

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

**ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ**

**Московский государственный университет
приборостроения и информатики**



кафедра высшей математики

Теория вероятностей и математическая статистика

Методические указания для студентов дневной формы обучения.

Москва 2007

Составители: д.ф.м.н., проф. А.В. Манжиров, к.ф.м.н. М.Н. Михин
УДК 517.

Теория вероятностей и математическая статистика: методические для студентов дневной формы обучения./МГУПИ.Сост. д.ф.м.н., проф. А.В. Манжиров, к.ф.м.н. М.Н. Михин.

Излагаются основные методы решения задач по теории вероятностей и математической статистике. Приведены примеры решения различных типов задач

Пособие предназначено для студентов, обучающихся по дневной форме обучения. Библиогр: .

Рецензент: проф. Головешкин В.А.

Введение

Цель данного пособия помочь студенту самостоятельно подготовиться к выполнению контрольных работ. При написании пособия авторы не ставили своей целью дать систематическое изложение теоретического материала. Перед каждой рассматриваемой задачей дается тот теоретический материал, который необходим для ее решения. Если студент ранее овладел необходимым теоретическим материалом, то вводную часть каждой задачи он может опустить и перейти непосредственно к решению задачи. Необходимо отметить, что приведенный теоретический материал достаточно полно охватывает курс указанного предмета. Авторы надеются, что пособие будет полезно студентам в овладении методами решения основных задач курса теории вероятностей и математической статистики.

Глава 1. Случайные события и их вероятности

§1. События. Действия с событиями

Определение. Элементарным событием называется простейший неделимый исход некоторого опыта.

Элементарные исходы будем обозначать символом ω .

Определение. Пространством элементарных исходов называется множество всех элементарных исходов, которое будем обозначать символом Ω .

Рассмотрим несколько примеров.

Пример 1. Описать пространство элементарных исходов при подбрасывании монеты.

Решение. Очевидно, что при подбрасывании монеты возможны два элементарных исхода:

ω_1 — появление «герба»;

ω_2 — появление «решки».

Таким образом, пространство элементарных исходов содержит два элемента

$$\Omega = \{\omega_1, \omega_2\}. \bullet$$

Пример 2. Описать пространство элементарных исходов при подбрасывании игрального кубика.

Решение. Очевидно, что при подбрасывании игрального кубика элементарными исходами является число, выпавших очков, т.е.

ω_i — выпало ровно i очков; $i = 1, 2, 3, 4, 5, 6$.

Таким образом, пространство элементарных исходов содержит шесть элементов

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\} \text{ или } \Omega = \{1, 2, 3, 4, 5, 6\}. \bullet$$

Пример 3. На отрезке $[0, 1]$ случайным образом отмечается точка. Описать пространство элементарных исходов.

Решение. В этом случае результатом является координата x , удовлетворяющая условию $0 \leq x \leq 1$. Очевидно, что координата x меняется непрерывно, пространство элементарных исходов имеет вид

$$\Omega = \{x, 0 \leq x \leq 1\}.$$

Пространство элементарных исходов имеет бесконечно много элементов. $\bullet$

Определение. Событие называется **достоверным**, если в результате опыта оно непременно должно произойти.

Достоверное событие обозначают символом Ω , так как оно состоит из тех же элементарных исходов, что и пространство элементарных исходов. Событие, состоящее в появлении менее 7 очков при бросании игрального кубика, является достоверным.

Определение. Событие называется **невозможным**, если в результате опыта оно не может произойти.

Невозможное событие обозначают символом $\emptyset$. Событие, состоящее в появлении 7 очков при бросании игрального кубика, является невозможным.

Определение. События $H_1, H_2, \dots, H_n$ называют **несовместными**, если при наступлении одного из событий, остальные $n - 1$ события в данном испытании наступить уже не могут.

Так, например, если при бросании игральной кости выпала грань «2», то это означает, что при том же бросании не могла появиться грань «4».

Определение. События $H_1, H_2, \dots, H_n$ образуют **полную группу событий**, если в результате опыта, одно из событий обязательно происходит.

Например, при однократном бросании игральной кости полная группа попарно несовместимых событий состоит из событий

$$H_1, H_2, \dots, H_6,$$

которые состоят соответственно в выпадении 1, 2, 3, 4, 5, 6 очков соответственно.

Наряду с элементарными событиями будем рассматривать случайные события.

Определение. **Случайным событием** A (или просто **событием** A) называется любое подмножество множества Ω .

Определение. Элементарные события, принадлежащие подмножеству A , называются **благоприятствующими** событию A .

Рассмотрим основные операции над событиями; они полностью соответствуют основным операциям над множествами.

Определение. **Суммой событий** A и B называется событие $C = A \cup B = A + B$, состоящее в наступлении хотя бы одного из событий A или B .

Элементарными исходами суммы событий $A + B$ являются элементарные исходы, принадлежащие хотя бы одному из событий A и B .

Определение. **Произведением событий** A и B называется событие $C = A \cap B = AB$, состоящее в совместном (одновременном) наступлении этих событий.

Элементарными исходами произведения событий AB являются те элементарные исходы, которые одновременно принадлежат событиям A и B .

Определение. **Разностью событий** A и B называется событие $C = A \setminus B = A - B$, состоящее в том, что событие A произошло, а событие B не произошло.

Элементарными исходами разности событий $A \setminus B$ являются те элементарные исходы события A , которые не принадлежат событию B .

Определение. Событие, состоящее в том, что событие A не происходит, называется **противоположным** событию A и обозначается $\bar{A}$.

Элементарными исходами противоположного события $\bar{A}$ являются те элементарные исходы, которые не принадлежат событию A .

Определение. Событие A **влечет** событие B (A является подмножеством множества B), если из того, что происходит событие A , следует, что происходит событие B ; записывают $A \subseteq B$.

Определение. Если одновременно $A \subseteq B$ и $B \subseteq A$, то в этом случае события A и B называют **равносильными**, при этом пишут $A = B$.

Пример 4. Если A — событие, состоящее в том, что взятое наудачу изделие первого сорта, а B — изделие качественное (не брак), то в том событие A влечет событие B : $A \subseteq B$.

Свойства операций над событиями:

- $A \cup B = B \cup A$, $A \cap B = B \cap A$ (коммутативность);
- $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$, $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ (дистрибутивность);
- $(A \cup B) \cup C = A \cup (B \cup C)$, $(A \cap B) \cap C = A \cap (B \cap C)$ (ассоциативность);
- $A \cup A = A$, $A \cap A = A$;
- $A \cup \Omega = \Omega$, $A \cap \Omega = A$;
- $A \cup \bar{A} = \Omega$, $A \cap \bar{A} = \emptyset$;
- $\bar{\emptyset} = \Omega$, $\bar{\Omega} = \emptyset$, $\bar{\bar{A}} = A$;
- $A \setminus B = A \cap \bar{B}$;
- $\overline{A \cup B} = \bar{A} \cap \bar{B}$, $\overline{A \cap B} = \bar{A} \cup \bar{B}$ (законы де Моргана).

§2. Общее определение и свойства вероятности

Определение. Вероятностью $P(A)$ события A называется функция, определенная на пространстве элементарных исходов и удовлетворяющая трем условиям:

- Для каждого события A
 $P(A) \geq 0$ (условие неотрицательности);
- Для достоверного события Ω
 $P(\Omega) = 1$ (условие нормировки);
- Если $B \cap C = \emptyset$, то
 $P(B \cup C) = P(B) + P(C)$ (теорема сложения для несовместных событий).

Свойства вероятности:

1. Вероятность события $\bar{A}$, противоположного событию A , равна
 $P(\bar{A}) = 1 - P(A)$.

Доказательство. Используем очевидное свойство суммы противоположных событий $A \cup \bar{A} = \Omega$. Тогда, используя условие нормировки и теорему сложения для несовместных событий, получим:

$$1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}),$$

Из двух последнего равенства следует, что

$$P(\bar{A}) = 1 - P(A). \blacksquare$$

2. Вероятность невозможного события равна нулю, т.е.
 $P(\emptyset) = 0$.

Доказательство. Используем очевидное свойство $\Omega = \Omega + \emptyset$ и теорему сложения для несовместных событий, получим:

$$P(\Omega) = P(\Omega + \emptyset) = P(\Omega) + P(\emptyset),$$

откуда и следует данное свойство. $\blacksquare$

3. Если событие A влечёт за собой событие B ($A \subseteq B$), то
 $P(A) \leq P(B)$.

Доказательство. Представим событие B в виде суммы двух несовместных событий $B = A + \bar{A}B$, $A \cap (\bar{A}B) = \emptyset$.

Используя теорему сложения для несовместных событий, получим:

$$P(B) = P(A \cup (\bar{A}B)) = P(A) + P(\bar{A}B) \geq P(A). \blacksquare$$

4. Для каждого события A , справедливо неравенство
 $0 \leq P(A) \leq 1$.

Доказательство. Данное свойство следует из условий нормировки и теоремы сложения для несовместных событий. $\blacksquare$

ГЛАВА 2. Классическая и геометрическая вероятности

§1. Классическое определение вероятности

Вернемся к монете. Пространство элементарных исходов Ω содержит два элементарных исхода:

ω_1 — появление «герба»;

ω_2 — появление «решки».

В силу того, что монета симметрична, нельзя предпочесть «гербу» «решке» (или наоборот). Следовательно, обоим элементарным исходам необходимо сопоставить одинаковую вероятность $P(\omega_1) = P(\omega_2)$. Далее очевидно, что

$$1 = P(\Omega) = P(\omega_1 + \omega_2) = P(\omega_1) + P(\omega_2).$$

Откуда получаем:

$$P(\omega_1) = P(\omega_2) = \frac{1}{2}.$$

Рассмотрим общий случай. Пусть пространство Ω состоит из n всевозможных равнозначных исходов $\omega_1, \dots, \omega_n$. Теперь каждому элементарному исходу $\omega_i (i = 1, \bar{n})$ поставим в соответствие вероятность $P(\omega_i) = \frac{1}{n}$.

Далее рассмотрим некоторое событие A , которому соответствует ровно m (благоприятных) элементарных исходов ω .

Положим

$$P(A) = \frac{m}{n}. \quad (2.1.1)$$

Таким образом, в классической схеме вероятность любого события A определяется как отношение числа m благоприятных для события A элементарных исходов к общему числу элементарных исходов n .

Пример 1. В урне находятся a белых и b черных шаров. Из урны вынимают наугад один шар. Найти вероятность того, что этот шар белый (событие A).

О Решение. Число всевозможных исходов равно

$$n = a + b.$$

Число благоприятных исходов равно

$$m = a.$$

Таким образом, используя классическое определение вероятности, получаем

$$P(A) = \frac{m}{n} = \frac{a}{a+b}. \bullet$$

Пример 2. Имеются две урны: в первой — a белых и b черных шаров; во второй — c белых и d черных шаров. Из каждой урны вынимается по шару. Найти вероятность того, что оба шара будут белыми (событие A).

О Решение. Каждый шар из первой урны может комбинировать с каждым шаром из второй урны. Следовательно, число всевозможных исходов:

$$n = (a+b)(c+d).$$

Аналогично, число благоприятных исходов:

$$m = ac.$$

Следовательно, используя классическое определение вероятности, получаем:

$$P(A) = \frac{ac}{(a+b)(c+d)} \cdot \bullet$$

Пример 3. Из колоды карт (36 листов) наудачу выбирается одна карта. Определить вероятность того, что она окажется тузом (событие А).

О Решение. Число всевозможных исходов равно:

$$n = 36.$$

Число благоприятных исходов равно числу тузов, т.е.

$$m = 4.$$

Таким образом, используя классическое определение вероятности, получаем:

$$P(A) = \frac{m}{n} = \frac{1}{9} \cdot \bullet$$

§2. Применение комбинаторного анализа

Теорема. Из m элементов $a_1, \dots, a_m$ и n элементов $b_1, \dots, b_n$ можно образовать mn пар (a_i, b_j) .

Доказательство. Составим из этих пар прямоугольную таблицу, состоящую из m строк и n столбцов, так, чтобы пара (a_i, b_j) стояла на пересечении i -ой строки и j -го столбца. В этом случае каждая пара появляется один и только один раз. Число элементов такой таблицы равно mn . ■

Пример 4. Найти число всевозможных исходов при бросании двух игральных костей.

О Решение. Очевидно, что каждый элемент пары принимает шесть значений. Следовательно, существует $6 \cdot 6 = 36$ возможных комбинаций. ●

Определение. Перестановкой из n различных элементов называется любой упорядоченный набор этих элементов.

Теорема. Число различных перестановок из n различных элементов вычисляется по формуле:

$$P_n = n!. \quad (2.2.1)$$

Доказательство. Первый элемент можно выбрать n способами, второй элемент можно выбрать $n-1$ способами (т.к. один элемент уже выбран), третий — $n-2$ способами и т.д. В итоге получим:

$$P_n = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1 = n!. \quad \blacksquare$$

Определение. Размещением из n различных элементов по m называется любой упорядоченный набор из m элементов, выбранных из общей совокупности в n элементов.

Теорема. Число различных размещений из n элементов по m вычисляется по формуле:

$$A_n^m = n \cdot (n-1) \cdot \dots \cdot (n-m+1) = \frac{n!}{(n-m)!}. \quad (2.2.2)$$

Доказательство. Данная теорема доказывается аналогично предыдущей теореме.

Определение. Сочетанием из n различных элементов по m называется любой неупорядоченный набор из m элементов, выбранных из общей совокупности в n элементов.

Теорема. Число сочетаний из n элементов по m вычисляется по формуле:

$$C_n^m = \frac{n!}{(n-m)!m!}. \quad (2.2.3)$$

Доказательство. Число сочетаний отличается от числа размещений только тем, что входящие в него элементы неупорядочены; m различных элементов можно упорядочить $m!$ способами. Следовательно, каждому размещению соответствует $m!$ сочетаний. Отсюда:

$$A_n^m = C_n^m m! \quad \text{или} \quad C_n^m = \frac{A_n^m}{m!} = \frac{n!}{(n-m)!m!}. \quad \blacksquare$$

Способ выбора, приводящий к перестановкам, размещениям и сочетаниям, называется выборкой без возвращения.

Рассмотрим выборку с возвращением. В этом случае каждый взятый элемент из общей совокупности возвращается обратно. Таким образом, один и тот же элемент может быть выбран несколько раз.

Теорема. Число выборок k элементов с возвращением из n различных элементов равно n^k .

Доказательство. Первый элемент может быть выбран n способами, второй также n способами и т.д. В итоге

$$\underbrace{n \cdot n \cdot \dots \cdot n}_k = n^k. \quad \blacksquare$$

Пример 5. (Гипергеометрическое распределение). Предположим, что имеются n шаров: n_1 красных и $n_2 = n - n_1$ черных. Случайным образом выбираются r шаров. Найти вероятность того, что выбранная группа будет содержать ровно k_1 красных и $k_2 = r - k_1$ черных шаров (событие A).

О Решение. Число способов, которыми можно выбрать, k_1 красных шаров из n_1 шаров равно $C_{n_1}^{k_1}$. Аналогично, число способов, которыми можно выбрать k_2 черных шаров из n_2 равно $C_{n_2}^{k_2} = C_{n-n_1}^{r-k_1}$. Так как любой выбор красных шаров может комбинировать (составлять пару) с любым выбором черных шаров, имеем число благоприятных исходов, равное $C_{n_1}^{k_1} C_{n-n_1}^{r-k_1}$.

Число всевозможных исходов равно C_n^r .

Используя классическое определение вероятности, получаем:

$$P(A) = \frac{C_{n_1}^{k_1} C_{n-n_1}^{r-k_1}}{C_n^r}. \quad \bullet$$

Теорема. Пусть $r_1, \dots, r_k$ — целые числа, такие, что $r_1 + r_2 + \dots + r_k = n$. Число способов, которыми множество из n элементов можно разделить на k упорядоченных подмножеств, из которых первое подмножество содержит r_1 элементов, второе — r_2 элементов и т.д., равно

$$\frac{n!}{r_1! r_2! \dots r_k!}. \quad (2.2.4)$$

Доказательство. Прежде чем доказывать теорему, заметим, что порядок подмножеств существенен в том смысле, что $(r_1 = 2, r_2 = 3)$ и $(r_1 = 3, r_2 = 2)$ представляет собой разные разбиения; однако, порядок элементов внутри групп игнорируется.

Перейдем к доказательству теоремы. Сначала необходимо выбрать r_1 элементов из n ; из оставшихся $n - r_1$ необходимо выбрать r_2 элементов и т.д. Получаем:

$$C_n^{r_1} C_{n-r_1}^{r_2} \cdot \dots \cdot C_{n-r_1-\dots-r_{k-1}}^{r_k} = \frac{n!}{r_1!(n-r_1)!} \frac{(n-r_1)!}{r_2!(n-r_1-r_2)!} \dots \frac{(n-r_1-\dots-r_{k-1})!}{r_k!(n-r_1-\dots-r_k)!} = \frac{n!}{r_1!r_2! \cdot \dots \cdot r_k!}. \blacksquare$$

Пример 6. Колода карт (52 листа) делится поровну между четырьмя игроками. Найти вероятность того, что каждый игрок имеет туза (событие A).

О Решение. Используя (2.2.4), найдем число всевозможных исходов:

$$n = \frac{52!}{13!13!13!13!} = \frac{52!}{(13!)^4}.$$

Найдем число благоприятных исходов. Четыре туза можно упорядочить $4!$ способами, и каждый порядок представляет одну возможность получения одного туза каждым игроком. Оставшиеся 48 карт, согласно (2.2.4), можно распределить $\frac{48!}{12!12!12!12!} = \frac{48!}{(12!)^4}$

способами. Таким образом, число благоприятных исходов равно $m = 4! \frac{48!}{(12!)^4}$.

Следовательно, искомая вероятность равна

$$P(A) = 4! \frac{48!}{(12!)^4} : \frac{52!}{(13!)^4} \approx 0,105. \bullet$$

Пример 7. Из полной колоды карт (52 листа) вынимаются сразу несколько карт. Какое минимальное число кар нужно вынуть, чтобы с вероятностью, большей чем 0,5, можно было утверждать, что среди них будут карты одной масти.

О Решение. Рассмотрим события A_k — среди вынутых карт есть хотя бы две карты одной масти. Пусть $k = 2$. В этом случае число всевозможных исходов равно $n = C_{52}^2$. Число благоприятных исходов получаем следующим образом: выбираем масть (4 способа), затем две карты этой масти (C_{13}^2), т.е. $m = 4C_{13}^2$.

Следовательно, используя классическое определение вероятности, получаем:

$$P(A_2) = \frac{4C_{13}^2}{C_{52}^2} = \frac{12}{51} < 0,5.$$

Пусть $k = 3$. В этом случае число всевозможных исходов равно $n = C_{52}^3$. Число благоприятных исходов получаем следующим образом: у нас либо две карты одной масти, либо три карты одной масти, т.е.

$$m = 4C_{13}^2 C_{39}^1 + 4C_{13}^3.$$

Следовательно, используя классическое определение вероятности, получаем:

$$P(A_3) = \frac{4C_{13}^2 C_{39}^1 + 4C_{13}^3}{C_{52}^3} = 0,602 > 0,5.$$

Таким образом, необходимо вынуть три карты. $\bullet$

§3. Геометрическое определение вероятности

Пусть теперь рассматривается непрерывная вероятностная схема, т.е. пространство элементарных исходов представляет собой некоторую ограниченную область (отрезок, круг, шар и т.д.) k -мерного пространства (прямой, плоскости, трёхмерного пространства и т.д.). В непрерывном случае число элементарных исходов бесконечно, следовательно, при использовании принципа равновероятности каждому элементарному исходу можно приписать только нулевую вероятность. Поэтому подойдём к определению геометрической

вероятности по-другому. Рассмотрим сначала отрезок $[0,1]$ и предположим, что идеальная частица равномерно бросается на данный отрезок. Каждому интервалу (a,b) ($0 \leq a < b \leq 1$) поставим в соответствие вероятность попадания частицы на этот интервал, равную его длине: $P(a,b) = b - a$.

В общем случае геометрическая вероятность определяется аналогично. Пусть Ω - некоторая область, имеющая меру $mes(\Omega)$ (длину, площадь, объём и т.д.) такую, что $0 < mes(\Omega) < \infty$. Пусть внутри области Ω находится область A .

Определение. Геометрической вероятностью называют отношение меры области A к мере области Ω :

$$P(A) = \frac{mes(A)}{mes(\Omega)}.$$

Пример 8. (Задача о встрече.) Два лица A и B договорились встретиться в определённом месте между 12 часами и часом. Пришедший первый ждёт другого в течение 20 минут, после чего уходит. Чему равна вероятность встречи лиц A и B , если приход каждого из них в течение указанного часа может произойти наудачу и моменты прихода независимы.

Решение. Обозначим момент прихода лица A через x , а момент прихода лица B через y . На плоскости; в качестве единицы масштаба выберем минуту. Всевозможные исходы изобразятся точками квадрата со сторонами 60. Для того, чтобы встреча произошла необходимо и достаточно, чтобы выполнялось неравенство $|x - y| \leq 20$.

Исходы, благоприятствующие встрече, изображены в заштрихованной области (рис. 2.1).

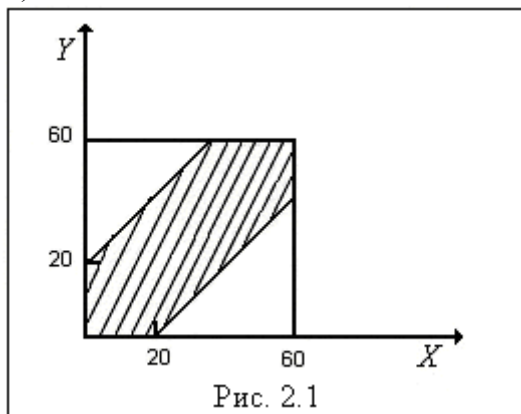
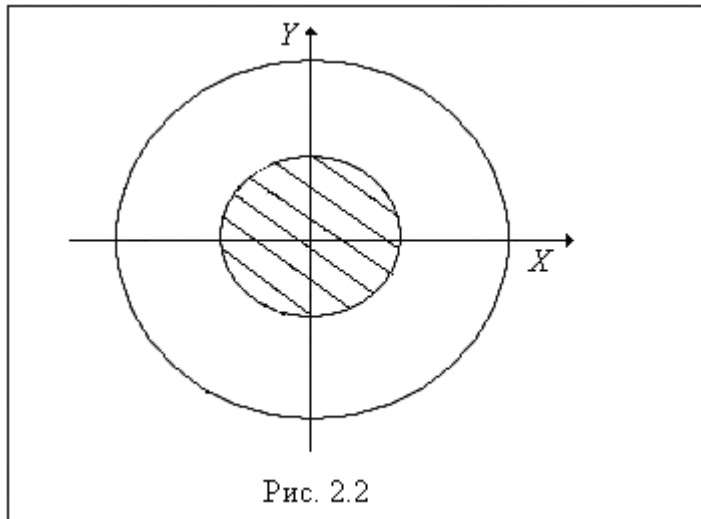


Рис. 2.1

Искомая вероятность равна отношению площади заштрихованной области к площади всего квадрата $p = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}$. ●

Пример 9. В круг радиуса $R = 1$ случайным образом бросается точка. Найти вероятность того, что точка попадёт в круг радиуса $r = \frac{1}{2}$ с тем же центром (рис. 2.2).



О Решение. Первый способ.

Пусть A — событие, состоящее в попадании точки в малый круг. Определим вероятность $P(A)$ как отношение площади малого круга к площади большего:

$$P(A) = \frac{\pi r^2}{\pi R^2} = \frac{1}{4}.$$

Второй способ. Рассмотрим полярную систему координат, в которой положение точки определяется углом φ между радиус-вектором точки и осью OX и расстоянием ρ от точки до начала координат. Поскольку точки, равностоящие от центра, все либо одновременно принадлежат меньшему кругу, либо нет, то вероятность попадания в этот круг равна отношению радиусов: $P(A) = \frac{r}{R} = \frac{1}{2}$.

Итак, мы получили в одной и той же задаче два разных ответа. Причина заключается в том, что понятие геометрической вероятности не инвариантно относительно преобразований рассматриваемой области Ω и зависит от того, как задана мера $mes(A)$.

Отметим, что для нас предпочтительнее первый способ решения. ●

Глава 3. Условная вероятность. Независимость событий. Формулы полной вероятности и Байеса

§1. Условная вероятность

Рассмотрим следующий пример. Бросаются две игральные кости. Найдем вероятность того, что сумма выпавших очков равна 8, если заранее известно, что сумма выпавших очков есть четное число. Всевозможные исходы запишем в виде таблицы 3.1:

Таблица 3.1

2 ^{ой} кубик 1 ^{ый} кубик	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Из таблицы видно, что число всевозможных исходов равно 36, но число исходов, удовлетворяющих условию, при которых сумма очков есть четное число, равно 18. Из них ровно в 5 исходах сумма очков равна 8. Пользуясь классическим определением вероятности, находим, что искомая вероятность равна $p = \frac{5}{18}$.

Заметим, что безусловная вероятность того, что сумма выпавших очков, равная 8, равна $\frac{5}{36}$, т.е. задание дополнительного условия может повлиять на вычисление вероятности.

Найдем условную вероятность $P(A|H)$ события A при условии, что событие H уже произошло. Для простоты рассмотрим классическую схему. Естественно положить, что данная вероятность есть отношение числа исходов m_{AH} , благоприятных совместному (одновременному) осуществлению событий A и H , к числу исходов, благоприятных событию H , т.е.

$$P(A|H) = \frac{m_{AH}}{m_H}.$$

Разделив числитель и знаменатель на число всевозможных исходов n , получим:

$$P(A|H) = \frac{m_{AH}/n}{m_H/n} = \frac{P(AH)}{P(H)}.$$

Последняя формула может служить общим определением условной вероятности при аксиоматическом подходе.

Определение. Условной вероятностью $P(A|H)$ события A при условии, что событие H уже произошло, называется отношение вероятности совместного (одновременного) осуществления событий A и H к вероятности события H :

$$P(A|H) = \frac{P(AH)}{P(H)}. \quad (3.1.1)$$

Пример 1. При трехкратном подбрасывании монеты выпало два «герба». Найти условную вероятность того, что при втором подбрасывании выпал «герб».

О Решение. Рассмотрим следующие события:

H — при трехкратном подбрасывании выпало два «герба»;

A — при втором подбрасывании выпал «герб».

Событию AH соответствует два исхода: $\Gamma - \Gamma - P, P - \Gamma - \Gamma$.

