внимательно прочитать, возможно многократно, материал, указанный в рубрике «Литература», осмыслить приведенные в данном подразделе базовые понятия теории вероятностей, их прямую связь с реальными явлениями жизни. Для этого рекомендуется кроме внимательного рассмотрения приведенных типовых примеров, иллюстрирующих образование пространства элементарных событий Ω и случайных событий, самостоятельно привести и зафиксировать в своем конспекте подобные два-три примера из сферы личных или профессиональных интересов.

Упражнения

1. Сколько элементарных событий может иметь место при двукратном подбрасывании игральной кости?
2. Сколько элементарных событий возникает при пятикратной стрельбе в цель?
3. Монета подбрасывается три раза подряд. Построить пространство Ω. Описать событие *A*, состоящее в том, что выпало не менее двух «гербов».

1.2. Операции над событиями. Алгебра событий

Операции над событиями и их запись аналогичны принятым в операциях над множествами.

Запись $\omega_i \in A$ означает, что ω_i есть благоприятный исход события A, а $\omega_i \notin A$ – ω_i не есть таковым.

Запись $A \subset B$ означает, что событие A влечет за собой событие B. Если $A \subset B$, а $B \subset A$, то события A и B называют равными (эквивалентными, равносильными) и записывают $A = B$.

Сумма событий A и B есть событие C, записывают $C = A + B$ или $C = A \cup B$, состоящее в наступлении либо A, либо B, либо обоих событий в одном опыте. Если события A и B определяются элементарными событиями, т.е. $A = \{\omega_i\}$, $B = \{\varphi_i\}$,то

$$C = \{\omega_i\} \cup \{\varphi_i\}.$$

Разность событий A и B есть событие $C = A - B$, состоящее в том, что A происходит, а B не происходит.

Произведение событий A и B есть событие $C = A\ B$ или $C = A \cap B$, состоящее в наступлении обоих событий. Если $A = \{\omega_i\}$, $B = \{\varphi_i\}$, то событие C определяется общими элементарными событиями, входящими и в A и в B.

Операции сложения и умножения событий обобщаются на произвольные (конечное или счетное) число событий $\{A_k\}, \left(k = \overline{1,n}\right)$:

$$C = A_1 + A_2 + \ldots + A_k = \sum_{k=1}^{n} A_k\ , \quad C = A_1 \cdot A_2 \cdot \ldots \cdot A_k = \prod_{k=1}^{n} A_k\ .$$

Событие $\overline{A}$ есть противоположное событию A. Это событие означает, что событие A не происходит.

Приведем ***правила выполнения операций*** над событиями.

1. $A + A = A$
2. $A + B = B + A$
3. $A + (B + C) = (A + B) + C$
4. $A\,B = B\,A$
5. $A\,A = A$
6. $A\,(B\,C) = (A\,B)\,C$
7. $A\,(B + C) = A\,B + A\,C$
8. $A + (B\,C) = (A + B)\,(A + C)$
9. $A + V = A$
10. $A + U = U$
11. $A \cdot V = V$
12. $A \cdot U = A$
13. $\overline{A + B} = \overline{A} + \overline{B}$
14. $\overline{A} + \overline{B} = \overline{AB}$
15. $\overline{U} = V$
16. $A + \overline{A} = U$
17. $A \cdot \overline{A} = V$
18. $\overline{\overline{A}} = A$

Эти правила легко проверяются с помощью кругов Эйлера.

Алгебра событий. Система событий S_A называется алгеброй событий, если выполняются следующие условия:

1) $\Omega \in S_A$.

2) Из того, что $A \in S_A$ и $B \in S_A$, следует, что $AB \in S_A$, $A + B \in S_A$, $A \setminus B \in S_A$.

Алгебра событий называется σ алгеброй, если из того, что $A_n \in S_A$, $n = 1, 2, \ldots$, следует $\bigcup\limits_{n=1}^{\infty} A_n \in S_A$, $\bigcap\limits_{n=1}^{\infty} A_n \in S_A$.

Литература [5, С. 11-14; 13, С. 18-27; 10, С. 21-26].

Типовые примеры

1. Пусть *A*, *B* и С – случайные события. Записать события, состоящие в том, что из *A*, *B*, *C* произошли (произошло):

а) все три события;

б) по крайней мере одно событие;

в) только одно событие *A*;

г) событие *A* и *B*, и не произошло событие *C* .

Решение

а) имеет место событие $D = A \cdot B \cdot C'$,

б) произошли или A, или B, или С , или любые два из них, или все вместе, т.е. имеет место сумма событий $D = A + B + C$;

в) A – наступило, B и С – нет, т.е. произошли $\overline{B}$ и $\overline{C}$, следовательно $D = A \cdot \overline{B} \cdot \overline{C}$;

г) аналогично, $D = A \cdot B \cdot \overline{C}$.

2. При каких событиях A и B возможно равенство $A + B = A$?

Решение

Указанное равенство означает, что событие A включает в себя событие B.

3. Упростить выражение $(A + B) \cdot (A + \overline{B})$, где A и B – случайные события.

Решение

Используя свойства операций над событиями, получим

$$(A + B) \cdot (A + \overline{B}) = AA + AB + A\overline{B} + B\overline{B} =$$
$$= A + A(B + \overline{B}) + V = A + A \cdot U + V = A + A + V = A.$$

Методические указания

Приведенные литературные источники достаточно полно освещают вопросы этого подраздела. Для лучшего усвоения данного материала рекомендуется самостоятельно придумать два-три примера, ориентируясь при этом на приведенные типовые примеры.

Упражнения

1. Наудачу подбрасывается игральная кость. Пусть A = «выпало четное число»; B = «выпало число меньше трех». Описать события: $\overline{A}$, $\overline{B}$, $A + B$, $A \cdot \overline{B}$, $\overline{A} + \overline{B}$, $A \cdot B$, $\overline{A} \cdot B$, $(\overline{A \cdot B})$.
2. Из 25 студентов 20 увлекаются спортом (событие A), 9 – музыкой (событие B) 6 – музыкой и спортом (событие $A \cdot B$). С помощью кругов Эйлера выяснить, в чем состоят события $A \cdot \overline{B}$, $\overline{A} \cdot B$, $\overline{A + B}$.
3. Упростить выражение для события

$$(A + B) \cdot (A + \overline{B}) \cdot (\overline{A} + B).$$

4. Совместны ли события A и $\overline{A \cdot B}$?

1.3. Относительная частота случайного события, ее устойчивость. Вероятность события, ее определение

Относительной частотой случайного события (статистической вероятностью) называется отношение числа (m) случав появления этого события к общему числу (n) произведенных испытаний – $\left(\frac{m}{n}\right)$.

В разных достаточно больших сериях испытаний отношение $\left(\frac{m}{n}\right)$, как показывают многочисленные опыты, обладает определенной устойчивостью, т.е. числовые значения указанного отношения группируются около некоторого постоянного числа. Например, английский математик Карл Пирсон подбрасывал монету 24000 раз. При этом герб выпал 12012 раз $\frac{m}{n} = \frac{12012}{24000} = 0{,}5005\ldots.$ Другие подобные испытания также дают значение отношения $\left(\frac{m}{n}\right)$ близкое к 0,5.

Устойчивость относительной частоты может быть объяснена, только проявлением объективного свойства случайного события, заключающегося в определенной степени его возможности. Такая возможность количественно определяется понятием классическая вероятность.

Классической вероятностью случайного события *A* (обозначается *P*(*A*)) называется отношение числа элементарных событий, благоприятствующих событию *A*, предположим таковых *m*, к общему числу событий – *n*. Тогда

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

<u>Геометрическая вероятность</u>. Эта вероятность рассматривается в случае сложных событий, когда пространство элементарных событий Ω содержит бесконечную последовательность исходов, а некое событие *A* интерпретируется как выбор наудачу точки из подмножества (A^*) некоторого ограниченного множества (Ω^*) *n*-мерного евклидова пространства. В частном случае, это может быть отрезок, часть плоскости (см. пример 2 п. 1.1), пространства. В общем случае под геометрической вероятностью понимается число

$$P(A)=\frac{\mu(A^*)}{\mu(\Omega^*)}, \qquad (1.2)$$

где $\mu(A^*)$, $\mu(\Omega^*)$ – меры множества A^* и Ω^* (на прямой – длины, на плоскости – площади, в пространстве – объемы).

Наиболее общее понятие вероятности, охватывающее приведенные выше, определенно при аксиоматическом построении теории вероятностей, осуществленное выдающимся математиком А.Н. Колмогоровым (1903-1987.)

Числовая функция *P*(*A*) определяется на алгебре событий S_σ, являющейся σ-алгеброй (см, п. 1.2) следующим образом.

Говорят, что на σ-алгебре S_σ задано распределение вероятностей, если каждому событию $A \in S_\sigma$ однозначно поставлено в соответствие число *P*(*A*), называемое ***вероятностью события*** *A* так, что выполняются следующие условия **(*аксиомы теории вероятностей*)**:

1. $0 \le P(A) \le 1$.

2. $P(U) = 1$.

3. Если события $A_i \in S_\sigma$ ($n = 1,2,...$) попарно несовместны ($A_i \cdot A_j = V$, $i \neq j$; $i,j = 1,2,...$), то

$$P\left(\sum_{i=1}^{\infty} A_1\right) = \sum_{i=1}^{\infty} P(A_i).$$

Совокупность трех объектов (Ω, S_σ, $P(A)$), в которой S_σ является σ-алгеброй функция $P(A)$ удовлетворяет аксиомам теории вероятностей, называется ***вероятностным пространством.***

Литература [5, С. 15-30; 13, С. 29-37; 10, С. 26-36].

Типовые примеры

1. Куб, все грани которого окрашены, распилен на 1000 кубиков одинакового размера. Полученные кубики тщательно перемешаны. Определить вероятность того, что кубик, извлеченный наудачу, будет иметь две окрашенные стороны.

Решение

Куб имеет 12 ребер, на каждом из которых, очевидно, по 8 кубиков с двумя окрашенными сторонами. Поэтому число исходов, благоприятствующих извлечению кубика с двумя окрашенными сторонами (событие A) равно $m = 8 \cdot 12 = 96$. Общее число исходов $n = 1000$. Отсюда

$$P(A) = \frac{m}{n} = \frac{96}{1000} = 0{,}096.$$

2. Студенческая группа, состоящая из пятнадцати девушек и двенадцати юношей, выбирает по жребию трех представителей на профсоюзную конференции. Какова вероятность того, что среди представителей окажутся две девушки и один юноша?

Решение

Общее число равновозможных исходов выбора по жребию $n = C_{27}^{3}$ (число сочетаний из 27 по 3). Событие A = «выбор двух девушек и одного юноши» содержит благоприятствующих исходов $m = C_{15}^2 \cdot C_{12}^1$, где C_{15}^2 – число сочетании из 15, по 2 определяет количество вариантов возможного выбора двух девушек, а $C_{12}^1 = 12$ – выбора юношей. Их произведение и дает число благоприятствующих событию *A* исходов. Отсюда

$$P(A) = \frac{m}{n} = \frac{C_{15}^2 \cdot C_{12}^1}{C_{27}^3} = 0{,}4308$$

3. Из промежутка [0, 2] наудачу выбраны два числа *x* и *y*. Найти вероятность того, что эти числа удовлетворяют неравенству $x^2 \le 4y \le 4x$.

Решение

Будем интерпретировать указанный выбор двух чисел как выбор наудачу точки *M*(*x*,*y*) из множества всех точек квадрата со стороной равной двум.

Рассмотрим множество A^*, представляющее собой точки квадрата (множество Ω^*), координаты которых удовлетворяют двойному неравенству $x^2 \le 4y \le 4x$. (рис. 1.1).

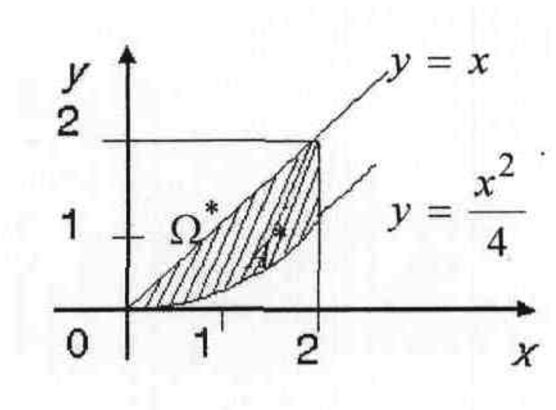


Рис. 1.1

Событие *A* = «два числа из [0, 2] удовлетворяют условию $x^2 \le 4y \le 4x$» происходит тогда и только тогда, когда точка *M*(*x*,*y*) принадлежит области A^*. По формуле (1.2) искомая вероятность равна отношению площади фигуры Ω^* к площади квадрата:

$$P(A) = \frac{\int_0^2 \left(x - \frac{x^2}{4}\right)dx}{4} = \frac{1}{3}.$$

Методические указания

Следует ясно представлять, что понятие вероятности является первичным, основным понятием, и в общем случае его нельзя определить через более простые понятия. Только в некоторых относительно простых схемах вероятность может быть вычислена непосредственно. Читателю рекомендуется самостоятельно, по аналогии с приведенными типовыми примерами придумать примеры, в которых вероятность события может быть вычислена по простой схеме.

Обратим внимание читателя на следующий интересный факт. Так, если $P(A)=1$, то при классическом определении вероятности событие A – достоверно. При геометрическом подходе это заключение неверно. Действительно, выделяя в рассматриваемой области Ω^* любое подмножество меры нуль (на отрезке конечное или счетное число точек, части плоскости – линию и т.д.), мера оставшейся части области Ω^* будет по-прежнему равна мере всей области Ω^*. Поэтому вероятность попадания точки в оставшуюся часть области будет равна единице. Однако, это событие не является достоверным, т.к. возможны попадания в выделенную часть области. Аналогично вероятность попадания точки в выделенную часть области Ω^* равна нулю, в то время как это событие является возможным. Этот факт следует учитывать при геометрическом подходе к определению вероятности.

Упражнения

1. Отдел технического контроля обнаружил в партии из 1000 изделий 20 бракованных. Найти частоту изготовления бракованного и небракованного изделия.

2. В начале координат расположена подвижная точка. Каждую секунду она с равной вероятностью смещается по вертикали на единицу вниз либо вверх. Сколько времени точка будет находиться на положительной полуоси координат, если наблюдать за ней 30 с?

3. Одновременно наудачу подбрасывают две игральные кости и следят за событиями: A = «сумма выпавших чисел равна 7» и B = «сумма выпавших чисел равна 8». Что вероятнее: событие A или B?

4. После бури на участке между 40-м и 70-м километрами линии электропередачи произошел обрыв провода. Какова вероятность того, что обрыв произошел между 45-м и 50-м километрами линии?

1.4. Условная вероятность. Независимость событий. Теорема умножения и сложения вероятностей

Условной вероятностью события A называется вероятность его появления, вычисленная в предположении, что имеет место событие B. Обозначается – $P(A/B)$. ***События*** A и B ***независимы,*** если $P(A/B) = P(A)$.

Вероятность произведения двух событий определяется по формуле

$$P(AB) = P(A)P(B / A) = P(B)P(A / B), \qquad (1.4)$$

которая обобщается на произведение n событий:

$$P\left(\prod_{i=1}^{n} A_i\right) = P(A_1)P(A_2 / A_1)P(A_3 / A_1A_2)...P\left(A_n / \sum_{i=1}^{n-1} A_i\right) \qquad (1.5)$$

События A_1, A_2,..., A_n независимы в совокупности, если для любых m (m = 2,3,..., n) и любых k_j (j = 1,2,...,n) $1 \le k_1 < k_2 < ... < k_m \le n$

$$P\left(\prod_{j=1}^{m} A_{k_j}\right) = \prod_{j=1}^{m} P\left(A_{k_j}\right).$$

Вероятность суммы двух событий определяется по формуле

$$P(A + B) = P(A) + P(B) – P(AB). \qquad (1.6)$$

Используя правила операций над событиями, нетрудно получить формулу для случая трех событий

$P(A + B + C) = P(A + (B + C)) = P(A) + P(B + C) – P(A(B + C)) =$

$= P(A) + P(B) + P(C) – P(BC) – P(AB + AC) =$

$= P(A) + P(B) + P(C) – P(BC) – P(AB) – P(AC) + P(ABC)$, которую легко обобщить на большее число событий.

Литература [5, С. 31-34; 13, С. 41-42; 10, С. 40-44].

Типовые примеры

1. Разрыв электрической цепи происходит в том случае, когда выходит из строя хотя бы один из трех последовательно соединенных элементов. Определить вероятность того, что не будет разрыва цепи, если элементы выходят из строя, соответственно, с вероятностью 0,3; 0,4 и 0,6. Как изменится искомая вероятность, если второй элемент не выходит из строя?

Решение

Искомая вероятность P равна вероятности того, что все три элемента не выйдут из строя. В силу независимости отказа в работе каждого из этих трех элементов следует

$$P = (1 - 0{,}3)(1 - 0{,}4)(1 - 0{,}6) = 0{,}7 \cdot 0{,}6 \cdot 0{,}4 = 0{,}168.$$

Если второй элемент не выходит из строя, то

$$\mathrm{P} = 0{,}7 \cdot 0{,}4 = 0{,}28.$$

2. Определить вероятность того, что выбранные наудачу изделие является первосортным, если известно, что 4% всей продукции есть брак, а 75% небракованных изделий - первый сорт.

Решение

Пусть событие A = «выбранное изделие небракованное», B = «выбранное изделие первосортное». Тогда очевидно

$P(A) = 1 - 0{,}04 - 0{,}96$; $P(A/B) = 0{,}75$.

Искомая вероятность $P = P(A\ B) = 0{,}96 \cdot 0{,}75 = 0{,}72$.

3. В ящике находится n изделий, из которых m стандартных. Найти вероятность того, что среди k наудачу извлеченных деталей хотя бы одна будет стандартной.

Решение

Обозначим через A искомое событие. $P(A) = 1 - P(\overline{A})$. $\overline{A}$ – среди извлеченных деталей нет ни одной стандартной. Нестандартных деталей n-m, из которых k различных нестандартных деталей можно выбрать C_{n-m}^{k} способами. Общее число способов выбора из n деталей k различных равно C_n^k. Отсюда

$$P\left(\overline{A}\right) = \frac{C_{n-m}^{k}}{C_n^k}; \quad P(A) = 1 - \frac{C_{n-m}^{k}}{C_n^k}.$$

Методические указания

Приведенные литературные источники достаточно полно освещают вопросы данной темы. При использовании формул умножения и сложения событий важно уяснить являются ли рассматриваемые события совместными или нет. Этот факт зачастую неявно выражаются в условиях задачи. Следует также внимательно разобрать приведенные типовые примеры и попытаться самостоятельно поварьировать условия в них.

Упражнения

1. Пусть события A и B таковы, что $P(A) = \frac{1}{2}$, $P(B) = \frac{2}{3}$.
Совместны ли они?

2. Вероятность выхода из строя *k-го* блока компьютера за время T равна p_k, (k = 1,2,...,n). Определить вероятность выхода из строя за указанный период хотя бы одного из n блоков, если работа всех блоков независима.

3. Вероятность появления события A в каждом опыте одинакова и равна 0,2. Опыты производятся последовательно до наступления события A. Определить вероятность того, что придется производить четвертый опыт.

4. Каждое из четырех несовместных событий может произойти соответственно с вероятностью 0,012; 0,010; 0,006 и 0,002.

Определить вероятность того, что в результате опыта произойдет хотя бы одно из этих событий.

5. Слово «ПАПАХА» составлено из букв разрезной азбуки. Затем карточки с буквами перемешаны и из них извлекаются по очереди и раскладываются в ряд какие-то четыре. Какова вероятность получить таким путем слово «ПАПА»?

1.5. Формула полной вероятности. Формула Байеса (формула вероятности гипотез)

Пусть H_i. ($i = \overline{1,n}$) – попарно несовместные события, составляющие полную группу событий и пусть известны $P(H_i)$, условные вероятности $P(A/H_i)$. Тогда имеет место ***формула полной вероятности.***

$$P(A)=\sum_{i=1}^{n}P(H_i)P(A/H_i). \qquad (1.7)$$

События H_i по отношению к событию A называются ***гипотезами.***

Пусть выполнены предыдущие условия и стало известно, что событие A произошло. Тогда для нахождения (уточнения) условных вероятностей $P(H_i/A)$ пользуются ***формулой Байеса.***

$$P(H_i/A)=\frac{P(H_i)\cdot P(A/H_i)}{P(A)} \qquad (1.8)$$

где $P(A)$ определяется по формуле (1.7).

Литература [5, С. 34-40; 13, С. 46-52; 10, С. 48-52].

Типовые примеры

1. Имеются три партии персональных компьютеров в количестве 20, 30 и

50 штук. Вероятности того, что компьютеры, приобретенные у разных фирм, проработают без ремонта заданное время, равны, соответственно, 0,7; 0,8; 0,9. Какова вероятность того, что выбранный наудачу один из ста данных компьютеров проработает без ремонта заданное время?

Решение

Событие A = «выбранный компьютер проработает заданное время без ремонта». Пусть гипотезы H_i (i = 1,2,3) означают, что выбран наудачу компьютер из i-*й* партии. Все гипотезы H_i попарно несовместны и образуют полную группу. По классическому определению вероятности находим

$$P(H_1) = \frac{20}{100} = 0{,}2, \quad P(H_2) = \frac{30}{100} = 0{,}3, \quad P(H_3) = \frac{50}{100} = 0{,}5.$$

Из условия задачи следует $P(A/H_1)$ = 0,7, $P(A/H_2)$ = 0,8, $P(A/H_3)$ =0,9.

По формуле (1.7) находим

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A/H_i) = 0{,}2 \cdot 0{,}7 + 0{,}3 \cdot 0{,}8 + 0{,}5 \cdot 0{,}9 = 0{,}83$$

3. Передаваемая информация состоит из «0» и «1». Статистические свойства помех таковы, что искажаются в среднем 2/5 цифры «0» и 1/3 цифры «1». Известно, что среди передаваемой информации «0» и «1» встречаются в отношении 5:3.

Определить вероятность того, что принята передаваемая информация, если: а) принята цифра «0»; б) принята цифра «1».