Число всевозможных исходов при трехкратном подбрасывании монеты $n = 2^3 = 8$. Отсюда находим:

$$P(AH) = \frac{2}{8} = \frac{1}{4}.$$

Аналогично, событию H соответствует три исхода, следовательно, вероятность условия равна

$$P(H) = \frac{3}{8}.$$

Далее, применяя (3.1.1), получаем искомую вероятность

$$P(A|H) = \frac{P(AH)}{P(H)} = \frac{2}{3}. \bullet$$

§2. Теоремы сложения и умножения вероятностей

Теорема сложения вероятностей для двух событий. Вероятность суммы двух событий A и B вычисляется по формуле:

$$P(A \cup B) = P(A) + P(B) - P(AB). \quad (3.2.1)$$

Доказательство. Воспользуемся очевидными равенствами:

$$A \cup B = A \cup (B \setminus A) \text{ и } B = (B \setminus A) \cup AB.$$

При этом $A(B \setminus A) = \emptyset$ и $(B \setminus A)AB = \emptyset$, т.е. события A и $(B \setminus A)$, а также $B \setminus A$ и AB , несовместные (непересекающиеся).

Таким образом, из аксиомы сложения находим:

$$P(A \cup B) = P(A) + P(B \setminus A) \text{ и } P(B) = P(AB) + P(B \setminus A).$$

Вычитая из первого равенства второе, получаем:

$$P(A \cup B) - P(B) = P(A) - P(AB).$$

Отсюда получаем:

$$P(A \cup B) = P(A) + P(B) - P(AB). \blacksquare$$

Теорема сложения вероятностей для нескольких событий. Вероятность суммы нескольких событий выражается формулой:

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{\substack{i \neq j \\ i, j=1}}^n P(A_i A_j) + \sum_{\substack{i \neq j \neq k \\ i, j, k=1}}^n P(A_i A_j A_k) - \dots + (-1)^{n+1} P(A_1 A_2 \dots A_n). \quad (3.2.2)$$

Теорема умножения вероятностей. Вероятность произведения двух событий равна вероятности одного из них, умноженной на условную вероятность другого, при наличии первого:

$$P(AH) = P(H)P(A|H) = P(A)P(H|A). \quad (3.2.3)$$

Доказательство. Для доказательства (3.2.3) достаточно умножить обе части равенства (3.1.1) на вероятность условия.

Теорема умножения вероятностей для нескольких событий.

$$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdot \dots \cdot P(A_n|A_1 A_2 \dots A_{n-1}). \quad (3.2.4)$$

Пример 2. Бросают две монеты. Рассматриваются два события:

A — выпадение «герба» на первой монете;

B — выпадение «герба» на второй монете.

Найти вероятность события $A \cup B$.

О Решение. Очевидно, что пространство элементарных исходов состоит из четырех исходов: «герб»-«герб», «герб»-«решка», «решка»-«герб», «решка»-«решка».

Применим теорему сложения вероятностей $P(A \cup B) = P(A) + P(B) - P(AB)$.

Очевидно, что $P(A) = P(B) = \frac{1}{2}$ и $P(AB) = \frac{1}{4}$, так как событию AB благоприятствует всего один исход, а число возможных исходов равно 4. Окончательно получим:

$$P(A \cup B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

Заметим, что задачу можно решить с помощью противоположного события. Рассмотрим событие $\overline{A \cup B}$ — выпадение пары «решка»-«решка», тогда

$$P(A \cup B) = 1 - P(\overline{A \cup B}) = 1 - \frac{1}{4} = \frac{3}{4}. \bullet$$

Пример 3. В урне a белых и b черных шаров. Из урны вынимаются два шара. Найти вероятность того, что оба шара будут белыми.

О Решение. Рассмотрим события:

A — первый шар белый;

B — второй шар белый.

Применяя теорему умножения вероятностей, получаем:

$$P(AB) = P(A)P(B|A) = \frac{a}{a+b} \cdot \frac{a-1}{a+b-1}.$$

$P(B|A) = \frac{a-1}{a+b-1}$, так как общее число шаров, а также число белых, уменьшилось на 1. $\bullet$

Пример 4. На семи карточках написаны буквы, образующие слово «телефон». После перестановки карточек наудачу последовательно берут пять из них, и прикладывают справа одну к другой. Найти вероятность образования слова «фенол».

О Решение. Применим теорему умножения вероятностей для нескольких событий. Вероятность $P(\Phi)$ того, что первой буквой будет «Ф», равна $\frac{1}{7}$. Вероятность $P(E|\Phi)$ того, что второй буквой будет «Е», при условии, что букву «Ф» уже взяли, равна $\frac{2}{6}$ и т.д. В итоге получаем:

$$P(\Phi E H O L) = P(\Phi)P(E|\Phi)P(H|\Phi E)P(O|\Phi E H)P(L|\Phi E H O) = \frac{1}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{1260}. \bullet$$

Пример 5. В урне a белых и b черных шаров. Из урны вынимаются два шара. Найти вероятность того, что эти шары будут разных цветов.

О Решение. Рассмотрим события:

A_1 — первый шар белый;

A_2 — второй шар белый;

B_1 — первый шар черный;

B_2 — второй шар черный;

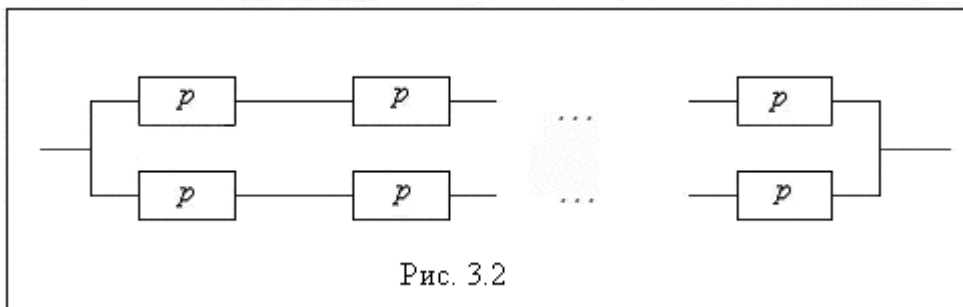
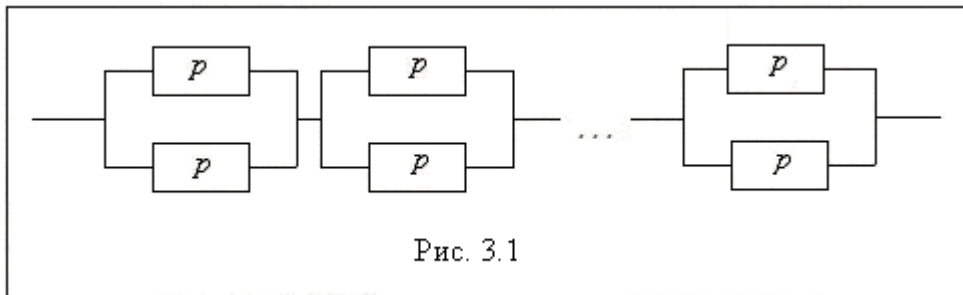
C — шары разных цветов.

Очевидно, что $C = A_1B_2 \cup B_1A_2$, причем события A_1B_2 и B_1A_2 несовместимы. По теореме сложения и по теореме умножения вероятностей для независимых событий:

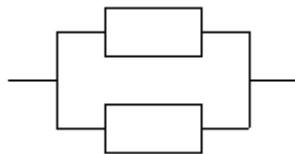
$$\begin{aligned} P(C) &= P(A_1B_2 \cup B_1A_2) = P(A_1B_2) + P(B_1A_2) = P(A_1)P(B_2) + P(B_1)P(A_2) = \\ &= \frac{a}{a+b} \frac{b}{a+b-1} + \frac{b}{a+b} \frac{a}{a+b-1} = \frac{2ab}{(a+b)(a+b-1)}. \end{aligned}$$

Пример 6. Техническая система состоит из n элементов, надежность каждого из них p . Выход из строя хотя бы одного влечет за собой выход всей системы. С целью повышения надежности системы производится дублирование, для чего выделено еще n таких же приборов. Определить, какой из способов дублирования надежнее:

- дублирование каждого элемента (рис. 3.1);
- дублирование всей системы (рис. 3.2).



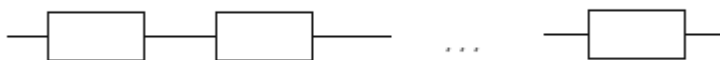
○ **Решение.** Найдем надежность блока:



Для этого найдем вероятность выхода из строя. Блок выходит из строя, если выходит из строя каждый элемент, т.е. $(1-p)^2$. Тогда надежность блока равна $1 - (1-p)^2$. Далее система работает надежно, если работает каждый блок, т.о.:

$$p_1 = (1 - (1-p)^2)^n = p^n (2-p)^n.$$

Надежность системы



равна p^n . Отсюда надежность системы (рис. 3.2)

$$p_2 = 1 - (1 - p^n)^2 = p^n(2 - p^n).$$

Сравним p_1 и p_2 , для этого нужно сравнить $(2 - p)^n$ и $(2 - p^n)$.

Докажем, что $(2 - p)^n > (2 - p^n)$.

Подставим $q = 1 - p$, т.е. $p = 1 - q$.

$$(1 + q)^n > 2 - (1 - q)^n,$$

$$(1 + q)^n + (1 - q)^n > 2.$$

Далее достаточно раскрыть скобки.

Таким образом, надежнее дублирование каждого элемента. ●

§3. Независимость событий

Определение. События A и B называются **независимыми**, если условная вероятность события A при условии B совпадает с безусловной вероятностью события A , т.е.

$$P(A|B) = P(A). \quad (3.3.1)$$

Можно сформулировать и другое определение независимых событий A и B .

Определение. События A и B независимы, если

$$P(AB) = P(A)P(B). \quad (3.3.2)$$

Очевидно, что данные два определения равносильны.

Пример 7. Пусть события A и B независимы. Доказать, что независимыми являются пары событий $\bar{A}$ и B , A и $\bar{B}$, $\bar{A}$ и $\bar{B}$.

О Решение. Применяя определение независимости событий, и используя вероятность противоположного события, имеем

$$P(\bar{B}|A) = 1 - P(B|A) = 1 - P(B) = P(\bar{B}), \text{ т.е. } P(\bar{B}|A) = P(\bar{B});$$

$$\begin{aligned} P(\bar{A}|\bar{B}) &= \frac{P(\bar{A} \cdot \bar{B})}{P(\bar{B})} = \frac{P(\overline{A \cup B})}{P(\bar{B})} = \frac{1 - P(A \cup B)}{1 - P(B)} = \frac{1 - P(A) - P(B) + P(AB)}{1 - P(B)} = \\ &= \frac{1 - P(A) - P(B) + P(A)P(B)}{1 - P(B)} = 1 - P(A) = P(\bar{A}), \text{ т.е. } P(\bar{A}|\bar{B}) = P(\bar{A}); \end{aligned}$$

$$P(\bar{A}|B) = 1 - P(A|B) = 1 - P(A) = P(\bar{A}), \text{ т.е. } P(\bar{A}|B) = P(\bar{A}). \quad \bullet$$

Пример 8. Зависимы или независимы несовместные события.

О Решение. Пусть события A и B несовместные, т.е. $A \cap B = \emptyset$, причем $P(B) \neq 0$. Тогда $P(A|B) = 0$, т.к. события не пересекаются. Следовательно A и B зависимы.

Таким образом, несовместные события зависимы. ●

Пример 9. Из полной колоды карт (52 листа) вынимается одна карта. Рассматриваются следующие события:

A — появление туза;

B — появление карты красной масти;

C — появление бубнового туза.

Зависимы или независимы следующие пары событий: 1) A и B , 2) B и C , 3) A и C ?

О Решение.

$$P(A) = \frac{4}{52} = \frac{1}{13}, \quad P(A|B) = \frac{2}{26} = \frac{1}{13}, \text{ следовательно, события независимы;}$$

$P(C) = \frac{1}{52}$, $P(C|B) = \frac{1}{26}$, следовательно, события зависимы;

$P(C) = \frac{1}{52}$, $P(C|A) = \frac{1}{4}$, следовательно, события зависимы. ●

Определение. События $A_1, A_2, \dots, A_n$ независимы в совокупности, если для всех

$1 \leq i_1 \leq i_2 \leq \dots \leq i_m \leq n$, $m \leq n$, выполнено равенство

$$P\left(\bigcap_{k=1}^m A_{i_k}\right) = \prod_{k=1}^m P(A_{i_k}).$$

Замечание. Из попарной независимости событий A_i и A_j ($i \neq j$, $i, j = 1, n$) не следует, что события $A_1, A_2, \dots, A_n$ независимы в совокупности.

Пример 10. Пусть эксперимент состоит в выборе одного из четырех шаров. Пусть три из них пронумерованы цифрами 1, 2, 3, а на четвертом шаре имеются все эти цифры. Обозначим через A_i ($i = 1, 2, 3$) событие, состоящее в том, что на выбранном шаре имеется цифра i . Зависимы ли события A_1 , A_2 и A_3 .

○ **Решение.** Так как, каждая цифра встречается дважды, то

$$P(A_1) = P(A_2) = P(A_3) = \frac{2}{4} = \frac{1}{2}.$$

Так как две различные цифры присутствуют только на одном шаре, то

$$P(A_1 A_2) = P(A_2 A_3) = P(A_1 A_3) = \frac{1}{4},$$

следовательно, события A_1 , A_2 и A_3 попарно независимы.

Все три различные цифры присутствуют только на одном шаре

$$P(A_1 A_2 A_3) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3) = \frac{1}{8}.$$

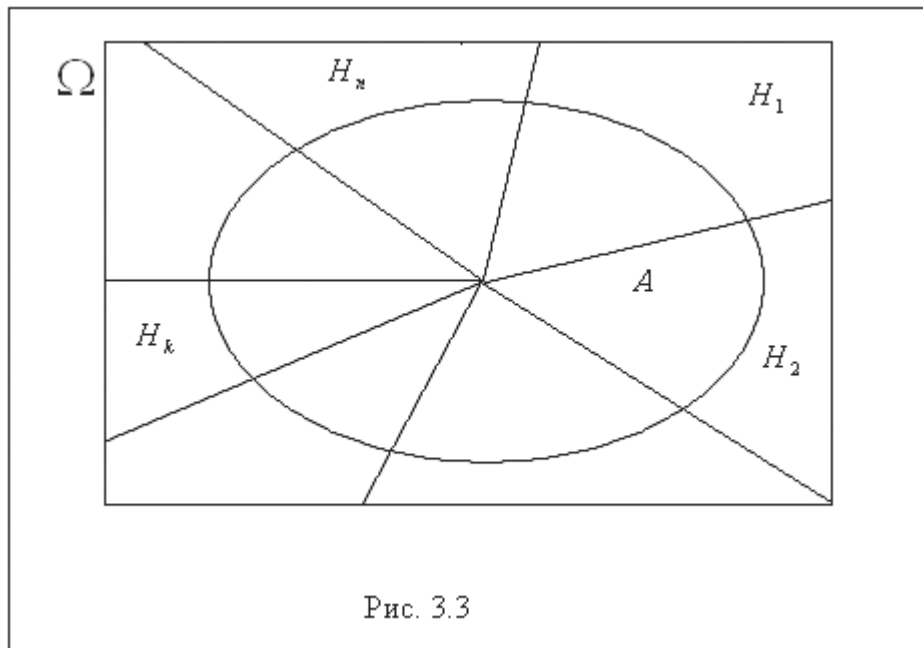
Таким образом, получаем, что события A_1 , A_2 и A_3 зависимы в совокупности, в то время как они являются попарно независимыми. ●

§4. Формула полной вероятности

Определение. События $H_1, \dots, H_n$ образуют полную группу несовместных событий (являются гипотезами), если они удовлетворяют двум требованиям:

- они попарно несовместны, т.е. $H_i H_j = \emptyset$ при $i \neq j$;
- в результате опыта одно из событий обязательно должно произойти, т.е. $H_1 \cup H_2 \cup \dots \cup H_n = \Omega$.

Пусть имеется некоторое событие A и известны вероятности $P(H_1), \dots, P(H_n)$ и условные вероятности $P(A|H_1), \dots, P(A|H_n)$. Найдем вероятность $P(A)$.



Событие A можно представить в виде (рис. 3.3):

$$A = A\Omega = A(H_1 \cup H_2 \cup \dots \cup H_n) = AH_1 \cup AH_2 \cup \dots \cup AH_n,$$

причем события $(AH_i)(AH_j) = \emptyset$ при $i \neq j$, т.е. события AH_i и AH_j несовместны.

Тогда по аксиоме сложения:

$$P(A) = P(AH_1) + P(AH_2) + \dots + P(AH_n).$$

Далее, применяя теорему умножения вероятностей $P(AH_i) = P(H_i)P(A|H_i)$, получаем:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + \dots + P(H_n)P(A|H_n). \quad (3.4.1)$$

Это и есть **формула полной вероятности**.

Пример 11. Имеются две урны: в первой a белых и b черных шаров; во второй c белых и d черных шаров. Из первой урны во вторую наудачу перекладывают один шар. После этого из второй урны берут один шар. Найти вероятность того, что этот шар будет белым.

Решение. Пусть искомое событие A — вынут белый шар. Рассмотрим следующие гипотезы:

H_1 — переложен белый шар;

H_2 — переложен черный шар.

Очевидно, что

$$P(H_1) = \frac{a}{a+b}, \quad P(H_2) = \frac{b}{a+b};$$

$$P(A|H_1) = \frac{c+1}{c+d+1}, \quad P(A|H_2) = \frac{c}{c+d+1}.$$

Теперь по формуле полной вероятности (3.4.1) получаем:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \frac{a}{a+b} \frac{c+1}{c+d+1} + \frac{b}{a+b} \frac{c}{c+d+1}. \bullet$$

Пример 12. В условиях предыдущей задачи из первой урны перекладывают сразу три шара (предполагается, что $a \geq 3$ и $b \geq 3$). Найти вероятность того, что шар, взятый из второй урны, будет белым.

О Решение. Пусть искомое событие A — вынут белый шар. Рассмотрим гипотезы:

H_1 — вынутый шар принадлежит 1-ой урне;

H_2 — вынутый шар принадлежит 2-ой урне.

Так как во второй урне 3 шара принадлежат 1-ой урне, а $c + d$ принадлежат 2-ой, то вероятности гипотез равны:

$$P(H_1) = \frac{3}{c+d+3}, \quad P(H_2) = \frac{c+d}{c+d+3}.$$

Вероятность появления белого шара из первой урны не зависит от того, вынимается ли этот шар непосредственно из первой урны или после перекладывания во вторую. Следовательно, условная вероятность появления белого шара при условии того, что он изначально находился в первой урне, равна:

$$P(A|H_1) = \frac{a}{a+b}.$$

Аналогично условная вероятность появления белого шара при условии того, что он изначально находился во второй урне, равна:

$$P(A|H_2) = \frac{c}{c+d}.$$

По формуле полной вероятности (3.4.1) получаем:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \frac{3}{c+d+3} \frac{a}{a+b} + \frac{c+d}{c+d+3} \frac{c}{c+d}. \bullet$$

Пример 13. Среди 30 экзаменационных билетов: 25 «хороших» и 5 «плохих». Какова вероятность, отвечая вторым, взять «хороший» билет?

О Решение. Пусть искомое событие A — второй отвечающий взял «хороший» билет. Рассмотрим следующие гипотезы:

H_1 — первый отвечающий взял «хороший» билет;

H_2 — первый отвечающий взял «плохой» билет.

Очевидно, что

$$P(H_1) = \frac{25}{30} = \frac{5}{6}, \quad P(H_2) = \frac{5}{30} = \frac{1}{6};$$

$$P(A|H_1) = \frac{24}{29}, \quad P(A|H_2) = \frac{25}{29}.$$

По формуле полной вероятности (3.4.1), получим:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \frac{5}{6} \cdot \frac{24}{29} + \frac{1}{6} \cdot \frac{25}{29} = \frac{5}{6}. \bullet$$

§5. Формула Байеса

Пусть имеется полная группа несовместных событий $H_1, \dots, H_n$. Требуется найти вероятность события H_i , если известно, что событие A произошло.

По определению условной вероятности (3.4.1), имеем:

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

Далее, применяя теорему умножения вероятностей $P(AH_i) = P(H_i)P(A|H_i)$, получаем

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}. \quad (3.5.1)$$

Последняя формула называется формулой Байеса или формулой гипотез (события $H_1, \dots, H_n$ называют еще гипотезами).

Если после опыта, который заканчивается появлением события A , производится еще один опыт, в котором появляется или не появляется событие B , то условная вероятность этого последнего события вычисляется по формуле полной вероятности, в которую подставлены не прежние вероятности гипотез $P(H_i)$, а новые $P(H_i|A)$:

$$P(B|A) = \sum_{i=1}^n P(H_i|A)P(B|H_iA). \quad (3.5.2)$$

Пример 14. Имеются три урны: в первой — a белых и b черных шаров; во второй — c белых и d черных шаров, в третьей — k белых шаров. Выбирается наугад урна и из нее вынимается шар. Этот шар оказался белым. Найти вероятность того, что этот шар вынут из первой, второй или третьей урны.

О Решение. Пусть искомое событие A — вынутый шар белый. Рассмотрим следующие гипотезы:

H_1 — выбрана первая урна;

H_2 — выбрана вторая урна;

H_3 — выбрана третья урна.

Очевидно, что:

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}.$$

Условные вероятности равны:

$$P(A|H_1) = \frac{a}{a+b}, \quad P(A|H_2) = \frac{c}{c+d}, \quad P(A|H_3) = 1.$$

По формуле полной вероятности (3.4.1), находим, что

$$P(A) = \sum_{i=1}^3 P(H_i)P(A|H_i) = \frac{1}{3} \frac{a}{a+b} + \frac{1}{3} \frac{c}{c+d} + \frac{1}{3} \cdot 1.$$

По формуле Байеса (3.5.1), находим:

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(A)} = \frac{\frac{1}{3} \frac{a}{a+b}}{\frac{1}{3} \left(\frac{a}{a+b} + \frac{c}{c+d} + 1 \right)} = \frac{\frac{a}{a+b}}{\frac{a}{a+b} + \frac{c}{c+d} + 1}.$$

Аналогично получаем, что

$$P(H_2|A) = \frac{\frac{c}{c+d}}{\frac{a}{a+b} + \frac{c}{c+d} + 1}, \quad P(H_3|A) = \frac{1}{\frac{a}{a+b} + \frac{c}{c+d} + 1}. \quad \bullet$$

Пример 15. Имеются две урны: в первой — a белых и b черных шаров; во второй — c белых и d черных шаров. Выбирается наугад одна из урн и из нее вынимается один шар. Этот шар оказался белым (событие A). найти вероятность того, что следующий шар, который мы вынимаем из той же урны, будет тоже белым (событие B).

О Решение. Рассмотрим следующие гипотезы:

H_1 — выбрана первая урна;

H_2 — выбрана вторая урна.

Очевидно, что вероятности выбора урн равны:

$$P(H_1) = P(H_2) = \frac{1}{2}.$$

Находим условные вероятности:

$$P(A|H_1) = \frac{a}{a+b}, \quad P(A|H_2) = \frac{c}{c+d}.$$

По формуле полной вероятности (3.4.1), получаем:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = \frac{1}{2} \left(\frac{a}{a+b} + \frac{c}{c+d} \right).$$

По формуле Байеса (3.5.1), получаем:

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(A)} = \frac{\frac{a}{a+b}}{\frac{a}{a+b} + \frac{c}{c+d}}, \quad P(H_2|A) = \frac{\frac{c}{c+d}}{\frac{a}{a+b} + \frac{c}{c+d}}.$$

Далее применяем (3.5.2):

$$P(B|A) = P(H_1|A)P(B|H_1A) + P(H_2|A)P(B|H_2A).$$

Условная вероятность появления второго белого шара при условии, что была выбрана первая урна, и из нее вынут белый шар:

$$P(B|H_1A) = \frac{a-1}{a+b-1}.$$

Аналогично:

$$P(B|H_2A) = \frac{c-1}{c+d-1}.$$

В итоге:

$$P(B|A) = \frac{1}{\frac{a}{a+b} + \frac{c}{c+d}} \left(\frac{a(a-1)}{(a+b)(a+b-1)} + \frac{c(c-1)}{(c+d)(c+d-1)} \right). \bullet$$

Глава 4. Схема независимых испытаний. Схема Бернулли

§1. Формула Бернулли

Определение. Повторные независимые испытания называются **испытаниями Бернулли**, если каждое испытание имеет только два исхода, и вероятности исходов остаются неизменными для всех испытаний.

Обычно эти две вероятности обозначаются через p и q , исход с вероятностью p называют «успехом» и обозначают символом 1, а второй – «неудачей» и обозначают символом 0. Очевидно, что p и q должны быть неотрицательными и должно выполняться равенство

$$p + q = 1. \quad (4.1.1)$$

Пространство элементарных исходов каждого отдельного испытания состоит из двух исходов 1 и 0. Очевидно, пространство элементарных исходов n испытаний Бернулли содержит 2^n последовательностей из n символов 1 и 0. Так как испытания независимы, то вероятности перемножаются, т. е. вероятность любой конкретной последовательности есть произведение, полученное при замене символов 1 и 0 вероятности на p и q соответственно. Таким образом, вероятность исхода (11001...0111) равна:

$$P(11001...0111) = p p q q r \dots q r r r r.$$

Но на практике нас, как правило, интересует не порядок появления успехов в последовательности n испытаний Бернулли, а их общее число.

Теорема. Вероятность $p_n(m)$ того, что в n испытаниях Бернулли число успехов равно m , вычисляется по формуле

$$p_n(m) = C_n^m p^m q^{n-m}, \quad (4.1.2)$$

где p — вероятность «успеха», а q — вероятность «неудачи».

Доказательство. Событие «в n испытаниях Бернулли число успехов равно m и число неудач — $n - m$ » содержит столько элементарных исходов, сколько существует способов размещения m символов на n местах, т.е. C_n^m . А так как вероятность конкретной последовательности, содержащей m символов 1, равна $p^m q^{n-m}$, то в итоге получаем:

$$p_n(m) = C_n^m p^m q^{n-m}. \blacksquare$$

Число успехов в n испытаниях обозначают через S_n , тогда $p_n(m) = P(S_n = m)$. Очевидно, что S_n есть случайная величина, а функция (4.1.2) является «распределением» этой случайной величины. Будем называть это распределение биномиальным. Слово биномиальное отражает тот факт, что (4.1.2) представляет собой m -й член биномиального разложения $(p + q)^n$. Отсюда следует, что

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

Пример 1. Стрелок попадает в мишень с вероятностью $p = 0,8$. Найти вероятность того, что в результате пяти независимых выстрелов стрелок попадает:

- а) ровно четыре раза;
- б) не менее трех раз.

Решение. Для решения данной задачи применим формулу (4.1.2), в которой: $n = 5$; $p = 0,8$; $q = 1 - p = 1 - 0,8 = 0,2$.

- а) Число успехов равно $m = 4$. Таким образом, искомая вероятность:

$$p_5(4) = C_5^4 (0,8)^4 (0,2)^{5-4} = \frac{5!}{4!1!} (0,8)^4 (0,2)^1 = 0,4096.$$

б) Обозначим $p_5(\geq 3)$ — вероятность попадания не менее трех раз из пяти.

$$\begin{aligned} p_5(\geq 3) &= p_5(3) + p_5(4) + p_5(5) = C_5^3 (0,8)^3 (0,2)^{5-3} + C_5^4 (0,8)^4 (0,2)^{5-4} + C_5^5 (0,8)^5 (0,2)^{5-5} = \\ &= \frac{5!}{3!2!} (0,8)^3 (0,2)^2 + \frac{5!}{4!1!} (0,8)^4 (0,2) + \frac{5!}{5!0!} (0,8)^5 = 0,9415. \bullet \end{aligned}$$

Пример 2. Сколько испытаний с вероятностью успеха $p = 0,01$ нужно произвести, чтобы вероятность хотя бы одного успеха была не меньше 0,5?

○ **Решение.** Рассмотрим следующие события:

A — в схеме Бернулли наблюдался хотя бы один успех;

$\bar{A}$ — в схеме Бернулли не наблюдалось ни одного успеха.

Для решения задачи используем формулу (4.1.2), согласно которой вероятность того, что успехов не будет (т.е. число успехов равно нулю), равна:

$$P(\bar{A}) = p_n(0) = C_n^0 p^0 q^{n-0} = q^n.$$

Используя свойство вероятности противоположного события, получаем, что вероятность того, что будет хотя бы один успех, равна:

$$P(A) = 1 - P(\bar{A}) = 1 - q^n = 1 - (1 - p)^n.$$

Остается найти наименьшее целое n , для которого выполнено неравенство:

$$1 - (1 - p)^n = 1 - (1 - 0,01)^n = 1 - 0,99^n > 0,5.$$

Решим последнее неравенство.

$$0,99^n < 0,5 \Leftrightarrow \ln(0,99^n) < \ln(0,5) \Leftrightarrow n \ln(0,99) < \ln(0,5).$$

Разделив последнее неравенство на $\ln(0,99) < 0$, получим

$$n > \frac{\ln(0,5)}{\ln(0,99)} = 69,968.$$

Наименьшим целым числом n , удовлетворяющим последнему неравенству, является $n = 70$. •

Пример 3. Что вероятнее выиграть у равносильного противника (ничейный исход партии исключен):

а) три партии из четырех или пять из восьми;

б) не менее трех партий из четырех или не менее пяти партий из восьми.

○ **Решение.** Так как противники равносильны и ничейный исход партии исключен, то вероятности выигрыша и проигрыша каждой партии одинаковы и $p = q = \frac{1}{2}$.

а) Вероятность выигрыша трех партий из четырех равна:

$$p_4(3) = C_4^3 p^3 q^{4-3} = \frac{4!}{3!1!} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right) = 4 \left(\frac{1}{2}\right)^4 = \frac{1}{4},$$

а вероятность выигрыша пяти партий из восьми равна:

$$p_8(5) = C_8^5 p^5 q^{8-5} = \frac{8!}{5!3!} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^3 = \frac{6 \cdot 7 \cdot 8}{2 \cdot 3} \left(\frac{1}{2}\right)^8 = \frac{7}{32}.$$

Так как $\frac{1}{4} > \frac{7}{32}$, то вероятнее выиграть три партии из четырех.

б) Вероятность выигрыша не менее трех партий из четырех равна:

$$p_4(\geq 3) = p_4(3) + p_4(4) = C_4^3 p^3 q^{4-3} + C_4^4 p^4 q^{4-4} = \\ = \frac{4!}{3!1!} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right) + \frac{4!}{4!0!} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0 = 4 \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^4 = \frac{1}{4} + \frac{1}{16} = \frac{5}{16},$$

а вероятность выигрыша не менее пяти партий из восьми равна:

$$p_8(\geq 5) = p_8(5) + p_8(6) + p_8(7) + p_8(8) = C_8^5 p^5 q^{8-5} + C_8^6 p^6 q^{8-6} + C_8^7 p^7 q^{8-7} + C_8^8 p^8 q^{8-8} = \\ = \frac{8!}{5!3!} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^3 + \frac{8!}{6!2!} \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^2 + \frac{8!}{7!1!} \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^1 + \frac{8!}{8!0!} \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^0 = \frac{93}{256}.$$

Так как $\frac{93}{256} > \frac{5}{16}$, то вероятнее выиграть не менее пяти партий из восьми. ●

§2. Формула Пуассона

При больших значениях числа испытаний n применение формулы Бернулли (4.1.2) затруднительно. Поэтому применяются простые, но достаточно точные приближенные формулы для вычисления $p_n(m)$. Пусть число испытаний n достаточно «велико», вероятность «успеха» p достаточно «мала». Пусть произведение

$$\lambda = np \quad (4.2.1)$$

и не мало, и не велико. В таких случаях удобно использовать для вероятности $p_n(m)$ предложенное Пуассоном приближение (формула Пуассона), которое мы сейчас выведем. По формуле Бернулли (4.1.2)

$$p_n(m) = C_n^m p^m (1-p)^{n-m} = \frac{n(n-1)\dots(n-m+1)}{m!} p^m (1-p)^{n-m} = \\ = \frac{(np)^m}{m!} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-m+1}{n} \left(1 - \frac{np}{n}\right)^{n-m} = \quad (4.2.2) \\ = \frac{\lambda^m}{m!} \frac{n-1}{n} \frac{n-2}{n} \dots \frac{n-m+1}{n} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-m}.$$

При $n \rightarrow \infty$ и сделанных выше допущениях очевидны следующие приближения:

$$\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \frac{n-1}{n} \approx 1, \dots, \frac{n-m+1}{n} \approx 1, \quad \left(1 - \frac{\lambda}{n}\right)^{-m} \approx 1.$$

Следовательно, (4.2.2) примет вид:

$$p_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda}, \quad (4.2.3)$$

а это и есть **формула Пуассона**.

Замечание. При выводе формулы Пуассона (4.2.3) использовалось то, что m мало.

Замечание. Формула Пуассона (4.2.3) зависит от m и λ . Значения функции (4.2.2) можно определить следующими способами:

- можно воспользоваться Приложением 1;
- используя функцию *ПУАССОН(х;среднее;интегральная)* из EXCEL; в которой аргумент x равен числу «успехов» m , аргумент «среднее» равен λ , аргумент «интегральная» должен равняться 0;
- используя функцию *dpois(k, l)* из MATHCAD, в которой $k = m$ и $l = \lambda$.

Пример 4. Найти вероятность того, что среди 1460 человек ровно трое родились 29 февраля.

О Решение. Вероятность того, что один конкретный человек родился 29 февраля, равна $p = \frac{1}{1461}$, т.к. 29 февраля бывает ровно 1 раз в 4 года.

Далее находим коэффициент λ :

$$\lambda = np = 1460 \frac{1}{1461} \approx 1.$$

Применяя (4.2.2), получаем:

$$p_{1460}(3) \approx \frac{\lambda^3}{3!} e^{-\lambda} = \frac{e^{-3}}{6}. \bullet$$

Пример 5. Вероятность попадания в цель при каждом выстреле равна 0,001. Найти вероятность того, что при 5000 выстрелах в цель попало не менее двух выстрелов.

О Решение. Рассмотрим два противоположных события:

A — при 5000 выстрелах в цель попало не менее двух выстрелов;

$\bar{A}$ — при 5000 выстрелах в цель попало менее двух выстрелов.

Найдем вероятность события $\bar{A}$:

$$P(\bar{A}) = p_{5000}(< 2) = p_{5000}(0) + p_{5000}(1).$$

В рассматриваемом примере

$$\lambda = np = 5000 \cdot 0,001 = 5.$$

Используя формулу Пуассона, получим

$$p_{5000}(0) \approx \frac{5^0}{0!} e^{-5} = e^{-5} = 0,006738, \quad p_{5000}(1) \approx \frac{5^1}{1!} e^{-5} = 5e^{-5} = 0,03369.$$

Используя свойство вероятности противоположного события, получим

$$P(A) = 1 - P(\bar{A}) = 1 - p_{5000}(0) - p_{5000}(1) = 1 - 0,006738 - 0,03369 = 0,9596. \bullet$$

§3. Формулы Муавра – Лапласа

Если в схеме Бернулли $n \rightarrow \infty$, $np \rightarrow \infty$, $nq \rightarrow \infty$,

(4.3.1)

то следует применять формулы Муавра – Лапласа: локальную или интегральную.

Локальная теорема Муавра-Лапласа (без доказательства). Если в схеме Бернулли $n \rightarrow \infty$, то для всех m справедлива локальная формула Муавра-Лапласа:

$$p_n(m) \approx \frac{1}{\sqrt{npq}} \varphi(x),$$

(4.3.2)

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

Значения функции $\varphi(x)$, которую называют **плотностью нормального распределения** с параметрами $(0,1)$, можно найти одним из следующих способов:

- можно воспользоваться Приложением 2;
- используя функцию *НОРМРАСП(х;среднее;стандартное_откл;интегральная)* из EXCEL; в которой «среднее» необходимо положить равным 0, аргумент «стандартное_откл» необходимо положить равным 1, аргумент «интегральная» должен равняться 0.
- используя функцию *dnorm(x, mu, sigma)* из MATHCAD, в которой $mu = 0$ и $sigma = 1$.

Очевидно, что функция $\varphi(x)$ является четной. Поэтому при определении $\varphi(x)$ для отрицательных x нужно воспользоваться равенством $\varphi(x) = \varphi(-x)$.

Интегральная теорема Муавра-Лапласа (без доказательства). Если в схеме Бернулли число испытаний $n \rightarrow \infty$, то для вероятности $P(m_1 \leq S_n \leq m_2)$ того, что число успехов S_n заключено в пределах от m_1 до m_2 , справедлива интегральная теорема Муавра-Лапласа:

$$P(m_1 \leq S_n \leq m_2) \approx \Phi(x_2) - \Phi(x_1),$$

$$\Phi(x) = \int_{-\infty}^x \varphi(y) dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy, \quad x_1 = \frac{m_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{m_2 - np}{\sqrt{npq}}. \quad (4.3.3)$$

Функция $\Phi(x)$, определенная формулой (4.3.3), называется функцией распределения нормального распределения с параметрами (0,1). Значения функции $\Phi(x)$ можно найти одним из следующих способов:

- можно воспользоваться Приложением 3;
- используя функцию *НОРМРАСП(x;среднее;стандартное_откл;интегральная)* из EXCEL; в которой «среднее» необходимо положить равным 0, аргумент «стандартное_откл» необходимо положить равным 1, аргумент «интегральная» должен равняться 1.
- используя функцию *pnorm(x, mu, sigma)* из MATHCAD, в которой $mu = 0$ и $sigma = 1$.

Функцию $\Phi(x)$ при отрицательных значениях переменной можно определить по формуле $\Phi(-x) = 1 - \Phi(x)$.

Замечание. Наряду с функцией $\Phi(x)$ используют функцию

$$\Phi_0(x) = \int_0^x \varphi(y) dy = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{y^2}{2}} dy. \quad (4.3.4)$$

Для нее справедливо равенство $\Phi_0(-x) = \Phi_0(x)$; она связана с функцией $\Phi(x)$ равенством $\Phi(x) = 0,5 + \Phi_0(x)$. (4.3.5)

Пример 6. Симметричную монету бросают 400 раз. Определить вероятность появления герба:

- а) от 185 до 210 раз;
- б) ровно 200 раз;
- в) не менее 200 раз.

○ **Решение.** Для решения задачи применим локальную и интегральную теоремы Муавра-Лапласа, для которых

$n = 400$, т.к. монету подбрасывали 400 раз, $p = q = \frac{1}{2}$, т.к. монета симметрична.

а) Используя интегральную теорему Муавра-Лапласа, получим

$$P(185 \leq S_n \leq 210) = \Phi(x_2) - \Phi(x_1) = \Phi\left(\frac{210 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{185 - np}{\sqrt{npq}}\right) =$$

$$= \Phi\left(\frac{210 - 200}{\sqrt{100}}\right) - \Phi\left(\frac{185 - 200}{\sqrt{100}}\right) = \Phi(1) - \Phi(-1,5) = \Phi(1) - 1 + \Phi(1,5) = 0,7745;$$

б) Используя локальную теорему Муавра-Лапласа, получим

$$P_{400}(200) = \frac{1}{\sqrt{npq}} \varphi(x) = \frac{1}{10} \varphi\left(\frac{200 - np}{\sqrt{npq}}\right) = \frac{1}{10} \varphi(0) = 0,03989;$$

в) Используя интегральную теорему Муавра-Лапласа, получим

$$\begin{aligned}
P_{400}(\geq 200) &= P(200 \leq S_n \leq 400) = \Phi\left(\frac{400 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{200 - np}{\sqrt{npq}}\right) = \\
&= \Phi\left(\frac{400 - 200}{\sqrt{100}}\right) - \Phi\left(\frac{200 - 200}{\sqrt{100}}\right) = \Phi(20) - \Phi(0) = 1 - 0,5 = 0,5. \bullet
\end{aligned}$$

Пример 7. Команда состоит из 10 отличных и 15 хороших стрелков. Каждый стрелок производит по своей мишени 5 независимых выстрелов. Отличный стрелок при каждом выстреле попадает в цель с вероятностью 0,9, хороший — с вероятностью 0,8. Определить вероятность того, что общее число попаданий будет не менее 110.

О Решение. Найдем вероятность попадания при одном выстреле для произвольного стрелка. Для этого воспользуемся формулой полной вероятности. Пусть искомое событие A — мишень поражена одним стрелком. Рассмотрим следующие гипотезы:

H_1 — стреляет отличный стрелок;

H_2 — стреляет хороший стрелок.

Очевидно, что:

$$P(H_1) = \frac{10}{10+15} = \frac{2}{5}, \quad P(H_2) = \frac{15}{10+15} = \frac{3}{5}, \quad P(A|H_1) = 0,9, \quad P(A|H_2) = 0,8.$$

Отсюда получаем:

$$p = P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = 0,84, \quad q = 1 - p = 1 - 0,84 = 0,16.$$

Заметим, что общее число выстрелов

$$N = 10 \cdot 5 + 15 \cdot 5 = 125.$$

Теперь найдем вероятность $p_{125}(\geq 110)$ того, что при 125 выстрелах число попаданий будет не менее 110. Для этого применим интегральную теорему Муавра-Лапласа:

$$\begin{aligned}
x_1 &= \frac{110 - np}{\sqrt{npq}} \approx 1,25, \quad x_2 = \frac{125 - np}{\sqrt{npq}} \approx 5, \\
p_{125}(\geq 100) &= P(110 \leq S_n \leq 125) = \Phi(x_2) - \Phi(x_1) = 1 - 0,89 = 0,11. \bullet
\end{aligned}$$

Пример 8. Вероятность попадания в мишень при одном выстреле $p = 0,6$. Найти наименьшее число выстрелов, которое надо произвести по мишени, чтобы с вероятностью 0,95 число попаданий было не менее 70.

О Решение. По условию задачи $0,95 = p_n(\geq 70)$. Для вычисления $p_n(\geq 70)$ применим интегральную теорему Муавра – Лапласа:

$$\begin{aligned}
0,95 &= p_n(\geq 70) = P(70 \leq S_n \leq n) = \Phi\left(\frac{n - np}{\sqrt{npq}}\right) - \Phi\left(\frac{70 - np}{\sqrt{npq}}\right) = \\
&= \Phi\left(\sqrt{\frac{2n}{3}}\right) - \Phi\left(\frac{70 - 0,6n}{\sqrt{0,24n}}\right) = 1 - \Phi\left(\frac{70 - 0,6n}{\sqrt{0,24n}}\right).
\end{aligned}$$

Заметим, что мы использовали то, что при больших значениях x ($x \geq 3$) $\Phi(x) \approx 1$.

Далее получаем

$$\Phi\left(\frac{70 - 0,6n}{\sqrt{0,24n}}\right) = 0,05.$$

Используя Приложение 3 находим, что

$$\frac{70 - 0,6n}{\sqrt{0,24n}} = -1,65.$$

Решая последнее уравнение для натуральных значений n , получаем, что $n=132$. $\bullet$

Глава 5. Случайные величины и их распределения

§1. Понятие случайной величины

Определение. Случайной величиной ξ называется функция, ставящая в соответствие каждому элементарному исходу ω число $\xi = \xi(\omega)$.

Пример 1. Рассмотрим схему Бернулли с вероятностью успеха p . Сопоставим каждому элементарному исходу $\omega = 10100\dots 0$ функцию $\xi(\omega)$, равную числу успехов данного исхода, т.е. числу символов 1, содержащихся в последовательности $10100\dots 0$.

Пример 2. Случайными величинами является число очков при бросании игральной кости, сумма очков при бросании нескольких игральных костей.

Пример 3. Случайной величиной является координата точки, упавшей на отрезок $[0;1]$. В этом случае $\xi(\omega) = \omega$.

В общем случае случайные величины делятся на дискретные и непрерывные.

Определение. Дискретной называется случайная величина, которая каждому элементарному исходу ω ставит в соответствие одно из конечного (или в общем случае счетного) набора чисел $x_1, x_2, \dots, x_n$ ($x_1, x_2, \dots, x_n, \dots$).

Определение. Непрерывной называется случайная величина, возможные значения которой непрерывно заполняют некоторую область.

Однако существуют случайные величины, не относящиеся ни к одному из этих типов. Простейшим примером такой случайной величины является время работы электроприбора. Купленный электроприбор может с ненулевой вероятностью оказаться бракованным, т.е. время его работы будет равно 0, и в этом смысле необходимо считать рассматриваемую случайную величину дискретной. Если же прибор окажется исправным, то время его работы необходимо считать непрерывной случайной величиной.

§2. Функция распределения случайной величины

Определение. Функцией распределения (вероятностей) случайной величины ξ называется функция $F(x)$, значения которой в точке x равно вероятности события $(\xi < x)$:

$$F(x) = P(\xi < x). \quad (5.2.1)$$

Свойства функции распределения.

1. Функция распределения является ограниченной, т.е.

$$0 \leq F(x) \leq 1.$$

Доказательство. Ограниченность функции распределения следует из того, что функция распределения является вероятностью. ■

2. Функция распределения является неубывающей, т.е. если $x_2 > x_1$, то

$$F(x_2) \geq F(x_1).$$

Доказательство. Если $x_2 > x_1$, то событие $(\xi < x_1)$ содержится в событии $(\xi < x_2)$, т.е. $(\xi < x_1) \subset (\xi < x_2)$. Отсюда, по свойству 3 вероятности, имеем

$$P(\xi < x_1) \leq P(\xi < x_2),$$

откуда, следует

$$F(x_1) \leq F(x_2). \quad \blacksquare$$

3. $F(x)$ обращается в ноль на минус бесконечности, т.е.

$$F(-\infty) = 0.$$

Доказательство. Событие $(\xi < -\infty)$ является невозможным событием, следовательно, $P(\xi < -\infty) = 0$. ■

4. $F(x)$ равна единице в плюс бесконечности, т.е.

$$F(+\infty) = 1.$$

Доказательство. Событие $(\xi < +\infty)$ является достоверным событием, следовательно, $P(\xi < +\infty) = 1$.

5. Вероятность попадания случайной величины ξ в интервал $[x_1, x_2)$ равна приращению ее функции распределения на этом промежутке, т.е.

$$P(x_1 \leq \xi < x_2) = F(x_2) - F(x_1).$$

Доказательство. При $x_1 < x_2$ событие $(\xi < x_2)$ можно представить как объединение двух непересекающихся событий $(\xi < x_1)$ и $(x_1 \leq \xi < x_2)$. Отсюда, используя аксиому сложения, получаем

$$P(\xi < x_2) = P(\xi < x_1) + P(x_1 \leq \xi < x_2),$$

$$P(x_1 \leq \xi < x_2) = P(\xi < x_2) - P(\xi < x_1)$$

или, используя определение функции распределения, получаем

$$P(x_1 \leq \xi < x_2) = F(x_2) - F(x_1). \blacksquare$$

6. $F(x)$ непрерывна слева, т.е.

$$\lim_{x \rightarrow x_0 - 0} F(x) = F(x_0).$$

С помощью функции распределения можно вычислить вероятность события $(\xi \geq x)$:

$$P(\xi \geq x) = 1 - F(x). \quad (5.2.2)$$

Иногда, чтобы подчеркнуть, кокой именно случайной величине принадлежит функция распределения $F(x)$, к функции распределения приписывают нижний индекс, обозначающий эту случайную величину, т.е.

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

Иногда функцией распределения называется вероятность события $(\xi \leq x)$. Такое определение ничего не меняет во всех рассуждениях. Единственное изменение касается свойства 6: функция $F(x)$ будет непрерывной справа.

§3. Дискретные случайные величины

Как уже говорилось, дискретной называется случайная величина, которая каждому элементарному исходу ω ставит в соответствие одно из конечного (или в общем случае счетного) набора чисел $x_1, x_2, \dots, x_n$ ($x_1, x_2, \dots, x_n, \dots$).

Дискретную случайную величину удобно характеризовать рядом распределения.

Определение. **Рядом распределения (вероятностей)** дискретной случайной величины называется таблица (таблица 5.1), состоящая из двух строк: в первой строке перечислены все возможные значения случайной величины, а во второй — вероятности $p_i = P(\xi = x_i)$ того, что случайная величина ξ примет значение x_i .

Таблица 5.1

ξ	x_1	x_2	$\dots$	x_i	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_i	$\dots$	p_n	$\dots$

При этом должно выполняться равенство

$$\sum_i p_i = 1. \quad (5.3.1)$$

Функцию распределения дискретной случайной величины можно определить по формуле

$$F(x) = \sum_{x_i < x} p_i. \quad (5.3.2)$$

Пример 4. Производится один опыт, в результате которого может произойти событие A с вероятностью p . Рассмотрим случайную величину

$$\xi = \begin{cases} 1, & \text{если событие } A \text{ произошло,} \\ 0, & \text{если событие } A \text{ не произошло.} \end{cases}$$

Для случайной величины ξ ряд распределения и найти ее функцию распределения.

○ **Решение.** Очевидно, что ряд распределения имеет вид:

ξ	0	1
P	q	p

где $q = 1 - p$.

Функцию распределения случайной величины ξ найдем по формуле (5.3.2).

Пусть $x \leq 0$. В этом случае событие $(\xi < x)$ является невозможным, так как случайная величина ξ не принимает значений меньших 0. Отсюда получаем:

$$F(x) = \sum_{x_i < x} p_i = 0.$$

Пусть $0 < x \leq 1$. В этом случае событие $(\xi < x)$ совпадает с событием $(\xi = 0)$, следовательно:

$$F(x) = \sum_{x_i < x} p_i = P(\xi = 0) = q.$$

Пусть $x > 1$. В этом случае событие $(\xi < x)$ является достоверным, следовательно

$$F(x) = \sum_{x_i < x} p_i = P(\xi = 0) + P(\xi = 1) = q + p = 1.$$

Таким образом, функция распределения примет вид:

$$F(x) = \begin{cases} 0, & \text{если } x \leq 0; \\ 1 - p, & \text{если } 0 < x \leq 1; \\ 1, & \text{если } x > 1. \end{cases}$$

Пример 5. Рассмотрим схему последовательных независимых испытаний, в каждом из которых с вероятностью p может произойти событие A . Рассмотрим случайную величину ξ — число испытаний, которое необходимо произвести, прежде чем событие A произойдет. Построить ряд распределения случайной величины ξ .

○ **Решение.** Случайная величина ξ может принимать значения $0, 1, \dots, n, \dots$. Случайная величина ξ принимает значение 0, если в первом же испытании произойдет событие A , следовательно:

$$P(\xi = 0) = p.$$

Случайная величина ξ принимает значение 1, если в первом испытании событие A не произошло, а во втором испытании событие A произойдет, следовательно,

$$P(\xi = 1) = qp.$$

Случайная величина ξ принимает значение 2, если в первых двух испытаниях событие A не произошло, а в третьем испытании событие A произойдет, следовательно,

$$P(\xi = 2) = qqp = q^2 p.$$

Продолжая аналогично данные рассуждения, получим ряд распределения:

ξ	0	1	2	...	n	...
P	p	qp	$q^2 p$	...	$q^n p$	...

Пример 6. На зачете студент получил $n = 4$ задачи. Вероятность решить правильно каждую задачу $p = 0,8$. Построить ряд распределения случайной величины ξ — числа правильно решенных задач.

О Решение. Случайная величина ξ может принимать значения 0, 1, 2, 3, 4. Для нахождения вероятности событий $(\xi = k)$ ($k = 0, 1, 2, 3, 4$) применим формулу Бернулли:

$$P(\xi = 0) = p_4(0) = C_4^0 p^0 (1-p)^4 = (1-0,8)^4 = 0,0016,$$

$$P(\xi = 1) = p_4(1) = C_4^1 p^1 (1-p)^3 = \frac{4!}{3!1!} (0,8)^1 (1-0,8)^3 = 0,0256,$$

$$P(\xi = 2) = p_4(2) = C_4^2 p^2 (1-p)^2 = \frac{4!}{2!2!} (0,8)^2 (1-0,8)^2 = 0,1536,$$

$$P(\xi = 3) = p_4(3) = C_4^3 p^3 (1-p)^1 = \frac{4!}{1!3!} (0,8)^3 (1-0,8)^1 = 0,4096,$$

$$P(\xi = 4) = p_4(4) = C_4^4 p^4 (1-p)^0 = 0,8^4 = 0,4096.$$

Таким образом, ряд распределения числа правильно решенных задач примет вид:

ξ	0	1	2	3	4
P	0,0016	0,0256	0,1536	0,4096	0,4096

§4. Непрерывные случайные величины

Как уже говорилось, непрерывной называется случайная величина, возможные значения которой непрерывно заполняют какую-то область. Дадим более строгое определение непрерывной случайной величины.

Определение. Случайная величина ξ называется **непрерывной**, если ее функцию распределения $F(x)$ можно представить в виде:

$$F(x) = \int_{-\infty}^x p(y) dy. \quad (5.4.1)$$

Определение. Функция $p(x)$, присутствующая в (5.4.1), называется **плотностью распределения** (вероятностей) случайной величины ξ .