Решение

Обозначим событие A = «принята цифра «0»», B = «принята цифра «1»». Имеет место две гипотезы: H_1 – передана цифра «0», H_2 – передана цифра «1». По условию задачи $P(H_1)/P(H_2)$ = 5 / 3. Кроме того $P(H_1) + P(H_2)$ = 1 – полная группа событий). Отсюда $P(H_1)$ = 5 / 8, $P(H_2)$ = 3 / 8. Из условия также следует $P(B / H_1)$ = 2 / 5 (искаженные помехами «1» воспринимаются как «0»), отсюда $P(A / H_2)$= 3 / 5. Аналогично

$$P(A/H_2) = 1/3 \quad \Rightarrow \quad P(B/H_2) = 2/3.$$

Вероятности событий A и B находим по формуле (1.7)

$$P(A)=\frac{5}{8}\cdot\frac{3}{5}+\frac{3}{8}\cdot\frac{1}{3}=\frac{1}{2};\quad P(B)=\frac{5}{8}\cdot\frac{2}{5}+\frac{3}{8}\cdot\frac{2}{3}=\frac{1}{2}.$$

Искомые вероятности, т.е. вероятность того, что принята цифра «0» при передачи «0» и принята «1» при передаче «1» будут равны

$$P(H_1/A)=\frac{P(H_1)P(A/H_1)}{P(A)}=\frac{\frac{5}{8}\cdot\frac{3}{5}}{1/2}=\frac{3}{4};$$

$$P(H_2/B)=\frac{P(H_2)P(B/H_2)}{P(B)}=\frac{\frac{3}{8}\cdot\frac{2}{3}}{1/2}=\frac{1}{2}.$$

Методические указания

Следует обратить внимание на практическую значимость формул (1.7) и (1.8), содержательный смысл которых состоит в следующем. До появления события A вероятности $P(H_i)$ ($i=\overline{1,n}$) гипотез H_i, непосредственно связанных с появлением события A, называют ***априорными*** или ***доопытными.***

После наступления события A вероятность гипотез $P(H_i)$ можно переоценить, т.е. вычислить условные вероятности $P(H_i / A)$ по формуле (1.8). Эти уточненные вероятности называются апостериорными (послеопытными).

Подобные задачи решаются по ***схеме***:

- выяснить, в чем состоит испытание;
- событие, вероятность которого ищется, обозначить, например, буквой A;
- составить множество попарно несовместных событий /гипотез $\{H_i\}$;
- вычислить вероятности $P(H_i)$ и $P(A / H_i)$ ($i=\overline{1,n}$);
- по формуле (1.7) найти $P(A)$. Если известно, что событие A уже произошло, то по формуле (1.8) определить $P(H_i / A)$.

Упражнения

1. **1.** В студенческой группе 20 девушек и 10 юношей. К практическому занятию по теории вероятности не подготовились 4 девушки и 3 юноши. Вызванный студент оказался не подготовленным. Какова вероятность, что отвечать был вызван юноша?
2. **2.** Из полного набора костей домино наугад берутся две кости. Найти вероятность того, что вторую кость можно приставить к первой?
3. **3.** Из партии в пять изделий наудачу взято одно изделие, оказавшееся бракованным. Количество бракованных изделий равновозможно любое. Какое предположение о количестве бракованных изделий наиболее вероятно?
4. **4.** Определить вероятность того, что среди 1000 мобильных телефонов нет ни одного неисправного, если из взятых наудачу 100 телефонов все оказались исправными. Предполагается, что число неисправных телефонов из 1000 равновозможно от 0 до 5.

1.6. Условная вероятность. Независимость событий. Теорема умножения и сложения вероятностей

В ***схеме Я. Бернулли*** (1654-1705) рассматривается серия n независимых испытаний, каждое из которых имеет лишь два исхода: наступления некоторого события A или его ненаступления. При этом известна вероятность $P(A) = p$ $(0 < p < 1)$. Она постоянна в серии испытаний. Числа n и p называются ***параметрами схемы Бернулли.***

Вероятность $P_n(m)$ $(0 \le m \le n)$ наступления события A равно m раз при n независимых испытаниях определяется по формуле Бернулли

$$P_n(m) = C_n^m p^m q^{n-m}, \qquad (1.9)$$

где $q = P(\overline{A}) = 1 - p$.

Вероятности $P_n(m)$ $(m = 0,1,2,...,n)$ называются биноминальными, т.к.

правая часть формулы (1.9) представляет собой общий член разложения бинома Ньютона

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

Отсюда следует, что сумма всех биноминальных вероятностей равна 1. При больших n и m вычисления по формуле (1.9) связаны с немалыми сложностями (растет объем вычислений). В этом случае для нахождения вероятностей $P_n(m)$ целесообразно пользоваться приближенными формулами, которые дают при достаточно больших n сколь угодно малую относительную погрешность вычислений.

Из ***локальной предельной теоремы Муавра-Лапласса*** следует

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \varphi(x), \qquad (1.10)$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$; $\quad x = \frac{m-np}{\sqrt{npq}}$.

Интегральная предельная теорема Муавра-Лапласа утверждает, что для схемы Бернулли при больших n и $p \in (0;1)$ справедливо приближенное равенство

$$P_n(m_1 \le m \le m_2) \approx \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{x^2}{2}} dx = \Phi_0(x_2) - \Phi_0(x_1) \qquad (1.11)$$

$$x_i = \frac{m_i - np}{\sqrt{npq}}, \quad i = 1,2.$$

Для вычислений по формуле (1.11) (интеграл в ней не вычисляется в элементарных функциях) используется функция Лапласа (интеграл вероятности).

$$\Phi_0(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt, \quad x \in R,$$

обладающая свойствами:

- ✓ $\Phi_0(x)$ – нечетная функция. $(\Phi(-x) = -\Phi(x))$;
- ✓ $\Phi_0(x)$ – возрастает на *R*;
- ✓ $\lim\limits_{x\to+\infty} \Phi_0(x) = 1$;
- ✓ $\lim\limits_{x\to-\infty} \Phi_0(x) = -1$;
- ✓ при $x \ge 4$ $\Phi_0(x)$ =0,9998…

Функция $\Phi_0(x)$ табулирована (см. табл. 1 приложения)

Из предельной теоремы Пуассона для схемы Бернулли, если $n \to \infty$, $p \to 0$ так, что $np \to \lambda$, $0 < \lambda < \infty$, следует

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}. \qquad (1.12)$$

Литература [5, С. 41-51; 13, С. 57-75; 10, С. 52-67].

Типовые примеры

1. Вероятность попадания стрелком мишень равна 0,6. Произведено 8 выстрелов. Найти: а) вероятность того, что при этом будет ровно два попадания; б) наивероятнейшее число попаданий и соответствующую вероятность.

Решение

Испытания (выстрелы по мишени) будем считать независимыми, а вероятность попадания в мишень неизменной. Тогда можно использовать в качестве модели схему Бернулли. По формуле (1.9) получаем

$$P_8(2) = C_8^2 (0{,}6)^2 (0{,}4)^2 \approx 0{,}041.$$

Для определения наиболее вероятного числа попаданий *m* используется формула $np - q \le m_0 \le np + q$. Откуда

$$8 \cdot 0{,}6 - 0{,}4 \le m_0 \le 8 \cdot 0{,}6 + 0{,}4 \quad \Rightarrow \quad m_0 = 5,$$

а соответствующая вероятность

$$P_8(5) = C_8^5 (0{,}6)^5 \cdot (0{,}4)^3 \approx 0{,}28.$$

2. В некотором производстве вероятность того, что отдельное изделие окажется бракованным постоянна и равна 0,005. Какова вероятность того, что в партии из 10 000 изделий бракованных будет не более 70?

Решение

Здесь можно воспользоваться формулой Бернулли. Тогда

$$P_{10000}(0 \le m \le 70) = \sum_{m=0}^{70} C_{10000}^m (0{,}005)^m (0{,}995)^{10000-m}$$

Очевидно, что вычисления по этой формуле достаточно трудоемки. Поэтому воспользуемся формулой (1.11). При этом $m_1 = 0$, $m_2 = 70$, $n = 10000$, $p = 0{,}005$, $q = 0{,}995$. Находим

$$x_1 = \frac{m_1 - np}{\sqrt{npq}} \approx 7{,}09, \quad x_2 = \frac{m_2 - np}{\sqrt{npq}} \approx 2{,}84.$$

Пользуясь свойствами функции $\Phi_0(x)$ и ее табличными значениями находим:

$$P_{10000}(0 \le m \le 70) \approx \frac{1}{2}\left(\Phi_0(2{,}84) + \Phi_0(7{,}09)\right) \approx 0{,}4977 + 0{,}5 \approx 0{,}9977.$$

3. Вероятность допустить ошибку при наборе некоторого текста, состоящего из 800 знаков, равна 0,005. Найти наиболее вероятное число сделанных ошибок в этом тексте и его вероятность.

Решение

Очевидно, можно применять схему Бернулли. Из формулы

$$np - q \le .\ m_0 \le np + q. \quad \Rightarrow \quad m_0 = 4.$$

По формуле (1.12) с $\lambda = np = 4$, получаем

$$P_{800}(4) \approx \frac{4^4}{4!} e^{-4} \approx 0{,}1954.$$

Заметим, что точная формула (1.9) дает $P_{800}(4) \approx 0{,}1954$.

Методические указания

Схема Бернулли – математическая модель реального явления. Для ее применения к решению задач необходимо выполнение ***условий***:

- проводимые испытания были независимы;
- каждое испытание имело только два исхода;
- вероятность появления интересующего события в каждом испытании была одна и та же.

При выборе расчетных формул следует руководствоваться такими соображениями:

1. Если число испытаний n достаточно велико и вероятность $p \in (0; 1)$ не мала, а число $npq > 10$, то для нахождения величин $P_n(m)$ и $P_n(m_1 \le m \le m_2)$ используется формулы (1.10) и (1.11) соответственно;

2. Если n достаточно велико, а вероятность p мала настолько, что число np не велико (обычно $p < 0{,}1$; $npq < 10$), то для нахождения $P_n(m)$ удобна формула (1.12). При этих же предположениях и малом числе слагаемых в сумме $\sum_{m=m_1}^{m_2} P_n(m)$ можно использовать формулу

$$P_n(m_1 \le m \le m_2) \approx e^{-\lambda} \sum_{m=m_1}^{m_2} \frac{\lambda^m}{m!}.$$

Упражнения

1. Вероятность выигрыша по одному лотерейному билету равна 0,2.

Какова вероятность того, что из шести приобретенных билетов два окажутся выигрышными?

2. Найти вероятность осуществления от двух до четырех разговоров по телефону при наблюдении пяти независимых вызовов, если вероятность того, что разговор состоится, равна 0,7.
3. Известно, что левши в среднем составляют 1% населения. Каковы шансы на то, что среди отобранных наудачу 200 человек окажутся ровно три левши?
4. Учебник издан тиражом 100 000 экз. Вероятность того, что некоторые экземпляры будут сброшюрованы неправильно равна 0,0001. Найти вероятность того, что тиражи содержат ровно пять бракованных книг.
5. Электростанция питает сеть из 10 000 потребителей. Вероятность того, что каждый потребитель в вечерние часы будет пользоваться электроэнергией равна 0,7. Определить вероятность того, что число одновременно подключенных к сети потребителей будет находится между 5 900 и 6 100.

1.7. Случайные величины

Дискретная случайная величина

Рассмотрим пространство элементарных событий Ω, элементами которого ω отвечает $P(\omega)$.

Определение. ***Дискретной случайной величины*** ξ называется функция ξ(ω), определенная на множестве Ω и принимающая дискретные вещественные или комплексные значения.

Таблица вида

$$\begin{pmatrix} a_1 & a_2 & \dots \\ P_1 & P_2 & \dots \end{pmatrix},$$

где в верхней строке стоят возможные значения случайной величины ξ, а в нижней строке под каждым значением случайной величины стоит вероятность $P_i = P(\xi = a_i)$ того, что ξ принимает это значение, называется ***распределением дискретной случайной величины***, причем

$$\sum_{a_i} P(\xi = a_i) = 1.$$

Литература [5, С. 58-65; 13, С. 64; 10, С. 73].

Типовые примеры

1. Пусть случайное событие A имеет вероятность $P(A) = p$. Назовем ***индикатором*** события A случайную величину I_A, которая принимает только два значения нуль и один.

I_A .= 1 – означает, что событие A «произошло».

I_A = 0 – означает, что событие A «не произошло».

Очевидно, что индикатор события A есть дискретная случайная величина с таблицей распределения.

I_A	0	1
$p(I_A)$	1-p	p

2. Из урны, в которой лежат 2 белых и 8 черных шаров, последовательно вынимают шары до тех пор, пока не появится черный шар. Число ξ вынутых при этом шаров есть дискретная случайная величина. Найти закон распределения ее вероятностей.

Решение

Возможными значениями величины ξ являются числа 1, 2, 3.

$p(\xi = 1)$ – вероятность того, что будет вынут только один шар, т.е. первый шар – черный, тогда

$$p(\xi = 1) = \frac{8}{10} = \frac{4}{5},$$

$p(\xi = 2)$ – вероятность того, что будет вынуты только два шара, т.е. первый шар – белый, а второй – черный, тогда

$$p(\xi = 2) = \frac{2}{10} \cdot \frac{8}{9} = \frac{8}{45},$$

$p(\xi = 3)$ – вероятность того, что первый и второй шар будут белые, тогда

$$p(\xi = 3) = \frac{2}{10} \cdot \frac{1}{9} = \frac{1}{45}.$$

Закон распределения количества вынутых шаров запишется

ξ	1	2	3
$p(\xi)$	$\frac{4}{5}$	$\frac{8}{45}$	$\frac{1}{45}$

3. Выполняется стрельба по мишени до первого промаха, причем вероятность попадания в мишень при каждом выстреле одна и та же $\frac{4}{5}$. ξ – количество выполненных выстрелов до первого промаха в мишень есть случайная величина. Записать ее закон распределения.

Решение

Стрельба заканчивается на *k*-ом выстреле, если *k*-1 первых выстрела были успешны, а при *k* выстреле допущен промах. Тогда $P(\xi = k) = \left(\frac{4}{5}\right)^{k-1} \cdot \frac{1}{5}$, $k = 1, 2, \ldots$. Полученная формула эквивалентна ряду распределения:

ξ	1	2	…	k	…
$p(\xi)$	$\frac{1}{5}$	$\frac{4}{5} \cdot \frac{1}{5}$	…	$\left(\frac{4}{5}\right)^{k-1} \cdot \frac{1}{5}$	…

Особенность данной задачи состоит в том, что теоретически число испытаний может быть бесконечно большим, однако вероятность такого события стремится к нулю.

4. Монета подбрасывается четыре раза. Найти закон распределения числа появлений герба, если вероятность его появления при каждом подбрасывании $\frac{1}{2}$.

Решение

ξ – количество появлений герба при бросании монеты четыре раза может принимать значения 0;1; 2; 3,4.

$p(\xi = m)$ вычисляется по формуле

$$P(\xi = m) = \frac{n!}{m!(n-m)!} p^m q^{n-m},$$

для нашего случая $n = 4$, $P = 0{,}5$, $q = 0{,}5$, $m = 0, 1, 2, 3, 4$.

Таблица распределения величины ξ выглядит следующим образом

ξ	0	1	2	3	4
$p(\xi)$	0,0625	0,25	0,375	0,25	0,0625

Методические указания

Обратите внимание на то, что сумма вероятностей всех возможных значений дискретной случайной величины равна единице, что может служить косвенным критерием правильности решения задачи на определение ее закона распределения. Кроме того, закон распределения можно изобразить графически. Для этого в прямоугольной системе координат строят точки $M_1(x_1, P_1), M_2(x_2, P_2), ..., M_n(x_n, P_n), ...,$ где x_i – возможные значения случайной величины, а P_i – соответствующие вероятности и соединяют их последовательно отрезками прямых. Полученную фигуру называют многоугольником распределения для дискретной случайной величины ξ.

Упражнения

1. Бросается игральная кость. Обозначим событие A = «число выпавших очков четно», B = «число выпавших очков делится на 3». Записать таблицу распределения I_A, I_B.

2. Из 28 костей домино случайно и равновероятно выбирается одна. Найти закон распределения суммы очков ξ на ее половинах.
3. Записать ряд распределения и построить многоугольник распределения числа попаданий мячом в корзину при двух бросках, если вероятность попадания при каждом броске 0,4.
4. Из партии в 25 изделий, среди которых имеются 6 бракованных, выбраны случайным образом 3 изделия для проверки их качества. Найти закон распределения случайного числа ξ количества бракованных изделий, содержащихся в выборке и изобразить его графически.

Непрерывные случайные величины

Определение. Величина ξ называется ***непрерывной случайной величиной***, если вероятность попадания ее значения в любой интервал (x_1, x_2) может быть представлена в виде

$$P\left(x_1 < \xi < x_2\right) = \int_{x_1}^{x_2} f\left(x\right) dx$$

при этом функция $f\left(x\right)$ должна быть неотрицательна и нормирована условием

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

Функция $f\left(x\right)$ называется ***плотностью распределения величины*** ξ.

Литература [5, С. 69-72; 13, С. 83-86; 10, С. 111-119].

Типовые примеры

1. Величина ξ имеет плотность распределения

$$f(x)=\begin{cases} 0- & \text{при} \quad x\le 0, \\ \frac{1}{2}\sin x - & \text{при} \quad 0<x\le\pi, \\ 0- & \text{при} \quad x>\pi. \end{cases}$$

Найти вероятность того, что в результате испытания величина ξ примет значения из интервала $\left(0;\frac{\pi}{4}\right)$.

Решение

Из определения непрерывной случайной величины имеем:

$$P\left(0<\xi<\frac{\pi}{4}\right)=\frac{1}{2}\int_0^{\frac{\pi}{4}}\sin x dx$$

или

$$P\left(0<\xi<\frac{\pi}{4}\right)=-\frac{\cos x}{2}\Bigg|_0^{\frac{\pi}{4}}=\frac{1}{2}\left(1-\frac{\sqrt{2}}{2}\right).$$

2. Плотность вероятности случайной величины ξ равна $f(x)=\dfrac{A}{e^x+e^{-x}}$, $(-\infty<x<+\infty)$. Найти а) коэффициент A; б) вероятность того, что в двух независимых наблюдениях ξ примет значение меньше единицы.

Решение

а) Из условия нормирования $f(x)$ имеем

$$A\int_{-\infty}^{+\infty}\frac{1}{e^x+e^{-x}}dx=1.$$

или

$$A\int_{-\infty}^{+\infty}\frac{1}{e^x+e^{-x}}dx = A arctg e^x \Big|_{-\infty}^{+\infty} = A\left(\frac{\pi}{2}-0\right),$$

но $\frac{A\pi}{2}=1$, откуда $A=\frac{2}{\pi}$.

б) Вычислим

$$P(-\infty<\xi<1)=\frac{2}{\pi}\int_{-\infty}^{1}\frac{e^x dx}{e^{2x}+1}=\frac{2}{\pi}arctg\ e.$$

$\frac{2}{\pi}arctg\ e$ – вероятность того, что величина ξ примет значение меньше единицы в одном испытании, тогда в двух независимых испытаниях вероятность этого события будет $\left(\frac{2}{\pi}arctg\ e\right)^2 \approx 0{,}6015$.

3. Случайная величина ξ называется ***распределенной нормально*** с параметрами a, σ ($\sigma > 0$), если ее плотность

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Вероятность попадания случайной величины ξ распределенной по нормальному закону вычисляется по формуле

$$P(x_1<\xi<x_2)=\frac{1}{2}\left(\Phi\left(\frac{x_2-a}{\sigma}\right)-\Phi\left(\frac{x_1-a}{\sigma}\right)\right),$$

где

$$\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}dt$$

Свойства функции $\Phi(x)$ приведены в гл. 1.6.

Вычислить вероятность попадания нормально распределенной случайной величины ξ с параметрами a, σ ($\sigma > 0$) в интервале $a-3\sigma<\xi<a+3\sigma$.

Решение

$$P(a-3\sigma<\xi<a+3\sigma)=\frac{1}{2}\left(\Phi\left(\frac{a+3\sigma-a}{\sigma}\right)-\Phi\left(\frac{a-3\sigma-a}{\sigma}\right)\right)=\Phi(3).$$

По таблице $\Phi(3)$ = 0,9973. Из этого соотношения практически достоверно следует, что значения случайной величины ξ распределенной нормально находятся в интервале ($a-3\sigma$; $a+3\sigma$). Указанное утверждение носит название правила трех сигм для нормального распределения.

Методические указания

Необходимо обратить особое внимание на изучение нормального закона распределения, его важность определяется тем, что к нему обычно приводят задачи связанные с распределением сумм большого числа случайных величин.

Упражнения

1. Плотность распределения ξ задана формулой

$$f(x)=\begin{cases} cx^{-\frac{3}{2}} & при \quad x\ge 1, \\ 0 \; - & при \quad x<1. \end{cases}$$

Найти *c*.

2. Пусть все возможные значения случайной величины ξ сосредоточены на интервале (α; β) и плотность распределения ее постоянна. Записать $f(x)$ и вероятность попадания величины ξ в интервал (x_1; x_2),если $x_1 \in$ (α; β),

$x_2 \in (\alpha;\beta)$.

3. Случайная величина имеет нормальное распределение с параметрами $a = 0$, $\sigma = 1$. Что больше

$$P(-0{,}5 \le \xi \le -0{,}1) \text{ или } P(1 \le \xi \le 2)?$$

1.8. Функция распределения вероятностей случайной величины

Определение. Функция $F(x)$ называется ***функцией распределения вероятностей случайной величины*** ξ, если ее значение в точке x равно вероятности того, что случайная величина ξ примет значение меньше x, т.е.

$$F(x) = P(\xi < x).$$

Для непрерывной случайной величины ξ

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

где $f(x)$ – плотность распределения вероятностей случайной величины ξ.

Из этой формулы следует

$$F(x) = f(x),$$

при условии кусочной непрерывности функции $f(x)$.

Для дискретной случайной величины ξ функция распределения вероятностей есть сумма вероятностей тех ее значений, которые меньше x.