Отметим, что все реально встречающиеся плотности распределения являются непрерывными (за исключением, быть может, конечного числа точек) функциями, и, следо-

вательно, для них плотность распределения $p(x)$ представляет собой производную функции распределения $F(x)$, т. е.

$$p(x) = F'(x). \quad (5.4.2)$$

Свойства плотности распределения:

1. Плотность является неотрицательной функцией, т.е.

$$p(x) \geq 0.$$

Доказательство. Функция распределения — неубывающая функция. Следовательно, ее производная, которая по (5.4.2.) является плотностью, есть неотрицательная функция. ■

2. Вероятность попадания случайной величины ξ в интервал $[x_1, x_2)$ равна определенному интегралу от ее плотности в пределах от x_1 до x_2 , т.е.

$$P(x_1 \leq \xi < x_2) = \int_{x_1}^{x_2} p(y) dy.$$

Доказательство. С одной стороны, по свойству 5 функции распределения

$$P(x_1 \leq \xi < x_2) = F(x_2) - F(x_1),$$

с другой стороны, в силу (5.4.2):

$$\int_{x_1}^{x_2} p(y) dy = \int_{x_1}^{x_2} dF(x) = F(x) \Big|_{x_1}^{x_2} = F(x_2) - F(x_1).$$

3. Условие нормировки: несобственный интеграл от плотности случайной величины в бесконечных пределах равен единице, т.е.

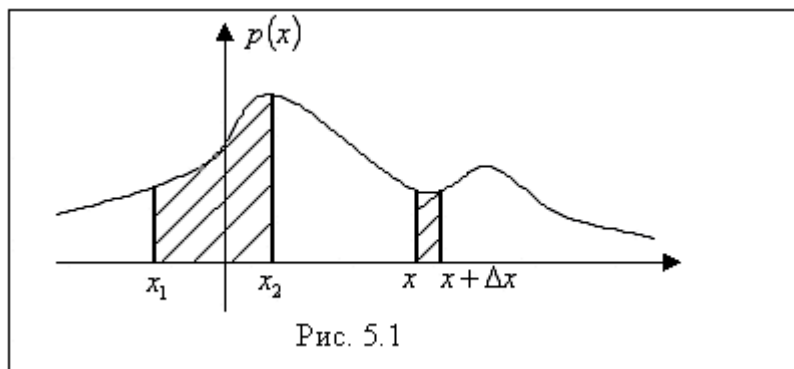
$$\int_{-\infty}^{+\infty} p(x) dx = 1.$$

Доказательство. В силу свойства 2, имеем

$$\int_{-\infty}^{+\infty} p(x) dx = F(+\infty) - F(-\infty) = 1. \quad \blacksquare$$

4. $P(x \leq \xi < x + \Delta x) \approx p(x) \Delta x$.

Доказательство. Из свойства 2 следует, что вероятность попадания случайной величины на интервал $[x_1, x_2)$ численно равна площади криволинейной трапеции (рис. 5.1).



Как видно из рис. 5.1, при $\Delta x \rightarrow 0$ вероятность попадания на интервал $[x, x + \Delta x)$ приближенно совпадает с площадью прямоугольника со сторонами Δx и $p(x)$. ■

5. Для непрерывных случайных величин $P(\xi = x) = 0$.

Доказательство. Достаточно применить свойство 4, где $\Delta x = 0$. ■

6. Для непрерывных случайных величин свойство 2 можно переписать в виде:

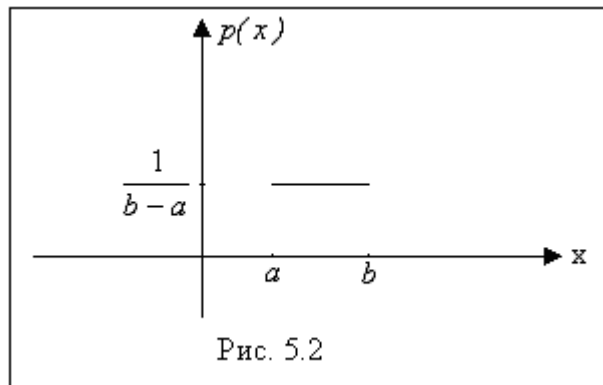
$$P(x_1 \leq \xi < x_2) = P(x_1 < \xi < x_2) = P(x_1 < \xi \leq x_2) = P(x_1 \leq \xi \leq x_2).$$

Пример 7. Случайная величина ξ имеет плотность

$$p(x) = \begin{cases} 0, & x \leq a; \\ \frac{1}{b-a}, & a < x \leq b; \\ 0, & x > b. \end{cases}$$

График функции $p(x)$ изображен на рис. 5.2.

Найти функцию распределения $F(x)$ случайной величины ξ и изобразить ее график.



О Решение. Для решения задачи применим формулу (5.4.1).

При $x \leq a$ получаем, что

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

При $a < x \leq b$ получаем:

$$F(x) = \int_{-\infty}^x p(t) dt = \int_{-\infty}^0 p(t) dt + \int_0^x p(t) dt = \int_{-\infty}^0 0 dt + \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}.$$

При $x > b$ имеем:

$$F(x) = \int_{-\infty}^x p(t) dt = \int_{-\infty}^a p(t) dt + \int_a^b p(t) dt + \int_b^x p(t) dt = 1.$$

Таким образом, получаем

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

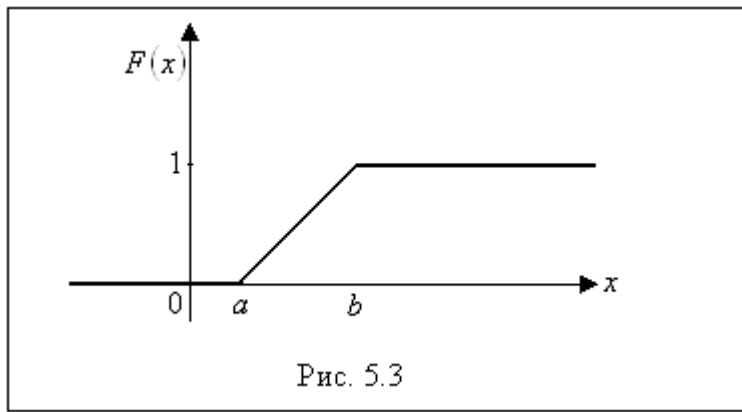


График функции распределения изображен на рис. 5.3. ●

Пример 8. Случайная величина ξ подчинена закону Симпсона («закону равнобедренного треугольника») на участке $[-a, a]$ (рис. 5.4).



Найти плотность распределения и функцию распределения.

● **Решение.** Используя свойство 3 плотности, найдем высоту треугольника:

$$1 = \int_{-\infty}^{+\infty} p(x) dx = \frac{1}{2} h 2a = ah \Rightarrow h = \frac{1}{a}.$$

Далее находим уравнения ребер:

$$p(x) = \frac{1}{a} \left(1 - \frac{x}{a} \right), \text{ при } x \in [0, a];$$

$$p(x) = \frac{1}{a} \left(1 + \frac{x}{a} \right), \text{ при } x \in [-a, 0]$$

или

$$p(x) = \frac{1}{a} \left(1 - \frac{|x|}{a} \right), \text{ при } x \in [-a, a].$$

Найдем функцию распределения $F(x)$.

Очевидно, что при $x \leq -a$ функция распределения равна нулю, т.е. $F(x) = 0$.

Далее, при $-a < x \leq 0$:

$$F(x) = \int_{-\infty}^x p(t) dt = \int_{-\infty}^{-a} p(t) dt + \int_{-a}^x p(t) dt = \int_{-a}^x \frac{1}{a} \left(1 + \frac{t}{a} \right) dt = \frac{1}{a} \left(x + \frac{x^2}{2a} \right) + \frac{1}{2}.$$

При $0 < x \leq a$:

$$F(x) = \int_{-\infty}^{-a} p(t)dt + \int_{-a}^0 p(t)dt + \int_0^x \frac{1}{a} \left(1 - \frac{t}{a}\right) dt = \frac{1}{2} + \frac{1}{a} \left(x - \frac{x^2}{2a}\right).$$

При $x > a$, получаем $F(x) = 1$.

Таким образом, получено

$$p(x) = \begin{cases} \frac{1}{a} \left(1 - \frac{|x|}{a}\right), & \text{при } x \in [-a, a] \\ 0, & \text{при } x \notin [-a, a] \end{cases}, \quad F(x) = \begin{cases} 0, & \text{при } x \leq -a; \\ \frac{1}{a} \left(x + \frac{x^2}{2a}\right) + \frac{1}{2}, & \text{при } -a < x \leq 0; \\ \frac{1}{2} + \frac{1}{a} \left(x - \frac{x^2}{2a}\right), & \text{при } 0 < x \leq a; \\ 1, & \text{при } x > a. \end{cases}$$

§5. Функция от случайных величин

Пусть ξ — случайная величина. Пусть задана функция $y = f(x)$. Каждому элементарному исходу ω поставим в соответствие число $\eta(\omega)$ по формуле $\eta(\omega) = f(\xi(\omega))$. Тем самым получим случайную величину η , называемую функцией $f(\xi)$ от случайной величины ξ .

Пусть ξ — дискретная случайная величина. Тогда случайная величина $\eta = f(\xi)$ также является дискретной случайной величиной, поскольку она не может принимать больше значений, чем случайная величина ξ . Очевидно, что ряд распределения случайной величины $\eta = f(\xi)$ имеет вид:

η	$f(x_1)$	$f(x_2)$	...	$f(x_n)$
P	p_1	p_2	...	p_n

При этом, если в верхней строке ряда распределения появляются одинаковые значения $f(x_i)$, то соответствующие столбцы необходимо объединить в один, приписав им суммарную вероятность.

Пример 9. Случайная величина ξ имеет ряд распределения:

ξ	-2	-1	0	1	2
P	0,1	0,2	0,3	0,3	0,1

Найти закон распределения случайной величины $\eta = |\xi|$.

О Решение. Составим ряд распределения случайной величины $|\xi|$:

$ \xi $	2	1	0	1	2
P	0,1	0,2	0,3	0,3	0,1

Объединив первый и пятый, второй и четвертый столбцы, получим:

η	0	1	2
P	0,3	0,5	0,2

Ряд распределения случайной величины $\eta = |\xi|$ получен. ●

Пусть ξ — непрерывная случайная величина. При этом случайная величина $\eta = f(\xi)$ может быть как непрерывной, так и дискретной. Будем рассматривать непрерывные функции $f(x)$. Пусть случайная величина ξ имеет плотность $p_\xi(x)$. Тогда

$$F_\eta(y) = P(\eta < y) = P(f(x) < y) = \int_{f(x) < y} p_\xi(x) dx. \quad (5.5.1)$$

Пример 10. Пусть случайная величина ξ имеет плотность

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

Найти распределение случайной величины $\eta = \xi^2$.

○ **Решение.** В данном случае $y = f(x) = x^2$. Согласно (5.5.1), получим

$$F_\eta(y) = \int_{x^2 < y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Очевидно, что при $y < 0$, функция распределения равна нулю, т.е. $F_\eta(y) = 0$. При $y > 0$ область $(x^2 < y)$ совпадает с областью $(-\sqrt{y} < x < \sqrt{y})$. Отсюда получаем

$$F_\eta(y) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_0^y \frac{e^{-\frac{t}{2}}}{\sqrt{t}} dt. \bullet$$

Выведем более удобные формулы для вычисления функции $F_\eta(y)$, где $\eta = f(x)$.

Теорема. Пусть ξ — непрерывная случайная величина с плотностью $p_\xi(x)$, а случайная величина η связана с ξ функциональной зависимостью

$$y = f(x),$$

где $f(x)$ — дифференцируемая функция, монотонная на всем участке возможных значений аргумента x .

Тогда плотность распределения случайной величины η выражается формулой

$$p_\eta(y) = p_\xi(\psi(y)) |\psi'(y)|, \quad (5.5.2)$$

где ψ — функция, обратная по отношению к функции $f(x)$.

Доказательство. Пусть $f(x)$ — монотонно возрастающая функция. Тогда

$$F_\eta(y) = P(\eta < y) = P(f(\xi) < y) = P(\xi < f^{-1}(y)) = P(\xi < \psi(y)) = F_\xi(\psi(y)).$$

Продифференцировав последнее равенство, получаем

$$p_\eta(y) = \frac{d}{dy}(F_\eta(y)) = \frac{d}{dy}(F_\xi(\psi(y))) = p_\xi(\psi(y)) \psi'(y). \quad (5.5.3)$$

Пусть $f(x)$ — монотонно убывающая функция. В этом случае $f'(x) < 0$ и, следовательно, $\psi'(y) = \frac{d}{dy} f^{-1}(y) < 0$. Отсюда получаем:

$$\begin{aligned} F_\eta(y) &= P(\eta < y) = P(f(\xi) < y) = P(\xi > f^{-1}(y)) = P(\xi > \psi(y)) = \\ &= 1 - P(\xi \leq \psi(y)) = 1 - F_\xi(\psi(y)). \end{aligned}$$

Продифференцировав последнее равенство, получаем

$$p_{\eta}(y) = \frac{d}{dy}(F_{\eta}(y)) = \frac{d}{dy}(F_{\xi}(\psi(y))) = -p_{\xi}(\psi(y))\psi'(y). \quad (5.5.4)$$

Учитывая свойства модуля, равенства (5.5.3) и (5.5.4) объединим в одно

$$p_{\eta}(y) = p_{\xi}(\psi(y))|\psi'(y)|,$$

что совпадает с (5.5.2). ■

Пример 11. Случайная величина ξ распределена равномерно в интервале $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$. Найти закон распределения случайной величины $\eta = \sin \xi$.

○ **Решение.** Функция $y = \sin x$ в интервале $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$ монотонна, следовательно,

можно применить формулу (5.5.2). Решение задачи удобно расположить в виде таблицы, содержащей три столбца: в первом запишем план решения задачи; во втором столбце запишем обозначения функций, принятые в общем случае; в третьем — конкретные функции, соответствующие данному примеру:

Плотность случайной величины ξ	$p_{\xi}(x)$	$\begin{cases} \frac{1}{\pi}, & \text{если } x \in \left(-\frac{\pi}{2}; \frac{\pi}{2}\right); \\ 0, & \text{если } x \notin \left(-\frac{\pi}{2}; \frac{\pi}{2}\right). \end{cases}$
Функциональная зависимость между случайными величинами ξ и η	$y = f(x)$	$y = \sin(x)$
Обратная функция	$x = \psi(y)$	$x = \arcsin(y)$
Модуль производной обратной функции	$ \psi'(y) $	$\frac{1}{\sqrt{1-y^2}}$
Плотность случайной величины η	$p_{\eta}(y) = p_{\xi}(\psi(y)) \psi'(y) $	$\begin{cases} \frac{1}{\pi} \frac{1}{\sqrt{1-y^2}}, & \text{если } y \in (-1; 1); \\ 0, & \text{если } y \notin (-1; 1). \end{cases}$

Интервал $(-1; 1)$, в котором лежат значения случайной величины η , определяется областью значений функции $y = \sin(x)$ для $x \in \left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$. ●

Следствие из теоремы. Если $f(x)$ — немонотонная функция, то обратная к ней функция неоднозначна, и плотность распределения случайной величины определяется в виде суммы столькох слагаемых, сколько значений (при данном значении y) имеет обратная функция:

$$p_{\eta}(y) = \sum_{i=1}^k p_{\xi}(\psi_i(y))|\psi_i'(y)|, \quad (5.5.5)$$

где $\psi_1(y), \dots, \psi_k(y)$ — значения обратной функции для данного y .

Пример 12. Случайная величина ξ распределена равномерно в интервале $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$. Найти плотность распределения случайной величины $\eta = \cos \xi$.

○ **Решение.** Функция $y = \cos x$ немонотонная в интервале $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$, ее значения лежат в интервале $(0,1)$. В данном случае для любого y обратная функция будет иметь два значения:

Плотность случайной величины ξ	$p_{\xi}(x)$	$\begin{cases} \frac{1}{\pi}, & \text{если } x \in \left(-\frac{\pi}{2}; \frac{\pi}{2}\right); \\ 0, & \text{если } x \notin \left(-\frac{\pi}{2}; \frac{\pi}{2}\right). \end{cases}$
Функциональная зависимость между случайными величинами ξ и η .	$y = f(x)$	$y = \cos(x)$
Обратная функция	$x = \begin{cases} \psi_1(y), \\ \psi_2(y). \end{cases}$	$x_1 = \arccos(y),$ $x_2 = -\arccos(y).$
Модуль производной обратной функции	$ \psi'(y) $	$\frac{1}{\sqrt{1-y^2}}$
Плотность случайной величины η	$p_{\eta}(y) = \sum_{i=1}^2 p_{\xi}(\psi_i(y)) \psi'_i(y) $	$\begin{cases} \frac{2}{\pi \sqrt{1-y^2}}, & \text{при } y \in (0,1); \\ 0, & \text{при } y \notin (0,1). \end{cases}$

$$p_{\eta}(y) = \sum_{i=1}^2 p_{\xi}(\psi_i(y)) |\psi'_i(y)| = \frac{1}{\pi} \frac{1}{\sqrt{1-y^2}} + \frac{1}{\pi} \frac{1}{\sqrt{1-y^2}} = \frac{2}{\pi \sqrt{1-y^2}}. \bullet$$

Глава 6. Числовые характеристики случайных величин

§1. Математическое ожидание случайной величины

Определение. Математическим ожиданием $M\xi$ дискретной случайной величины ξ называется выражение, вычисляемое по формуле:

$$M\xi = \sum_i x_i p_i, \quad (6.1.1)$$

где $x_1, x_2, \dots$ — значения случайной величины, $p_1, p_2, \dots$ — соответствующие им вероятности, которые определяются равенством $p_i = P(\xi = x_i)$.

Для существования математического ожидания необходимо, чтобы ряд (6.1.1) сходился абсолютно, т.е.

$$\sum_i |x_i| p_i < \infty, \quad (6.1.2)$$

в противном случае говорят, что математическое ожидание не существует.

Пример 1. Пусть ξ — случайная величина, равная числу выпавших очков при бросании игрального кубика. Найти математическое ожидание случайной величины ξ .

О Решение. Случайная величина ξ имеет следующий ряд распределения:

x_i	1	2	3	4	5	6
P_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Применяя формулу (6.1.1), получим

$$M\xi = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2}.$$

Таким образом, математическое ожидание числа выпавших очков при бросании игрального кубика равно 3,5. ●

Пример 2. Найти математическое ожидание случайной величины ξ — число успехов в схеме Бернулли.

О Решение. Как известно, распределение случайной величины ξ задается формулой

$$P(\xi = i) = C_n^i p^i q^{n-i}, \quad (i = 1, \dots, n)$$

где p — вероятность «успеха», $q = 1 - p$, n — количество испытаний в схеме Бернулли.

Используя формулу (6.1.1), получим

$$\begin{aligned} M\xi &= \sum_{i=0}^n i p_n(i) = \sum_{i=0}^n i C_n^i p^i q^{n-i} = \sum_{i=0}^n i \frac{n!}{i!(n-i)!} p^i q^{n-i} = \sum_{i=1}^n n p \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} q^{n-i} = \\ &= n p \sum_{i=0}^{n-1} C_{n-1}^i p^i q^{n-1-i} = n p \sum_{i=0}^{n-1} p_{n-1}(i) = n p. \end{aligned}$$

Таким образом, математическое ожидание числа успехов в схеме Бернулли равно np . ●

Определение. Математическим ожиданием $M\xi$ непрерывной случайной величины ξ называется интеграл:

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx . \quad (6.1.3)$$

Условием существования математического ожидания непрерывной случайной величины является абсолютная сходимость интеграла

$$\int_{-\infty}^{+\infty} |x|p(x)dx < \infty . \quad (6.1.4)$$

Пример 3. Найти математическое ожидание случайной величины ξ , плотность которой имеет вид:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

Решение. Используя (6.1.3), получим

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx = \int_b^a \frac{x}{b-a}dx = \frac{a+b}{2} . \bullet$$

Пример 4. Найти математическое ожидание случайной величины ξ , плотность которой имеет вид:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, \quad x \in (-\infty; \infty).$$

Решение. Используя (6.1.3), получим

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx = \int_{-\infty}^{+\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx . \quad (6.1.5)$$

Делаем замену $t = \frac{x-m}{\sigma}$ или $x = m + t\sigma$. В этом случае (6.1.5) примет вид:

$$\int_{-\infty}^{+\infty} \frac{t\sigma + m}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{+\infty} \frac{t\sigma}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + \int_{-\infty}^{+\infty} \frac{m}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}} dt + m \int_{-\infty}^{+\infty} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt . \quad (6.1.6)$$

Первое слагаемое равно нулю, т.к. равен нулю интеграл

$$\int_{-\infty}^{+\infty} te^{-\frac{t^2}{2}} dt = -e^{-\frac{t^2}{2}} \Big|_{-\infty}^{+\infty} = 0 .$$

Интеграл во втором слагаемом равен 1, т.к. этот интеграл равен функции распределения нормального закона с параметрами (0,1) при значении аргумента равным $x = +\infty$, т.е.

$$\int_{-\infty}^{+\infty} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt = \Phi(+\infty) = 1 .$$

Таким образом, математическое ожидание равно $M\xi = m$. $\bullet$

Пример 5. Случайная величина ξ имеет плотность Коши:

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad x \in (-\infty; +\infty). \quad (6.1.7)$$

Проверить, имеет ли случайная величина ξ математическое ожидание.

Решение. Проверим условие (6.1.4) существования математического ожидания

$$\int_{-\infty}^{+\infty} |x| p(x) dx = \int_{-\infty}^{+\infty} \frac{|x|}{\pi(1+x^2)} dx = \infty.$$

Математическое ожидание случайной величины ξ , имеющей плотность Коши, не существует, т.к. условие существования математического ожидания не выполнено. ●

§2. Математическое ожидание функции от случайной величины. Свойства математического ожидания

Пусть $\eta = f(\xi)$ — функция от случайной величины. Определим математическое ожидание $M\eta = Mf(\xi)$. Это возможно сделать двумя способами. Первый способ состоит в том, что сначала строится распределение случайной величины η , затем уже находим $M\eta$. Мы рассмотрим другой способ. Пусть сначала ξ — дискретная случайная величина, принимающая значения $x_1, \dots, x_n$. Тогда случайная величина $\eta = f(\xi)$ принимает значения $f(x_1), \dots, f(x_n)$ с теми же вероятностями $p_i = P(\xi = x_i) = P(f(\xi) = f(x_i))$. В этом случае математическое ожидание определяется по формуле

$$M\xi = Mf(\xi) = \sum_{i=1}^n f(x_i) p_i. \quad (6.2.1)$$

В случае, если случайная величина ξ принимает счетное число значений, то математическое ожидание случайной величины ξ определяется по формуле

$$M\eta = Mf(\xi) = \sum_{i=1}^{\infty} f(x_i) p_i. \quad (6.2.2)$$

При этом условие существования математического ожидания (6.1.4) примет вид:

$$\sum_i |f(x_i)| p_i < \infty. \quad (6.2.3)$$

Пример 6. Случайная величина ξ имеет ряд распределения:

ξ	4	16	100
P	0,7	0,1	0,2

Найти математическое ожидание математической величины: $\eta = \xi + \sqrt{\xi}$.

О Решение. Для решения задачи применим формулу (6.2.1).

$$\begin{aligned} M\eta &= M(\xi + \sqrt{\xi}) = \sum_{i=1}^3 (X_i + \sqrt{X_i}) p_i = \\ &= 0,7(4 + \sqrt{4}) + 0,1(16 + \sqrt{16}) + 0,2(100 + \sqrt{100}) = 4,2 + 2 + 22 = 28,2. \end{aligned}$$

Таким образом, математическое ожидание математической величины $\eta = \xi + \sqrt{\xi}$ равно 28,2. ●

Пусть ξ — непрерывная случайная величина, имеющая плотность распределения $p_{\xi}(x)$. Пусть функция $y = f(x)$ непрерывная (за исключением, быть может, счетного числа точек). Тогда математическое ожидание случайной величины $\eta = f(\xi)$ определяется по формуле

$$M\eta = Mf(\xi) = \int_{-\infty}^{+\infty} f(x)p_{\xi}(x)dx. \quad (6.2.4)$$

Условие существования математического ожидания случайной величины $\eta = f(\xi)$ имеет вид:

$$\int_{-\infty}^{+\infty} |f(x)|p_{\xi}(x)dx < \infty. \quad (6.2.5)$$

Пример 7. Случайная величина ξ имеет нормальное распределение с параметрами $(0,1)$, т.е. ее плотность имеет вид:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in (-\infty; +\infty).$$

Найти математическое ожидание случайной величины $\eta = a\xi + b$.

О Решение. Используя формулу (6.2.4), получаем:

$$M\eta = \int_{-\infty}^{+\infty} (ax + b) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = a \int_{-\infty}^{+\infty} \frac{x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + b \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = a \cdot 0 + b = b. \bullet$$

Пример 8. Случайная величина ξ распределена равномерно в интервале $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$, т.е.

$$p(x) = \begin{cases} \frac{1}{\pi}, & x \in \left(-\frac{\pi}{2}; \frac{\pi}{2}\right), \\ 0, & x \notin \left(-\frac{\pi}{2}; \frac{\pi}{2}\right) \end{cases}$$

Найти математическое ожидание случайной величины $\eta = \sin(\xi)$.

О Решение. Используя формулу (6.2.4.), получаем:

$$M\eta = \int_{-\infty}^{+\infty} f(x)p(x)dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{\pi} \sin x dx = 0. \bullet$$

Свойства математического ожидания.

1. Математическое ожидание постоянной равно самой этой постоянной, т.е.

$$MC = C, \text{ где } C = const.$$

Доказательство. Постоянную C можно рассматривать как случайную величину, принимающую только одно значение C с вероятностью 1. Следовательно,

$$MC = CP(\xi = C) = C \cdot 1 = C. \blacksquare$$

2. $M(a\xi + b) = aM\xi + b$.

Доказательство. Пусть ξ — непрерывная случайная величина. Тогда для случайной величины $\eta = a\xi + b$ по формуле (6.2.4.) получаем:

$$M\eta = M(a\xi + b) = \int_{-\infty}^{+\infty} (ax + b)p_{\xi}(x)dx = \int_{-\infty}^{+\infty} xp_{\xi}(x)dx + b \int_{-\infty}^{+\infty} p_{\xi}(x)dx = aM\xi + b.$$

Аналогично доказывается и для дискретной случайной величины. $\blacksquare$

3. Для любых случайных величин ξ и ψ математическое ожидание их суммы случайных величин равно сумме их математических ожиданий, т.е.

$$M(\xi + \psi) = M\xi + M\psi.$$

Доказательство. Пусть случайные величины ξ и ψ — дискретные. Случайная величина ξ принимает значения $X_1, \dots, X_n$, а случайная величина ψ принимает значения $Y_1, \dots, Y_m$. Рассмотрим случайную величину $\eta = \xi + \psi$ ($g(x, y) = x + y$). Случайная величина η принимает значения $g(x_i, y_j)$ с вероятностями $p_{ij} = P(\xi = X_i, \psi = Y_j)$. Тогда:

$$\begin{aligned} M\eta &= M(\xi + \psi) = \sum_{i=1}^n \sum_{j=1}^m (x_i + y_j) p_{ij} = \sum_{i=1}^n \sum_{j=1}^m x_i p_{ij} + \sum_{i=1}^n \sum_{j=1}^m y_j p_{ij} = \sum_{i=1}^n x_i \sum_{j=1}^m p_{ij} + \sum_{j=1}^m y_j \sum_{i=1}^n p_{ij} = \\ &= \sum_{i=1}^n x_i p_{\xi i} + \sum_{j=1}^m y_j p_{\psi j} = M\xi + M\psi. \end{aligned}$$

При доказательстве воспользовались тем, что

$$\sum_{j=1}^m p_{ij} = p_{\xi i} \text{ и } \sum_{i=1}^n p_{ij} = p_{\psi j}.$$

Действительно, учитывая, что

$$\sum_{j=1}^m (\xi = X_i, \psi = Y_j) = (\xi = X_i) \sum_{j=1}^m (\psi = Y_j) = (\xi = X_i) \Omega = (\xi = X_i),$$

то

$$p_{\xi i} = P(\xi = X_i) = P\left(\sum_{j=1}^m (\xi = X_i, \psi = Y_j)\right) = \sum_{j=1}^m P(\xi = X_i, \psi = Y_j) = \sum_{j=1}^m p_{ij},$$

аналогично доказывается, что

$$\sum_{i=1}^n p_{ij} = p_{\psi j}.$$

Аналогично доказывается и для непрерывной случайной величины. ■

4. Если ξ и ψ независимые случайные величины, то математическое ожидание произведения случайных величин равно произведению их математических ожиданий, т.е.

$$M(\xi\psi) = M\xi M\psi.$$

Доказательство. Пусть величины ξ и ψ — дискретные. В силу независимости случайных величин имеет место равенство:

$$p_{ij} = P(\xi = X_i, \psi = Y_j) = P(\xi = X_i)P(\psi = Y_j) = p_{\xi i} p_{\psi j}. \text{ Тогда}$$

$$\begin{aligned} M(\xi\psi) &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j p_{ij} = \sum_{i=1}^n \sum_{j=1}^m x_i y_j P(\xi = X_i)P(\psi = Y_j) = \\ &= \sum_{i=1}^n x_i p_{\xi i} \sum_{j=1}^m y_j p_{\psi j} = M\xi M\psi. \end{aligned}$$

Аналогично доказывается и для дискретной случайной величины. ■

Заметим, что свойство 3 допускает обобщение на сумму любого числа слагаемых, а свойство 4 допускает обобщение на произведение любого числа независимых (в совокупности) сомножителей.

Пример 9. Найти математическое ожидание случайной величины ξ — число успехов в схеме Бернулли.

О Решение. Представим число успехов ξ в схеме Бернулли из n испытаний в виде $\xi = \mathcal{G}_1 + \dots + \mathcal{G}_n$, где $\mathcal{G}_i$ — число успехов в i -ом испытании. Очевидно, что $M\mathcal{G}_i = 0(1-p) + 1p = p$. По свойству 3 математического ожидания, получаем

$$M\xi = M(\mathcal{G}_1 + \dots + \mathcal{G}_n) = M\mathcal{G}_1 + \dots + M\mathcal{G}_n = \sum_{i=1}^n p = np.$$

Этот результат совпадает с результатом примера 3, но получен более легкими вычислениями. ●

Пример 10. Производится ряд независимых опытов, в каждом из которых может появиться событие А. Вероятность события А в каждом опыте равна p . Опыты производятся до первого появления события А, после чего они прекращаются. Случайная величина ξ — число произведенных опытов. Найти $M\xi$.

○ **Решение.** Рассмотрим событие $(\xi = 1)$, в этом случае событие А произошло при первом опыте, т.о. $P(\xi = 1) = p$. Перейдем к событию $(\xi = 2)$, в этом случае событие А при первом опыте не произошло, но произошло при втором опыте, т.о. $P(\xi = 2) = (1 - p)p$. Производя аналогичные рассуждения, получаем ряд распределения:

ξ	1	2	3	...	i
P	p	$(1 - p)p$	$(1 - p)^2 p$	...	$(1 - p)^{i-1} p$

Используя формулу (6.1.1), получим:

$$\begin{aligned} M\xi &= \sum_{i=1}^{\infty} iP(\xi = i) = \sum_{i=1}^{\infty} i(1 - p)^{i-1} p = \sum_{i=1}^{\infty} iq^{i-1} p = p \sum_{i=1}^{\infty} iq^{i-1} = \\ &= p \frac{d}{dq} \left(\sum_{i=1}^{\infty} q^i \right) = p \frac{d}{dq} \left(\frac{q}{1 - q} \right) = p \frac{1}{(1 - q)^2} = \frac{1}{p}. \end{aligned}$$

Таким образом, математическое ожидание равно $\frac{1}{p}$. ●

Пример 11. Независимые случайные величины ξ и η заданы своими рядами распределений:

ξ	-1	0	1
P	1/4	1/4	1/2

η	2	1	0
P	1/3	1/2	1/6

Для случайной величины $Z = \xi + \eta$ найти математическое ожидание двумя способами:

- 1) по определению математического ожидания;
- 2) по свойствам математического ожидания.

○ **Решение.** Рассмотрим 1 способ нахождения математического ожидания. Для этого составим ряд распределения случайной величины Z . Для этого удобно воспользоваться таблицей сумм $\xi + \eta$ и соответствующих им вероятностей. Например:

$$P((\xi = -1) \cap (\eta = 2)) = P(\xi = -1)P(\eta = 2) = \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{12}.$$

$\xi \backslash \eta$	2	1	0
-1	1 1/12	0 1/8	-1 1/24
0	2 1/12	1 1/8	0 1/24
1	3 1/6	2 1/4	1 1/12

Далее очень легко получить ряд распределения случайной величины Z .

Z	3	2	1	0	-1
P	1/6	1/3	7/24	1/6	1/24

Естественно, можно было бы обойтись и без таблицы. Например:

$$\begin{aligned}
 P(\xi + \eta = 1) &= P([(\xi = -1) \cap (\eta = 2)] \cup [(\xi = 0) \cap (\eta = 1)] \cup [(\xi = 1) \cap (\eta = 0)]) = \\
 &= P(\xi = -1)P(\eta = 2) + P(\xi = 0)P(\eta = 1) + P(\xi = 1)P(\eta = 0) = \frac{1}{4} \cdot \frac{1}{3} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{6} = \\
 &= \frac{1}{12} + \frac{1}{8} + \frac{1}{12} = \frac{7}{24}.
 \end{aligned}$$

Используя ряд распределения, находим математическое ожидание

$$MZ = 3 \cdot \frac{1}{6} + 2 \cdot \frac{1}{3} + 1 \cdot \frac{7}{24} + 0 \cdot \frac{1}{6} + (-1) \cdot \frac{1}{24} = \frac{17}{12}.$$

Перейдем ко второму способу нахождения математического ожидания случайной величины $Z = \xi + \eta$. Используя свойство 3, получим:

$$MZ = M(\xi + \eta) = M\xi + M\eta = \left(-1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2}\right) + \left(2 \cdot \frac{1}{3} + 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{6}\right) = \frac{1}{4} + \frac{7}{6} = \frac{17}{12}. \bullet$$

§3. Дисперсия. Моменты высших порядков

Определение. Моментом k -го порядка называется выражение, вычисляемое по формулам:

а) для дискретной случайной величины:

$$M\xi^k = \sum_i x_i^k p_i, \quad (6.3.1)$$

б) для непрерывной случайной величины:

$$M\xi^k = \int_{-\infty}^{+\infty} x^k p(x) dx. \quad (6.3.2)$$

Для существования момента k -ого порядка необходимо:

а) для дискретной случайной величины — абсолютная сходимость ряда

$$\sum_i |x_i^k| p_i < \infty, \quad (6.3.3)$$

б) для непрерывной случайной величины — абсолютная сходимость интеграла

$$\int_{-\infty}^{+\infty} |x^k| p(x) dx < \infty. \quad (6.3.4)$$

Поскольку $|\xi^{k-1}| \leq |\xi^k| + 1$, то из существования момента k -го порядка вытекает существование момента $(k-1)$ -го порядка и, следовательно, всех моментов меньших порядков.

Определение. Момент k -го порядка величины $\xi^0 = \xi - M\xi$ называется **центральным моментом k -го порядка**.

Особую роль играет второй центральный момент, который называется дисперсией и обозначается $D\xi$.

Определение. Дисперсией случайной величины ξ называется математическое ожидание квадрата отклонения случайной величины от своего математического ожидания:

$$D\xi = M(\xi - M\xi)^2. \quad (6.3.5)$$

Свойства дисперсии.

1. Дисперсия постоянной величины равна нулю, т.е.

$$DC = M(C - MC)^2 = M(C - C)^2 = M0 = 0. \blacksquare$$

2. Дисперсия является неотрицательной величиной, т.е. $D\xi \geq 0$.

3. Для любых действительных чисел a и b справедливо равенство

$$D(a\xi + b) = a^2 D\xi + b.$$

Доказательство. Из свойства 2 математического ожидания получаем:

$$\begin{aligned} D(a\xi + b) &= M(a\xi + b - M(a\xi + b))^2 = M(a\xi + b - aM\xi - b)^2 = \\ &= M(a\xi - aM\xi)^2 = a^2 M(\xi - M\xi)^2 = a^2 D\xi. \end{aligned}$$

■

$$4. D\xi = M\xi^2 - (M\xi)^2.$$

Доказательство. Используя определение дисперсии и свойства 2 и 3 математического ожидания, получаем:

$$\begin{aligned} D\xi &= M(\xi - M\xi)^2 = M(\xi^2 - 2\xi M\xi + (M\xi)^2) = M\xi^2 - M(2\xi M\xi) + M((M\xi)^2) = \\ &= M\xi^2 - 2(M\xi)^2 + (M\xi)^2 = M\xi^2 - (M\xi)^2. \end{aligned}$$

■

5. Если ξ и η независимые случайные величины, дисперсия суммы случайных величин равна сумме дисперсий случайных величин, т.е.

$$D(\xi + \eta) = D\xi + D\eta.$$

Доказательство. Если ξ и η независимые случайные величины, то и случайные величины $\xi^0 = \xi - M\xi$ и $\eta^0 = \eta - M\eta$ будут независимыми. Тогда, используя свойства 2-4 математического ожидания, получаем:

$$\begin{aligned} D(\xi + \eta) &= M(\xi + \eta - M(\xi + \eta))^2 = M((\xi - M\xi) + (\eta - M\eta))^2 = M(\xi - M\xi)^2 + \\ &+ 2M((\xi - M\xi)(\eta - M\eta)) + M(\eta - M\eta)^2 = D\xi + 2M(\xi - M\xi)M(\eta - M\eta) + D\eta = \\ &= D\xi + D\eta. \end{aligned}$$

■

Очевидно, что дисперсия $D\xi$ имеет размерность квадрата случайной величины ξ . Для практических же целей удобно иметь числовую характеристику, размерность которой совпадает с размерностью ξ .

Определение. Средним квадратичным отклонением случайной величины ξ называется выражение, вычисляемое по формуле:

$$\sigma = \sqrt{D\xi}. \quad (6.3.6)$$

Пример 12. Найти дисперсию случайной величины ξ , плотность которой имеет вид (равномерно распределенной на отрезке $[a, b]$):

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

○ **Решение.** Используя определение дисперсии $D\xi = M(\xi - M\xi)^2$, формулу для вычисления дисперсии непрерывной случайной величины (6.3.2) и результат примера 4, получаем:

$$D\xi = M(\xi - M\xi)^2 = \int_a^b \left(x - \frac{b+a}{2}\right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}. \bullet$$

Пример 13. Найти дисперсию случайной величины ξ , распределенной по нормальному закону с параметрами (m, σ^2) (см. Пример 4).

Решение. Используя определение дисперсии $D\xi = M(\xi - M\xi)^2$, формулу для вычисления дисперсии непрерывной случайной величины (6.3.2) и результат примера 5, получаем:

$$D\xi = M(\xi - M\xi)^2 = \int_{-\infty}^{+\infty} (x-m)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx. \quad (6.3.5)$$

Делаем замену $t = \frac{x-m}{\sigma}$ или $x = m + t\sigma$, при этом $dx = \sigma dt$. В этом случае выражение (6.3.5) примет вид:

$$= \sigma \int_{-\infty}^{+\infty} \frac{t^2}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \sigma^2 \int_{-\infty}^{+\infty} \frac{t}{\sqrt{2\pi}} d\left(-e^{-\frac{t^2}{2}}\right) = \sigma^2 \left. \frac{-t}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \right|_{-\infty}^{+\infty} + \sigma^2 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 0 + \sigma^2 = \sigma^2. \bullet$$

Пример 14. При условии примера 11 найти дисперсию случайной величины Z .

Решение. Рассмотрим 1 способ нахождения дисперсии, используя ряд распределения случайной величины Z , найденный в примере 12,

Z	3	2	1	0	-1
P	1/6	1/3	7/24	1/6	1/24

и свойство 4 дисперсии, получим:

$$MZ^2 = 3^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{7}{24} + 0^2 \cdot \frac{1}{6} + (-1)^2 \cdot \frac{1}{24} = \frac{19}{6},$$

$$DZ = MZ^2 - (MZ)^2 = \frac{19}{6} - \left(\frac{17}{12}\right)^2 = \frac{167}{144}.$$

Напомним, что математическое ожидание $MZ = \frac{17}{12}$ было найдено в примере 12.

Перейдем ко второму способу нахождения математического ожидания случайной величины $Z = \xi + \eta$. Используя свойство 5 дисперсии случайной величины, получим:

$$M\xi = \frac{1}{4}, \quad M\xi^2 = 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} = \frac{3}{4},$$

$$M\eta = \frac{7}{6}, \quad M\eta^2 = 4 \cdot \frac{1}{3} + 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{6} = \frac{11}{6},$$

$$D\xi = M\xi^2 - (M\xi)^2 = \frac{3}{4} - \frac{1}{16} = \frac{11}{16}, \quad Dz = D(\xi + \eta) = D\xi + D\eta = \frac{11}{16} + \frac{17}{36} = \frac{167}{144}. \bullet$$

Глава 7. Элементы математической статистики

§1. Основные понятия и основные задачи математической статистики

В математической статистике исследуются способы получения выводов на основе эмпирических (опытных) данных. Основными понятиями математической статистики являются: генеральная совокупность, выборка, теоретическая функция распределения.

Определение. Генеральной совокупностью называются все возможные результаты наблюдений, которые могут быть сделаны при данном комплексе условий.

В некоторых задачах генеральную совокупность рассматривают как случайную величину X . Примером генеральной совокупности может быть все население страны. В этой совокупности нас могут интересовать, например, возраст жителей. Другим примером генеральной совокупности являются детали, изготовленные на данном станке. Эти детали могут быть качественными и бракованными.

Определение. Выборочной совокупностью (выборкой) называется множество результатов, случайно отобранных из генеральной совокупности.

Выборка должна быть репрезентативной, т.е. правильно отражать пропорции генеральной совокупности. Это достигается случайностью отбора, когда все объекты генеральной совокупности имеют одинаковую вероятность быть отобранными.

Определение. Число объектов в совокупности (генеральной или выборочной) называется ее **объемом**.

Объем генеральной совокупности обозначим символом N , а объем выборочной совокупности обозначим символом n . При этом подразумевается, что $n \ll N$.

Заметим, что сам процесс выбора можно осуществлять разными способами: выбрав объект и определив его значение, изымать объект и не допускать к последующим испытаниям (выборка без возвращения); после определения его значения объект возвращается в генеральную совокупность (выборка с возвращением). Очевидно, что при достаточно большом объеме генеральной совокупности исчезает различие между выборками с возвращением и без возвращения. Будем рассматривать случай бесконечно большого объема генеральной совокупности.

Основные задачи математической статистики:

1. Оценка значения неизвестной вероятности случайного события;
2. Определение неизвестной теоретической функции распределения;
3. Определение неизвестных параметров распределения теоретической функции распределения;
4. Проверка статистических гипотез;
5. Оценка зависимости.

§2. Простейшие статистические преобразования

Для обоснованных статистических выводов необходимо иметь выборку достаточно большого объема n . Очевидно, что использование и хранение такой выборки весьма затруднительно. Чтобы избавиться от данных проблем, используют понятие статистики.

Определение. Статистикой $S = (S_1, S_2, \dots, S_n)$ называется произвольная k -мерная функция от выборки $X_1, X_2, \dots, X_n$:

$$S_1 = S_1(X_1, X_2, \dots, X_n),$$

...

$$S_k = S_k(X_1, X_2, \dots, X_n).$$

Как функция от случайного вектора $(X_1, X_2, \dots, X_n)$ статистика $S = (S_1, S_2, \dots, S_n)$ также будет случайным вектором.

Определение. Вариационным рядом $X_1^*, X_2^*, \dots, X_n^*$ является выборка $X_1, X_2, \dots, X_n$, элементы которой расположены в порядке возрастания элементов: $X_1^* \leq X_2^* \leq \dots \leq X_n^*$.

Очевидно, что данное преобразование не приводит к потере информации относительно теоретической функции распределения.

Для величин X_1^* и X_n^* употребляют название «крайние члены вариационного ряда».

Определение. Размахом варьирования называется разность между крайними членами вариационного ряда, т.е.

$$R = X_n^* - X_1^*. \quad (7.2.1)$$

Если среди элементов выборки $X_1, X_2, \dots, X_n$ имеются одинаковые, что происходит при наблюдении дискретной случайной величины, то целесообразно произвести группировку данных.

Определение. Значение выборки $X_1, X_2, \dots, X_n$, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных, называется **вариантой**, а численность отдельной группы сгруппированного ряда наблюдаемых данных, называется **частотой** или **весом** варианты.

Если i — индекс варианты, то n_i — число значений i -ой варианты.

Определение. Отношение частоты n_i к общей сумме всех частот $\sum n_i = n$ называется **относительной частотой** варианты и обозначается $p_i^* = \frac{n_i}{n}$.

Определение. Статистическим рядом называется расположенная по возрастанию совокупность различных вариантов Z_i , представляющих выборку $X_1, X_2, \dots, X_n$, с соответствующими им частотами или относительными частотами.

При наблюдении непрерывной случайной величины используют интервальный ряд. В этом случае весь возможный интервал, которому принадлежат значения выборки, разбивают на конечное число частичных интервалов и подсчитывают частоту попадания элементов выборки в каждый частичный интервал.

Определение. Интервальным рядом называется упорядоченная последовательность интервалов с соответствующими им частотами или относительными частотами попадания элементов выборки в каждый из этих интервалов.

Пример 1. В городе А для определения сроков гарантированного обслуживания проведено исследование величины среднего пробега автомобилей, находящихся в эксплуатации в течение двух лет. Получены следующие результаты (тыс. км.):

3,0; 25,0; 18,6; 12,1; 10,6; 18,0; 17,3; 29,1; 20,0; 18,3; 21,5; 26,7; 12,2; 14,4; 7,3; 9,1; 2,9; 5,4; 40,1; 16,8; 11,2; 9,9; 25,3; 4,2; 29,6.

Составить интервальный ряд.

О Решение. Очевидно, что величина среднего пробега автомобилей, находящихся в эксплуатации в течение двух лет, является непрерывной случайной величиной. Полученные данные представляют собой выборку из $n = 25$ наблюдений. Найдем сначала минимальное и максимальное значения случайной величины (т.е. крайние члены вариацион-

ного ряда): $X_{\min} = 2,9$ и $X_{\max} = 40,1$. Размах варьирования будет равен $R = X_n^* - X_1^* = 40,1 - 2,9 = 37,2$.

Возьмем число частичных интервалов $l = 6$. В этом случае длина частичного интервала равна

$$h = \frac{R}{n} = \frac{X_n^* - X_1^*}{n} = \frac{37,2}{6} = 6,2.$$

Соответствующий интервальный ряд приведен в таблице 7.1.

Таблица 7.1

Номер интервала i	Средний пробег автомобилей (интервалы) $x_i < X \leq x_{i+1}$	Частота n_i	Относительная частота $p_i^* = \frac{n_i}{n}$
1	2,9 — 9,1	6	0,24
2	9,1 — 15,3	6	0,24
3	15,3 — 21,5	7	0,28
4	21,5 — 27,7	3	0,12
5	27,7 — 33,9	2	0,08
6	33,9 — 40,1	1	0,04

Пример 2. Наблюдается число выигрышей в мгновенной лотерее. В результате наблюдения получены следующие значения выигрышей (тыс. руб.):

0, 1, 0, 0, 5, 0, 10, 0, 1, 0, 0, 1, 5, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 5, 0, 5, 0, 0, 1, 1, 1, 5, 10, 0, 1, 1, 0, 5, 0, 0, 0, 0, 1, 0, 1, 0, 5, 0, 0, 0, 0, 1, 0.

Требуется составить статистический ряд случайной величины X — выигрыша в мгновенной лотерее.

О Решение. Случайная величина X принимает 4 различных значения: 0, 1, 5 и 10. Для каждого значения подсчитаем частоту и относительную частоту. Результаты задачи представим в таблице 7.2.

Таблица 7.2

№	i	1	2	3	4
Выигрыш в мгновенной лотерее	Z_i	0	1	5	10
Частота	n_i	31	14	7	2
Относительная частота	p_i^*	31/54	7/27	7/54	1/27

§3. Эмпирическая функция распределения

По статистическом ряду, приведенному в таблице 7.3,

Таблица 7.3

№	1	2	...	L
Варианты Z_i	Z_1	Z_2	...	Z_L
Относительная частота p_i^*	$p_1^* = \frac{n_1}{n}$	$p_2^* = \frac{n_2}{n}$	...	$p_L^* = \frac{n_L}{n}$

можно построить эмпирическую (выборочную) функцию распределения.

Определение. Эмпирической (выборочной) функцией распределения называется функция $F^*(x)$, задающая для каждого значения x относительную частоту события $X < x$.

Следовательно, по определению

$$F^*(x) = \frac{n_x}{n}, \quad (7.3.1)$$

где n_x — число элементов выборки, значения которых меньше x .

Очевидно, что для нахождения функции распределения можно использовать формулу

$$F^*(x) = \sum_{Z_i < x} p_i^*. \quad (7.3.2)$$

Эмпирическую функцию распределения можно задать таблично или графически. Построим эмпирическую функцию распределения по данным, приведенным в таблице 7.2.

Объем выборки по условию примера $n = 54$. Наименьшая варианта равна 0, следовательно, $n_x = 0$ при $x \leq 0$. Тогда $F^*(x) = 0/54 = 0$ при $x \leq 0$. Если $0 < x \leq 1$, то неравенство $X < x$ выполняется для варианты $z_1 = 0$, которая встречается 31 раз, поэтому $n_x = 31$ и $F^*(x) = 31/54$. Если $1 < x \leq 5$, то неравенство $X < x$ выполняется для вариантов $z_1 = 0$ и $z_2 = 1$, которые встречаются 31 и 14 раз соответственно, поэтому, $n_x = 31 + 14 = 45$, $F^*(x) = 45/54$ и т.д. Результаты вычисления $F^*(x)$ приведем в таблице 7.4

Таблица 7.4

x	$F^*(x)$
$x \leq 0$	0
$0 < x \leq 1$	$F^*(x) = p_1^* = \frac{31}{54}$
$1 < x \leq 5$	$F^*(x) = p_1^* + p_2^* = \frac{31+14}{54} = \frac{45}{54}$
$5 < x \leq 10$	$F^*(x) = p_1^* + p_2^* + p_3^* = \frac{31+14+7}{54} = \frac{52}{54}$
$x > 10$	$F^*(x) = p_1^* + p_2^* + p_3^* + p_4^* = \frac{31+14+7+2}{54} = 1$

График этой функции приведен на рис. 7.1.

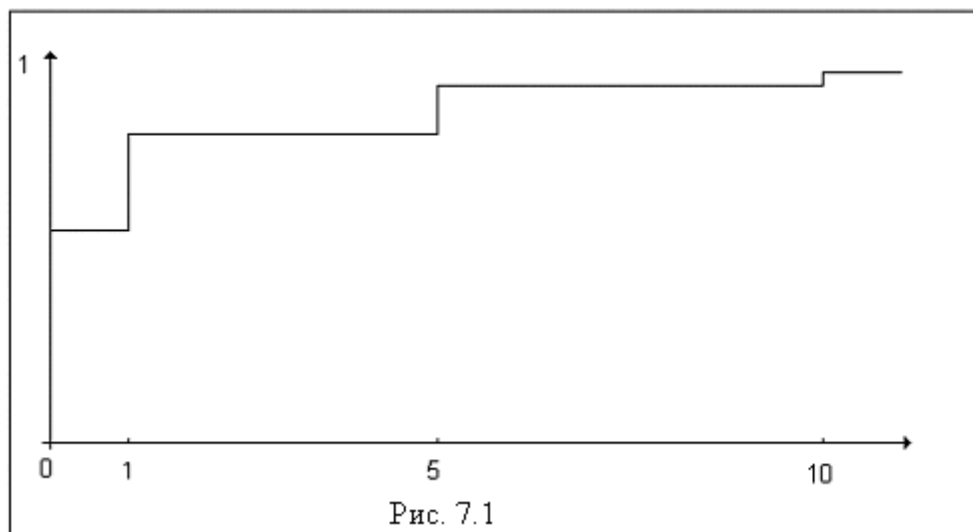


Рис. 7.1

В случае интервального ряда значения эмпирической функции $F^*(x)$ подсчитывают на концах частичных интервалов.

Эмпирическая функция $F^*(x)$ применяется для оценивания теоретической функции распределения генеральной совокупности.

§4. Полигон и гистограмма

Определение. Полигоном частот (многоугольником распределения) называется ломаная линия, проходящая через точки с координатами (Z_i, n_i) , где Z_i — варианты статистического ряда, а n_i — соответствующие им частоты.

Если ломаная линия строится по точкам (Z_i, p_i^*) , где p_i^* — относительные частоты, то получаем **полигон относительных частот**.

Построим полигон относительных частот для выборки из примера 2. Используя статистический ряд, представленный в таблице 7.2, получаем полигон относительных частот, изображенный на рис. 7.2.