Функция распределения имеет ***свойства***:

1. $0 \le F(x) \le 1$.
2. $P(a \le \xi < b) = F(b) - F(a)$.
3. $F(x_1) \le F(x_2)$, если $x_1 < x_2$.
4. $F(-\infty) = 0$, $F(+\infty) = 1$.

Литература [5, С. 65-72; 13, С. 77-86; 10, С. 111-119].

Типовые примеры

1. Закон распределения дискретной случайной величины ξ задан таблицей

ξ	1	2	3
$P(\xi)$	0,3	0,5	0,2

Записать $F(x)$ случайной величины ξ.

Решение

Из определения функции распределения имеем

$$F(x)=\begin{cases}0, & \text{при} \quad \xi \le 1,\\ 0{,}3, & \text{при} \quad 1<\xi \le 2,\\ 0{,}5, & \text{при} \quad 2<\xi\le 3,\\ 1, & \text{при} \quad \xi>3.\end{cases}$$

Для удобства записи таких функций введем функцию

$$\eta(x)=\begin{cases}0, & \text{при} \quad x\le 0,\\ 1, & \text{при} \quad x>0\end{cases}$$

$$F(x)=0{,}3\eta(x-1)+0{,}5\eta(x-2)+0{,}2\eta(x-3)$$

или

$$F(x)=\sum_i P_i\eta(x-x_i)$$

2. Рассмотреть функцию распределения для стандартного нормального закона $(a=0,\ \sigma=1)$ записать ее через интеграл вероятности.

Решение

По определению

$$F(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

или

$$F(x)=\frac{1}{2}\left(\Phi(x)-\Phi(-x)\right)$$

где $\Phi(x)=\frac{2}{\sqrt{2\pi}}\int\limits_{0}^{x}e^{-\frac{t^2}{2}}dt;\quad \Phi(-\infty)=-1.$

Тогда $F(x)=\frac{1}{2}\left(\Phi(x)+1\right).$

Методические указания

Следует обратить внимание на то, что функция распределения случайной величины является ее исчерпывающей характеристикой.

Из определения функции распределения следует, что с ее помощью можно задавать как закон распределения дискретной случайной величины, так и непрерывной величины.

Упражнения

1. Плотность вероятности случайной величины ξ равна $f(x)=Ax^2e^{-\lambda x}$, $(x>0,\quad 0\le x<+\infty)$. Найти: а) коэффициент A; б) вычислить вероятность попадания ξ в интервал $\left(0;\frac{1}{\lambda}\right)$.

2. Функция распределения случайной величины ξ задана формулой

$$F(x) = A + B\,arctg\,x, \quad (-\infty < x < +\infty).$$

Найти:

а) постоянные *A* и *B*;

б) плотность вероятности $f(x)$;

в) вероятность того, что величина ξ попадет в отрезок [–1;1].

3. Бросают три монеты

а) Построить закон распределения случайной величины ξ по количеству выпавших при этом «решек», если вероятность появления «решки» на каждой 1 монете $\frac{1}{2}$.

б) Записать функцию распределения $F(x)$ величины ξ, построить ее график.

4. Функция распределения случайной величины ξ имеет вид:

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

а) Найти плотность вероятности случайной величины ξ.

б) Построить график функции распределения $F(x)$ и ее плотности $f(x)$.

1.9. Закон распределения функции от одной случайной величины

Под ***функцией*** $f(x)$ *случайной величины* ξ понимают такую случайную величину η, которая принимает значение $y = f(\xi)$, когда величина ξ принимает значение *x*. Основной задачей здесь является нахождение закона распределения вероятностей функции $y = f(\xi)$ по закону распределения ее аргумента.

Если случайная величина ξ имеет закон распределения

ξ	x_1	x_2	…
$P(\xi)$	P_1	P_2	…

и $y = \varphi(\xi)$ монотонная функция от действительного аргумента x, тогда дискретная случайная величина $\eta = \varphi(\xi)$ имеет закон распределения

η	$\varphi(x_1)$	$\varphi(x_2)$	…
$P(\eta)$	P_1	P_2	…

Если же функция $y = \varphi(\xi)$ не монотонная функция, то среди ее значений, могут быть равные. В этом случае столбцы с равными, значениями например, $a = \varphi(x_i)$, объединяют в один столбец и для вычисления $P(\varphi(x_i) = a)$ складывают вероятности, которые в распределении величины ξ записаны в столбцах под x_i.

Литература [5, С. 73-74; 13, С. 48-102; 9, С. 63-64].

Типовые примеры

1. Дискретная случайная величина ξ имеет закон распределения

ξ	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$
$P(\xi)$	0,2	0,7	0,1

Построить закон распределения случайной величины $\eta = \sin \xi$.

Решение

η	$\sin\frac{\pi}{4}$	$\sin\frac{\pi}{2}$	$\sin\frac{3\pi}{4}=\sin\frac{\pi}{4}$
$P(\eta)$	0,2	0,7	0,1

η	$\frac{\sqrt{2}}{2}$	1
$P(\eta)$	0,3	0,7

Для ξ непрерывной случайной величины ниже рассмотрим несколько примеров, после чего сформулируем правило нахождения закона распределения функции по известному закону распределения непрерывной величины ξ.

2. $y = ax + b$.

Положим $\eta = a\xi + b$, если $a > 0$, то по определению функции распределения

$$F_\eta(x) = P(\eta < x) = P(a\xi + b < x) \underset{a>0}{=} P\left(\xi < \frac{x-b}{a}\right) = F_\xi\left(\frac{x-b}{a}\right).$$

Для вычисления $f_\eta(x)$ имеем

$$f_\eta(x) = F_\eta'(x) = \left(F_\xi\left(\frac{x-b}{a}\right)\right)_x' = F_\xi'\left(\frac{x-b}{a}\right)\frac{1}{a} = f_\xi\left(\frac{x-b}{a}\right)\frac{1}{a}.$$

Если $a < 0$, то

$$F_\eta(x) = P(\eta < x) = P(a\xi + b < x) \underset{a<0}{=} P\left(\xi > \frac{x-b}{a}\right) = 1 - F_\xi\left(\frac{x-b}{a}\right).$$

и

$$f_\eta(x) = F_\eta'(x) - \left(F_\xi\left(\frac{x-b}{a}\right)\right)_x' = \frac{-1}{a} f_\xi\left(\frac{x-b}{a}\right).$$

Объединяя эти два случая, имеем

$$f_\eta(x)=\frac{1}{|a|}\,f_\xi\left(\frac{x-b}{a}\right).$$

3. Рассмотрим монотонно возрастающую функцию

$$f(x)=x^3.$$

Положим $\eta=\xi^3$. Имеем по определению функции распределения

$$F_\eta(x)=P(\eta<x)=P(\xi^3<x)\underset{a<0}{=}P\left(\xi<x^{\frac{1}{3}}\right)=F_\xi\left(x^{\frac{1}{3}}\right).$$

Тогда

$$f_\eta(x)=\left(F_\xi\left(x^{1/3}\right)\right)'_x=F'_\xi\left(x^{1/3}\right)\cdot\frac{1}{3}x^{-2/3}=\frac{1}{3\sqrt[3]{x^2}}\,f_\xi\left(x^{1/3}\right)$$

4. Рассмотрим монотонно убывающую функцию

$$f(x)=e^{-x}.$$

При $x>0$

$$F_\eta(x)=P(\eta<x)=P\left(e^{-\xi}<x\right)=$$
$$=P(\xi>-\ln x)=1-P(\xi<-\ln x)=1-F_\xi(-\ln x),$$

откуда

$$f_\eta(x)=F'_\eta(x)=-\left(F'_\xi(-\ln x)\right)'_x=\frac{1}{x}\,f_\xi(-\ln x).$$

При $x<0$

$$F_\eta(x)=P(\eta<x)=0,\ \text{т.к.}\ \eta=e^{-\xi}>0.$$

Поэтому на отрицательной полуоси $f_\eta(x)=F'_\eta(x)=0$.

5. Рассмотрим немонотонную функцию $y = x^2$. Найдем $F_\eta(x)$ и $f_\eta(x)$ случайной величины $\eta = \xi^2$.

При $x > 0$ имеем

$$F_\eta(x) = P(\eta < x) = P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) =$$
$$= P(\xi < \sqrt{x}) - P(\xi < -\sqrt{x}) = F_\xi(\sqrt{x}) - F_\xi(-\sqrt{x}),$$

откуда

$$f_\eta(x) = F'_\eta(x) = F'_\xi(\sqrt{x})\frac{1}{2\sqrt{x}} - F'_\xi(-\sqrt{x})\left(-\frac{1}{2\sqrt{x}}\right) =$$
$$\frac{1}{2\sqrt{x}}\left(F'_\xi(\sqrt{x}) + F^1_\xi(-\sqrt{x})\right) = \frac{1}{2\sqrt{x}}\, f_\xi((\sqrt{x}) + f(-\sqrt{x})).$$

При $x \le 0$ имеем $F_\eta(x) = 0$ и $f_\eta(x) = 0$.

После рассмотренных примеров можно указать следующее правило записи закона распределения непрерывной функции от одной случайной величины по известному распределению ее аргумента.

Пусть ξ непрерывная случайная величина с функцией распределения $F_\xi(x)$ и плотностью $f_\xi(x)$ и функция $y = \varphi(x)$ монотонно возрастает, $x = \varphi^{-1}(y)$ – обратная функция. Тогда

$$P(\eta < x) = P(\varphi(\xi) < x) = P(\xi < \varphi^{-1}(y)) = F_\xi(\varphi^{-1}(y));$$
$$f_\eta(x) = \left(F_\xi(\varphi^{-1}(y))\right)'_x.$$

Если $y = \varphi(x)$ – монотонно убывающая функция, то

$$1 - F'_\xi(\varphi^{-1}(y)),\; f_\eta(x) = -\left(F_\xi(\varphi^{-1}(y))\right)'_x.$$

Литература [5, С. 73-74; 9, С. 63-64; 13, С. 98-102].

Упражнения

1. Дискретная случайная величина ξ имеет закон распределения

ξ	–2	–1	0	1	2
$P_\xi(x)$	0,1	0,2	0,3	0,3	0,1

Построить законы распределения случайных величин $\eta = \xi^2 + 1$ и $\xi = |\xi|$.

2. Случайная величина распределена по нормальному закону с плотностью вероятности

$$P_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

Найти закон распределения обратной ей величины $\eta = \dfrac{1}{\xi}$.

3. Непрерывная случайная величина ξ имеет плотность $f_\xi(x)$. Найти плотность вероятности случайной величины $\eta = 3\xi$.

4. Случайная величина ξ распределена равномерно в интервале $\left[-\frac{\pi}{2}; \frac{\pi}{2}\right]$. Найти закон распределения случайной величины $\eta = \sin\xi$.

1.10. Числовые характеристики случайных величин

Основными характеристиками случайной величины являются математическое ожидание и дисперсия.

Математическое ожидание декретной случайной величины

Пусть на пространстве элементарных событий Ω задана дискретная случайная величина $\xi(\omega)$ ($\omega \in$) таблицей распределения

$\xi(\omega)$	a_i	a_1	a_2	...
$P(\xi(\omega))$	P_i	P_1	P_2	...

причем ряд $\sum_{\omega\in\Omega}\xi(\omega)P(\xi(\omega))$ сходится абсолютно, тогда число

$$\mathbf{M}\,\xi = \sum_{\omega\in\Omega}\xi(\omega)P(\xi(\omega))$$

называется ***математическим ожиданием дискретной случайной величины*** ξ. При абсолютной сходимости ряда $\sum_{\omega\in\Omega}\xi(\omega)P(\xi(\omega))$ можно показать абсолютную сходимость $\sum_i a_i P_i$ и равенство чисел $\sum_{\omega\in\Omega}\xi(\omega)P(\xi(\omega)) = \sum_i a_i P_i$.

И наоборот, из абсолютной сходимости $\sum_i a_i P_i$ следует абсолютная сходимость ряда $\sum_{\omega\in\Omega}\xi(\omega)P(\xi(\omega))$ и равенство чисел

$$\sum_i a_i P_i = \sum_{\omega\in\Omega}\xi(\omega)P(\xi(\omega)).$$

Поэтому правомочно следующее определение: ***математическим ожиданием дискретной случайной величины*** ξ называется сумма произведений всех ее возможных значений (a_1, a_2,...) на их вероятности (P_1, P_2,...) при условии абсолютной сходимости ряда $\sum_i a_i P_i$.

Литература [5, С. 85-89; 9, С. 108-110; 13, С. 105-113].

Математическое ожидание непрерывной случайной величины

Математическое ожидание **M** ξ ***непрерывной случайной величины*** называется интеграл от произведения ее значений x на плотность ее распределения $f(x)$.

$$\mathbf{M}\,\xi = \int_{-\infty}^{+\infty} x f(x)\,dx \qquad (1.16)$$

где несобственный интеграл предполагается абсолютно сходящимся.

Литература [5, С. 96-98; 9, С. 110-114; 13, С. 105-115].

Математическое ожидание функции от одной случайной величины

Можно показать справедливость формул вычисления для математического ожидания от функции случайной величины, при абсолютной сходимости соответствующего ряда и интеграла.

Пусть $\eta = \varphi(x)$ тогда

$$\mathbf{M}\varphi(\xi) = \sum_i \varphi(x_i)P_i \qquad (1.17)$$

где ξ – дискретная случайная величина; x_i – ее значения и ряд $\sum_i \varphi(x_i)P_i$ сходится абсолютно.

Если ξ непрерывная случайная величина, то

$$\mathbf{M}\varphi(\xi) = \int_{-\infty}^{+\infty} \varphi(x) f(x)dx \qquad (1.18)$$

где $f(x)$ – плотность распределения случайной величины ξ и интеграл сходится абсолютно.

Свойства математического ожидания

1. $\mathbf{M}c$ = с, где c – *const*.
2. $\mathbf{M}(c\,\xi) = c\mathbf{M}\,\xi$.
3. $\mathbf{M}(\xi \pm \eta) = \mathbf{M}\,\xi \pm \mathbf{M}\,\eta$, где ξ и η – любые случайные величины.
4. $\mathbf{M}(\xi \cdot \eta) = \mathbf{M}\xi \cdot \mathbf{M}\eta$, если ξ и η независимые случайные величины.

Случайные величины ξ и η называются ***независимыми***, если для любых x и y

$$P(\xi < x, \eta < y) = P(\xi < x) \cdot P(\eta < y).$$

Литература [5, С. 87-88; 9, С. 118-120; 13, С. 113-117].

Дисперсия и среднее квадратичное отклонение случайной величины

Дисперсией случайной величины $\xi = \xi\ (\omega)$, $(\omega \in \Omega)$ называется число

$$\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 \qquad (1.19)$$

т.е. дисперсия равна математическому ожиданию квадрата отклонения случайной величины от своего математического ожидания.

Число $\sqrt{\mathbf{D}\xi}$ называется ***среднее квадратичное отклонение*** случайной величины ξ.

Из определения дисперсии для дискретной величины имеем

$$\mathbf{D}\xi = \sum_i (x_i - \mathbf{M}\xi)^2 P_i,$$

где x_i – значение случайной величины ξ.

Для непрерывной величины ξ

$$\mathbf{D}\xi = \int_{-\infty}^{+\infty} (x - \mathbf{M}\xi)^2 f(x) dx$$

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

Для вычисления дисперсии удобна формула

$$\mathbf{D}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2.$$

Свойства дисперсии

1. $\mathbf{D}c$ = с, где c – *const*.
2. $\mathbf{D}\ (c\ \xi) = c^2\ \mathbf{D}\ \xi$, где c – *const*.
3. $\mathbf{D}\ (\xi \pm \eta) = \mathbf{D}\ \xi \pm \mathbf{D}\ \eta$, где ξ и η – независимые случайные величины.

Литература [5, С. 100-106; 9, С. 120-125; 13, С. 117-121].

Типовые примеры

1. Вычислить математическое ожидание, дисперсию индикатора события *A*.

Решение

Закон распределения I_A имеет вид

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

где *P* вероятность «появления» события *A* в одном испытании.

$$\mathbf{M}\, I_A = 0(1 - P) + 1 \cdot P = P,$$

$$\mathbf{D}\, I_A = \mathbf{M}\,(I_A - \mathbf{M}\,(I_A))^2 = (0 - P)^2\,(1 - P) + (1 - P)^2\, P =$$
$$= (1 - P)(P^2 + P - P^2) = P(1 - P),$$

или

$$\mathbf{D}\, I_A = p \cdot q$$

где $q = 1 - P$,

2. Вычислить **D**ξ и **M**(ξ), если случайная величина ξ распределена по биноминальному закону, т.е.

$$P(\xi = m) = C_n^m p^m q^{n-m},$$

где $n \geq m$ – целое положительное число; p – вероятность «появления» события *A* вероятность успеха в каждом единичном испытании; $q = 1 - p$ – «не появления» события *A* отсутствие успеха в каждом единичном испытании.

Решение

Представим случайную величину ξ числа успехов («появлений» события A) при n кратном повторении в виде суммы индикаторов успехов.

$$\xi = I(A_1) + I(A_2) + \ldots + I(A_n).$$

$I(A_k) = 1$ – в случае успеха при k испытании и 0 – в противном случае. Поэтому их сумма есть случайная величина ξ равная числу успехов при n кратном повторении испытания.

Из свойства (3) математического ожидания и результата решения примера 1 имеем

$$\mathbf{M}\xi = \sum_{k=1}^{n} \mathbf{M}\left(I_{A_k}\right) = nP \text{ (т.к. } \mathbf{M}\left(I_{A_k}\right) = P\text{)}.$$

Случайные величины $I(A_1)$, $I(A_2)$, ... независимые, поэтому по свойству (3) дисперсии $\mathbf{D}\xi = \sum_{k=1}^{n} \mathbf{D}\left(I_{A_k}\right)$, но $\mathbf{D}\left(I_{A_k}\right) = pq$, $k = 0,1,\ldots,n$.

Поэтому $\mathbf{D}\xi = npq$.

3. Вычислить математическое ожидание случайной величины ξ распределенной по закону Пуассона

$$P(\xi = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \; k = 0,1,\ldots,n.$$

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

т.к.

$$1 + \frac{\lambda}{1!} + \ldots + \frac{\lambda^k}{k!} + \ldots = e^{\lambda}.$$

$\mathbf{D}\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2$, вместе с тем $\mathbf{M}\xi^2 = \mathbf{M}\xi\,(\xi - 1) + \mathbf{M}\xi$.

Вычислим

$$\mathbf{M}\xi(\xi-1)=\sum_{k=0}^{\infty}k(k-1)\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\cdot\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}=\lambda^2.$$

Поэтому

$$\mathbf{M}\xi^2=\lambda^2+\lambda \text{ и } \mathbf{D}\xi=\lambda^2+\lambda-\lambda^2=\lambda.$$

4. Пусть случайная величина ξ распределена равномерно на $[a;\ b]$. Вычислить $\mathbf{M}(\xi)$ и $\mathbf{D}\xi$.

Решение

$$f(x)=\frac{1}{b-a},\quad x\in[a;b].$$

$$f(x)=0,\quad x\notin[a;b].$$

Поэтому

$$\mathbf{M}\xi=\int_a^b\frac{xdx}{b-a}=\frac{x^2}{2(b-a)}\Big|_a^b=\frac{b+a}{2}.$$

Для вычисления $\mathbf{D}\,\xi$ воспользуемся формулой $\mathbf{D}\,\xi=\mathbf{M}\,\xi^2-(\mathbf{M}\,\xi)^2$.

$$\mathbf{M}\xi^2=\int_a^b\frac{x^2dx}{b-a}=\frac{x^3}{3(b-a)}\Big|_a^b=\frac{b^2+ab+a^2}{3}.$$

$$\mathbf{D}\xi=\frac{b^2+ab+a^2}{3}-\frac{(a+b)^2}{4}=\frac{(b-a)^2}{12}.$$

5. Вычислить математическое ожидание и дисперсию случайной величины ξ распределенной по показательному закону, т.е.

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

Р е ш е н и е

По формуле

$$\mathbf{M}\xi = \int_0^{+\infty} x\lambda e^{-\lambda x}dx = \frac{1}{\lambda} \text{ (интеграл вычисляется по частям).}$$

$$\mathbf{D}\xi = \int_0^{+\infty} x^2\lambda e^{-\lambda x}dx - \frac{1}{\lambda^2} = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

$$\sigma_\xi = \sqrt{\mathbf{D}\xi} = \frac{1}{\lambda}.$$

Интеграл $\int_0^{+\infty} x^2\lambda e^{-\lambda x}dx$ вычисляется как и предыдущий по частям.

Методические указания

Следует обратить внимание на содержательный смысл числовых характеристик случайной величины.

Математическое ожидание называют средним значением случайной величины ξ, подчеркивая тем самым статистический смысл этого понятия.

Вместе с тем математическое ожидание называют центром распределения, центром группирования, центром рассеяния случайной величины ξ, когда хотят подчеркнуть, что математическое ожидание применяется в качестве характеристики положения случайной величины. Математическое ожидание имеет размерность одинаковую со случайной величиной ξ. Наряду с характеристиками положения случайной величины большую роль как в теории, так и в приложениях имеют значения рассеяния значений случайной величины около центра ее распределения.

В качестве меры рассеяния применяется среднее квадратичное отклонение $\sigma = \sqrt{\mathbf{D}\xi}$ случайной величины ξ. Причем размерность σ совпадает с размерностью случайной величины ξ.

Упражнения

1. Случайно берется одна кость домино. Найти математическое ожидание суммы очков на ее половинах ξ и η. Найти **M**(ξ + η).
2. Математическое ожидание и дисперсия случайной величины ξ равны, соответственно, 2 и 10. Найти математическое ожидание и дисперсию величины 2ξ + 5.
3. Плотность распределения ξ равна

$$f(x) = \begin{cases} \dfrac{c}{x}, & x \in [1;2], \\ 0, & x \notin [1;2]. \end{cases}$$

Найти *c*, **M**ξ, **D**ξ

4. Случайная величина ξ имеет плотность вероятности

$$f(x) = \begin{cases} \dfrac{\pi}{2}\cos^2 x, & при \quad |x| < \dfrac{\pi}{2}, \\ 0, & при \quad |x| > \dfrac{\pi}{2}. \end{cases}$$

Найти **M**ξ, **D**ξ.

1.11. Неравенства Чебышева и закон больших чисел

Для любой случайной величины ξ, имеющей конечную дисперсию и для любого числа ξ > 0

$$P(|\xi - \mathbf{M}\xi| \geq \xi) \leq \frac{\mathbf{D}\xi}{\varepsilon^2}.$$

Это неравенство можно записать в другом виде

$$P\left(\left|\xi - \mathbf{M}\xi\right| < \xi\right) \geq 1 - \frac{\mathbf{D}\xi}{\varepsilon^2}.$$

Эти два неравенства называются неравенствами Чебышева.

Теорема Чебышева

Если. ξ_1, ξ_2,...ξ_n последовательность случайных величин таких, что

1). они попарно независимы;
2). имеют конечные дисперсии, ограниченные одной и той же постоянной $c > 0$ то, како бы ни было $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n} M\xi_i\right| < \varepsilon\right) = 1.$$

Теорема Бернулли. Пусть μ_n – число успехов в n испытаниях Бернулли и p вероятность успеха в каждом отдельном испытании. Тогда для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\mu_n}{n} - p\right| < \varepsilon\right) = 1.$$

Литература [5, С. 115-119; 9, С. 146-154; 13, С. 130-133].

Типовые примеры

1. Используя неравенство Чебышева, оценить вероятность того, что любая случайная величина ξ отклонится от своего математического ожидания менее чем на три средних квадратичных отклонения этой величины.

Решение

Из условия задачи **D** $\xi = \sigma^2$, $\varepsilon = 3\sigma$, тогда из неравенства Чебышева имеем

$$P\left(\left|\xi - \mathbf{M}\xi\right| < 3\sigma\right) \geq 1 - \frac{\sigma^2}{9\sigma^2}, \text{ т.е.}$$

$$P\left(\left|\xi - \mathbf{M}\xi\right| < 3\sigma\right) \geq \frac{8}{9} \approx 0{,}89:$$

Полученное соотношение носит название «**правило трех сигм**».
Сравните с «правилом трех сигм» для нормального распределения (п.1.7).
2. Дискретная случайная величина ξ задана законом распределения

ξ	0,3	0,6
P	0,2	0,8

Используя неравенство Чебышева, оценить вероятность того, что

$$\left|\xi - \mathbf{M}\xi\right| < 0{,}2$$

Решение

Вычислим **M**ξ, $\mathbf{M}\xi^2$ и $\mathbf{D}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2$.

$\mathbf{M}\xi = 0{,}3 \cdot 0{,}2 + 0{,}6 \cdot 0{,}8 = 0{,}06 + 0{,}48 = 0{,}54$,

$\mathbf{M}\xi^2 = 0{,}09 – 0{,}2 + 0{,}36 \cdot 0{,}8 = 0{,}018 + 0{,}288 = 0{,}306$,

$\mathbf{D}\xi = 0{,}306 – 0{,}2916 = 0{,}0144$.

Из неравенства Чебышева получим

$$P\left(\left|\xi - 0{,}54\right| < 0{,}2\right) \geq 1 - \frac{0{,}0144}{0{,}04} = 0{,}64.$$

Методические указания

Неравенство Чебышева позволяет оценить вероятность уклонения случайной величины ξ от своего математического ожидания по известной $\mathbf{D}\xi$.

Это неравенство весьма полезно при различных теоретических исследованиях. Содержание теоремы Чебышева состоит в том, что среднее арифметическое случайных величин (попарно независимых и с ограниченной дисперсией) при достаточно большом n будет как угодно мало отличаться от числа

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{M}\xi_i .$$

Упражнения

1. Длина изготавливаемых изделий представляет случайную величину, среднее значение которой (математическое ожидание) равно 90 см. Дисперсия этой величины равна 0,0225. используя неравенство Чебышева, оценить вероятность того, что а) отклонение длины изготавливаемого изделия от ее среднего значения по абсолютной величине не превысит 0,4; б) длина изделия выразится числом, заключенным между 89,7 и 90,3.
2. Дана последовательность независимых случайных величин ξ_1, ξ_2, ... ξ_n. Случайная величина ξ (n = 1,2,...) может принимать только три значения: $\sqrt{n}$, 0, $\sqrt{n}$ с вероятностями, соответственно, $\frac{1}{n}, 1-\frac{2}{n}, \frac{1}{n}$. Применима ли к этой последовательности теорема Чебышева?

РАЗДЕЛ 2. ОСНОВНЫЕ ПОНЯТИЯ И МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

2.1. Генеральная совокупность и выборка. Вариационный ряд.

Совокупность объектов одной природы или совокупность значений некоторого признака объектов называется ***генеральной совокупностью***. Основной задачей математической статистики является определение вероятных характеристик генеральной совокупности: закона распределения, математического ожидания какого-либо признака, его дисперсии и др. как правило, полное исследование генеральной совокупности оказывается невозможным или нецелесообразно. Например, при оценке некоторого показателя качества продукции массового производства. Поэтому из генеральной совокупности случайным образом выбирают часть объектов, которая называется ***выборкой.*** Число объектов выборочной или генеральной совокупности называется ***объектом совокупности.***

По результатам обслуживания выборки делается вывод о вероятных характеристиках генеральной совокупности. При этом пользуются статистической вероятностью. Выборка называется ***представительной*** (***репрезентативной***), если: каждый элемент генеральной совокупности с одинаковой вероятностью имеет возможность попасть в выборку; все объекты выборки выбраны случайно. Различают ***повторные*** и ***бесповторные выборки***. В первом случае каждый выбранный объект после обследования возвращается в генеральную совокупность перед выбором следующего объекта. Во втором случае все выбранные объекты выборки в генеральную совокупность после обследования не возвращаются. Генеральная совокупность может быть разделена на группы и выборка осуществляется, например, по одному объекту из группы, серии.

Под генеральной совокупностью может пониматься и исследуемая случайная величина. При этом выборкой будет служить совокупность ее значений, полученных в результате экспериментов с постоянными условиями.

Измерения значения некоторого признака объекта выборки называется ***статистическим наблюдением.*** В общем случае измерения могут осуществляться и по нескольким признакам объектов выборки. Пусть ξ – случайная величина, представляющая некий признак генеральной совокупности; $F(x)$ – ее функция распределения x_1, x_2, … x_n – результаты независимых наблюдений случайной величины ξ. Эти результаты, записанные в порядке их получения образуют ***статистический ряд*** или выборку объема n для случайной величины ξ. Выборка будет наглядней, если все результаты упорядочить по возрастанию или убыванию. Однако в выборке некоторые значения могут повторяться несколько раз. Поэтому целесообразно результаты записать в виде таблицы – в первом столбце (строке) которого записываются значения x_i в порядке их возрастания, а во втором – числа n_i – частоты появления i-го значения. Такая таблица называется ***вариационным рядом***. Если выборка объема n содержит m различных значений и значение y_i $\left(i=\overline{1,m}\right)$ встречается n_i раз, то отношение $\omega_i = n_i / n$ называется относительной частотой результата y_i. Очевидно $\sum_{i=1}^{m} n_i = n$. $\sum_{i=1}^{m} \omega_i = 1$.

Если число m слишком велико или близко к n, то целесообразно составить так называемый ***интервальный вариационный ряд***.

При этом первый столбец (строка), указанный выше таблицы, заполняется интервалами значений генеральной совокупности. Для построения интервального вариационного ряда необходимо определить длину интервала, их число и соответственно сгруппировать результаты наблюдений. Для ориентировочной оценки числа интервалов k рекомендуется пользоваться формулой Стерджеса.

$$k \approx 1 + 3{,}322 \ln n$$

Тогда длина интервала (все одинаковой длины) определиться по формуле

$$h = \frac{x_{\max} - x_{\min}}{k}$$

где $x_{\max}$; $x_{\min}$ – соответственно максимальное и минимальное значения выборки. Эти значения должны войти соответственно в первый и последующий интервалы.

Графики вариационных рядов. Используется два вида графиков: ***полигон*** *и* ***гистограмму***. Если вариационный ряд составлен по значениям результатов наблюдений, то полигон строят из отрезков, соединяющие точки с координатами (x_i, ω_i). Гистограмма служит для изображения только интервального вариационного ряда.

При построении гистограммы над каждым интервалом (x_i, x_{i+1}), строят прямоугольник, высота которого пропорциональна значению ω_i или n_i.

2.2. Эмпирическая функция распределения

Генеральная совокупность имеет некую функцию распределения $F(x) = P(\xi < x)$, которая, как правило, неизвестна. На основании закона больших чисел можно найти эмпирическую функцию распределения $F^*(x)$.

Формула для определения $F^*(x)$ аналогична формуле функции распределения $F(x)$ дискретной случайной величины, в которой вместо вероятности p_i принята случайная величина значения x_i используется относительная частота n_i / n, т.е.

$$F^*(x) = \sum_{x_i < x} \frac{n_i}{n}. \tag{2.1}$$

Формула (2.1) используется и для интервального вариационного ряда. При этом в качестве представителя интервала $[x_i, x_{i+1}]$ берется его середина. Иначе, особенно если все значения случайной величины в каждом интервале распределены равномерно, в качестве представителя интервала берут его правый конец.

Выборочная средняя и дисперсия

В качестве выборочной средней используется среднее арифметическое выборки x_1, x_2, … x_n, объема n:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Если по выборке построен вариационный ряд, то

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} y_i n_i , \tag{2.2}$$

где y_1, y_2, … y_m, – m различных значений случайной величины ξ в выборке объема n; n_i $\left(i=\overline{1,m}\right)$ – частота появления y_i.

В случае интервального вариационного ряда в формуле (2.2) y_i есть среднее значение всех выборочных значений, вошедших в i-й інтервал.

Несмещенная, т.е. не содержащая систематической ошибки, выборочная дисперсия определяется по формуле

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right).$$

Для выборки, представленной вариационным рядом, используется формула

$$S^2=\frac{1}{n-1}\left(\sum_{i=1}^{m}n_i y_i^2-n\bar{x}^2\right). \tag{2.3}$$

Формула (2.3) применяется и для случая интервального вариационного ряда с замечаниями аналогичными в отношении формулы (2.2).

Литература [5, С. 120-125; 10, С. 181-187].

Типовые примеры

1. Для случайной величины X получен статистический ряд (выборка):

1, 9, 6, 7, 7, 3, 5, 6, 6, 2, 4, 7, 8, 0, 9, 7, 5, 3, 5, 2, 6, 5, 4, 6, 9, 3, 10, 4.
1, 6, 7, 2, 4, 5, 4, 6, 9, 4, 2, 3, 5, 2, 10, 7, 2, 4, 5, 8, 8, 4. Объем выборки n = 50.

Записать соответствующий данной выборке вариационный ряд и построить полигон частот выборки.

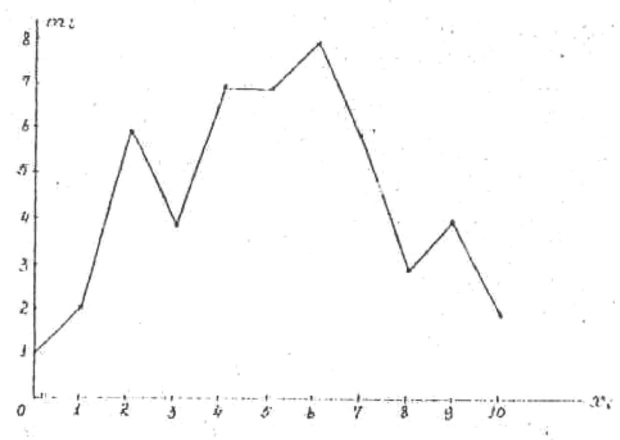
Решение

Выборка содержит 11 различных значений (m = 11):

0, 1, 2, 2, 4, 5, 6, 7, 8, 9, 10. Записывая их в таблицу с соответствующими частотами, получим вариационный ряд

y_i	0	1	2	3	4	5	6	7	8	9	10
n_i	1	2	6	4	7	7	8	6	3	4	2

Полигон частот имеет вид:



2. При исследовании непрерывной случайной, величины получена выборка:

7,59;	7,48;	7,46;	7,40;	7,24;	7,41;	7,34;	7,43;	7,38;	7,60;
7,26;	7,43;	7,37;	7,55;	7,42;	7,41;	7,30;	7,14;	7,42;	7,52;
7,46;	7,39;	7,35;	7,32;	7,18;	7,30;	7,54;	7,38;	7,37;	7,34;
7,50;	7,61;	7,42;	7,32;	7,36;	7,40;	7,67;	7,31;	7,40;	7,36;
7,28;	7,58;	7,38;	7,58;	7;26;	7,37;	7,28;	7,39	7,32;	7,20;
7,43;	7,34;	7,45;	7,33;	7,41;	7,33;	7,45;	7,31;	7,45;	7,39.

Объем выборки n = 60.

Требуется записать выборку в виде интервального вариационного ряда и построить гистограмму выборки.

Решение

Оценим число интервалов по формуле Стерджеса $1 + 3{,}32\lg 60 \approx 6{,}90$.

Примем число интервалов $r = 7$. Для данной выборки $x_{\min} = 7{,}14$; $x_{\max} = 7{,}61$; поэтому ширина интервала оценивается как $h = (7{,}61 - 7{,}14) : 7 = 0{,}067$. Поскольку исходные данные определены с точностью 0,01, нет смысла в более точном вычислений h. Округлив и несколько увеличив найденное значение (чтобы не уменьшить диапазон значений исследуемой случайной величины), примем $h = 0{,}08$. Если принять в качестве левого конца первого интервала $x_{н1} = 7{,}10$, то в семи интервалах разместится весь диапазон выборочных значений, так как правый конец последнего (седьмого) интервала будет

$$x_{вт} = x_{н1} + 7 \cdot h = 7{,}10 + 7 \cdot 0{,}08 = 7{,}66 > x_{\max} = 7{,}61.$$

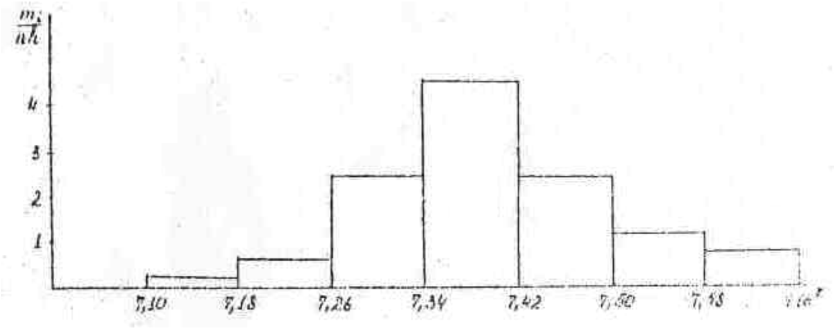
При этом $x_{\min}$ и $x_{\max}$ попадают примерно в середины соотвественно первого и последнего интервалов. Именно такое расположение интервалов обычно рекомендуется.

Составим таблицу в соответствии с принятыми значениями $x_{н1}$, h, и r.

Четвертая строка добавлена для размещения величин $\frac{m_i}{nh}$ необходимых для построения гистограммы.

Интервалы ($x_{н1}$ и $x_в$)	(7,10;7,18)	(7,18;7,26)	(7,26;7,34)	(7,34;7,42)	(7,42;7,50)	(7,50;7,58)	(7,58;7,69)
Средин-ные значения x_i	7,14	7,22	7,30	7,36	7,46	7,54	7,62
Частоты	1	3	12	22	12	6	4
$\frac{m_i}{nh}$	0,21	0,63	2,5	4,59	2,50	1,25	0,83

Выбрав подходящий масштаб на координатных осях, строим гистограмму.



3. Вычислить выборочное среднее и выборочную дисперсию для выборки примера 1.

Решение

Пользуясь формулой (2.2), получим

$$\overline{x} = \frac{1}{50}\left(0\cdot 1 + 2\cdot 1 + 6\cdot 2 + 4\cdot 3 + 7\cdot 4 + 7\cdot 5 + 8\cdot 6 + 6\cdot 7 + 3\cdot 8 + 4\cdot 9 + 2\cdot 10\right) =$$

$$= \frac{1}{50}\cdot 259 = 5{,}18$$

Аналогично по формуле (2.3) вычисляем выборочную дисперсию

$$S^2 = \frac{1}{49}(1\cdot 0 + 2\cdot 1 + 6\cdot 2^2 + 4\cdot 3^2 + 7\cdot 4^2 + 7\cdot 5^2 + 8\cdot 6^2 +$$

$$+ 6\cdot 7^2 + 3\cdot 8^2 + 4\cdot 9^2 + 2\cdot 10^2) = \frac{1}{19}\cdot 305{,}38 \approx 6{,}23.$$

Методические указания

Предполагается самостоятельно построить эмпирическую функцию распределения для выборки примера 1 и убедиться визуально в выполнении ее очевидных свойств.

К числу интервалов, определенных по формуле Стерджеса, следует относиться как к рекомендации.

В силу наглядности и простоты приведенных примеров, иллюстрирующих содержание данного подраздела, нет необходимости приводить на эту тему упражнения. Их придумать в достаточном количестве самостоятельно.

2.3. Точечная оценка неизвестных параметров по выборке

Предположим, что по значениям выборки объема n требуется оценить неизвестный параметр θ некой вероятной характеристики генеральной совокупности.

Точечной оценкой неизвестного параметра θ назовем произвольную функцию

$$\tilde{\theta}_n = \tilde{\theta}_n\left(X_1, X_2, ..., X_n\right).$$

Значения функции $\tilde{\theta}_n\left(x_1, x_2, ..., x_n\right)$ при конкретных в результате выборки значениях $X_1 = x_1$, $X_2 = x_2$, … $X_n = x_n$, будем рассматривать как приближенное значение параметра θ.

Укажем свойства оценок, которые обеспечивают их близость к оцениваемому параметру θ.

Оценка $\tilde{\theta}_n$ параметра θ называется ***несмещенной***, если $\mathbf{M}\tilde{\theta}_n = \theta$.

Оценка $\tilde{\theta}_n$ параметра θ называется ***состоятельной***, если при $n \to \infty$ $P\left(\left|\tilde{\theta}_n - \theta\right| < \varepsilon\right) \to 1$ для любого $\varepsilon > 0$.

Оценка называется ***эффективной***, если она имеет наименьшую возможную дисперсию.

Литература [5, С. 126-130; 10, С. 202-208].

Типовые примеры

1. Рассмотрим выборку $x_1, x_2, ..., x_n$ соответствующую случайной величине ξ с $F_\xi(x) = P(\xi \le x)$. Пусть нас интересует параметр $a = \mathbf{M}\xi$. Предположим, что нам известна величина $\mathbf{D}\xi = \sigma^2$ (или известно, что она существует и конечна). Обычно в качестве оценки параметра a выбирают среднее арифметическое $\overline{X}$ значений выборки, т.е. полагают

$$\tilde{a}_n = \overline{X} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Согласно определению выборки случайные величины $X_1, X_2, \ldots, X_n$ независимы и каждая имеет распределение совпадающее с ξ.

Поэтому $\mathbf{M}X_i = a$, $\mathbf{D}X_i = \sigma^2$.

Из определения $\tilde{a}_n$ и свойства линейности для математического ожидания и дисперсии, получим

$$\mathbf{M}\tilde{a}_n = \mathbf{M}X = \frac{1}{n}\left(\mathbf{M}X_1 + \mathbf{M}X_2 + \ldots + \mathbf{M}X_n\right) = \frac{na}{n} = a,$$

тем самым $\tilde{a}_n = \overline{X}$ есть несмещенная оценка математического ожидания случайной величины ξ.

Далее

$$\mathbf{D}\tilde{a}_n = \frac{1}{n^2}\left(\mathbf{D}X_1 + \mathbf{D}X_2 + \ldots + \mathbf{D}X_n\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Тогда на основании неравенства Чебышева получим

$$P\left(\left|\tilde{a}_n - a\right| < \varepsilon\right) \ge 1 - \frac{\mathbf{D}\tilde{a}_n}{\varepsilon^2} = 1 - \frac{\sigma^2}{n\varepsilon^2},$$

при $n \to \infty$ $P\left(\left|\tilde{a}_n - a\right| < \varepsilon\right) \to 1$, т.е. оценка $\tilde{a}_n = \overline{X}$ является состоятельной.

Если в качестве оценки параметра σ^2 взять $\sigma^2 = S^2$, где

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_1 - \overline{X}\right)^2,$$

то можно показать, что S^2 является несмещенной оценкой σ^2 и если предположить, что существует $\mathbf{M}\xi^4$, то при $n \to \infty$ $\mathbf{D}S^2 = 0\left(\frac{1}{n}\right)$, после чего, используя неравенство Чебышева получим состоятельность оценки S^2.

Методические указания

Следует обратить внимание, что свойство несмещенности означает, что оценка не имеет систематической ошибки.

Свойство состоятельности обеспечивает сближение оценки с измеряемым параметром при увеличении числа измерений.

Эффективная оценка имеет наименьшее рассеяние значений оценки в последовательных выборках около истинного значения оцениваемого параметра.

Упражнения

1. По данному распределению

ξ	1	2	3	4
n_i	20	15	10	5

Найти а) выборочную среднюю (среднее арифметическое); б) несмещенную оценку дисперсии S^2.

2. В результате пяти измерений длины стержня одним прибором (без систематических погрешностей) получена следующие результаты (в мм): $x_1 = 92$; $x_2 = 94$; $x_3 = 103$; $x_4 = 105$; $x_5 = 106$. Найти а) выборочную среднюю длины стержня; б) несмещенную оценку дисперсии ошибок прибора.

2.4. Интервальные оценки неизвестных параметров

Сущность интервальной оценки состоит в определении (по данным наблюдения, выборки) такого числового интервала, относительно которого, с заранее установленной вероятностью правильности суждения, можно заключить, что внутри этого интервала находится истинное значение оцениваемого параметра.

Наиболее часто применяется на практике вариант интервальной оценки, который носит название доверительного интервала.

Определение. ***Доверительным интервалом*** $\left[\theta_n^1; \theta_n^2\right]$ для параметра θ называется такой интервал, относительно которого можно с заранее определенной, близкой к единице вероятностью утверждать, что он содержит неизвестное значение параметра.

Литература [5, С. 133-136; 10, С. 233-236; 13, С. 176-180].