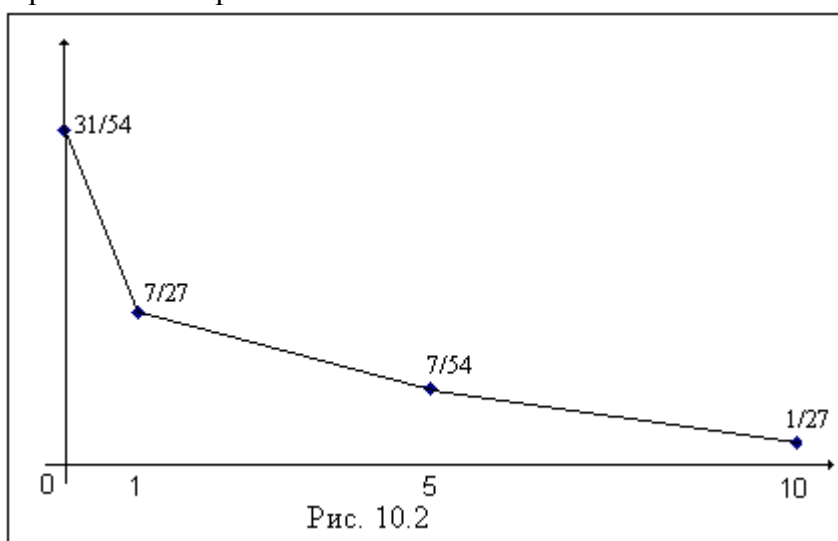


Рис. 10.2

В случае непрерывной случайной величины выборку преобразуют следующим образом. Всю ось абсцисс делят на интервалы (x_i, x_{i+1}) длины $\Delta_i = x_{i+1} - x_i$ и определяют функцию $f^*(x)$, которая на i -м интервале принимает значение

$$f^*(x) = \frac{p^*(x)}{\Delta_i} = \frac{n_i}{n\Delta_i}, \quad (7.4.1)$$

где n_i — число элементов выборки, попавших в интервал.

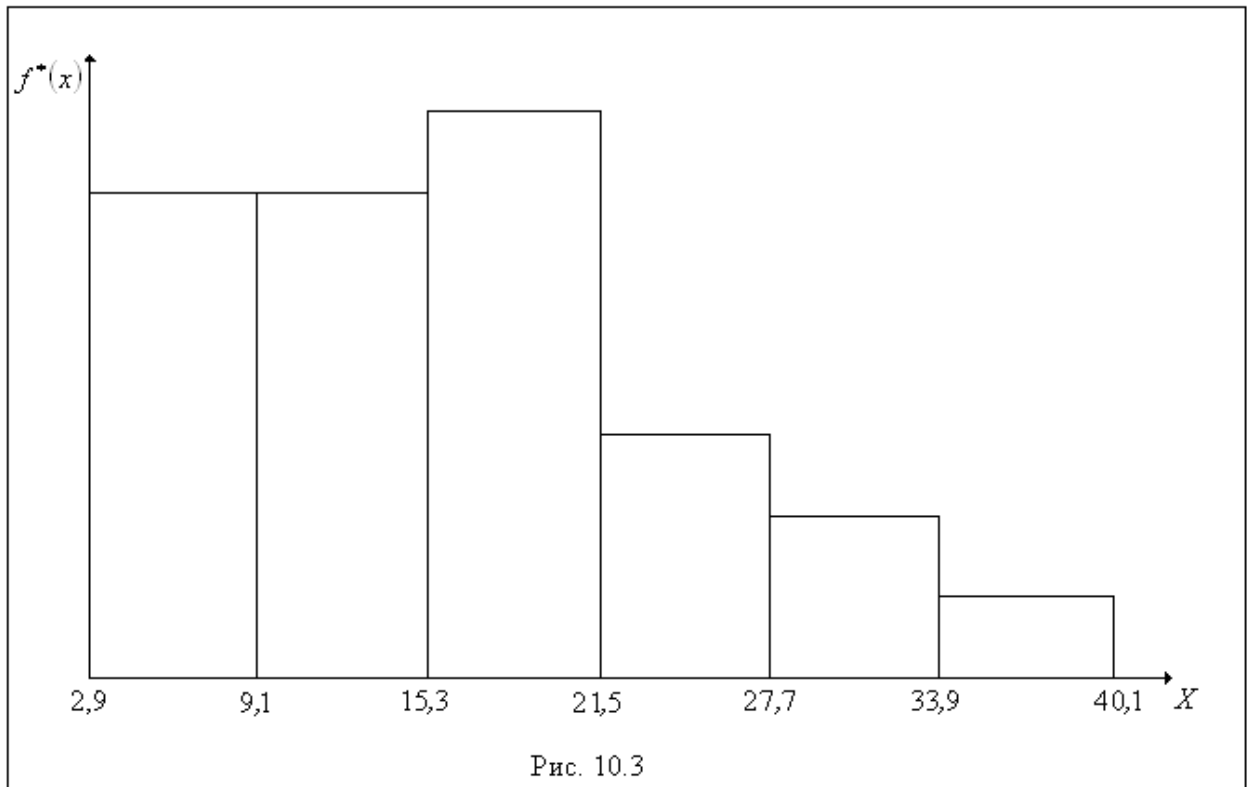
Определение. Функция $f^*(x)$, определенная соотношением (7.4.1), называется **гистограммой**.

Гистограмма является выборочной оценкой плотности вероятности.

Построим гистограмму по данным, приведенным в примере 1. Длина каждого интервала равна $\Delta_i = 6,2$. Подсчитаем значения $\frac{n_i}{n\Delta_i}$:

$x_i < X \leq x_{i+1}$	2,9 — 9,1	9,1 — 15,3	15,3—21,5	21,5—27,7	27,7—33,9	33,9—40,1
$\frac{n_i}{n\Delta_i}$	0,038710	0,038710	0,045161	0,019355	0,012903	0,006452

На рис. 7.3 представлена гистограмма примера 1.



Графическое изображение статистических рядов в виде полигона и гистограммы позволяет получить первоначальное представление о закономерностях, имеющих место в совокупности наблюдений.

Глава 8. Статистическое оценивание

§1. Точечные оценки. Выборочная средняя и выборочная дисперсия

Пусть θ — некоторый параметр закона распределения генеральной совокупности.

Определение. Точечной оценкой θ_n^* параметра θ называется произвольная функция $\theta_n^*(X_1, \dots, X_n)$ случайной выборки $X_1, \dots, X_n$.

Статистическая оценка является случайной величиной и меняется в зависимости от выборки. Опишем свойства, которым должна удовлетворять оценка θ_n^* .

Определение. Оценка θ_n^* называется **несмещенной**, если ее математическое ожидание равно оцениваемому параметру, т.е.

$$M(\theta_n^*) = \theta. \quad (8.1.1)$$

Определение. Оценка θ_n^* называется **состоятельной**, если с ростом объема выборки она сходится к оцениваемому параметру. Можно рассматривать сходимость различных типов: по вероятности, с вероятностью равной единице, в среднем квадратичном и т.д. Как правило, рассматривается сходимость по вероятности, т.е. состоятельной называется оценка θ_n^* , которая для каждого $\varepsilon > 0$ при всех возможных значениях неизвестного параметра θ удовлетворяет соотношению

$$\lim_{n \rightarrow \infty} P\{|\theta_n^* - \theta| > \varepsilon\} = 0. \quad (8.1.2)$$

Определение. Несмещенная оценка θ_n^* называется **эффективной**, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра θ , вычисленных по случайным выборкам одного и того же объема.

Если оценка не является несмещенной, то она будет либо завышать значение θ , либо занижать его. В обоих случаях это приводит к систематическим ошибкам одного знака в оценке параметра θ . Состоятельность оценки обосновывает увеличение объема случайной выборки, так как при этом становится менее вероятной возможность большой ошибки в оценке параметра θ .

Замечание. В дальнейшем вместо обозначения θ_n^* будем использовать θ^* .

Определение. Выборочной средней называется среднее арифметическое полученных по выборке значений

$$m^* = \frac{1}{n} \sum_{i=1}^n X_i. \quad (8.1.3)$$

Определение. Вторым выборочным моментом называется среднее арифметическое квадратов, полученных по выборке значений

$$m_2^* = \frac{1}{n} \sum_{i=1}^n X_i^2. \quad (8.1.4)$$

Определение. Выборочной дисперсией называется среднее арифметическое квадратов отклонений значений в случайной выборке от выборочной средней

$$\sigma^{2*} = \frac{1}{n} \sum_{i=1}^n (X_i - m^*)^2 = m_2^* - (m^*)^2. \quad (8.1.5)$$

Определение. Исправленной выборочной дисперсией называется произведение выборочной дисперсии на величину $\frac{n}{n-1}$, т.е.

$$S^{2*} = \frac{n}{n-1} \sigma^{2*} = \frac{1}{n-1} \sum_{i=1}^n (X_i - m^*)^2. \quad (8.1.6)$$

Пример 1. Проверить, является ли второй выборочный момент m_2^* несмещенной оценкой второго теоретического момента.

О Решение. Найдем математическое ожидание оценки m_2^* .

$$M(m_2^*) = \frac{1}{n} M(X_1^2 + X_2^2 + \dots + X_n^2) = \frac{1}{n} (MX_1^2 + MX_2^2 + \dots + MX_n^2) = \frac{1}{n} \sum_{i=1}^n m_2 = m_2. \bullet$$

Пример 2. Проверить, является ли оценка σ^{2*} несмещенной.

О Решение.

$$\begin{aligned} M\sigma^{2*} &= M\left(\frac{1}{n} \sum_{i=1}^n (X_i - m^*)^2\right) = M\left(\frac{1}{n} \sum_{i=1}^n (X_i^2 - 2m^* X_i + (m^*)^2)\right) = \\ &= m_2 - \frac{2}{n} \sum_{i=1}^n M(m^* X_i) + M(m^*)^2 = m_2 - \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n M(X_i X_j) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n M(X_i X_j) = \\ &= m_2 - \frac{1}{n^2} \sum_{i=1}^n M X_i^2 - \frac{1}{n^2} \sum_{\substack{i,j \\ i \neq j}} M(X_i X_j) = m_2 - \frac{1}{n} m_2 - \frac{n-1}{n} m^2 = \frac{n-1}{n} \sigma^2 \neq \sigma^2. \end{aligned}$$

Таким образом, σ^{2*} является смещенной оценкой дисперсии σ^2 . Очевидно, что S^{2*} будет уже несмещенной оценкой дисперсии σ^2 . •

§2. Метод моментов

Пусть имеется выборка $X_1, \dots, X_n$, произведенная из генеральной совокупности с теоретической функцией распределения $F(x)$, зависящей от k параметров $\theta_1, \dots, \theta_k$, которые нужно оценить. Зная функцию распределения, можно найти первые k теоретических моментов, которые будут зависеть от параметров $\theta_1, \dots, \theta_k$:

$$\begin{cases} m_1 = MX = m_1(\theta_1, \dots, \theta_k), \\ m_2 = MX^2 = m_2(\theta_1, \dots, \theta_k), \\ \dots \\ m_k = MX^k = m_k(\theta_1, \dots, \theta_k), \end{cases} \quad (8.2.1)$$

где X — случайная величина, имеющая функцию распределения $F(x)$.

Метод моментов состоит в том, что в системе (8.2.1) при большом объеме выборки n теоретические моменты $m_1, \dots, m_k$ заменяются на выборочные $m_1^*, \dots, m_k^*$, а затем, решая эту систему относительно $\theta_1, \dots, \theta_k$, находят оценки неизвестных параметров. Таким образом, в методе моментов оценки $\theta_1^*, \dots, \theta_k^*$ неизвестных параметров $\theta_1, \dots, \theta_k$ определяются из системы уравнений

$$\begin{cases} m_1^* = m_1(\theta_1^*, \dots, \theta_k^*), \\ m_2^* = m_2(\theta_1^*, \dots, \theta_k^*), \\ \dots \\ m_k^* = m_k(\theta_1^*, \dots, \theta_k^*). \end{cases} \quad (8.2.2)$$

Метод моментов был предложен в 1894 г. К. Пирсоном. Оценки, полученные методом моментов, как правило, являются состоятельными.

Пример 3. Выборка $X_1, \dots, X_n$ произведена из генеральной совокупности с теоретической функцией распределения, имеющей плотность показательного закона

$$p(x) = p(x, \theta) = \theta e^{-\theta x} \quad (x \geq 0).$$

Найти оценку параметра θ .

○ **Решение.** Математическое ожидание случайной величины X , имеющей плотность показательного закона, задаётся формулой

$$m_1 = MX = \int_0^{\infty} xp(x)dx = \int_0^{\infty} x\theta e^{-\theta x} dx = \frac{1}{\theta}.$$

Используя систему (8.2.2), получаем $m_1^* = \frac{1}{\theta^*}$.

Откуда окончательно получаем $\theta^* = \frac{1}{m^*}$. ●

§3. Метод максимального правдоподобия

Метод максимального правдоподобия является наиболее распространенным методом нахождения оценок. Метод максимального правдоподобия опирается на использование условий экстремума функций одной или нескольких случайных величин. В качестве такой функции используется функция правдоподобия.

Определение. Функцией правдоподобия называется функция

$$L(X_1, \dots, X_n) = L(X_1, \dots, X_n; \theta) = P(X_1; \theta) \cdots P(X_n; \theta) \quad (8.3.1)$$

в дискретном случае и

$$L(X_1, \dots, X_n) = L(X_1, \dots, X_n; \theta) = p(X_1; \theta) \cdots p(X_n; \theta) \quad (8.3.2)$$

в непрерывном случае.

В функции правдоподобия $L(X_1, \dots, X_n; \theta)$ элементы выборки $X_1, \dots, X_n$ являются фиксированными параметрами, а θ — аргументом.

Определение. Оценкой максимального правдоподобия называется такое θ^* , для которого

$$L(X_1, \dots, X_n; \theta^*) = \max_{\theta} L(X_1, \dots, X_n; \theta). \quad (8.3.3)$$

Поскольку L и $\ln(L)$ принимают максимум при одном и том же значении аргумента θ , то при практической реализации метода максимального правдоподобия удобно использовать не саму функцию правдоподобия, а ее логарифм.

Определение. Уравнением правдоподобия называется уравнение

$$\frac{\partial}{\partial \theta} \ln(L(X_1, \dots, X_n; \theta)) = 0. \quad (8.3.4)$$

В случае, когда теоретическая функция распределения $F(X_1, \dots, X_n; \theta_1, \dots, \theta_k)$ зависит от нескольких параметров $\theta_1, \dots, \theta_k$, при применении метода максимального правдоподобия вместо уравнения (8.3.4) необходимо использовать систему уравнений

$$\begin{cases} \frac{\partial}{\partial \theta_1} \ln(L(X_1, \dots, X_n; \theta_1, \dots, \theta_k)) = 0, \\ \dots \\ \frac{\partial}{\partial \theta_k} \ln(L(X_1, \dots, X_n; \theta_1, \dots, \theta_k)) = 0. \end{cases} \quad (8.3.5)$$

Пример 4. Найти оценки параметров m (математическое ожидание) и σ^2 (дисперсия) нормального закона распределения.

● **Решение.** Нормальный закон распределения характеризуется плотностью распределения

$$p(x, m, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Функция правдоподобия примет вид:

$$\begin{aligned} L(X_1, \dots, X_n; m, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(X_1-m)^2}{2\sigma^2}} \dots \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(X_n-m)^2}{2\sigma^2}} = \\ &= \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-\frac{1}{2\sigma^2}((X_1-m)^2 + \dots + (X_n-m)^2)}. \end{aligned} \quad (8.3.6)$$

Логарифмируя выражение (8.3.6), получим

$$\ln(L(X_1, \dots, X_n; m, \sigma^2)) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} ((X_1 - m)^2 + \dots + (X_n - m)^2). \quad (8.3.7)$$

Подставляя (8.3.7) в систему (8.3.5), получим

$$\begin{cases} \frac{\partial}{\partial m} \ln(L(X_1, \dots, X_n; m, \sigma^2)) = \frac{1}{\sigma^2} \left(\sum_{i=1}^n X_i - nm \right) = 0, \\ \frac{\partial}{\partial \sigma^2} \ln(L(X_1, \dots, X_n; m, \sigma^2)) = \frac{1}{\sigma^2} \left(\sum_{i=1}^n (X_i - m)^2 - n\sigma^2 \right) = 0. \end{cases} \quad (8.3.8)$$

Из первого уравнения системы (8.3.8) получаем

$$m^* = \frac{1}{n} \sum_{i=1}^n X_i, \quad (8.3.9)$$

а из второго уравнения системы (8.3.8), с учетом (8.3.9), получаем

$$\sigma^{2*} = \frac{1}{n} \sum_{i=1}^n (X_i - m^*)^2.$$

Следует отметить, что полученная оценка является смещенной.

Читателю предлагается самостоятельно показать, что m^* и σ^{2*} доставляют максимум функции правдоподобия $L(X_1, \dots, X_n; m, \sigma^2)$. ●

§4. Интервальные оценки (доверительные интервалы)

Статистическая оценка θ^* некоторого параметра распределения θ наблюдаемой случайной величины X , будучи функцией случайной выборки (статистикой), сама является случайной величиной, имеющей свой закон распределения и числовые характеристики (параметры) распределения. При малом числе наблюдений могут возникнуть следующие задачи:

- к каким ошибкам может привести замена параметра θ его точечной оценкой θ^* ;
- с какой степенью надежности можно ожидать, что получаемые ошибки не выйдут за известные пределы.

Определение. Интервальной оценкой (доверительным интервалом) называется числовой интервал (θ_L^*, θ_R^*) , определяемый по результатам выборки, относительно которого можно утверждать с определенной, близкой к единице, вероятности, что он заключает в себе значение оцениваемого параметра генеральной совокупности, т.е.

$$P(\theta_L^* \leq \theta \leq \theta_R^*) = \gamma, \quad (8.4.1)$$

где θ_L^* и θ_R^* — нижняя и верхняя (левая и правая) границы доверительного интервала параметра θ , γ — **доверительная вероятность**.

Доверительная вероятность и уровень значимости связаны соотношением

$$\alpha + \gamma = 1. \quad (8.4.2)$$

По заданной оценке θ^* и заданной доверительной вероятности γ доверительные интервалы можно построить различными способами. На практике обычно используются два типа доверительных интервалов:

- двусторонние;
- односторонние.

Ограничимся нахождением двусторонних доверительных интервалов. Односторонние доверительные интервалы находятся аналогично.

Для каждой доверительной вероятности γ можно указать такое значение Δ , что

$$P(|\theta^* - \theta| \leq \Delta) = P(\theta^* - \Delta \leq \theta \leq \theta^* + \Delta) = \gamma. \quad (8.4.3)$$

Определение. Величину Δ , равную половине ширины доверительного интервала, называют **точностью оценки**.

Определение. Квантилем уровня γ некоторого распределения называется такое число t_γ , при котором значение соответствующей функции распределения равно γ , т.е.

$$\gamma = F(t_\gamma). \quad (8.4.4)$$

Теорема. Пусть плотность случайной величины X симметрична относительно оси OY , и пусть $P(|X| < t) = \gamma$. Тогда t — квантиль уровня $\frac{\gamma+1}{2}$ распределения случайной величины X .

Доказательство.

$$\gamma = P(|X| < t) = 1 - P(|X| \geq t) = 1 - P(X \geq t) - P(X \leq -t) = 1 - 2P(X \geq t),$$

т.к. ввиду симметричности закона распределения относительно оси OY выполнено равенство $P(X \geq t) = P(X \leq -t)$. Отсюда следует, что

$$P(|X| \geq t) = \frac{1-\gamma}{2} = 1 - P(X < t) \quad \text{и} \quad P(X < t) = \frac{1+\gamma}{2}. \blacksquare$$

Рассмотрим теперь правила построения доверительных интервалов для некоторых параметров распределения.

Доверительный интервал для математического ожидания генеральной совокупности, имеющей нормальное распределение при известной дисперсии.

Пусть из генеральной совокупности X , имеющей нормальный закон распределения при известной дисперсии σ^2 и неизвестном математическом ожидании m , произведена случайная выборка $X_1, \dots, X_n$. Для оценки математического ожидания используем статистику

$m^* = \frac{1}{n} \sum_{i=1}^n X_i$, которая имеет нормальное распределение с параметрами $\left(m, \frac{\sigma^2}{n}\right)$.

Тогда статистика $\frac{m^* - m}{\sigma} \sqrt{n}$ имеет нормальное распределение с параметрами $(0,1)$. Найдем вероятность отклонения $|m^* - m|$:

$$\gamma = P\left(\left|\frac{m^* - m}{\sigma} \sqrt{n}\right| < t\right) = P\left(|m^* - m| < t \frac{\sigma}{\sqrt{n}}\right) = P\left(m^* - t \frac{\sigma}{\sqrt{n}} < m < m^* + t \frac{\sigma}{\sqrt{n}}\right). \quad (8.4.5)$$

Интервал $\left(m^* - t \frac{\sigma}{\sqrt{n}}; m^* + t \frac{\sigma}{\sqrt{n}} \right)$, определенный по (8.4.5), представляет собой доверительный интервал для математического ожидания m при известной дисперсии σ^2 . Осталось указать, как, зная доверительную вероятность γ , выбрать в (8.4.5) значение t . Ответом на этот вопрос является доказанная ранее теорема, т.е. t является квантилем уровня $\frac{\gamma+1}{2}$ нормального распределения с параметрами $(0,1)$.

Таким образом, доверительный интервал для математического ожидания m при известной дисперсии σ^2 строится следующим образом:

1. Вычисляем оценку m^* , которая является средним арифметическим элементов выборки $X_1, \dots, X_n$.

2. Вычисляем квантиль $t_{\frac{\gamma+1}{2}}$ нормального распределения с параметрами $(0,1)$.

3. Строим доверительный интервал, который имеет вид:

$$\left(m^* - \frac{\sigma}{\sqrt{n}} t_{\frac{\gamma+1}{2}}; m^* + \frac{\sigma}{\sqrt{n}} t_{\frac{\gamma+1}{2}} \right). \quad (8.4.6)$$

Очевидно, что точность оценки равна

$$\Delta = t_{\frac{\gamma+1}{2}} \frac{\sigma}{\sqrt{n}}. \quad (8.4.7)$$

Замечание. Квантиль $t_{\frac{\gamma+1}{2}}$ нормального распределения с параметрами $(0,1)$ можно определить разными способами, например:

- используя таблицы функции $\Phi(t)$ нормального распределения с параметрами $(0,1)$ (Приложение 3);
- используя функцию *НОРМСТОБР(вероятность)* из EXCEL, при этом аргумент «вероятность» данной функции должен быть равен $\frac{\gamma+1}{2}$;
- используя функцию *qnorm(v, m, σ)* из MATHCAD, при этом переменные данной функции должны быть равны $v = \frac{\gamma+1}{2}$, $m = 0$, $\sigma = 1$.

Доверительный интервал для математического ожидания генеральной совокупности, имеющей нормальное распределение при неизвестной дисперсии.

Пусть из генеральной совокупности X , имеющей нормальный закон распределения при неизвестной дисперсии σ^2 и неизвестном математическом ожидании m , произведена случайная выборка $X_1, \dots, X_n$. Для оценки математического ожидания используем статистику

$$T = \frac{m^* - m}{S^*} \sqrt{n-1}, \quad (8.4.8)$$

имеющую t -распределение (распределение Стьюдента) с $\nu = n-1$ числом степеней свободы.

Доверительный интервал для математического ожидания при неизвестной дисперсии имеет вид:

$$\left(m^* - \frac{S^*}{\sqrt{n}} t_{\frac{\gamma+1}{2}}; m^* + \frac{S^*}{\sqrt{n}} t_{\frac{\gamma+1}{2}} \right). \quad (8.4.9)$$

Очевидно, что точность оценки равна

$$\Delta = t_{\frac{\gamma+1}{2}} \frac{S^*}{\sqrt{n}}. \quad (8.4.10)$$

Таким образом, доверительный интервал для математического ожидания m при неизвестной дисперсии σ^2 строится следующим образом:

1. Вычисляем оценку m^* , которая является средним арифметическим элементов выборки $X_1, \dots, X_n$, и исправленную выборочную дисперсию

$$S^{2*} = \frac{1}{n-1} \sum_{i=1}^n (X_i - m^*)^2.$$

2. Вычисляем квантиль $t_{\frac{\gamma+1}{2}}$ t -распределения (распределения Стьюдента) с $v = n - 1$ числом степеней свободы.
3. Используя формулу (8.4.9), получаем доверительный интервал.

Замечание. Квантиль $t_{\frac{\gamma+1}{2}}$ t -распределения (распределения Стьюдента) с $l = n - 1$

числом степеней свободы можно определить разными способами, например:

- используя таблицы t -распределения (распределения Стьюдента) с $l = n - 1$ числом степеней;
- используя функцию `СТЮДРАСПОБР(вероятность; степени_свободы)` из EXCEL, при этом аргумент «вероятность» данной функции должен быть равен удвоенному уровню значимости, а аргумент «степени_свободы» должен быть равен $n - 1$;
- используя функцию `qt(p, d)` из MATHCAD, при этом переменные данной функции должны быть равны $p = \gamma$, $d = n - 1$.

Замечание. При достаточно большом объеме n выборки различия между доверительными интервалами, определенными по формулам (8.4.6) и (8.4.9), мало, т.к., при $n \rightarrow \infty$, распределение Стьюдента стремится к нормальному распределению.

Доверительный интервал для дисперсии генеральной совокупности, имеющей нормальное распределение при известном математическом ожидании.

Предположим, что выборка $X_1, \dots, X_n$ произведена из нормальной генеральной совокупности с неизвестной дисперсией при известном математическом ожидании, равном m . В качестве оценки неизвестной дисперсии σ^2 возьмем выборочную дисперсию

$$\sigma^{2*} = \frac{1}{n} ((X_1 - m)^2 + \dots + (X_n - m)^2). \quad (8.4.11)$$

В этом случае статистика $\chi^2 = \frac{n\sigma^{2*}}{\sigma^2}$ будет иметь χ^2 -распределение с n степенями свободы.

Доверительный интервал для дисперсии при известном математическом ожидании имеет вид:

$$\frac{n\sigma^{2*}}{r_{\frac{1+\gamma}{2}}} < \sigma^2 < \frac{n\sigma^{2*}}{r_{\frac{1-\gamma}{2}}}, \quad (8.4.12)$$

где r_τ — квантиль уровня τ χ^2 -распределения с n степенями свободы.

Таким образом, доверительный интервал для дисперсии при известном математическом ожидании строится следующим образом:

1. Используя выборку $X_1, \dots, X_n$, по формуле (8.4.11) вычисляем оценку дисперсии.
2. Вычисляем квантили $r_{\frac{\gamma+1}{2}}$ и $r_{\frac{1-\gamma}{2}}$ χ^2 -распределения с n степенями свободы.
3. По формуле (8.4.12) получаем искомый доверительный интервал.