Типовые примеры

1. Пусть требуется построить доверительный интервал для арифметической средней

$$\overline{X} = \frac{1}{n}\left(X_1 + X_2 + ... + X_n\right)$$

в генеральной совокупности, при этом предположим, что $X_i \;\; i = \overline{1,n}$ распределены нормально с параметрами (a, σ), тогда

$$\mathbf{M}\overline{X} = \mathbf{M}X_i = a, \quad \mathbf{D}\overline{X} = \frac{\sigma^2}{n}.$$

Поэтому случайная величина $\frac{\overline{X} - a}{\sigma}\sqrt{n}$ имеет нормальное распределение с параметрами (0;1) и по таблице 1 можно найти число для которого выполняется соотношение

$$P\left(-u_\gamma \le \frac{\overline{X} - a}{\sigma}\sqrt{n} \le u_\lambda\right) = \gamma. \qquad (2.4)$$

Например, при γ = 0,95 из таблицы имеем u_γ = 1,96, а при γ = 0,99 – u_γ = 2,58.

Соотношение (2.4) можно переписать

$$P\left(\overline{X} - u_\gamma \frac{\sigma}{\sqrt{n}} \le a \le \overline{X} + u_\lambda \frac{\sigma}{\sqrt{n}}\right) = \gamma .$$

Если наблюдаемая случайная величина ξ имеет нормальное распределение с параметрами (*a*, σ), но ее среднее квадратичное отклонение σ неизвестно, то можно построить доверительный интервал для *a*, опираясь на распределение Стьюдента. Из таблиц распределения Стьюдента с (*n*-1) степенями свободы (значение *k*) находим величину t_γ для которой справедливо равенство

$$P\left(\overline{X} - t_\gamma \frac{S}{\sqrt{n-1}} \le a \le \overline{X} + t_\lambda \frac{S}{\sqrt{n-1}}\right) = \gamma . \qquad (2.5)$$

где $S^2 = \frac{1}{n-1}\sum_{i-1}^{n}\left(X_i - \overline{X}\right)^2$.

2. Новый сорт пшеницы был высеян на 10 опытных участках. После сбора урожая получены следующие данные об урожайности в ц на 1 га:

Номер участка	1	2	3	4	5	6	7	8	9	10
Урожай	25,4	28	20,1	27,4	25,6	23,9	24,8	26,4	27,0	25,4

Требуется построить доверительный интервал для среднего значения *a* урожайности нового сорта пшеницы, приняв γ = 0,95.

Решение

Т.к. уровень урожайности обычно имеет нормальное распределение, а параметр σ нам неизвестен, то воспользуемся соотношением (2.5). Для нашего случая $\overline{X}$ = 25,4; *S* = 2,12. Из таблицы 2 распределения Стьюдента находим, что при k = 1 степенях свободы t_γ =2,26, поэтому

$$t_\gamma \cdot \frac{S}{\sqrt{n-1}} = 2{,}26 \cdot \frac{2{,}12}{\sqrt{9}} = 1{,}6 .$$

Откуда доверительный интервал запишется

25,4 – 1,6 $\le a \le$ 25,4 + 1,6 или 23,8 $\le a \le$ 27.

3. Случайная величина ξ распределена нормально, σ_ξ = 1. Найти доверительный интервал для a при $\overline{X} = 2$; n = 16; γ = 0,99.

Решение

По таблице 1 находим u_γ = 2,58, тогда $u_\gamma \frac{\sigma}{\sqrt{n}}$ для нашего случая равно

$2{,}58 \cdot \frac{1}{4}$ = 0,65. Поэтому доверительный интервал имеет вид

2 – 0,65 $\le a \le$ 2 + 0,65 или 1,35 $\le a \le$ 2,65.

Литература [5, С. 133-136; 10, С. 233-236].

Методические указания

Обратим внимание на следующий практически важный аспект.

Обозначим половину длины доверительного интервала Δ. В теории доверительных интервалов все три величины γ, Δ, n обычно тесно взаимосвязаны и приняв определенное значение двух из них, мы можем вычислить величину третьего элемента. На практике наиболее важна зависимость n от γ и Δ. Избрав значение γ доверительной вероятности, Δ «половинки» длины доверительного интервала, обычно вычисляют, каков должен быть объем выборки. Однако иногда встречаются случаи, когда заданы величины n, γ и вычисляется соответствующая им величина Δ.

Упражнения

1. Станок-автомат штампует валики. По выборке объема n = 100 вычислено выборочное среднее диаметров изготовленных валиков. Найти с доверительной вероятностью 0,95 точность, с которой выборочное среднее оценивает математическое ожидание диаметров изготавливаемых валиков, зная, что их среднее квадратичное отклонение $\sigma = 2$ мм.

2. Из генеральной совокупности извлечена выборка объема $n = 10$.

ξ	-2	1	2	3	4	5
n_i	2	1	2	2	2	1

Оценить с доверительной вероятностью 0,95 $\mathbf{M}\xi = a$ при неизвестном σ нормально распределенного признака генеральной совокупности по выборочному среднему (арифметической средней) с помощью доверительного интервала.

2.5. Проверка статистических гипотез

Статистической гипотезой называют такую гипотезу, которая относится или к виду или к параметрам распределения случайной переменной и которую можно проверить статистически, т.е. опираясь на результаты наблюдений в случайной выборке. Проверку статистических гипотез выполняют с помощью статистического критерия. Он устанавливает правило: при каких результатах случайной выборки проверяемая гипотеза принимается или отвергается?

Проверяемую гипотезу обозначим H_0 и будем ее называть основной, противопоставляя ее множеству $\{H_\lambda\}$ альтернативных гипотез (λ может пробегать любое множество). Далее имеется опыт, результат которого x есть элемент некоторого множества X, называемого выборочным пространством.

Связь между гипотезами H_λ и результатом опыта x состоит в следующем. Предполагается, что в X выделен достаточно широкий класс подмножеств $A \subseteq X$ таких, что при любой верной гипотезе H_λ определены вероятности $P(x \in A / H_\lambda)$, т.е. вероятности того, что x попадет в A, если на самом деле верна гипотеза H_λ.

Формально процесс проверки гипотезы H_0 состоит в том, что выбирается (определяется) некоторое множество S (называемое критическим для гипотезы H_0) и проводится опыт. Если результат опыта $x \in S$, то гипотеза отвергается. Для построения множества S необходимо выбрать заранее «уровень значимости», т.е. положительное число $\alpha > 0$ и потребовать, что бы $P\left(x \in S/H_0 \right) \le \alpha$.

Если мы дорожим гипотезой H_0 и не хотим ее отвергнуть понапрасну, то α должно быть малым, обычно выбирают одно из значений 0,05; 0,01 или 0,001.

Для определения множества S используют специально подобранную случайную величину, точное или приближенное распределение которой известно, ее и называют статистическим критерием. Из сказанного следует, то $P\left(x \in S/H_0 \right) \le \alpha$ есть вероятность отвергнуть гипотезу H_0 (когда она верна). Такая ошибка называется ошибкой первого рода. Из $P\left(x \in S/H_0 \right) \le \alpha$ следует, что вероятность ошибки первого рода не превосходит уровня значимости α. Кроме ошибки первого рода возможна еще ошибка второго рода, которая состоит в том, что гипотеза H_0 не отвергается, когда она на самом деле не верна, а верна одна из гипотез H_λ. Вероятность этой ошибки есть

$$\beta(\lambda) = P\left(\overline{x \in} S/H_\lambda \right) = 1 - P\left(x \in S/H_\lambda \right) = P\left(\bar{S}/H_\lambda \right).$$

Функция $1 - \beta(\lambda) = P\left(x \in S/H_\lambda \right)$, равна вероятности отвергнуть гипотезу H_0, если на самом деле верна гипотеза H_λ, называется функцией мощности статистического критерия. Поэтому желательно, чтобы среди S, удовлетворяющих условию $P\left(x \in S/H_\lambda \right) \le \alpha$ выбрать такое $\tilde{S}$, при котором

$$P\left(\tilde{S}/H_\lambda \right) = \max P\left(S/H_\lambda \right),$$

где max берется по всем S удовлетворяющих условию $P\left(x \in S/H_0 \right) \le \alpha$, но

$\tilde{S} = S_\lambda$ будет зависеть от λ. Если S_λ не зависит от λ, то $\tilde{S} = S_\lambda$ есть наилучший критерий, если он известен, то он описан во всех учебниках математической статистики. На практике исследователю задачу выбора $\tilde{S}$ приходится решать самому.

Типовые примеры

Статистическая проверка гипотезы о законе распределения. Критерий χ^2 (хи квадрат).

Постановка задачи

Пусть гипотеза H_0 состоит в том, что выборка $x_1, x_2, \ldots x_n$ объема n соответствует случайной величине X с функцией распределения $F_0(x)$.

Требуется выполнить статистическую проверку этой гипотезы.

Решение

Для статистической проверки гипотезы H_0 используется критерий χ^2, суть которой следующая [7,10]. Разобьем числовую ось на k непересекающихся интервалов $h_1, h_2, \ldots, h_k$, $\bigcup\limits_{i=1}^{k} h_i = (-\infty; +\infty)$. Положим $P_i = P(x \in h_i)$, $i = 1,2,\ldots,k$.

Эти вероятности вычисляются по известной функции распределения $F_0(x)$. Обозначим через n_i число элементов x_i выборки попавших в интервал h_i. За меру отклонения распределения выборки от предполагаемого распределения $F_0(x)$ принимается величина

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - nP_i)^2}{nP_i}, \quad n = \sum_{i=1}^{k} n_i .$$

К. Пирсон доказал, что в случае справедливости гипотезы H_0 распределение случайной величины χ^2 при $n \to \infty$ сходится к распределению χ^2 с $(k - 1)$ степенями свободы.

Зададим уровень значимости α (α = 0,05 или α = 0,01 или 0,001). По таблице распределения χ^2 с (k – 1) степенью свободы (см. табл.3) находим такое значение

$$\chi^2_{\alpha,k-1}, \text{ что } P\left(\chi^2 > \chi^2_{\alpha,k-1}\right) = \alpha .$$

Далее вычислим по имеющейся выборке значение

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i} .$$

Если при этом окажется $\chi^2 > \chi^2_{\alpha,k-1}$, то такое отклонение значимо и мы с уровнем значимости α отвергаем гипотезу H_0 как не согласующуюся с опытными данными. Если же в результате вычислений получим $\chi^2 \le \chi^2_{\alpha,k-1}$, то это означает, что данная выборка согласуется с гипотезой H_0.

На практике интервалы $h_i\left(i = 1,2,...,k\right)$ выбирают так, чтобы число элементов выборки в каждом из них было не очень маленьким, например, $np_i \ge 7$. Если распределение $F_0(x)$ зависит от неизвестных параметров, то вероятности $P_i = P\left(x \in h_i\right)$ вычисляют, заменяя параметры их оценками. В этом случае число (k - 1) должно быть уменьшено на число неизвестных параметров.

Числовой пример 1.

Пусть задана выборка *A*.

2, 4, 2, 4, 3, 3, 3, 2, 0, 6, 1, 2, 3, 2, 2, 4, 3, 3, 5, 1
0, 2, 4, 3, 2, 2, 3, 3, 1, 3, 3, 3, 1, 1, 2, 3, 1, 4, 3, 1
7, 4, 3, 4, 2, 3, 2, 3, 3, 1, 4, 3, 4, 4, 5, 3, 4, 2, 4, 5
3, 6, 4, 1, 3, 2, 4, 1, 3, 1, 0, 0, 4, 6, 4, 7, 4, 1, 3.

N = 79. Начало первого интервала: 0. Длина интервала 7.

Требуется с уровнем значимости α =0,01, проверить гипотезу о распределении Пуассона генеральной совокупности по данным выборки *A*.

Решение

За меру отклонения распределения выборки от предлагаемого распределения Пуассона принимается величина χ^2. Параметром распределения Пуассона является $\lambda = \mathbf{M}\mathrm{X} = \mathbf{D}\mathrm{X}$, его значение неизвестно, поэтому по данным выборки вычислим оценку $\mathbf{M}\mathrm{X}$ по формуле:

$$\tilde{x} = \frac{\sum_{i=1}^{m} \frac{x_i - c}{k}}{n} k + c$$

и оценку $\mathbf{D}\mathrm{X}$ по формуле:

$$S^2 = \frac{n}{n-1} \overline{S}^2,$$

где

$$\overline{S}^2 = \frac{\sum_{i=1}^{m} \left(\frac{x_i - c}{k} \right)^2 n_i}{n} k^2 - \left(\overline{x} - c \right)^2.$$

$\overline{S}^2$ – выборочная дисперсия;

S^2 – несмещенная оценка дисперсии.

При этом

x_i – варианты случайной величины,

n_i – соответствующие частоты,

m – количество вариантов,

n – объем выборки,

k – шаг таблицы,

с – произвольное число (ложный нуль, для простоты вычислений следует выбрать в качестве ложного нуля вариант, имеющий максимальную частоту). Вычисления $\overline{x}$, S^2 выполним при помощи таблицы 1. В нашем случае $c = 3$, $k = 1$.

Таблица 1

x_i	n_i	$\frac{x_i - c}{k}$	$\frac{x_i - c}{k} n_i$	$\left(\frac{x_i - c}{k}\right)^2$	$\left(\frac{x_i - c}{k}\right)^2 n_i$
1	**2**	**3**	**4**	**5**	**6**
0	4	–3	–12	9	36
1	13	–2	–26	4	52
2	14	–1	–14	1	14
3	24	0	0	0	0
4	16	1	16	1	16
5	3	2	6	4	12
6	3	3	9	9	27
7	2	4	8	16	32
$\sum$	79	–	–13	–	189

Получим

$$\bar{x} = \frac{-13}{79} \cdot 1 + 3 = -0{,}16 + 3 = 2{,}84\,.$$

$$\bar{S}^2 = \frac{189}{79} \cdot 1 - (2{,}84 - 3)^2 = 2{,}3924 - 0{,}0256 = 2{,}3668\,.$$

Вычислим несмещенную оценку дисперсии.

$$S^2 = \frac{79}{78} \cdot 2{,}3668 = 2{,}3971\,.$$

Оценки среднего значения случайной величины и дисперсии близки по значению, но не равны.

В таблице распределения Пуассона ближайшими к ним значениями λ являются 2 и 3.

Проверим гипотезу при $\lambda = 3$.

Частоты последних значений вариационного ряда малы, поэтому объединим их в один интервал (≥ 5). Количество интервалов k = 6. Следовательно статистика χ^2 имеет распределение с числом степеней свободы

6 – 1 – 1 = 4 (один неизвестный параметр распределения (λ)).

Критическая область определяется неравенством

$$P\left(\chi^2 > \chi^2_{0,01,4}\right) = 0,01.$$

Выполним необходимые вычисления при помощи таблицы 2. Вероятность P_i получим по таблице распределения Пуассона.

Таблица 2

x_i	n_i	P_i	nP_i	$n_i - nP_i$	$(n_i - nP_i)^2$	$\frac{(n_i - nP_i)^2}{nP_i}$
1	**2**	**3**	**4**	**5**	**6**	**7**
0	4	0,0498	3,9	0,1	0,01	0,0026
1	13	0,1494	11,8	1,2	1,44	0,1220
2	14	0,2240	17,7	–3,7	13,69	0,7734
3	24	0,2240	17,7	6,3	36,69	0,2424
4	16	0,1680	13,3	2,7	7,29	0,5481
≥ 5	8	0,1847	14,6	–6,6	43,56	2,9836
$\sum$	79	0,9999	79	–	–	6,6721

По таблице распределения χ^2 получаем

$$\chi^2_{0,01;4} = 13,277$$

т.е. критическая область определяется неравенством $\chi^2 \geq 13,277$. Значение статистики по выборочным данным 6,6721, т.е. $\chi^2_{выб}$ не принадлежит критической области, поэтому нет оснований отвергать гипотезу о распределении генеральной совокупности по закону Пуассона.

Решение числового примера 2.

Пусть задана выборка *B*.

65	71	67	68	73	68	72	68	67	70	78	74	79	65	72
65	71	70	69	69	76	71	63	77	75	70	74	65	71	68
74	69	69	71	66	69	73	74	80	69	73	76	69	69	67
67	74	68	60	74	70	66	70	68	64	75	78	71	70	69
73	75	74	80	72	72	69	69	71	70	73	65	66	67	69
71	70	72	72	76	73	64	74	71	76	68	69	75	76	73
74	78	66	72	75	69	68	63	70	70	78	76	73	73	67
71	66	66	69	72	71	71	68	72	69	73	73	66	72	73
70	69	74	69	72	74	70	74	72	76	71	66	62	69	74
76	74	69	75	64	71	76	68	68	78	71	71	68	67	74
68	81	72	72	68	71	71	71	69	61	74	66	70	72	65
67	73	78	71	73	75	73	71	72	68	67	69	69	77	63
71	74	67	69	68	74	69	67	74	66	74	74	69	75	70
73	63	77	75	74										

N = 200. Начало первого интервала 59. Длина интервала 2.

Требуется с уровнем значимости $\alpha = 0{,}1$ проверить гипотезу о распределении генеральной совокупности по нормальному закону, используя данные выборки В.

За меру отклонения распределения генеральной совокупности от предполагаемого нормального распределения принимается величина χ^2.

Параметры нормального распределения a и σ,
где

$$a = \mathbf{M}(x), \text{ а } \quad \sigma = \sqrt{\mathbf{D}(x)}.$$

Для получения оценки параметров нормального распределения составим таблицу 3.

Таблица 3

Интервал	Середина интервал	n_i	$\frac{x_i - c}{k}$	$\left(\frac{x_i - c}{k}\right)^2$	$\frac{x_i - c}{k} n_i$	$\left(\frac{x_i - c}{k}\right)^2 n_i$
1	**2**	**3**	**4**	**5**	**6**	**7**
59-61	60	1	–5	25	–5	25
61-63	62	2	–4	16	–8	32
63-65	64	7	–3	9	–21	63
65-67	66	16	–2	4	–32	64
67-69	68	27	–1	1	–27	27
69-71	70	40	0	0	0	0
71-73	72	38	1	1	38	38
73-75	74	38	2	4	76	152
75-77	76	18	3	9	54	162
77-79	78	9	4	16	36	144
79-81	80	3	5	25	15	75
81-83	82	1	6	36	6	36
$\sum$	–				132	818

Причем $c = 70$, $k = 2$. Тогда

$$\bar{x} = \frac{132}{200} \cdot 2 + 70 = 1{,}32 + 70 = 71{,}32.$$

$$\bar{S}^2 = \frac{818}{200} \cdot 4 - (71{,}32 - 70)^2 = 16{,}36 - 1{,}7424 = 14{,}6176.$$

$\bar{x}$, $\bar{S}^2$ оценки $\mathbf{M}(x)$, $\mathbf{D}(x)$, которые вычислялись по формулам:

$$\bar{x} = \frac{\sum_{i=1}^{m} \frac{x_i - c}{k}}{n} k + c,$$

$$\bar{S}^2 = \frac{\sum_{i=1}^{m} \left(\frac{x_i - c}{k}\right)^2 n_i}{n} k^2 - (\bar{x} - c)^2,$$

где m – количество интервалов оценка σ есть $\overline{S} = \sqrt{14{,}6176} = 3{,}82$.

В вариационном ряде (см. табл. 3) частоты первых и последних интервалов малы, поэтому объединим их с соседними интервалами. В результате остается m = 8 интервалов.

Статистика χ^2 имеем 8 – 2 – 1 = 5 степеней свободы (два параметра (σ, a) были неизвестны).

Оформим вычисления для проверки гипотезы о нормальном распределении генеральной совокупности в виде таблицы 4.

Таблица 4

Интервалы	n_i	P_i	nP_i	$n_i - nP_i$	$(n_i - nP_i)$	$\frac{(n_i - nP_i)^2}{nP_i}$
1	**2**	**3**	**4**	**5**	**6**	**7**
$-\infty$-65	10	0,04846	9,7	0,3	0,03	0,0093
65-67	16	0,08078	16,2	–0,2	0,04	0,0025
67-69	27	0,14501	29,0	–2,0	4,00	0,1379
69-71	40	0,19387	38,8	1,2	1,44	0,0371
71-73	38	0,20552	41,1	–3,1	9,61	0,2338
73-75	38	0,16034	32,1	5,9	34,81	1,0844
75-77	18	0,09921	19,8	–1,8	3,24	0,1636
77-∞	13	0,06681	13,4	–0,4	0,16	0,0119
$\sum$	200	1,0000	200,1			1,6805

Вычисление P_i выполнялось по формуле

$$P_i = \Phi\left(\frac{x_i - 71{,}32}{3.83}\right) - \Phi\left(\frac{x_{i-1} - 71{,}32}{3.83}\right),$$

где $x_{i\text{-}1}$ – начало i-го интервала, x_i – конец i-го интервала, $\overline{i = 1{,}8}$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$$ – функция распределения нормального закона.

Критическая область определяется неравенством

$$P\left(\chi^2 > \chi^2_{0,1;5}\right) = 0{,}1.$$

По таблице распределения $\chi^2_{0,1;5} = 9{,}236$ т.е. критическая область определяется неравенством $\chi^2 \geq 9{,}236$.

Значение статистики по выборочным данным 1,6805, т.е. $\chi^2_{выб}$ не принадлежит критической области, поэтому нет оснований отвергать гипотезу о распределении генеральной совокупности по нормальному закону.

Методические указания

Понятие статистической значимости (уровня значимости) и понятие функции мощности статистического критерия, или, что тоже самое понятия ошибок первого и второго рода относятся к основным понятиям математической статистики. Причем понятие ошибки второго рода требует указать, кроме основной гипотезы H_0, множество альтернативных гипотез $\{H_\lambda,\ \lambda \neq 0\}$ и набор множеств $A \subset \mathrm{X}$, где Х – выборочное пространство и возможность вычисления $P\left(x \in A/H_\alpha\right)$. В некоторых случаях выполнить это затруднительно. В таких случаях учитывается лишь уровень значимости.

Литература [7, С. 63-65, С. 69; 1, С. 48; 10, С. 244-254].

Упражнения

Задача 1.

В книге «Математические методы статистики» Г. Крамер приводит следующие данные распределения толщины (ξ) 12000 бобов.

Номер интервала	1	2	3	4	5	6	7	8	9
частоты	32	103	239	624	1187	1650	1883	1930	1638
Номер интервала	10	11	12	13	14	15	16		
частоты	1130	737	417	221	110	57	32		

В таблице первый интервал – значения ξ меньше 7,00 мм,
второй 7,00–7,25
третий 7,25–7,5 и т.д.

Проверить согласуется ли толщина бобов в выборке, предполагая, что этот признак распределен в интервале совокупности по нормальному закону. Уровень значимости принять равным 0,01.

Задача 2.

Через равные промежутки времени в тонком слое раствора золота регистрировалось число частиц золота, попавших в поле зрения микроскопа. В результате наблюдений было получено следующее эмпирическое распределение.

ξ_i	0	1	2	3	4	5	6	7
n_i	112	168	130	68	32	5	1	1

В первой строке приведено ξ_i число частиц золота, а во второй строке – частота n_i, т.е. число интервалов времени, в течение которых в поле зрения попало ровно ξ_i частиц; Объем выборки

$$n = \sum n_i = 517 .$$

Проверить, согласуется ли эмпирическое распределение частиц золота в растворе, с законом распределения Пуассона, принять за уровень значимости $\alpha = 0,05$.

2.6. Дисперсионный анализ

Характер распределения случайной величины определяется различными факторами. В связи с этим возникает вопрос об оценке влияния факторов на значения случайной величины. Числовой характеристикой рассеяния значений случайной величины около ее центра группирования (математическое ожидание), как известно, является дисперсия. Поэтому естественно при исследовании влияния факторов на рассеяние значения случайной величины выделить в дисперсии независимые слагаемые (факторная дисперсия), каждое из которых характеризовало бы влияние одного из факторов, на рассеяние. Затем провести сравнительный анализ факторной дисперсии с остаточной, т.е. той, которая характеризует влияние на рассеяние значений случайной величины остальных факторов. В этом состоит суть дисперсионного анализа.

Проиллюстрируем применение дисперсионного анализа на следующем относительно простом случае.

Предположим, что имеется m нормальных (распределенных по нормальному закону) совокупностей с одинаковыми дисперсиями σ^2, тогда как средние значения в этих совокупностях могут быть различными.

Образуем выборку X следующим образом: Из каждой такой совокупности взяты случайные выборки, состоящие из n_j (j = 1, 2, …, m) независимых наблюдений. Необходимо проверить гипотезу H_0, что средние значения во всех m совокупностях одинаковы.

Пусть x_{ij} обозначаем i-ое наблюдение случайной величины x при образовании выборки из j-ой совокупности. $\bar{x}_j$ есть средняя в выборке из j-ой совокупности. $\bar{x}$ – общая средняя по всем выборкам, т.е.

$$\bar{x}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij} ,$$

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{m} \bar{x}_j \bar{n}_j ,$$

где $n = n_1 + n_2 + \ldots n_m$.

В качестве критерия используем дисперсионное отношение

$$F = \frac{S_1^2}{S_2^2},$$

где S_1^2 – факторная дисперсия,

S_2^2 – остаточная дисперсия.

С учетом явных выражений S_1^2 и S_2^2 получим:

$$F = \frac{\frac{1}{m-1}\sum_{j=1}^{m}\left(\bar{x}_j - \bar{x}\right)^2 n_j}{\frac{1}{n-m}\sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}_j\right)^2}$$

Случайная величина F имеет F – распределение (Фишера-Снедекора) со степенями свободы $(m – 1)$ и $(n – m)$ (табл. 4).

Для проверки гипотезы H_0 (о равенстве средних значений) строим критическое множество, определяемое условием:

$$P\left(F > f_\alpha\right) = \alpha,$$

где α уровень значимости; число f_α – определяется по таблице 4 распределения. Если значение статистики F больше числа f_α, то гипотезу H_0 о равенстве средних значений на всех уровнях фактора отвергаем, т.е. считаем влияние исследуемого фактора значимым. В противном случае принимаем гипотезу H_0, т.е. считаем, что значимость влияния фактора не установлена.

Литература [7, С. 83-86, 10, С. 294-306].

Типовые примеры

Пример 1.

Для проверки того, влияет ли день недели на объем торгового оборота в каком либо универсальном магазине, было проведено обследование в выбранные наугад дни. Полученные данные в таблице.

Оборот универсального магазина в разные дни недели (в условных единицах).

Понедельник	Вторник	Среда	Четверг	Пятница	Суббота
1,2	1,1	1,5	1,6	1,2	1,5
1,0	1,4	1,0	1,4	1,2	2,0
0,8	1,1	1,1	1,5	1,8	1,9

Нетрудно убедиться $\bar{x}_1 = 1$; $\bar{x}_2 = \bar{x}_3 = 1{,}2$; $\bar{x}_4 = 1{,}5$; $\bar{x}_6 = 1{,}4$; $\bar{x}_6 = 1{,}8$. Общая средняя $\bar{x} = 1{,}35$.

В рассматриваемом случае $n_j = 3$, j = 1,6; n = 18 и m = 6.

Числитель статистики равен 0,2730, а ее знаменатель – 0,0567. Получим числовое значение характеристики 4,81.

Примем уровень значимости α = 0,05. Из таблицы 4 находим, что при этом уровне существенности для 5 и 12 степеней свободы $f_\alpha = 3{,}11$. Полученное ранее числовое значение характеристики 4,81, таким образом, оно принадлежит области, и, следовательно, гипотезу, что день недели не влияет на объем торгового оборота следует, отвергнуть. Риск ошибки первого рода составляет лишь 5%.

Замечание.

При практических вычислениях величины F удобно поступить следующим образом.

Найти сумму квадратов отклонений элементов выборки относительно общего среднего арифметического

$$Q = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}\right)^2 .$$

Эту сумму можно разделить на два независимых слагаемых

$$Q_1 = \sum_{j=1}^{m}\frac{1}{n_j}\left(\sum_{i=1}^{n_j} x_{ij}\right)^2 - \frac{1}{n}\left(\sum_{j=1}^{m}\sum_{i=1}^{n_j} x_{ij}\right)^2$$

$$Q_2 = \sum_{j=1}^{m}\sum_{i=1}^{n_j}\left(x_{ij} - \overline{x}_j\right)^2,$$

причем $Q = Q_1 + Q_2$.

После чего вычислить соответствующие дисперсии

$$S^2 = \frac{Q}{n-1}, \quad S_1^2 = \frac{Q_1}{m-1}, \quad S_2^2 = \frac{Q_2}{n-m}$$

Две последние дисперсии используют при вычислении статистики F. Обычно по выборке вычисляют Q и Q_1, а Q_2, определяют по формуле

$$Q_2 = Q - Q_1.$$

Далее, для упрощения вычислений из каждого наблюдавшегося значения x_{ij} вычитаем одно и тоже число C, которое целесообразно выбрать близким к общему среднему (например: C – целая часть числа $\left(\overline{x} + 0{,}5\right)$), введем новую переменную $y_{ij} = x_{ij} - C$. Путем элементарных преобразований можно получить более удобные формулы для вычисления сумм Q и Q_1.

$$Q = \sum_{j=1}^{m} P_j - \frac{1}{n}\left(\sum_{j=1}^{m} T_j\right)^2,$$

$$Q_1 = \sum_{j=1}^{m} \frac{1}{n_j}\left(T_j\right)^2 - \frac{1}{n}\left(\sum_{j=1}^{m} T_j\right)^2,$$

где $P_j = \sum_{i=1}^{n_j} y_{ij}^2$, $\quad T_j = \sum_{i=1}^{n_j} y_{ij}$.

Числовой пример 2.

По выборке E:

	F_1	F_2	F_3	F_4	F_5	F_6	F_7
1	75	104	96	92	76	92	89
2	86	89	88	89	89	87	85
3		92	105		90	88	93
4		90	90		77	82	
5		81	91		75	90	
6						86	

Проверить с уровнем значимости α = 0,01 гипотезу о равенстве средних значений

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7.$$

Для выборки E имеет $m = 7$ и $n_1 = n_4 = 2$, $n_2 = n_3 = n_5 = 5$, $n_6 = 6$, $n_7 = 3$, откуда $n_7 = 28$.

Для вычисления суммы по группам, общей суммы и общего среднего $\bar{x}$ составим таблицу 1.

Таблица 1

№ испытания	Уровни фактора							
	F_1	F_2	F_3	F_4	F_5	F_6	F_7	$\sum$
1	75	104	96	92	76	92	89	
2	86	89	88	89	89	87	85	
3		92	105		90	88	93	
4		90	90		77	82		
5		81	91		75	90		
6						86		
$\sum$	161	456	470	181	407	525	267	2467

$$\bar{x} = \frac{2467}{28} = 88{,}1$$

Положим C = 88 и составим таблицу 2

Таблица 2

i \ j	F_1		F_2		F_3		F_4		F_5		F_6		F_7		$\sum$
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	y_{i4}	y_{i4}^2	y_{i5}	y_{i5}^2	y_{i6}	y_{i6}^2	y_{i7}	y_{i7}^2	
1	–13	169	16	256	8	64	4	16	–12	144	4	16	1	1	
2	–2	4	1	1	0	0	1	1	1	1	–1	1	–3	9	
3			4	16	17	289			2	4	0	0	5	25	
4			2	4	2	4			–11	121	–6	36			
5			–7	49	3	9			–13	169	2	4			
6											–2	4			
T_j	–15		16		30		5		–33		–3				3
T_j^2/n_j	112,5		51,2		180		12,5		217,8		1,5				578,5
P_j		173		326		366		17		439		61		35	1417

Вычислим $Q = \sum_{j=1}^{m} P_j - \frac{1}{n}\left(\sum_{j=1}^{m} T_j\right)^2$.

Для нашего случая $Q = 1417 - \frac{9}{28} \approx 1416{,}7$;

$$Q_1 = \sum_{j=1}^{m} \frac{1}{n_j} T_j^2 - \frac{1}{n}\left(\sum_{j=1}^{m} T_j\right)^2.$$

$$Q_1 = 578{,}5 - \frac{9}{28} \approx 578{,}2;$$

$$Q_2 = 1416{,}7 - 578{,}2 = 838{,}5.$$

Далее $S_1^2 = \frac{578{,}2}{7-1} = 96{,}4$; $S_2^2 = \frac{838{,}5}{28-7} = 39{,}9$;

$$F = \frac{96{,}4}{39{,}9} = 2{,}42.$$

По степеням свободы $m - 1 = 6$, $n - m = 21$ и уровню значимости $\alpha = 0{,}01$ по таблице 4 F распределения определяем критическую область $F > 3{,}81$. Значение статистики не входит в критическую область и поэтому нет основания отвергать гипотезу H_0 о равенстве средних значений по уровням, т.е. влияние фактора незначимо.

Пример 3.

По выборке *Д* при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о равенстве средних значений

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

Выборка *Д*

	F_1	F_2	F_3	F_4
1	–19	–31	–35	–31
2	–28	–33	–32	–27
3	–39	–35	–26	–28
4	–36	–25	–35	–35
5	–44	–28	–30	–40
6	–39	–31	–17	–41

Для выборки *Д* имеем $m = 4$, $n_j = 6$, $j = \overline{1,4}$, $n = 24$.

Таблица 3

№ испытания	Уровни фактора			
	F_1	F_2	F_3	F_4
1	–19	–31	–35	–31
2	–28	–33	–32	–27
3	–39	–35	–26	–28
4	–36	–25	–35	–35
5	–44	–28	–30	–40
6	–39	–31	–17	–41
$\bar{x}_j$	–34,2	–30,5	–29,2	–32,0

Вычислим групповые средние (табл.3) и общее среднее

$$\bar{x} = \frac{-4{,}2 - 0{,}5 + 0{,}8 - 2.0}{4} - 30 = -31{,}5.$$

Положим $C = -31$.

Таблица 4

$i \backslash j$	F_1		F_2		F_3		F_4		$\sum$
	y_{i1}	y_{i1}^2	y_{i2}	y_{i2}^2	y_{i3}	y_{i3}^2	y_{i4}	y_{i4}^2	
1	12	144	0	0	–4	16	0	0	
2	3	9	–2	4	–1	1	4	16	
3	–8	64	–4	16	5	25	3	9	
4	–5	25	6	36	–4	16	–4	16	
5	–13	169	3	9	1	1	–9	81	
6	–8	64	0	0	14	196	0	0	
T_j	–19		3		11		–6		–11
T_j^2/n_j	60,2		1,5		20,2		6		87,9
P_j		475		65		255		122	917

Вычислим $Q = 917 - \frac{121}{24} \approx 917 - 5 = 912$;

$$Q_1 = 87{,}9 - \frac{121}{24} \approx 87{,}9 - 5 = 82{,}9;$$

$$Q_2 = 912 - 82{,}9 = 829{,}1$$

$$S_1^2 = \frac{82{,}9}{4-1} = 27{,}63;$$

$$S_2^2 = \frac{829{,}1}{24} = 41{,}46;$$

$$F = \frac{27{,}63}{41{,}46} = 0{,}67$$

По таблице 4 распределения статистики F при степенях свободы $m - 1 = 3$, $n - m = 20$ и уровню значимости $\alpha = 0{,}05$ находим критическую область $F > 3{,}10$.

Значение статистики не входит в критическую область, поэтому нет оснований отвергать гипотезу H_0 о равенстве средних значений по уровням, т.е. влияние фактора незначимо.

Методические указания

Если факторная дисперсия окажется меньше остаточной, то уже отсюда следует справедливость гипотезы о равенстве групповых средних а, значит, нет надобности прибегать к критерию *F*.

Если нет уверенности в справедливости предположения о равенстве дисперсий рассматриваемых *m* совокупностей, то это предположение следует предварительно проверить.

Необходимо понимать, что если существенность расхождений между средними не обнаружена и нулевая гипотеза получила подтверждение, следует помнить что это сделано лишь по наличному опытному материалу и, главное, при той группировке материала, которой мы придерживались в данном случае.

Упражнения

1). По выборке *А* при уровне значимости $\alpha = 0{,}05$ проверить гипотезу о равенстве средних значений

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

Выборка *Д*.

	F_1	F_2	F_3	F_4
1	48	45	41	49
2	33	41	41	46
3	53	49	34	41
4	43	42	38	41
5	38	43	50	47
6	47	41	45	47

2). По выборке *B* проверить с уровнем значимости $\alpha = 0{,}01$ гипотезу о равенстве средних значений

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7.$$

Выборка *B*

	F_1	F_2	F_3	F_4	F_5	F_6
1	89	92	91	107	103	113
2	106	105	101	99	104	109
3	88	94		76	74	89
4	102	92		101	97	99
5	107	100		101	103	92
6	89			96		100
7	101					

2.7. Элементы корреляции.

Между случайными величинами могут существовать зависимости (связи) различные по своей природе. С одной из них функциональной связью мы ознакомились ранее (1.9).

Но между случайными величинами может существовать и связь другого рода, которая называется стохастической (вероятностной). Последняя возникает тогда, когда между случайными величинами имеются сбщие случайные факторы наряду с другими неодинаковыми. Например, если случайная величина ξ представляет некоторую функцию от случайных величин

X_1, X_2, Y_1, Y_2,:

$$\xi = f\,(X_1, X_2, Y_1, Y_2,),$$

а η представляет функцию от тех же случайных величин X_1, X_2 и других Z_1, Z_2:

$$\eta = \varphi\,(X_1, X_2, Z_1, Z_2),$$

то величины ξ и η будут между собой стохастически связаны. Изучение такой связи между случайными величинами занимается математическая теория корреляции, которая позволяет по результатам массовых наблюдений составить прогноз значения одной случайной величины в предположении, что другие случайные величины примут определенные значения с указанием вероятности того, что ошибка прогноза не выйдет за определенные границы. Результат массовых наблюдений (выборка) оформляется в виде корреляционной таблицы. Например для 2$^{\underline{x}}$ случайных величин X, Y она составляется следующим образом.

Варианты или интервалы значений одной случайной величины X записывают в первую строку, а варианты или интервалы значений другой случайной величины Y записывают в первый столбец. По каждой паре значений (x; y) в клетку, которая находится на пересечении значений x; y записывают частоту $n_{x,y}$. Далее суммируем $n_{x,y}$ по строкам и столбцам и находим частоты n_x и n_y. Объем выборки n можно получить с помощью одной из сумм

$$n = \sum_x n_x\ ,\ \ n = \sum_y n_y \ \text{ или }\ n = \sum_{xy} n_{xy}\ .$$

В корреляционной таблице могут остаться пустые клетки, это означает, что в выборке отсутствуют соответствующие сочетания значений x и y.

Инструментом измерения статистической зависимости между случайными величинами служит функция регрессии. Для понимания дальнейшего понадобится понятие условного распределения одной случайной величины по отношению к другой. Это понятие в некотором смысле аналогично понятию условной вероятности.

Определение: Условным распределением переменной Y при условии $X = x$ называется распределение случайной переменной Y в предположении, что случайная переменная X приняла значение x.

Точно так же, условным распределением случайной переменной X при условии $Y = y$ называется распределение случайной переменной X в предположении, что случайная переменная Y приняла значение y.

Для указанных распределений можно, очевидно, вычислить математическое ожидание $\mathbf{M}(Y/X = x)$ и $\mathbf{M}(X/Y = y)$.

Определение: Условное математическое ожидание $\mathbf{M}(Y/X = x)$ случайной переменной Y, трактуемое как функция x называется функцией регрессии случайной переменной Y относительно случайной переменной X. Точно так же, условное математическое ожидание $\mathbf{M}(X/Y = y)$ случайной переменной X, рассматриваемая как функция y, называется функцией регрессии случайной переменной X относительно Y.

Т.е.

$$f(x) = \mathbf{M}(Y/X = x)$$

функция регрессии случайной переменной Y относительно X,

$$g(y) = \mathbf{M}(X/Y = y)$$

есть функция регрессии случайной переменной X, относительно Y.

Обычно конкретный вид функции регрессии неизвестен и определяется по двумерной выборке. Для чего на координатном поле в виде точек изображают пары чисел: значения одной случайной величины и условные средние арифметические другой случайные величины, используя следующие формулы

$$\bar{x}_{Y=y} = \frac{1}{n_y} \sum_{i=1}^{n} y x_i n_{x_i y} ,$$

$$\bar{y}_{X=x} = \frac{1}{n_x} \sum_{i=1}^{n} x y_i n_{x_i y} ,$$

причем необходимые данные берут из корреляционной таблицы случайных переменных X, Y.

После чего, соединяя точки плавной линией, получаем график эмпирической регрессии.

Кроме графика эмпирической регрессии, в качестве числовой меры степени зависимости между случайными переменными X, Y вводится понятие коэффициента корреляции.

Определение: Коэффициентом корреляции r_{XY} переменных X и Y называют отношение математического ожидания произведения отклонений этих величин к произведению средних квадратичных отклонений этих величин:

$$r_{XY} = \frac{\mathbf{M}\left(\left(X - \mathbf{M}X\right)\left(Y - \mathbf{M}Y\right)\right)}{\sigma_x \cdot \sigma_y}.$$

Коэффициент корреляции имеет свойства:

1). Если случайные величины X и Y независимы, то коэффициент корреляции r_{XY} равен нулю. Противоположное утверждение неверно.

2). Коэффициент корреляции принимает значения в промежутке [–1;1].

3). Если между случайными величинами X и Y существует линейная функциональная зависимость, т.е.

$$Y = aX + b,$$

то

$$r_{XY} = \begin{cases} +1, & a > 0. \\ -1, & a < 0. \end{cases}$$

и наоборот, т.е. $r_{XY}^2 = 1$, то имеется линейная функциональная зависимость

$$Y = aX + b,$$

4). Если $r_{XY} \approx 1$, то это означает, что уравнение линейной зависимости $Y = aX + b$, хотя точно не соблюдается, но в приближении имеет место.

Если $|r_{XY}| < 0{,}5$, то обычно не используют линейную функцию как приближенную функциональную зависимость между случайными величинами X и Y.

В этом случае можно попробовать найти приближение функциональной зависимости в виде нелинейной функции. Если же функция регрессии линейна, т.е.

$$y=\alpha_1 x+\alpha_0, \quad x=\beta_1 y+\beta_0.$$

Коэффициент α_0, α_1, β_0, β_1 определяют по методу наименьших квадратов и получают формулы:

$$\alpha_0=\overline{y}-\alpha_1\overline{x}, \quad \alpha_1=r\cdot\frac{\overline{S}_Y}{\overline{S}_X}.$$

$$\beta_0=\overline{x}-\beta_1\overline{y}, \quad \beta_1=r\frac{\overline{S}_X}{\overline{S}_Y},$$

где

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i, \quad \overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i,$$

$$\overline{xy}=\frac{1}{n}\sum_{i=1}^{n}x_i y_i, \quad r=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{\overline{S}_x\cdot\overline{S}_y},$$

$$\overline{S}_x=\sqrt{\overline{S}_x^2}, \quad \overline{S}_y=\sqrt{\overline{S}_y^2},$$

$$\overline{S}_x^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\left(\overline{x}\right)^2, \quad \overline{S}_y^2=\frac{1}{n}\sum_{i=1}^{n}y_i^2-\left(\overline{y}\right)^2$$

Литература: [7, С. 76-83, 10, С. 313-333].

Типовые примеры

Пусть требуется по данным корреляционной таблицы один вычислить выборочный коэффициент корреляции.

Вычисления значительно упрощается, если перейти к «условным» вариантам

$$u_i=\frac{x_i-c_1}{h_1} \text{ и } v_j=\frac{y_j-c_2}{h_2}$$

где c_1, c_2, – ложный нуль,

h_1, h_2, – шаг, т.е. разность между двумя соседними первоначальными вариантами.

Частоты $n_{xy} = n_{uv}$, $n_x = n_u$, $n_y = n_v$.

В этом случае выборочный коэффициент корреляции вычисляется по формуле:

$$r_в = \left(\sum n_{uv} uv - n\bar{u}\bar{v} \right) / n\sigma_u \sigma_v .$$

Для вычисления $\sum n_{uv} uv$ используются формулы

$$\sum n_{uv} uv = \sum vU ,$$

где $U = \sum n_{uv} u$,

$$\sum n_{uv} uv = \sum vV ,$$

где $V = \sum n_{uv} v$.