Замечание. Квантили $r_{\frac{\gamma+1}{2}}$ и $r_{\frac{1-\gamma}{2}}$ χ^2 -распределения с n степенями свободы можно определить разными способами, например:

- используя таблицы χ^2 -распределения с n степенями свободы;
- используя функцию *ХИ2ОБР (вероятность; степени_свободы)* из EXCEL, при этом аргумент «вероятность» данной функции должен быть равен уровню значимости, а аргумент «степени_свободы» должен быть равен n ;
- используя функцию *qchisq(p, d)* из MATHCAD, при этом переменные данной функции должны быть равны: $p = \gamma$, $d = n$.

Аналогично строится доверительный интервал для дисперсии генеральной совокупности, имеющей нормальное распределение при неизвестном математическом ожидании.

В качестве оценки неизвестной дисперсии σ^2 возьмем выборочную дисперсию

$$S^{2*} = \frac{1}{n-1} \left((X_1 - m^*)^2 + \dots + (X_n - m^*)^2 \right). \quad (8.4.13)$$

В этом случае статистика $\chi^2 = \frac{(n-1)S^{2*}}{\sigma^2}$ будет иметь χ^2 -распределение с $n-1$ степенью свободы.

Доверительный интервал для дисперсии при неизвестном математическом ожидании имеет вид:

$$\frac{(n-1)S^{2*}}{r_{\frac{1+\gamma}{2}}} < \sigma^2 < \frac{(n-1)S^{2*}}{r_{\frac{1-\gamma}{2}}}, \quad (8.4.14)$$

где r_τ — квантиль уровня τ χ^2 -распределения с $n-1$ степенью свободы.

Таким образом, доверительный интервал для дисперсии при неизвестном математическом ожидании строится следующим образом:

1. Вычисляем оценку m^* , которая является средним арифметическим элементов выборки $X_1, \dots, X_n$, и по формуле (8.4.13) вычисляем выборочную оценку дисперсии.
2. Вычисляем квантили $r_{\frac{\gamma+1}{2}}$ и $r_{\frac{1-\gamma}{2}}$ χ^2 -распределения с $n-1$ степенью свободы.
3. По формуле (8.4.14) получаем искомый доверительный интервал.

Пример 5. Найти доверительный интервал с доверительной вероятностью 0,95 для оценки математического ожидания, нормально распределенной случайной величины X , если известны ее дисперсия $\sigma^2 = 16$, выборочная средняя $m^* = 16$ и объем выборки $n = 16$.

Решение. По условию задачи $\gamma = 0,95$. Найдем квантиль $t_{\frac{\gamma+1}{2}} = t_{\frac{0,95+1}{2}} = t_{0,975}$ нормального распределения с параметрами $(0,1)$, используя, например EXCEL: применяя функцию *НОРМСТОБР(0,975)*, получим $t_{0,975} = 1,96$.

Применив (8.4.6), получим

$$m^* - \frac{\sigma}{\sqrt{n}} t_{\frac{\gamma+1}{2}} < m < m^* + \frac{\sigma}{\sqrt{n}} t_{\frac{\gamma+1}{2}},$$

$$16 - \frac{4}{\sqrt{16}} 1,96 < m < 16 + \frac{4}{\sqrt{16}} 1,96,$$

$$14,04 < m < 17,96. \bullet$$

Пример 6. Случайная величина X распределена по нормальному закону. Статистический ряд выборки представлен таблицей:

Z_i	3	5	7	8	10	12	14
n_i	3	7	4	6	7	5	8

Требуется:

- 1) построить доверительный интервал для оценки математического ожидания, если доверительная вероятность равна 0,97;
- 2) построить доверительный интервал для оценки дисперсии, если доверительная вероятность равна 0,95.

О Решение. Для оценки математического ожидания будем использовать формулу (8.4.9), т.к. дисперсия неизвестна. Вычислим все величины, присутствующие в (8.4.9).

Объем выборки $n = 3 + 7 + 4 + 6 + 7 + 5 + 8 = 40$.

Оценка математического ожидания:

$$m^* = \frac{1}{n} \sum_{i=1}^7 n_i Z_i = \frac{1}{40} (3 \cdot 3 + 7 \cdot 5 + 4 \cdot 7 + 6 \cdot 8 + 7 \cdot 10 + 5 \cdot 12 + 8 \cdot 14) = \frac{362}{40} = 9,05.$$

Исправленная выборочная дисперсия:

$$S^{2*} = \frac{1}{n-1} \sum_{i=1}^7 n_i (Z_i - m^*)^2 = \frac{493,9}{39} = 12,664.$$

Корень квадратный исправленной выборочной дисперсии:

$$S^* = \sqrt{S^{2*}} = \sqrt{12,664} = 3,559.$$

Для доверительной вероятности $\gamma = 0,97$ найдем квантиль $t_{\frac{\gamma+1}{2}} = t_{\frac{0,97+1}{2}} = t_{0,985}$

t -распределения) с $\nu = n - 1 = 40 - 1 = 39$ числом степеней свободы. Для этого используем, например, функцию $qt(p, d)$ из MATHCAD: $qt(0.985, 39) = 2.252$.

Доверительный интервал имеет вид:

$$9,05 - \frac{3,559}{\sqrt{40}} 2,252 < m < 9,05 + \frac{3,559}{\sqrt{40}} 2,252$$

и окончательно

$$7,77 < m < 10,33.$$

Перейдем ко второй части задачи. Для оценки дисперсии будем использовать формулу (8.4.14), т.к. математическое ожидание неизвестно. Для данной формулы необходимо вычислить квантили $r_{\frac{\gamma+1}{2}} = r_{\frac{0,95+1}{2}} = r_{0,975}$ и $r_{\frac{1-\gamma}{2}} = r_{\frac{1-0,95}{2}} = r_{0,025}$. Используя функцию

$qchisq(p, d)$ из MATHCAD, получим: $r_{0,975} = qchisq(0.975, 39) = 58.12$ и $r_{0,025} = qchisq(0.025, 39) = 23.654$.

Доверительный интервал имеет вид:

$$\frac{39 \cdot 12,664}{58,12} < \sigma^2 < \frac{39 \cdot 12,664}{23,654}$$

и окончательно

$$8,50 < \sigma^2 < 20,88. \bullet$$

Глава 9. Проверка статистических гипотез

§1. Основные понятия

Определение. Статистической гипотезой называется любое предположение H_0 (гипотеза) о виде закона распределения генеральной совокупности или о числовых значениях параметров закона распределения.

Определение. Правило, по которому гипотеза H_0 принимается или отвергается, называется **статистическим критерием**.

Проверяемую гипотезу H_0 называют нулевой, а противоположную ей гипотезу H_1 называют альтернативной.

Схема проверки нулевой гипотезы:

1. Используя проверочные данные $X_1, \dots, X_n$ и учитывая условия задачи, принимают нулевую гипотезу H_0 и альтернативную гипотезу H_1 .

2. По случайной выборке $X_1, \dots, X_n$ определяется функция $S(X_1, \dots, X_n)$, называемая статистикой, для которой будет известен точный или приближённый закон распределения.

3. По заранее выбранной малой вероятности α определяется критическая область W_k , для которой $P(S(X_1, \dots, X_n) \in W_k) = \alpha$. И если величина $S(X_1, \dots, X_n)$, вычисляется при конкретной выборке $X_1, \dots, X_n$, окажется вне критической области W_k , то гипотеза H_0 принимается, а если она окажется в области W_k , то гипотеза H_0 отвергается (или принимается гипотеза H_1). При этом возможны 4 случая, которые представлены в таблице 9.1

Таблица 9.1

	Принимается H_0	Отвергается H_0
Верна гипотеза H_0	Правильное решение	Ошибка первого рода
вероятность	$1 - \alpha$	α
Верна H_1	Ошибка второго рода	Правильное решение
вероятность	β	$1 - \beta$

Определение. Вероятность α допустить ошибку первого рода называется **уровнем значимости** критерия.

Определение. Вероятность $1 - \beta$ не допустить ошибку второго рода называется **мощностью** критерия.

Если использовать терминологию качества продукции, то α — это «риск поставщика», связанный с забраковкой по результату выборки всей партии товара, соответствующей стандарту, а β — «риск потребителя», связанный с принятием по результатам выборки партии товара, не соответствующей стандарту.

Возможны три варианта расположения критической области:

1. Правосторонняя критическая область (рис 9.1), состоящая из интервала $(t_{кр}^R, \infty)$, где $t_{кр}^R$ определяется из условия:

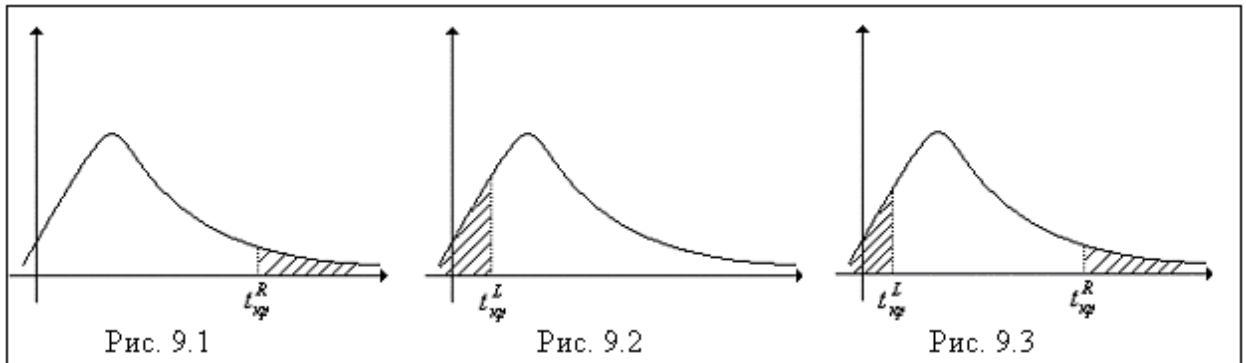
$$P(S(X_1, \dots, X_n) > t_{кр}^R) = \alpha. \quad (9.1.1)$$

2. Левосторонняя критическая область (рис 9.2), состоящая из интервала $(-\infty; t_{кр}^L)$, где $t_{кр}^L$ определяется из условия:

$$P(S(X_1, \dots, X_n) < t_{кр}^L) = \alpha. \quad (9.1.2)$$

3. Двусторонняя критическая область (рис 9.3), состоящая из интервалов $(-\infty; t_{кр}^L)$ и $(t_{кр}^R; \infty)$, где точки $t_{кр}^L$ и $t_{кр}^R$ определяется из условий:

$$P(S(X_1, \dots, X_n) < t_{кр}^L) = \frac{\alpha}{2} \text{ и } P(S(X_1, \dots, X_n) > t_{кр}^R) = \frac{\alpha}{2}. \quad (9.1.3)$$



В следующих параграфах рассмотрим несколько конкретных практических примеров.

§2. Проверка гипотезы о значении математического ожидания

Пусть из генеральной совокупности X , распределенной по нормальному закону с параметрами (m, σ^2) при неизвестном математическом ожидании m и неизвестной дисперсии σ^2 , взята случайная выборка $X_1, \dots, X_n$ объемом n и вычислена выборочная средняя арифметическая

$$m^* = \frac{1}{n} \sum_{i=1}^n X_i,$$

а m_0 и m_1 — определенные значения параметра m . Для проверки нулевой гипотезы $H_0 : m = m_0$ при конкурирующей гипотезе $H_1 : m = m_1$ используют статистику

$$T = \frac{m^* - m}{S^*} \sqrt{n-1}, \quad (9.2.1)$$

которая при выполнении нулевой гипотезы H_0 имеет t -распределение (распределение Стьюдента) с $\nu = n - 1$ степени свободы.

Учитывая соотношения (9.1.1) — (9.1.3), определяем границы $t_{кр}^L$ и $t_{кр}^R$ критической области по таблицам t -распределения (распределения Стьюдента) или используя функцию *СТЬЮДРАСПОБР(вероятность; степени_свободы)* из EXCEL, или используя функцию *qt(p, d)* из MATHCAD.

При проверке нулевой гипотезы $H_0 : m = m_0$ при известной дисперсии σ^2 используют статистику

$$T = \frac{m^* - m}{\sigma} \sqrt{n}, \quad (9.2.2)$$

которая при выполнении нулевой гипотезы H_0 имеет нормальное распределение с параметрами $(0,1)$.

Учитывая соотношения (9.1.1) — (9.1.3), определяем границы $t_{кр}^L$ и $t_{кр}^R$ критической области по таблицам нормального распределения с параметрами $(0,1)$ или используя функцию *НОРМСТОБР(вероятность)* из EXCEL, или используя функцию *qnorm(v, m, σ)* из MATHCAD.

Пример 1. Поставщик двигателей утверждает, что средний срок их службы равен 800 ч. Для выборки из 17 двигателей средний срок службы оказался равным 865 ч. При «исправленном» среднем квадратичном отклонении 120 ч. Проверить гипотезу о том, что значение 800 является математическим ожиданием при 1% уровне значимости.

○ **Решение.** Предположим, что случайная величина среднего времени службы подчинена нормальному закону о числовом значении математического ожидания нормально распределённой величины (генеральной средней) при неизвестной генеральной дисперсии. В этом случае в качестве критерия выбирают функцию

$$T = \frac{m^* - m}{S^*} \sqrt{n-1},$$

где m^* — выборочная средняя, m — математическое ожидание, S^* — «исправленное» выборочное среднее квадратичное отклонение. Случайная величина T имеет t - распределение (распределение Стьюдента) с $\nu = n - 1$ степени свободы.

Требуется найти критическую область для нулевой гипотезы $H_0 : m = 800$ при альтернативной гипотезе $H_1 : m > 800$.

Критическая область правосторонняя, $t_{кр}^R$ находим из условия $P(T > t_{кр}^R) = \alpha$.

При $\alpha = 0,01$ и $\nu = 17 - 1 = 16$ в таблице t - распределения находим $t_{кр}^R = 1,746$. Таким образом, критическая область $W_{кр} = (1,746 ; +\infty)$. Найдем расчетное значение $t_{расч}$, полагая $m = 800$:

$$t_{расч} = \frac{m^* - m}{S^*} \sqrt{n-1} = \frac{865 - 800}{120} \sqrt{16} = \frac{65}{30} \approx 2,0167.$$

Очевидно, что найденное значение $t_{расч} \approx 2,0167$ попадает в критическую область. Таким образом, нулевая гипотеза не подтверждается. ●

§3. Проверка гипотезы о равенстве математических ожиданий двух генеральных совокупностей

Пусть X и Y — две генеральные совокупности с известными дисперсиями σ_X^2 и σ_Y^2 и неизвестными математическими ожиданиями m_X и m_Y . Из генеральных совокупностей взяты две независимые выборки $X_1, \dots, X_n$ и $Y_1, \dots, Y_k$ и вычислены выборочные математические ожидания (средние) m_X^* и m_Y^* . Для проверки гипотезы о равенстве математических ожиданий $H_0 : m_X = m_Y$ используют статистику

$$T = \frac{m_X^* - m_Y^*}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{k}}}, \quad (9.3.1)$$

которая при выполнении нулевой гипотезы H_0 имеет нормальное распределение с параметрами $(0,1)$.

При неизвестных дисперсиях генеральных совокупностей либо требуется достаточно большой объем выборки для надежной и точной оценки, либо требуется, чтобы эти дисперсии были одинаковы, в противном случае известные критерии малоэффективны. Если дисперсии генеральных совокупностей равны $\sigma_X^2 = \sigma_Y^2$, то для проверки нулевой гипотезы $H_0: m_X = m_Y$, используют статистику

$$T = \frac{m_X^* - m_Y^*}{\sqrt{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right) \cdot \frac{S_X^{2*}(n_X - 1) + S_Y^{2*}(n_Y - 1)}{n_X + n_Y - 2}}}, \quad (9.3.2)$$

имеющую t - распределение (распределение Стьюдента) с $\nu = n_X + n_Y - 2$ степенями свободы.

Выбор критической области зависит от вида альтернативной гипотезы:

- при альтернативной гипотезе $H_1: m_X > m_Y$ необходимо выбрать правостороннюю критическую область;
- при альтернативной гипотезе $H_1: m_X < m_Y$ необходимо выбрать левостороннюю критическую область;
- при альтернативной гипотезе $H_1: m_X \neq m_Y$ необходимо выбрать двустороннюю критическую область.

Пример 2. Средний ежедневный объем продаж за I квартал 2004 года для 17 торговцев района А составляет 15 тысяч рублей при «исправленном» среднем квадратичном отклонении 2,5 тысяч рублей, а для 10 торговцев района В — 13 тысяч рублей при «исправленном» среднем квадратичном отклонении 3 тысячи рублей. Существенно ли различие в объеме продаж в районах А и В при 5%-м уровне значимости.

О Решение. Решим задачу, предположив, что ежедневный объем продаж подчинён нормальному закону при неизвестных параметрах распределения. Предположим, что дисперсии объемов продаж одинаковы. Найдём критическую область для нулевой гипотезы $H_0: m_X = m_Y$ при альтернативной гипотезе $H_1: m_X \neq m_Y$, где m_X — математическое ожидание объема продаж для района А, а m_Y — для района В.

В качестве критерия необходимо использовать функцию

$$T = \frac{m_X^* - m_Y^*}{\sqrt{\left(\frac{1}{n_X} + \frac{1}{n_Y}\right) \cdot \frac{S_X^{2*}(n_X - 1) + S_Y^{2*}(n_Y - 1)}{n_X + n_Y - 2}}},$$

где n_X и n_Y объемы выборок, а m_X^* и m_Y^* оценки параметров m_X и m_Y .

Функция T имеет t - распределение (распределение Стьюдента) с $\nu = n_X + n_Y - 2 = 17 + 10 - 2 = 25$ степенями свободы.

По таблице t - распределения для $l = 25$ и 5%-го уровня значимости (для двусторонней критической области) находим:

$$P(T < t_{кр}^L) = \frac{0,05}{2} \Rightarrow t_{кр}^L = -2,06, \quad P(T > t_{кр}^R) = \frac{0,05}{2} \Rightarrow t_{кр}^R = 2,06.$$

Это означает, что критической областью является $W_k = (-\infty; -2,06) \cup (2,06; \infty)$.

Вычислим $t_{расч}$:

$$t_{расч} = \frac{15 - 13}{\sqrt{\left(\frac{1}{17} + \frac{1}{10}\right) \cdot \frac{2,5^2 \cdot 16 + 3^2 \cdot 9}{25}}} = 1,86.$$

Полученное значение критерия $t_{расч}$ не принадлежит критической области, следовательно, гипотеза $H_0 : m_X = m_Y$ принимается. ●

§4. Проверка гипотезы о значении дисперсии генеральной совокупности

Пусть X — генеральная совокупность, распределенная по нормальному закону с неизвестной дисперсией σ^2 . Из генеральной совокупности взята случайная выборка $X_1, \dots, X_n$ и вычислена выборочная дисперсия σ^{2*} .

Требуется проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2$, где σ_0^2 — заданное значение дисперсии генеральной совокупности. Для проверки нулевой гипотезы используют статистику

$$T = \frac{n\sigma^{2*}}{\sigma_0^2}, \quad (9.4.1)$$

которая при выполнении нулевой гипотезы H_0 имеет χ^2 -распределение с $\nu = n - 1$ степенями свободы.

Границы $t_{кр}^L$ и $t_{кр}^R$ критической области определяем по таблицам χ^2 -распределения с $\nu = n - 1$ степенями свободы или используя функцию *ХИ2ОБР* (вероятность; степени_свободы) из EXCEL, или используя функцию *qchisq(p, d)* из MATHCAD.

Пример 3. Точность работы автоматической линии проверяют по дисперсии контролируемого размера, которая не должна превышать $0,1 \text{ мм}^2$. По результатам выборочного контроля получены следующие данные:

Контролируемый размер, Z_i , мм ²	43,0	43,5	43,8	44,4	44,6
Частота n_i	3	7	10	8	2

Требуется проверить при уровне значимости 0,01, обеспечивает ли линия требуемую точность.

○ **Решение.** Задача состоит в проверке гипотезы о значении дисперсии $H_0 : \sigma^2 = 0,1$ при альтернативной гипотезе $H_1 : \sigma^2 > 0,1$. Таким образом, необходимо построить правостороннюю критическую область. Расчетное значение критерия вычисляем по формуле $t_{расч} = \frac{n\sigma^{2*}}{\sigma_0^2}$, следовательно, по данным статистического ряда необходимо вычислить выборочную дисперсию. Дальнейшие вычисления удобно оформить в виде таблицы:

№	Z_i	n_i	$n_i Z_i$	$Z_i - m^*$	$(Z_i - m^*)^2$	$n_i (Z_i - m^*)^2$
1	43,0	3	129,0	-0,863	0,745	2,236
2	43,5	7	304,5	-0,363	0,132	0,924
3	43,8	10	438,0	-0,063	0,004	0,040
4	44,4	8	355,2	0,537	0,288	2,304
5	44,6	2	89,2	0,737	0,543	1,085
Итого		30	1315,9			6,590
Среднее			43,863			0,220

$$m^* = \frac{\sum n_i Z_i}{\sum n_i} = \frac{1315,9}{30} = 43,863; \quad \sigma^{2*} = \frac{\sum n_i (Z_i - m^*)^2}{\sum n_i} = \frac{6,59}{30} = 0,22;$$

$$t_{расч} = \frac{n \sigma^{2*}}{\sigma_0^2} = \frac{30 \cdot 0,22}{0,1} = 66.$$

Из условия $P(T > t_{кр}^R) = 0,01$, при числе степеней свободы равным $\nu = n - 1 = 30 - 1 = 29$ находим границу правосторонней критической области. По таблице χ^2 - распределения получаем $t_{кр}^R = 49,5878$.

Заметим, что аналогичный результат можно получить с помощью EXCEL или MATHCAD: $t_{кр}^R = \text{ХИ2ОБР}(0,01; 29) = 49,5878$ или $t_{кр}^R = qchisq(0.99, 29) = 49,5878$.

Это означает, что критической областью является $W_k = (49,5878; \infty)$. Очевидно, что расчетное значение $t_{расч} = 66$ находится внутри критической области, следовательно, нулевая гипотеза отвергается, т.е. автоматическая линия не обеспечивает заданную точность и требуется ее регулировка. ●

§5. Проверка гипотезы о равенстве дисперсий двух генеральных совокупностей

Пусть X и Y — две генеральные совокупности, распределенные по нормальному закону с неизвестными дисперсиями σ_X^2 и σ_Y^2 . Из генеральных совокупностей взяты две независимые выборки $X_1, \dots, X_n$ и $Y_1, \dots, Y_k$ и вычислены исправленные выборочные дисперсии S_X^{2*} и S_Y^{2*} .

Требуется проверить нулевую гипотезу $H_0: \sigma_X^2 = \sigma_Y^2$. В данном случае используется статистика

$$T = \frac{S_X^{2*}}{S_Y^{2*}}, \quad (9.5.1)$$

которая имеет F -распределение (распределение Фишера) с $\nu_1 = n - 1$ и $\nu_2 = k - 1$ степенями свободы, если $S_X^{2*} > S_Y^{2*}$, и

$$T = \frac{S_X^{2*}}{S_Y^{2*}}, \quad (9.5.2)$$

с числом степеней свободы $\nu_1 = k - 1$ и $\nu_2 = n - 1$, если $S_X^{2*} < S_Y^{2*}$.

Если задаться уровнем значимости α , то можно построить критические области для проверки нулевой гипотезы при двух альтернативных гипотезах:

а) $H_1: \sigma_X^2 > \sigma_Y^2$, если $S_X^{2*} > S_Y^{2*}$, или $H_1: \sigma_X^2 < \sigma_Y^2$, если $S_X^{2*} < S_Y^{2*}$. В этом случае критическая область правосторонняя $(t_{кр}^R; +\infty)$. Граница $t_{кр}^R$ критической области определяется из условия $P(T(\nu_1, \nu_2) > t_{кр}^R) = \alpha$;

б) $H_1: \sigma_X^2 \neq \sigma_Y^2$. В этом случае критическая область двусторонняя. Однако можно использовать только правостороннюю область $(t_{кр}^R; +\infty)$, где граница $t_{кр}^R$ определяется из

условия $P(T(v_1 = n-1, v_2 = k-1) > t_{кр}^R) = \frac{\alpha}{2}$, если $S_X^{2*} > S_Y^{2*}$, и из условия $P(T(v_1 = k-1, v_2 = n-1) > t_{кр}^R) = \frac{\alpha}{2}$, если $S_X^{2*} < S_Y^{2*}$.

Пример 4. При измерении производительности двух агрегатов получены следующие результаты в кг вещества за час работы:

№ замера	1	2	3	4	5
Агрегат А	14,1	10,1	14,7	13,7	14,0
Агрегат В	14	14,5	13,7	12,7	14,1

При уровне значимости $\alpha = 0,7$ проверить гипотезу о равенстве дисперсий.

О Решение. Проверим нулевую гипотезу $H_0 : \sigma_X^2 = \sigma_Y^2$ при альтернативной гипотезе $H_1 : \sigma_X^2 \neq \sigma_Y^2$. Вычислим «исправленные» выборочные дисперсии S_X^{2*} и S_Y^{2*} . Для этого сначала найдем выборочные средние m_X^* и m_Y^* :

$$m_X^* = \frac{14,1 + 10,1 + 14,7 + 13,7 + 14}{5} = 13,32; \quad m_Y^* = \frac{14 + 14,5 + 13,7 + 12,7 + 14,1}{5} = 13,8.$$

Тогда

$$S_X^{2*} = \frac{1}{4} \left((14,1 - 13,32)^2 + (10,1 - 13,32)^2 + (14,7 - 13,32)^2 + (13,7 - 13,32)^2 + (14 - 13,32)^2 \right) = 3,372;$$

$$S_Y^{2*} = \frac{1}{4} \left((14 - 13,8)^2 + (14,5 - 13,8)^2 + (13,7 - 13,8)^2 + (12,7 - 13,8)^2 + (14,1 - 13,8)^2 \right) = 0,46.$$

Учитывая, что $S_X^{2*} > S_Y^{2*}$, определяет $t_{расч}$:

$$t_{расч} = \frac{S_X^{2*}}{S_Y^{2*}} = \frac{3,372}{0,46} = 7,3304.$$

Критическое значение $t_{кр}^R$ находим из условия

$$P(T(v_1 = 5-1, v_2 = 5-1) > t_{кр}^R) = \frac{\alpha}{2} = 0,05$$

По таблице F -распределения (распределения Фишера) с $v_1 = 4$ и $v_2 = 4$ степенями свободы определяем $t_{кр}^R = 6,39$.