Для контроля целесообразно выполнить расчеты по обоим формулам и сравнить результаты; их совпадение подтверждает правильность выполненных вычислений.

Таблица 1

Y	X						
	10	20	30	40	50	60	n_y
15	5	7	–	–	–	–	12
25	–	20	23	–	–	о–	43
35	–	–	30	47	2	–	49
45	–	–	10	11	20	6	47
55	–	–	–	9	7	3	19
n_x	5	27	63	67	29	9	n = 200

Решение

Перейдем к условным вариантам $u_i = \frac{(x_i - c_1)}{h_1}$, $v_j = \frac{(y_j - c_2)}{h_2}$. В качестве ложного нуля c_1 возьмем вариант x = 40, а c_2 = 35, каждая из которых расположена примерно в середине вариационного ряда. h_1 равен разности между двумя соседними вариантами: h_1 =20 – 10 = 10, аналогично h_2 =25 – 15 = = 10.

Получим формулы $u_i = \frac{(x_i - 40)}{10}$, $v_j = \frac{(y_j - 35)}{10}$.

Составим корреляционную таблицу 3 в условных вариантах.

Таблица 2

V	U						
	–3	–2	–1	0	1	2	n_v
–2	5	7	–	–	–	–	12
–1	–	20	23	–	–	–	43
0	–	–	30	47	2	–	79
1	–	–	10	11	20	6	47
2	–	–	–	9	7	3	19
n_u	5	27	63	67	29	9	n = 200

Для вычисления искомой суммы $\sum n_{uv}uv$ составим расчетную таблицу 3. пояснения к составлению таблицы 3.

1). В каждой клетке, в которой частота $n_{uv} \neq 0$ записываем в правом верхнем углу произведение частоты n_{uv} на варианту u. Например, в правых верхних углах клеток первой строки записаны произведения 5(–3) = –15; 7 · (–2) = –14.

2). Складываем все числа, помещенные в правых верхних углах клеток одной строки и их сумму запишем в клетку этой же строки столбца *U*.

3). Умножаем варианту *V* на *U* и полученное произведение записываем в последнюю клетку той же строки, т.е. в клетку столбца *VU*.

Например, в первой строке таблицы *V* = –2, *U* = –29, следовательно, *V* · *U* = (–2) · (–29) = 58.

4). Складывая все числа столбца $V \cdot U$, получим сумму $\sum_{v} vU$, которая равна искомой сумме $\sum n_{uv}uv$. Например для таблицы 3 имеем $\sum_{v} vU$ = 169; тем самым искомая сумма $\sum n_{uv}uv$ = 169.

Таблица 3

V	U						$U = \sum n_{uv}u$	VU
	–3	–2	–1	0	1	2		
–2	5	7	–	–	–	–	–29	58
–1	–	20	23	–	–	–	–63	63
0	–	–	30	47		–	–28	0
1	–	–	10	11	20	6	22	22
2	–	–	–	9	7	3	13	26
$V = \sum n_{uv}v$	–10	–34	–13	29	34	12		$\sum_{v} vU = 169$
UV	30	68	13	0	34	24	$\sum_{v} uV = 169$	контроль

Для контроля аналогичные вычисления выполняем по столбцам: произведения $n_{uv}v$ записываем в левый нижний угол клетки, которая содержит частоту $n_{uv} \neq 0$; все числа помещенные в левых нижних углах клеток одного столбца складывают и их сумму записывают в строку V; далее умножают каждую варианту u на V и результат записывают в клетках последней строки.

После вычисления $\sum n_{uv}uv$ вычислим выборочный коэффициент корреляции.

Для чего по данным таблицы 3 вычислим $\overline{u}$ и $\overline{v}$, исходя из определения средней, а $\overline{\sigma}_u$ и $\overline{\sigma}_v$ по формулам

$$\overline{\sigma}_u = \sqrt{\overline{u^2} - (\overline{u})^2}, \quad \overline{\sigma}_v = \sqrt{\overline{v^2} - (\overline{v})^2}.$$

$$\overline{u} = \frac{\left(\sum n_u u\right)}{n} = \frac{\left(5\cdot(-3) + 27\cdot(-2) + 63(-1) + 29\cdot 1 + 9\cdot 1\right)}{200} = -0{,}425$$

$$\overline{v} = \frac{\left(\sum n_v v\right)}{n} = \frac{\left(12\cdot(-2) + 43(-1) + 47\cdot 1 + 19\cdot 2\right)}{200} = 0{,}09$$

$$\overline{u^2} = \frac{\left(\sum n_u u^2\right)}{n} = \frac{\left(5\cdot 9 + 27\cdot 4 + 63\cdot 1 + 29\cdot 1 + 9\cdot 4\right)}{200} = 1{,}405$$

$$\overline{\sigma_u} = \sqrt{\overline{u^2} - \left(\overline{u}\right)^2} = \sqrt{1{,}405 - (0{,}425)^2} = 1{,}106.$$

Аналогично получили $\overline{\sigma_v} = 1{,}209$. Вычислим искомый выборочный коэффициент корреляции, учитывая, что ранее уже вычислена сумма $\sum n_{uv}uv = 169$.

$$r_в = \left(\sum n_{uv}uv - n\overline{u}\overline{v}\right)\left(n\sigma_{\overline{u}}\sigma_{\overline{v}}\right) = \left(169 - 200(-0{,}425)\cdot 0{,}09\right)/\left(200\cdot 1{,}106\cdot 1{,}209\right) = 0{,}603$$

Итак $r_в = 0{,}603$.

Пример 2

Найти выборочное уравнение прямой линии регрессии Y на X по данным таблицы 2 предыдущего примера.

Решение

Искомое уравнение в общем виде запишется

$$\overline{y}_x - \overline{y} = r_в \frac{\overline{\sigma}_y}{\overline{\sigma}_x}\left(x - \overline{x}\right).$$

Для получения $\overline{\sigma}_x$, $\overline{\sigma}_y$, $\overline{x}$, $\overline{y}$ воспользуемся формулами

$$\overline{\sigma}_x = h_1\overline{\sigma}_u,\quad \overline{\sigma}_y = h_{21}\overline{\sigma}_v,\quad \overline{x} = \overline{u}h_1 + c_1,\quad \overline{y} = \overline{v}h_2 + c_2,$$

т.к. $\overline{u}$, $\overline{v}$, $\overline{\sigma}_u$, $\overline{\sigma}_v$ были найдены правильно при вычислении $r_в$, то

$$\overline{x} = \overline{u}\cdot h_1 + c_1 = -0{,}425\cdot 10 + 40 = 35{,}75.$$

$$\overline{y} = \overline{v}h_2 + c_2 = 0{,}09\cdot 10 + 35 = 35{,}9.$$

$$\overline{\sigma}_x = \overline{\sigma}_u h_1 = 1{,}106\cdot 10 = 11{,}06;$$

$$\overline{\sigma}_y = \overline{\sigma}_v h_2 = 1{,}209\cdot 10 = 12{,}09.$$

Подставим найденные величины в уравнение прямой линии регрессии *Y* на *X* получим

$$\overline{y}_x - 35{,}9 = 0{,}603\frac{12{,}09}{11{,}06}\left(x - 35{,}75\right),$$

или окончательно

$$\overline{y}_x = 0{,}603x + 12{,}34.$$

Методические указания

Из определения функции регрессии следует, что она выражает математическое ожидание переменной *Y* (или *X*), когда другая переменная *X* (*Y*) примет значение *x* (*y*). Поэтому ее можно использовать для прогноза величины одной из случайных переменных, если известно значение другой переменной. Порядок точности такого прогноза например переменной Y определяется $\overline{\sigma}_{y/X=x}$ условного распределения Y. $\overline{\sigma}_{y/X=x}$ показывает насколько значения Y будут отклоняться в среднем от значений функции регрессии, когда $X = x$. Точно так же $\overline{\sigma}_{x=/Y=y}$ показывает насколько значения X будут отклоняться в среднем от значений функции регрессии, тогда $Y = y$. Среднее квадратические отклонения условных распределений могут иметь постоянную величину, но могут и зависеть от значения принимаемого другой случайной величиной. В первом случае условное распределение называется гомоскедастичным, т.е. имеющим постоянную условную дисперсию, а во втором – гетероскедастичным, т.е. имеющим дисперсию, существенно зависящую от другой случайной величины.

Литература: [7, С. 76-83, 1, 45; 10, С. 313-333].

Упражнения

1). Дана таблица результатов наблюдений:

X	*2*	4	6	8	10
Y	3,5	6	7	6	7,5

Найти выборочный коэффициент корреляции.

2). Через ρ_1 обозначен предел текучести стали, через ρ_2 предел прочности стали; X – процентное содержание углерода в стали. В результате 79 опытов получена корреляционная таблица величины $Y = \frac{\rho_1}{\rho_2}$ и X :

Y \ X	0,5	0,6	0,7	0,8
0,5	0	2	0	8
0,6	0	4	2	9
0,7	2	12	3	1
0,8	21	14	0	0
0,9	1	0	0	0

Целые числа приведенные в таблице, показывают в каком количестве опытов наблюдалось значение (x, y), так например число 14 показывает, что значение (0,6; 0,8) наблюдалось в 14 опытах. Требуется определить выборочный коэффициент корреляции и уравнение прямой линии регрессии X на Y.

2.8. Понятие случайного процесса. Некоторые типы случайных процессов

Пусть имеется случайная величина $\mathrm{X}(t)$ зависящая от параметра t (например времени) так, что при каждом фиксированном значении t (из некоторой области изменения T) это будет случайная величина с определенным (зависящим от параметра t) законом распределения вероятностей. Пусть, кроме того, закон распределения случайной величины зависит от тех числовых значений, которые фактически приняли случайные величины $\mathrm{X}(t_v)$ для какого либо набора значений t_v, меньших t («моменты времени, более ранние чем t»). Тогда $\mathrm{X}(t)$ называется случайной функцией аргумента t, или случайным процессом. Пусть, осуществляя некоторый эксперимент, мы отмечаем для каждого момента t значение, фактически принятое процессом $\mathrm{X}(t)$ в этот момент. Тогда мы получим неслучайную функцию $x_p(t)$, называемую реализацией случайного процесса $\mathrm{X}(t)$ и описывающую одно из возможных течений этого процесса, а именно то, которое наблюдалось в данном эксперименте. Если представить себе независимое повторение в неизменных условиях неограниченного множества таких экспериментов, то мы получим множество реализаций $\{x_p(t)\}$ нашего процесса. Таким образом, случайный процесс можно рассматривать как совокупность всех его реализаций, т.е. всех возможных путей течения этого процесса. Пусть *t* фиксировано. Тогда закон распределения случайной величины $\mathrm{X}(t)$ – сечения этого процесса – называется одномерным законом распределения процесса $\mathrm{X}(t)$. Одномерный закон распределения задается одномерной функцией распределения деления

$$F_1(x;t) = P(\mathrm{X}(t) < x)$$

или одномерной плотностью

$$f_1(x;t) = \frac{dF_1(x;t)}{dx}.$$

Для процессов с дискретными распределениями всех сечений этот одномерный закон распределения задается, как обычно, таблицей всех возможных значений данного сечения с их вероятностями

$$x_1(t), \dots, x_n(t), \dots$$

$$P_1(t), \dots, P_n(t), \dots$$

$$P_i(t) = P(\mathrm{X}(t) = x_i(t)), \quad \sum_i P_i(t) = 1.$$

Пусть $t_1, t_2 \dots, t_n$ – набор значений аргумента *t*. Образуем случайный вектор $(\mathrm{X}(t_1), \mathrm{X}(t_2), \dots, \mathrm{X}(t_n))$ компонентами которого являются сечения процесса $\mathrm{X}(t)$, соответствующие указанным значениям t_i. Закон распределения этого *n* мерного вектора называется *n* – мерным законом распределения процесса $\mathrm{X}(t)$.

Обычно *n* – мерный закон распределения процесса задается *n* – мерной функцией распределения

$$F_n(x_1, x_2, \dots, x_n,; t_1, t_2, \dots, t_n,) =$$
$$= P(\mathrm{X}(t_1) < x_1, \mathrm{X}(t_2) < x_2, \dots, \mathrm{X}(t_n) < x_n),$$

или *n* – мерной плотностью

$$f_n(x_1, x_2, \dots, x_n,; t_1, t_2, \dots, t_n,) = \\ = \frac{dF_n(x_1, x_2, \dots, x_n,; t_1, t_2, \dots, t_n,)}{dx_1 dx_2 \dots dx_n}.$$

Очевидно n – мерная функция распределения и n – мерная плотность случайного процесса $\mathrm{X}(t)$ зависит от n – параметров – моментов $t_1, t_2, \dots, t_n$. Для полной вероятностной характеристики произвольного случайного процесса $\mathrm{X}(t)$ необходимо задание всех n – мерных законов распределения этого процесса, т.е. для n = 1, 2, … , n при каждом n – для любого набора моментов $t_1, t_2, \dots, t_n$.

Указанный метод описания случайных процессов универсален и позволяет в принципе выяснить все особенности поведения процесса во времени.

Однако этот способ очень громоздок. Именно поэтому для получения более глубоких результатов приходится идти по иному пути: выделять важные частные типы случайных процессов и для них искать аналитический аппарат, приспособленный к расчетами построения математических моделей изучаемых явлений.

Цепи Маркова

Среди различных типов, классов случайных процессов особое значение приобрели **марковские процессы**, названные так в честь выдающегося русского математика конца XIX-начала XX столетия А.А. Маркова. Марков ввел в рассмотрение и первый стал систематически изучать свойства так называемых цепных зависимостей, явившихся основой для построения понятия и теории марковских случайных процессов. Предположим, что процесс $\mathrm{X}(t)$ обладает следующим свойством. Для любых множеств времени t_0 и t, $t_0 < t$, вероятность перейти из состояния x_0 в момент времени t_0 в состояние x (или в одно из состояний) принадлежащих некоторому множеству А) в момент t зависит только от t_0, x_0, t, и x, (или А), и дополнительное значение состояний, в которых был процесс в предшествующие моменты времени ее не изменяет.

Для таких процессов вся история их развития как бы концентрируется в момент t_0 состоянии x_0 и только через x_0 влияет на последующее его развитие. Именно такие процессы и называются **марковскими.** Другими словами для марковских процессов однозначно определяется лишь вероятность перехода в то или иное состояние за заданный промежуток времени.

Иначе: Пусть некоторая физическая система находится в одном из состояний E_1, E_2, … E_n, … В фиксированные моменты времени t_1, t_2, … t_n, … система под влиянием случайных факторов может переходить из одного состояния в другое, причем в любой момент времени t_n вероятность системе оказаться в наперед заданном состоянии E_j, вполне определяется тем состоянием, в котором она находилась непосредственно перед скачком, и не зависит от всех остальных состояний, в которых эта система находилась до момента t_{n-1} тогда говорят, что поведение этой системы описывается (или управляется) простой цепью Маркова (или простым марковским процессом.

$$P_{ij}^{(n)} = P\left(E_j, t_n \big/ E_i, t_{n-1} \right)$$

вероятность перехода из состояния E_i (в момент t_{n-1}) в состояние E_j в момент t_n называется переходной вероятностью этой цепи.

Цепь Маркова задается матрицей перехода π, элементами которой являются переходные вероятности $P_{ij}^{(n)}$ и вероятностями всех состояний системы в начальный момент $t_0 = 0$.

Эти вероятности P_{01}, P_{02}, … P_{0k}, … называются начальными вероятностями состояний системы.

Если все переходные вероятности $P_{ij}^{(n)}$ не зависят от времени (т.е. $P_{ij}^{(n)} = P_{ij},\ n = 1,3,...$), то цепь Маркова называется однородной. Тогда матрица π будет иметь вид

$$\pi = \begin{pmatrix} P_{11} & P_{12} & ... & P_{1k} & ... \\ P_{21} & P_{22} & ... & P_{2k} & ... \\ P_{k1} & P_{k2} & ... & P_{kk} & ... \\ - & - & - & - & - \end{pmatrix}, \quad \sum_j P_{ij} = 1$$

при любом фиксированном i. Обозначим через $P_{ij}(m)$ вероятность системе, управляемой однородной цепью Маркова, из состояния E_i через m «шагов» оказаться в состоянии E_j. Матрица перехода через m шагов

$$\pi = \begin{pmatrix} P_{11}(m) & P_{12}(m) & \ldots & P_{1k}(m) & \ldots \\ P_{21}(m) & P_{22}(m) & \ldots & P_{2k}(m) & \ldots \\ - & - & - & - & - \end{pmatrix},$$

Тогда $\pi_m = \pi^m$. Вероятность $P_{ij}(m)$ удовлетворяет уравнению Маркова

$$P_{ij}(m) = \sum_l P_{il}(S) P_{lj}(m - S),$$

где S – может принимать любое целое значение из отрезка [0;m].

Безусловная вероятность $P_j(n) = \sum_i P_{0i} P_{ij}(n)$ называется абсолютной вероятностью системе оказаться в момент t_n в состоянии E_j.

Простейший поток событий

При изучении многих практически важных или же интересных в познавательном отношении процессов (прибытия судов в морской порт, отказов в работе сложного устройства, замены перегоревших лампочек, регистрации моментов распада атомов радиоактивного вещество и т.д.) были выявлены общие закономерности, которые позволили объединить их в один класс случайных процессов, который носит название «Простейший поток событий».

Предположим, что изучаемый поток обладает следующими тремя свойствами:

1). **Стационарностью.** Под этим понимается следующее; для любой группы конечного числа непересекающихся интервалов времени вероятность появления в них соответственно k_1, k_2, … k_n, … событий зависит лишь от этих чисел и длин промежутков времени. В частности, вероятность появления k требований в промежутке (T, T+t) не зависит от T и является функцией лишь k и t.

2). **Отсутствие последействия.** Это свойство означает, что вероятность поступления k событий потока в продолжении промежутка времени (T, T+t) не зависит от того, сколько событий и как поступили до этого промежутка. Это требование означает марковость изучаемого потока.

3). **Ординарность.** Это условие выражает собой практическую невозможность появления двух или большего числа событий за очень малый промежуток времени.

Поток событий, удовлетворяющий этим трем условиям, называется простейшим. Можно доказать, что простейший поток полностью определяется равенством

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad (1)$$

где $P_k(t)$ – есть вероятность появления k событий потока за промежуток времени t,

λ – положительная постоянная, характеризующая собой «интенсивность» наступления событий потока. В частности, вероятность того, что за промежуток времени t не наступит на одного события потока равна

$$P_0(t) = e^{-\lambda t} \qquad (2)$$

Простейший поток можно определить и иначе, а именно, как поток моментов времени, расстояния между которыми случайны, причем вероятность того, что расстояние между соседними моментами окажется большим чем t, определяется формулой (2). Такое определение простейшего потока также часто используется при решении многих вопросов прикладного и теоретического характера.

Непосредственная проверка наличия трех перечисленных условий – стационарности, отсутствия последствия и ординарности – нередко трудно выполнима, поэтому очень важно найти иные условия, которые позволяли бы делать вывод о том, что поток событий простейший или близок к простейшему. Такое условие было найдено в работах ряда исследователей и доказано в общей форме А.Я. Хинчиным. Оно состоит в следующем.

Предположим, что интересующий нас поток является суммой очень большого числа независимых между собой стационарных потоков, каждый из которых мало влияет на сумму. **Суммарный поток** при дополнительном ограничении арифметического характера, гарантирующем ординарность **суммарного потока** оказывается близким к простейшему.

Указанная теорема имеет принципиальное значение для приложений. Так например поток грузовых судов, прибывающих в данный морской порт, составлен из большого числа потоков, отправляемых из различных других портов, следовательно, поток судов также должен быть близок к простейшему. Число примеров, имеющих иное реальное содержание, можно продолжить и далее.

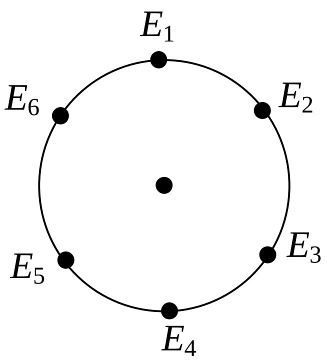
Типовые примеры.

Пример 1.

На окружности расположены шесть точек E_1, E_2, … , E_6, равноотстоящих друг от друга. Частица движется из точки в точку следующим образом. Из данной точки она перемещается в одну из ближайших соседних точек с вероятностью $1/4$ или в диаметрально противоположную точку с вероятностью $1/2$. Выписать матрицу вероятностей перехода для этого процесса.

Р е ш е н и е

Пусть E_i, $i=\overline{1,6}$ означает, что частица находится в точке *i*. E_{ij}, $i=\overline{1,6}$, $j=\overline{1,6}$ означает, что она переместилась из точки *i* в *j*. Из условия задачи имеем: $E_{ii}=0$, $i=\overline{1,6}$. $E_{12}=\frac{1}{4}$, $E_{13}=0$, $E_{14}=\frac{1}{2}$, $E_{15}=0$, $E_{16}=\frac{1}{4}$. Аналогично $E_{21}=\frac{1}{4}$, $E_{23}=\frac{1}{4}$, $E_{24}=0$, $E_{25}=\frac{1}{2}$, $E_{26}=0$ и т.д. Тогда матрица вероятностей перехода для этого процесса запишется:



$$\begin{matrix} & \\ E_1 \\ E_2 \\ E_3 \\ E_4 \\ E_5 \\ E_6 \end{matrix}\begin{pmatrix} E_1 & E_2 & E_3 & E_4 & E_5 & E_6 \\ 0 & 1/4 & 0 & 1/2 & 0 & 1/4 \\ 1/4 & 0 & 1/4 & 0 & 1/2 & 0 \\ 0 & 1/4 & 0 & 1/4 & 0 & 1/2 \\ 1/2 & 0 & 1/4 & 0 & 1/4 & 0 \\ 0 & 1/2 & 0 & 1/4 & 0 & 1/4 \\ 1/4 & 0 & 1/2 & 0 & 1/4 & 0 \end{pmatrix}$$

Пример 2.

Задана матрица перехода $P = \begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}3 & 0{,}7 \end{pmatrix}$ из состояния в состояние за один шаг. Найти матрицу перехода из состояния в состояние за два шага.

Решение

$$P_2 = P^2, \text{ т.е. } P_2 = \begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}3 & 0{,}7 \end{pmatrix} \begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}3 & 0{,}7 \end{pmatrix} = \begin{pmatrix} 0{,}34 & 0{,}66 \\ 0{,}33 & 0{,}67 \end{pmatrix}.$$

Пример 3.

Группа их трех 3^x подводных лодок должна форсировать l противолодочных рубежей. Вероятность уничтожения на данном рубеже каждой подводной лодки, не уничтоженной на предшествующих рубежах, не зависит от числа лодок, форсирующих рубеж и равна g.

Определить матрицу перехода из состояния в состояние за один шаг. Определить распределение числа лодок в группе после форсирования каждого рубежа.

Решение

К данной задаче применима схема однородной цепи Маркова, если в качестве системы рассматривать группу лодок, а под номером состояния понимать число лодок в группе. Начальное состояние системы определяется очевидным образом $P_0(0)=P_1(0)=P_2(0)=0$, $P_3(0)=1$ или в матричном виде $P_0=(0\,0\,0\,1)$. Очевидно, что система не может переходить из состояний с меньшими номерами в состояние с большими номерами.

Если $i \geq j$, то вероятность перехода из состояния *i* в состоянии *j* равна вероятности уничтожения (*i* – *j*) лодок, которая задается биноминальным распределением

$$P_{ij}=c_i^{i-j}g^{i-j}(1-g)^j.$$

Тогда матрица перехода из состояния в соответствие за один шаг имеет вид

$$P=\begin{matrix} \\ E_0 \\ E_1 \\ E_2 \\ E_3 \end{matrix}\begin{pmatrix} E_0 & E_1 & E_3 & E_4 \\ 1 & 0 & 0 & 0 \\ g & 1-g & 0 & 0 \\ g^2 & 2g(1-g) & (1-g)^2 & 0 \\ g^3 & 3g^2(1-g) & 3g(1-g)^2 & (1-g)^3 \end{pmatrix}.$$

$P_l=P_{l-1}P$, $l=1,2,...$ Например:

$$P_1=(0\,0\,0\,1)\begin{pmatrix} E_0 & E_1 & E_3 & E_4 \\ 1 & 0 & 0 & 0 \\ g & 1-g & 0 & 0 \\ g^2 & 2g(1-g) & (1-g)^2 & 0 \\ g^3 & 3g^2(1-g) & 3g(1-g)^2 & (1-g)^3 \end{pmatrix}=$$

$$=\left(g^3\;3g^2(1-g)\;\;3g(1-g)^2\;\;(1-g)^3\right).$$

Пример 4. Фирма кожизделия открывает производство дамских сумок. Все множество состояний фирмы можно разделить условно на два основных состояния:

1). Модели, которые она выпускает, находят спрос у покупателя

2). Эти модели не находят спроса.

Предположим, что матрица перехода из состояния в состояние за один шаг имеет вид:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix}.$$

Предположим далее, что Фирма начинает с удачной модели сумки. Тогда матрица начальных состояний $P_0 = (1\,0)$. Пусть переходы системы из состояния в состояние происходят в дискретные моменты времени (например, через неделю). Тогда спустя одну неделю состояние системы

$$P(1) = P(0)\cdot P = (1\,0)\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{2}{5} & \frac{3}{5} \end{pmatrix} = \left(\frac{1}{2}\,\frac{1}{2}\right).$$

Аналогично, через n недель состояние фирмы определится

$$P(n) = P(n-1)P = P(0)P^n$$

Вычислим

$$P(n) = (P_1(n)P_2(n))$$

и запишем в таблицу

n	0	1	2	3	4	5
$P_1(n)$	1	0,5	0,45	0,445	0,4445	0,44445
$P_2(n)$	0	0,5	0,55	0,555	0,5555	0,55555

Далее вычислим вероятности состояний фирмы через n недель, если она начинает с неудачной модели, т.е. $P(0) = (0\,1)$.

Вычислим $P(n)$, $n = \overline{0,5}$ и занесем результаты в таблицу.

n	0	1	2	3	4	5
$P_1(n)$	0	0,4	0,44	0,444	0,4444	0,44444
$P_2(n)$	1	0,5	0,56	0,556	0,5556	0,55556

Анализируя данные таблиц, замечаем, что для каждой из них при $n \to \infty$; $P_1(n) \to \frac{4}{9}$; $P_2(n) \to \frac{5}{9}$, т.е. предельные вероятности состояний данной системы не зависят от начальных состояний. Марковские процессы, обладающие этим свойством, называются эргодическими.

Упражнения.

Пример 1.

В учениях участвуют два корабля, которые одновременно производят выстрелы друг в друга через равные промежутки времени. При каждом обмене выстрелами корабль А поражает корабль В с вероятностью $\frac{1}{2}$, а корабль В поражаем корабль А с вероятностью $\frac{3}{8}$.

Предполагается, что при любом попадании корабль выходит из строя. Рассматриваются результаты серии выстрелов. Найти матрицу перехода, если состояниями цепи являются комбинации кораблей, оставшихся в строю: E_1 – оба корабля в строю. E_2 – в строю корабль А, E_3 – в строю корабль В, E_4 – оба корабля поражены.

Пример 2.

В двух отделениях ящика находятся три шара. Каждую секунду выбирается случайным образом один из трех шаров и перекладывается из одного отделения в другое. В качестве состояния марковской цепи рассматривается число шаров в первом отделении. Выписать матрицу перехода из состояния в состояние за один шаг.

Литература: [10, С. 420-437; 13, С. 163-173; С. 190-193].

ПРИЛОЖЕНИЯ

Таблица 1

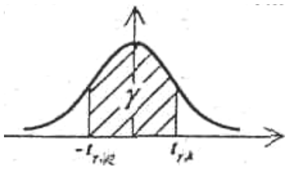
Значения функции Лапласа $\Phi(x) = \frac{2}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} \partial t$

Целые и десятые доли *x*	Сотые доли *x*									
	0	1	2	3	4	5	6	7	8	9
0,0	0,000	0080	0160	0238	0319	0399	0478	0558	0638	0717
0,1	0797	0876	0955	1034	1113	1192	1271	1350	1428	1507
0,2	1585	1663	1741	1819	1897	1974	2051	2128	2205	2282
0,3	2358	2434	2510	2586	2661	2737	2812	2886	2960	3035
0,4	3108	3182	3255	3328	3401	3473	3545	3616	3688	3759
0,5	3829	3899	3969	4039	4108	4177	4245	4313	4381	4448
0,6	4515	4581	4647	4713	4778	4843	4907	4971	5035	5098
0,7	5161	5223	5285	5346	5407	5467	5527	5587	5646	5705
0,8	5763	5821	5878	5935	5991	6047	6102	6157	6211	6265
0,9	6319	6372	6424	6476	6528	6579	6629	6679	6729	6778
1,0	6827	6875	6923	6970	7017	7063	7109	7154	7199	7243
1,1	7287	7330	7373	7415	7457	7499	7540	7580	7620	7660
1.2	7699	7737	7775	7813	7850	7887	7923	7959	7994	8029
1,3	8064	8098	8132	8165	8198	8230	8262	8293	8324	8355
1,4	8385	8415	8444	8473	8501	8529	8557	8584	8611	8638
1,5	8664	8690	8715	8740	8764	8789	8812	8836	8859	8882
1,6	8904	8926	8948	8969	8990	9011	9031	9051	9070	9090
1,7	9109	9127	9146	9164	9181	9189	9216	9233	9249	9265
1,8	9281	9297	9312	9327	9342	9357	9371	9385	9399	9412
1,9	9426	9439	9451	9464	9476	9488	9500	9512	9523	9534
2,0	9545	9556	9566	9576	9586	9596	9606	9616	9625	9634
2,1	9643	9651	9660	9668	9676	9684	9692	9700	9707	9715
2,2	9722	9729	9736	9743	9749	9756	9762	9768	9774	9780
2,3	9786	9791	9797	9802	9807	9812	9817	9822	9827	9832

Продолжение таблицы 1

2,4	9836	9841	9845	9849	9853	9857	9861	9865	9869	9872
2,5	9876	9879	9883	9886	9889	9892	9895	9898	9901	9904
2,6	9907	9910	9912	9915	9917	9920	9922	9924	9926	9928
2,7	9931	9933	9935	9937	9938	9940	9942	9944	9946	9947
2,8	9949	9951	9952	9953	9955	9956	9958	9959	9960	9961
2,9	9963	9964	9965	9966	9967	9968	9969	9970	9971	9972
3,0	9973	9974	9975	9976	9976	9977	9978	9979	9979	9980
3,1	9981	9981	9982	9983	9983	9984	9984	9985	9985	9986
3,2	9985	9987	9987	9988	9988	9989	9989	9989	9990	9990
3,3	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,4	9993	9994	9994	9994	9994	9994	9995	9995	9995	9995
3,5	9995	9996	9996	9996	9996	9996	9996	9996	9997	9997
3,6	9997	9997	9997	9997	9997	9997	9997	9998	9998	9998
3,7	9998	9998	9998	9998	9998	9998	9998	9998	9998	9998
3,8	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
3,9	9999	9999	9999	9999	9999	9999	9999	9999	9999	9999
4,0	0,999936									
4,5	0,999994									
5,0	0,99999994									

Таблица 2



Значения $t_{\gamma,k}$ – критерия Стьюдента

k	Вероятность												
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99	0,999
1	0,16	0,32	0,51	0,73	1,00	1,38	1,96	3,08	6,31	12,71	31,82	63,66	636,6
2	14	29	44	62	0,82	06	34	1,89	2.92	4,30	6,96	9,92	31,6
3	14	28	42	58	76	0,98	25	64	35	3,18	4,54	5,84	12,94
4	13	27	41	57	74	94	19	53	13	2,78	3,75	4,60	8,61
5	13	27	41	56	73	92	16	48	01	57	36	03	6,86
6	0,13	0,26	0,40	0,55	1,72	1,91	1,13	1,44	1,94	2,45	3,14	3,71	5,96
7	13	26	40	55	71	90	12	41	89	36	00	50	5,41
8	13	26	40	55	70	89	11	40	86	31	2,90	35	5,04
9	13	26	40	54	70	88	10	38	83	26	82	25	4,78
10	13	26	40	54	70	88	09	37	81	23	76	17	4,59
11	0,13	0,26	0,40	0,54	0,70	0,88	1,09	1,36	1,80	2,20	2,72	3,11	4,49
12	13	26	39	54	69	87	08	36	78	18	68	05	4,32
13	13	26	39	54	69	87	08	35	77	16	65	01	4,22
14	13	26	39	54	69	87	08	34	76	14	62	2,98	4,14
15	13	26	39	54	69	87	07	34	75	13	60	95	4,07
16	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,34	1,75	2,12	2,58	2,92	4,02
17	13	26	39	53	69	86	07	33	74	11	57	90	3,97
18	13	26	39	53	69	86	07	33	73	10	55	88	3,92
19	13	26	39	53	69	86	07	33	73	09	54	86	3,88
20	13	26	39	53	69	86	06	32	72	09	53	84	3,85
21	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,72	2,08	2,52	2,83	3,82
22	13	26	39	53	69	86	06	32	72	07	51	82	3,79
23	13	26	39	53	68	86	06	32	71	07	50	81	3,77
24	13	26	39	53	68	86	06	32	71	06	49	80	3,75
25	13	26	39	53	68	86	06	32	71	06	48	79	3,73

Продолжение таблицы 2

26	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,31	1,71	2,06	2,48	2,78	3,71
27	13	26	39	53	68	85	06	31	70	05	47	77	3,69
28	13	26	39	53	68	85	06	31	70	05	47	76	3,67
29	13	26	39	53	68	85	05	31	70	04	46	76	3,66
30	13	26	39	53	68	85	05	31	70	04	46	75	3,65
40	0,13	0,25	0,39	0,53	0,68	0,85	1,05	1,30	1,68	2,02	2,42	2,70	3,55
60	13	25	39	53	68	85	05	30	67	00	39	66	3,46
120	0,13	0,25	0,39	0,53	0,68	0,84	1,04	1,29	1,66	1,98	2,36	2,62	3,37
∞	13	25	38	52	67	84	04	28	64	96	33	58	3,29

Таблица 3

χ^2 – распределение (распределение Пирсона)

$$P\left(\chi^2 > x_\alpha\right) = \alpha$$

k \ α	0,99	0,98	0,95	0,90	0,80	0,20	0,10	0,05	0,02	0,01
1	0,00016	0,0063	0,393	0,0158	0,6642	1,642	2,706	3,841	5,412	6,635
2	0,0201	0,0404	0,103	0,211	0,446	3,219	4,605	5,991	7,824	9,210
3	0,115	0,185	0,352	0,584	1,005	4,642	6,251	7,815	9,837	11,341
4	0,297	0,429	0,711	1,064	1,649	5,989	7,779	9,488	11,668	13,277
5	0,554	0,752	1,145	1,610	2,343	7,289	9,236	11,070	13,388	15,086
6	0,872	1,134	1,635	2,204	3,070	8,558	10,645	12,592	15,033	16,812
7	1,239	1,564	2,167	2,833	3,822	9,803	12,017	14,067	16,622	18,475
8	1,646	2,032	2,733	3,490	4,594	11,030	13,362	15,507	18,168	20,090
9	2,088	2,532	3,325	4,168	5,380	12,242	14,684	16,919	19,679	21,666
10	2,558	3,059	3,940	4,865	6,179	13,442	15,987	18,307	21,161	23,209
11	3,053	3,609	4,575	5,578	6,989	14,631	17,275	19,675	22,618	24,725
12	3,571	4,178	5,226	6,304	7,807	15,812	18,549	21,026	24,054	26,217
13	4,107	4,765	5,892	7,042	8,634	16,985	19,812	22,362	25,472	27,688
14	4,660	5,368	6,571	7,790	9,467	18,151	21,064	23,685	26,873	29,141
15	5,229	5,985	7,261	8,547	10,307	19,311	22,307	24,996	28,259	30,578
16	5,812	6,614	7,962	9,312	11,152	20,465	23,542	26,296	29,633	32,000
17	6,408	7,255	8,672	10,085	12,002	21,615	24,769	27,587	30,995	33,409
18	7,015	7,906	9,390	10,865	12,857	22,760	25,989	28,869	32,346	34,805
19	7,633	8,567	10,117	11,651	13,716	23,900	27,204	30,144	33,687	36,191
20	8,260	9,237	10,851	12,443	14,578	25,038	28,412	31,410	35,020	37,566
21	8,897	9,915	11,591	13,240	15,445	26,171	29,615	32,671	36,343	38,932
22	9,542	10,600	12,338	14,041	16,314	27,301	30,813	33,924	37,659	40,289
23	10,196	11,293	13,091	14,848	17,187	28,429	32,007	35,172	38,968	41,638
24	10,856	11,992	13,848	15,659	18,062	29,553	33,196	36,415	40,270	42,980
25	11,524	12,697	14,611	16,473	18,940	30,675	34,382	37,652	41,566	44,314
26	12,198	13,409	15,379	17,292	19,820	31,795	35,563	38,885	42,856	45,642
27	12,879	14,125	16,151	18,114	20,703	32,912	36,741	40,112	44,140	46,963
28	13,565	14,847	16,928	18,939	21,588	34,027	37,916	41,337	45,419	48,278
29	14,256	15,574	17,708	19,768	22,475	35,139	39,087	42,557	46,693	49,588
30	14,953	16,306	18,493	20,599	23,364	36,250	40,256	43,773	47,962	50,892

Продолжение таблицы 3

$\alpha = 0,01$

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98,49	99,00	99,17	99,25	99,30	99,33	99,36	99,42	99,46	99,50
3	34,12	30,81	29,46	28,71	28,24	27,91	27,49	27,05	26,60	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,80	14,37	13,93	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,29	9,89	9,47	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,10	7,72	7,31	6,83
7	12,25	9,55	8,45	7,85	7,46	7,19	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,06	4,71	4,33	3,91
11	9,65	7,20	6,22	5,67	5,32	5,07	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,30	3,96	3,59	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,14	3,80	3,43	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,79	3,45	3,08	2,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,71	3,37	3,00	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,51	3,17	2,80	2,36
22	7,94	5,72	4,82	4,31	3,99	3,76	3,45	3,12	2,75	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,86	3,63	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,23	2,90	2,52	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,20	2,87	2,49	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,84	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,66	2,29	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,50	2,12	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,66	2,34	1,95	1,38
∞	6,64	4,60	3,78	3,32	3,02	2,80	2,51	2,18	1,79	1,00

Таблица 4

F – распределение (Фишера)

$P(F > f_{\alpha}) = \alpha \quad \alpha = 0{,}05$

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,52
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,83	1,61	1,25
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

Таблица 5

Функция плотности вероятности нормального распределения

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

f(x)	0	1	2	3	4	5	6	7	8	9
0,0	39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29430	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	2468,1	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20327	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014
4,0	00013	00009	00006	00004	00002	00002	00001	00001	00000	00000

Таблица 6

Функция распределения нормального закона

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

$f(x)$	0	1	2	3	4	5	6	7	8	9
0,0	50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0,1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0,2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0,3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0,4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793
0,5	69146	69497	69847	70194	70540	70884	71226	71566	71904	72240
0,6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0,7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0,8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0,9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1,0	84134	84375	84614	84850	85083	85314	85543	85769	85993	86214
1,1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1,2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1,3	90320	90490	90658	90824	90988	91149	91308	91466	91621	91774
1,4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1,5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1,6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1,7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1,8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1,9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670
2,0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2,1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574
2,2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
2,3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
2,4	99180	99202	99224	99245	99266	99286	99305	99324	99343	99361
2,5	99379	99396	99413	99430	99446	99461	99477	99492	99506	99520
2,6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2,7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
2,8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2,9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
3,0	99865	99869	99874	99878	99882	99886	99889	99893	99896	99900
3,1	99903	99906	99910	99913	99916	99918	99921	99924	99926	99929
3,2	99931	99934	99936	99938	99940	99942	99944	99946	99948	99950
3,3	99952	99953	99955	99957	99958	99960	99961	99962	99964	99965
3,4	99966	99968	99969	99970	99971	99972	99973	99974	99975	99976
3,5	99977	99978	99978	99979	99980	99981	99981	99982	99983	99983
3,6	99984	99985	99985	99986	99986	99987	99987	99988	99988	99989
3,7	99989	99990	99990	99990	99991	99991	99992	99992	99992	99992
3,8	99993	99993	99993	99994	99994	99994	99994	99995	99995	99995
3,9	99995	99995	99996	99996	99996	99996	99996	99996	99997	99997
4,0	99997	99998	99999	99999	99999	–	–	–	–	–

Таблица 7

Распределение Пуассона

$$P(k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	9048	8187	7408	6703	6065	5488	4966	4493	4066
1	0905	1638	2222	2681	3033	3293	3476	3595	3659
2	0045	0164	0333	0536	0758	0988	1217	1438	1647
3	0002	0019	0033	0072	0126	0198	0284	0383	0494
4		0001	0002	0007	0016	0030	0050	0077	0111
5				0001	0002	0004	0007	0012	0020
6							0001	0002	0003

k \ λ	1	2	3	4	5	6	7	8	9	10
0	3679	1353	0498	0183	0067	0025	0009	0003	0001	0000
1	3679	2707	1494	0733	0337	0149	0064	0027	0011	0005
2	1839	2707	2240	1465	0842	0446	0223	0107	0050	0023
3	0613	1804	2240	1954	1404	0892	0521	0286	0150	0076
4	0153	0902	1680	1954	1766	1339	0912	0572	0337	0189
5	0031	0361	1008	1563	1755	1606	1277	0916	0607	0378
6	0005	0120	0504	1042	1462	1606	1490	1221	0911	0631
7	0001	0037	0216	0595	1044	1377	1490	1396	1171	0901
8		0009	0081	0298	0653	1033	1304	1396	1318	1126
9		0002	0027	0132	0363	0688	1014	1241	1318	1251
10			0008	0053	0181	0413	0710	0993	1186	1251
11			0002	0019	0082	0225	0452	0722	0970	1137
12			0001	0006	0034	0126	0263	0481	0728	0948
13				0002	0013	0052	0142	0296	0504	0729
14				0001	0005	0022	0071	0169	0324	0521
15					0002	0009	0033	0090	0194	0347
16						0003	0014	0045	0109	0217
17						0001	0006	0021	0058	0128
18							0002	0009	0029	0071
19							0001	0004	0014	0037
20								0002	0006	0019
21								0001	0003	0009
22									0001	0004
23										0002
24										0001

ИСПОЛЬЗОВАННАЯ И РЕКОМЕНДОВАННЯ ЛИТЕРАТУРА

1. ***Агапов Г.И.*** Задачник по теории вероятностей – М.: Высшая школа. 1986 г.
2. ***Боровков А.А.*** Теория вероятностей – М.: Наука, 1976.
3. ***Володин Б.Г. и др.*** Сборник задач по теории вероятностей, математической статистики и теории случайных функций. Под ред. А.А. Свешникова. – М.: Наука, 1970.
4. ***Биденко Б.В.*** Курс теории вероятностей. М.: Наука, 1965.
5. ***Захаров В.К.,*** Севасттьянов Б.А., Чистяков В.П. Теория вероятностей. – М.: Наука, 1983 г.
6. ***Колмогоров А.Н.*** Основные понятия теории вероятностей. – М.: Наука, 1974 г.
7. ***Колде Я.К.*** Практикум по теории вероятностей и математической статистики. – М.: Высшая школа, 1991.
8. ***Розанов Ю.А.*** Случайные процессы. – 2^{е} издание. – М.: Наука, 1979 г.
9. ***Рушинский Л.З.*** Элементы теории вероятностей. – М.: Наука, 1973.
10. ***Смирнов Н.В***., Дунин-Барновский И.В. Краткий курс теории вероятностей и математической статистики для технических приложений. – М.: Физматгиз, 1959.
11. ***Нейман Ю***. Вводный курс теории вероятностей и математическая статистика. – М.: Финансы и статистика, 1982.
12. ***Феллер В***. Введение в теорию вероятностей и ее приложения, Т.1. / Пер. с английского – М.: Мир, 1967.
13. ***Гитянов В.П.*** Курс теории вероятностей. – 2^{е} изд. – М.: Наука, 1982.