Так как число $t_{расч} = 7,3304$ попадает в критическую область $(6,39; +\infty)$, то гипотезу о равенстве дисперсий отвергаем. ●

Замечание. Границу критической области было можно определить и не используя таблицы, например:

- используя функцию $F_{РАСПОБ}$ (вероятность; степени_свободы1; степени_свободы2) из EXCEL. При этом задаваемый уровень значимости используется как аргумент «вероятность». В рассматриваемом примере получаем $t_{кр}^R = F_{РАСПОБ}(0,04; 4; 4) = 6,39$;
- используя функцию $qF(P, d1, d2)$ из MATHCAD, где P — доверительная вероятность $P = 1 - \alpha$, $d1$ и $d2$ степени свободы. В рассматриваемом примере получаем $t_{кр}^R = qF(0,95, 4, 4) = 6,39$.

§6. Проверка гипотезы о распределении. Критерий Пирсона

Пусть $X_1, \dots, X_n$ — выборка, произведенная из генеральной совокупности X , с известной функцией распределения $F(x)$. Проверяется нулевая гипотеза H_0 , утверждающая, что генеральная совокупность распределена по закону, имеющему функцию распределения $F(x)$, равную функции $F_0(x)$, т.е. проверяется нулевая гипотеза $H_0 : F(x) = F_0(x)$. При этом альтернативной гипотезой является $H_0 : F(x) \neq F_0(x)$.

Наибольшее применение при проверке согласования закона распределения, т.е. проверке нулевой гипотезы $H_0 : F(x) = F_0(x)$, является критерий Пирсона или критерий χ^2 (хи-квадрат).

Наблюдаемое значение критерия (статистика) вычисляется по следующей формуле:

$$\chi^2 = \sum_{i=1}^L \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^L \frac{n_i^2}{np_i} - n. \quad (9.6.1)$$

Согласно теореме К. Пирсона, при $n \rightarrow \infty$ статистика (9.6.1) имеет χ^2 -распределение с $\nu = L - r - 1$ степенями свободы, где L — число групп (интервалов) выборки, r — число параметров предполагаемого распределения.

Схема проверки нулевой гипотезы $H_0 : F(x) = F_0(x)$.

1. По выборке $X_1, \dots, X_n$ строят статистический ряд; он может быть как дискретным, так и интервальным. Рассмотрим для определенности интервальный ряд:

$[x_i, x_{i+1})$	$[x_0, x_1)$	$[x_1, x_2)$	...	$[x_{L-2}, x_{L-1})$	$[x_{L-1}, x_L)$
n_i	n_1	n_2	...	n_{L-1}	n_L

2. По данным предыдущих исследований или по предварительным данным делают предположение (выдвигают гипотезу) о модели закона распределения генеральной совокупности X .

Для предположения модели закона распределения генеральной совокупности X можно использовать графическое представление выборки.

3. Используя выборочные данные, строят оценки параметров выбранной модели закона распределения.

4. Определяют теоретические вероятности попадания случайной величины в интервалы $[x_i, x_{i+1})$ по формуле

$$p_i = F(x_{i+1}) - F(x_i), \quad (9.6.1)$$

где $F(x)$ — предполагаемая функция распределения генеральной совокупности X .

4. Определяют расчетное значение критерия К. Пирсона по формуле

$$\chi^2_{\text{расч}} = \sum_{i=1}^L \frac{(n_i - np_i)^2}{np_i}$$

или

$$\chi^2_{\text{расч}} = \sum_{i=1}^L \frac{n_i^2}{np_i} - n.$$

5. Выбрав уровень значимости α , находят критическую область (она всегда правосторонняя) $(t_{\text{кр}}^R; \infty)$. Границу критической области можно найти одним из следующих:

- используя таблицы χ^2 -распределения с $\nu = L - r - 1$ степенями свободы;

- используя функцию *ХИ2ОБР* (*вероятность; степени_свободы*) из EXCEL, при этом аргумент «вероятность» данной функции должен быть равен уровню значимости, а аргумент «степени_свободы» должен быть равен $\nu = L - r - 1$;
- используя функцию *qchisq(p, d)* из MATHCAD, при этом переменные данной функции должны быть равны: $p = 1 - \alpha$, $d = L - r - 1$.

6. Если расчетное значение $\chi^2_{расч}$ попадает в критическую область $(t_{кр}^R, \infty)$, то нулевая гипотеза $H_0 : F(x) = F_0(x)$ отвергается и принимается альтернативная гипотеза.

Замечание. При применении критерия К. Пирсона в каждом интервале должно быть не менее 5 элементов выборки (т.е. $n_i \geq 5$). Если это условие не выполняется, то число интервалов надо уменьшить путем объединения соседних интервалов.

Пример 5. Получены значения случайной величины X .

0,54	0,7	0,66	0,55	0,57	0,56	0,62	0,65	0,5	0,67
0,66	0,58	0,57	0,53	0,53	0,53	0,46	0,72	0,41	0,55
0,6	0,61	0,45	0,55	0,63	0,6	0,53	0,51	0,59	0,57
0,59	0,64	0,66	0,56	0,59	0,58	0,55	0,55	0,53	0,53
0,41	0,54	0,45	0,61	0,46	0,52	0,59	0,63	0,4	0,69
0,53	0,48	0,41	0,57	0,6	0,57	0,56	0,56	0,42	0,6
0,54	0,59	0,56	0,5	0,46	0,62	0,57	0,42	0,72	0,47
0,47	0,44	0,52	0,62	0,64	0,41	0,51	0,4	0,49	0,67
0,52	0,58	0,65	0,42	0,68	0,53	0,65	0,59	0,43	0,61
0,56	0,62	0,7	0,49	0,7	0,55	0,5	0,52	0,71	0,48

Необходимо:

1. Найти выборочные характеристики: выборочное среднее и исправленную выборочную дисперсию.
2. Построить доверительной вероятностью 0,95 доверительные интервалы для математического ожидания и дисперсии.
3. Построить гистограмму.
4. Проверить гипотезу о нормальном распределении случайной величины X при уровне значимости $\alpha = 0,05$.

О Решение. Учитывая, что количество значений равно 100, определяем выборочное среднее и исправленную выборочную дисперсию:

$$m^* = \frac{1}{100} (0,54 + 0,7 + 0,66 + 0,55 + \dots + 0,71 + 0,48) = \frac{55,69}{100} = 0,5569.$$

$$S^{2*} = \frac{1}{100-1} ((0,54 - 0,5569)^2 + (0,7 - 0,5569)^2 + \dots + (0,48 - 0,5569)^2) = \frac{0,64839}{99} = 0,00655.$$

Замечание. Выборочное среднее можно определить используя функцию *СРЗНАЧ(число1; число2; ...)* из EXCEL, а исправленную выборочную дисперсию можно определить используя функцию *ДИСП(число1; число2; ...)* из EXCEL.

Перейдем к построению доверительных интервалов. При построении доверительного интервала для математического ожидания считаем, что при этом дисперсия не известна. Как известно из предыдущей главы, необходимо использовать формулу

$$m^* - \frac{S^*}{\sqrt{n-1}} t_{\frac{\gamma+1}{2}} < m < m^* + \frac{S^*}{\sqrt{n-1}} t_{\frac{\gamma+1}{2}},$$

в которой неизвестна только величина $t_{\frac{\gamma+1}{2}}$, являющаяся квантилем t -распределения с $\nu = n - 1$ числом степеней свободы. Для доверительной вероятности $\gamma = 0,95$ найдем квантиль $t_{\frac{\gamma+1}{2}} = t_{\frac{0,95+1}{2}} = t_{0,975}$ t -распределения с $\nu = n - 1 = 100 - 1 = 99$ числом степеней свободы. Для этого используем, например, функцию $qt(p, d)$ из MATHCAD: $qt(0,975,99) = 1,984$ или функцию *СТЮДРАСПОБ(вероятность; степени_свободы)* из EXCEL: $СТЮДРАСПОБ(2 \cdot (1 - 0,975); 99) = 1,984$ или Приложение 5. Определяем точность оценки

$$\frac{S^*}{\sqrt{n-1}} t_{\frac{\gamma+1}{2}} = \frac{\sqrt{S^{2*}}}{\sqrt{100-1}} t_{0,975} = \frac{\sqrt{0,00655}}{\sqrt{99}} 1,984 = 0,0161.$$

Таким образом, получаем доверительный интервал $0,5569 - 0,0161 < m < 0,5569 + 0,0161 \Rightarrow 0,5408 < m < 0,5730$.

Построим доверительный интервал для дисперсии при неизвестном математическом ожидании. В этом случае необходимо использовать формулу

$$\frac{(n-1)S^{2*}}{r_{\frac{1+\gamma}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^{2*}}{r_{\frac{1-\gamma}{2}}},$$

где r_ν — квантиль уровня ν χ^2 -распределения с $n - 1$ степенью свободы.

Для данной формулы необходимо вычислить квантили $r_{\frac{\gamma+1}{2}} = r_{\frac{0,95+1}{2}} = r_{0,975}$ и

$r_{\frac{1-\gamma}{2}} = r_{\frac{1-0,95}{2}} = r_{0,025}$. Используя функцию $qchisq(p, d)$ из MATHCAD, получим:

$r_{0,975} = qchisq(0,975,99) = 128,422$ и $r_{0,025} = qchisq(0,025,99) = 73,361$. Аналогичный результат можно получить используя функцию *ХИ2ОБР(вероятность; степени_свободы)* из EXCEL: $r_{0,975} = ХИ2ОБР(1 - 0,975, 99) = 128,422$, $r_{0,025} = ХИ2ОБР(1 - 0,025, 99) = 73,361$.

Таким образом, получаем доверительный интервал

$$\frac{(n-1)S^{2*}}{r_{0,975}} < \sigma^2 < \frac{(n-1)S^{2*}}{r_{0,025}},$$

$$\frac{99 \cdot 0,00655}{128,422} < \sigma^2 < \frac{99 \cdot 0,00655}{73,361} \Rightarrow 0,0051 < \sigma^2 < 0,0088.$$

Перейдем к построению гистограммы. Построим интервальный ряд. Найдем сначала минимальное и максимальное значения случайной величины (т.е. крайние члены вариационного ряда): $X_{min} = 0,4$ и $X_{max} = 0,72$. Для нахождения минимального и максимального значений случайной величины можно использовать функции *МИН(число1; число2; ...)* и *МАКС(число1; число2; ...)* из EXCEL. Размах варьирования будет равен $R = X_n^* - X_1^* = 0,72 - 0,4 = 0,32$.

Возьмем число частичных интервалов $l = 8$. В этом случае длина частичного интервала равна

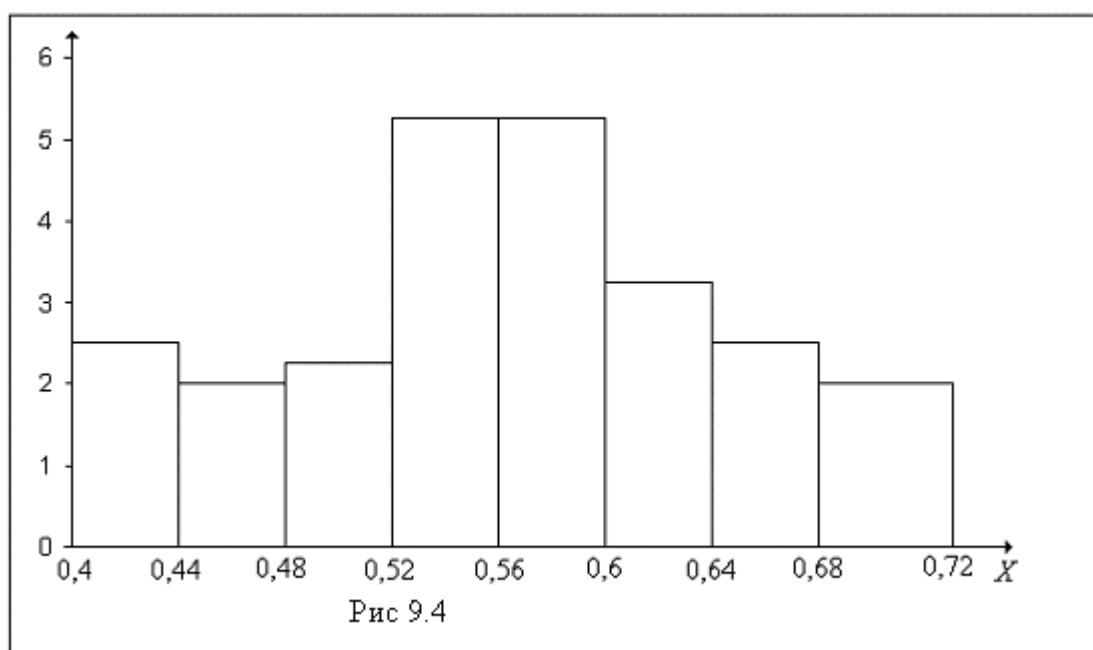
$$h = \frac{R}{n} = \frac{X_n^* - X_1^*}{n} = \frac{0,32}{8} = 0,04.$$

Соответствующий интервальный ряд приведен в таблице 9.2.

Таблица 9.2

Номер интервала i	Средний пробег автомобилей (интервалы) $x_i < X \leq x_{i+1}$	Частота n_i	Относительная частота $f_i^* = \frac{n_i}{nh}$
1	0,4 — 0,44	10	2,5
2	0,44 — 0,48	8	2
3	0,48 — 0,52	9	2,25
4	0,52 — 0,56	21	5,25
5	0,56 — 0,6	21	5,25
6	0,6 — 0,64	13	3,25
7	0,64 — 0,68	10	2,5
8	0,68 — 0,72	8	2

Гистограмма приведена на рис 9.4.



Проверим гипотезу о распределении случайной величины X . Найдем теоретические вероятности попадания случайной величины в интервалы $[x_i, x_{i+1})$ по формуле $p_i = F(x_{i+1}) - F(x_i)$, где $F(x)$ — функция распределения нормального закона. Так как случайная величина X , подчиненная нормальному закону распределения, определена на $(-\infty; \infty)$, то крайние интервалы в ряде распределения следует заменить на $(-\infty; 0,4)$ и $(0,72; \infty)$. Тогда

$$\begin{aligned}
 p_1 &= P(-\infty < X < 0,4) = \Phi\left(\frac{0,44 - m^*}{\sqrt{S^{2*}}}\right) - \Phi(-\infty) = \Phi\left(\frac{0,44 - 0,5569}{\sqrt{0,00655}}\right) - 0 = \\
 &= \Phi(-1,44) = 1 - \Phi(1,44) = 1 - 0,9251 = 0,0749.
 \end{aligned}$$

$$\begin{aligned}
p_2 &= P(0,4 < X < 0,44) = \Phi\left(\frac{0,48 - m^*}{\sqrt{S^{2*}}}\right) - \Phi\left(\frac{0,44 - m^*}{\sqrt{S^{2*}}}\right) = \Phi\left(\frac{0,48 - 0,5569}{\sqrt{0,00655}}\right) - \\
&- \Phi\left(\frac{0,44 - 0,5569}{\sqrt{0,00655}}\right) = \Phi(-0,95) - \Phi(-1,44) = 1 - \Phi(0,95) - (1 - \Phi(1,44)) = \\
&= 1 - 0,8289 - (1 - 0,9251) = 0,9251 - 0,8289 = 0,0962. \\
p_3 &= P(0,48 < X < 0,52) = \Phi\left(\frac{0,52 - m^*}{\sqrt{S^{2*}}}\right) - \Phi\left(\frac{0,48 - m^*}{\sqrt{S^{2*}}}\right) = \Phi\left(\frac{0,52 - 0,5569}{\sqrt{0,00655}}\right) - \\
&- \Phi\left(\frac{0,48 - 0,5569}{\sqrt{0,00655}}\right) = \Phi(-0,46) - \Phi(-0,95) = 1 - \Phi(0,46) - (1 - \Phi(0,95)) = \\
&= 1 - 0,6772 - (1 - 0,8289) = 0,8289 - 0,6772 = 0,1517. \\
p_4 &= P(0,52 < X < 0,56) = \Phi\left(\frac{0,56 - m^*}{\sqrt{S^{2*}}}\right) - \Phi\left(\frac{0,52 - m^*}{\sqrt{S^{2*}}}\right) = \Phi\left(\frac{0,56 - 0,5569}{\sqrt{0,00655}}\right) - \\
&- \Phi\left(\frac{0,52 - 0,5569}{\sqrt{0,00655}}\right) = \Phi(0,04) - \Phi(-0,46) = \Phi(0,04) - (1 - \Phi(0,46)) = \\
&= 0,516 - (1 - 0,6772) = 0,516 + 0,6772 - 1 = 0,1932. \\
p_5 &= P(0,56 < X < 0,6) = \Phi\left(\frac{0,6 - m^*}{\sqrt{S^{2*}}}\right) - \Phi\left(\frac{0,56 - m^*}{\sqrt{S^{2*}}}\right) = \Phi\left(\frac{0,6 - 0,5569}{\sqrt{0,00655}}\right) - \\
&- \Phi\left(\frac{0,56 - 0,5569}{\sqrt{0,00655}}\right) = \Phi(0,53) - \Phi(0,04) = 0,7019 - 0,516 = 0,1859. \\
p_6 &= P(0,6 < X < 0,64) = \Phi\left(\frac{0,64 - m^*}{\sqrt{S^{2*}}}\right) - \Phi\left(\frac{0,6 - m^*}{\sqrt{S^{2*}}}\right) = \Phi\left(\frac{0,64 - 0,5569}{\sqrt{0,00655}}\right) - \\
&- \Phi\left(\frac{0,6 - 0,5569}{\sqrt{0,00655}}\right) = \Phi(1,03) - \Phi(0,53) = 0,8485 - 0,7019 = 0,1466. \\
p_7 &= P(0,64 < X < 0,68) = \Phi\left(\frac{0,68 - m^*}{\sqrt{S^{2*}}}\right) - \Phi\left(\frac{0,64 - m^*}{\sqrt{S^{2*}}}\right) = \Phi\left(\frac{0,68 - 0,5569}{\sqrt{0,00655}}\right) - \\
&- \Phi\left(\frac{0,64 - 0,5569}{\sqrt{0,00655}}\right) = \Phi(1,52) - \Phi(1,03) = 0,9357 - 0,8485 = 0,0872. \\
p_8 &= P(0,68 < X < \infty) = \Phi(\infty) - \Phi\left(\frac{0,68 - m^*}{\sqrt{S^{2*}}}\right) = 1 - \Phi\left(\frac{0,68 - 0,5569}{\sqrt{0,00655}}\right) = \\
&= 1 - \Phi(1,52) = 1 - 0,9357 = 0,0643.
\end{aligned}$$

Дальнейшие вычисления удобно оформить в виде таблицы (табл. 9.3).

Таблица 9.3

Номер интервала i	средний пробег ав- томобилей (интервалы) $x_i < X \leq x_{i+1}$	Частота n_i	Теоретиче- ские частоты p_i	np_i	$\frac{n_i^2}{np_i}$
1	0,4 — 0,44	10	0,0749	7,49	13,3511
2	0,44 — 0,48	8	0,0962	9,62	6,6528
3	0,48 — 0,52	9	0,1517	15,17	5,3395
4	0,52 — 0,56	21	0,1932	19,32	22,8261
5	0,56 — 0,6	21	0,1859	18,59	23,7224
6	0,6 — 0,64	13	0,1466	14,66	11,5280
7	0,64 — 0,68	10	0,0872	8,72	11,4679
8	0,68 — 0,72	8	0,0643	6,43	9,9533
Итого		100	1	100	104,8411

Определяем расчетное значение критерия К. Пирсона:

$$\chi^2_{расч} = \sum_{i=1}^8 \frac{n_i^2}{np_i} - n = 13,3511 + 6,6528 + 5,3395 + 22,8261 + 23,7224 + 11,528 + \\ + 11,4679 + 9,9533 - 100 = 104,8411 - 100 = 4,8411.$$

Находим число степеней свободы. По выборке были рассчитаны два параметра, значит, $r = 2$. Количество интервалов 8, т.е. $L = 8$. Следовательно, $\nu = 8 - 2 - 1 = 5$. Зная, что $\alpha = 0,05$ и $\nu = 5$, находим границу правосторонней критической области $t_{кр}^R = 11,07$ (см. Приложение 4). Таким образом, критической областью является интервал $(11,07; \infty)$.

Так как расчетное значение критерия К. Пирсона $\chi^2_{расч} = 4,8411$ не попадает в критическую область, то нет оснований отвергнуть проверяемую гипотезу о нормальном законе распределения. ●

Приложения

Приложение 1

Таблица значений функции $p_n(m) = \frac{\lambda^m}{m!} e^{-\lambda}$

m	Значения λ						
	0,1	0,2	0,3	0,4	0,5	0,6	0,7
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812	0,496585
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287	0,347610
2	0,004524	0,016375	0,033337	0,053626	0,075816	0,098786	0,121663
3	0,000151	0,001092	0,003334	0,007150	0,012636	0,019757	0,028388
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964	0,004968
5		0,000002	0,000015	0,000057	0,000158	0,000356	0,000696
6			0,000001	0,000004	0,000013	0,000036	0,000081
7					0,000001	0,000003	0,000008
8							0,000001

m	Значения λ						
	0,8	0,9	1	2	3	4	5
0	0,449329	0,406570	0,367879	0,135335	0,049787	0,018316	0,006738
1	0,359463	0,365913	0,367879	0,270671	0,149361	0,073263	0,033690
2	0,143785	0,164661	0,183940	0,270671	0,224042	0,146525	0,084224
3	0,038343	0,049398	0,061313	0,180447	0,224042	0,195367	0,140374
4	0,007669	0,011115	0,015328	0,090224	0,168031	0,195367	0,175467
5	0,001227	0,002001	0,003066	0,036089	0,100819	0,156293	0,175467
6	0,000164	0,000300	0,000511	0,012030	0,050409	0,104196	0,146223
7	0,000019	0,000039	0,000073	0,003437	0,021604	0,059540	0,104445
8	0,000002	0,000004	0,000009	0,000859	0,008102	0,029770	0,065278
9			0,000001	0,000191	0,002701	0,013231	0,036266
10				0,000038	0,000810	0,005292	0,018133
11				0,000007	0,000221	0,001925	0,008242
12				0,000001	0,000055	0,000642	0,003434
13					0,000013	0,000197	0,001321
14					0,000003	0,000056	0,000472
15					0,000001	0,000015	0,000157
16						0,000004	0,000049
17						0,000001	0,000014
18							0,000004
19							0,000001

Плотность стандартного нормального распределения

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

t	0	1	2	3	4	5	6	7	8	9
0,0	0,39894	0,39892	0,39886	0,39876	0,39862	0,39844	0,39822	0,39797	0,39767	0,39733
0,1	0,39695	0,39654	0,39608	0,39559	0,39505	0,39448	0,39387	0,39322	0,39253	0,39181
0,2	0,39104	0,39024	0,38940	0,38853	0,38762	0,38667	0,38568	0,38466	0,38361	0,38251
0,3	0,38139	0,38023	0,37903	0,37780	0,37654	0,37524	0,37391	0,37255	0,37115	0,36973
0,4	0,36827	0,36678	0,36526	0,36371	0,36213	0,36053	0,35889	0,35723	0,35553	0,35381
0,5	0,35207	0,35029	0,34849	0,34667	0,34482	0,34294	0,34105	0,33912	0,33718	0,33521
0,6	0,33322	0,33121	0,32918	0,32713	0,32506	0,32297	0,32086	0,31874	0,31659	0,31443
0,7	0,31225	0,31006	0,30785	0,30563	0,30339	0,30114	0,29887	0,29659	0,29431	0,29200
0,8	0,28969	0,28737	0,28504	0,28269	0,28034	0,27798	0,27562	0,27324	0,27086	0,26848
0,9	0,26609	0,26369	0,26129	0,25888	0,25647	0,25406	0,25164	0,24923	0,24681	0,24439
1,0	0,24197	0,23955	0,23713	0,23471	0,23230	0,22988	0,22747	0,22506	0,22265	0,22025
1,1	0,21785	0,21546	0,21307	0,21069	0,20831	0,20594	0,20357	0,20121	0,19886	0,19652
1,2	0,19419	0,19186	0,18954	0,18724	0,18494	0,18265	0,18037	0,17810	0,17585	0,17360
1,3	0,17137	0,16915	0,16694	0,16474	0,16256	0,16038	0,15822	0,15608	0,15395	0,15183
1,4	0,14973	0,14764	0,14556	0,14350	0,14146	0,13943	0,13742	0,13542	0,13344	0,13147
1,5	0,12952	0,12758	0,12566	0,12376	0,12188	0,12001	0,11816	0,11632	0,11450	0,11270
1,6	0,11092	0,10915	0,10741	0,10567	0,10396	0,10226	0,10059	0,09893	0,09728	0,09566
1,7	0,09405	0,09246	0,09089	0,08933	0,08780	0,08628	0,08478	0,08329	0,08183	0,08038
1,8	0,07895	0,07754	0,07614	0,07477	0,07341	0,07206	0,07074	0,06943	0,06814	0,06687
1,9	0,06562	0,06438	0,06316	0,06195	0,06077	0,05959	0,05844	0,05730	0,05618	0,05508
2,0	0,05399	0,05292	0,05186	0,05082	0,04980	0,04879	0,04780	0,04682	0,04586	0,04491
2,1	0,04398	0,04307	0,04217	0,04128	0,04041	0,03955	0,03871	0,03788	0,03706	0,03626
2,2	0,03547	0,03470	0,03394	0,03319	0,03246	0,03174	0,03103	0,03034	0,02965	0,02898
2,3	0,02833	0,02768	0,02705	0,02643	0,02582	0,02522	0,02463	0,02406	0,02349	0,02294
2,4	0,02239	0,02186	0,02134	0,02083	0,02033	0,01984	0,01936	0,01888	0,01842	0,01797
2,5	0,01753	0,01709	0,01667	0,01625	0,01585	0,01545	0,01506	0,01468	0,01431	0,01394
2,6	0,01358	0,01323	0,01289	0,01256	0,01223	0,01191	0,01160	0,01130	0,01100	0,01071
2,7	0,01042	0,01014	0,00987	0,00961	0,00935	0,00909	0,00885	0,00861	0,00837	0,00814
2,8	0,00792	0,00770	0,00748	0,00727	0,00707	0,00687	0,00668	0,00649	0,00631	0,00613
2,9	0,00595	0,00578	0,00562	0,00545	0,00530	0,00514	0,00499	0,00485	0,00470	0,00457

Функция распределения стандартного нормального закона

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (\Phi(-x) = 1 - \Phi(x))$$

t	0	1	2	3	4	5	6	7	8	9
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91308	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574
2,2	0,98610	0,98645	0,98679	0,98713	0,98745	0,98778	0,98809	0,98840	0,98870	0,98899
2,3	0,98928	0,98956	0,98983	0,99010	0,99036	0,99061	0,99086	0,99111	0,99134	0,99158
2,4	0,99180	0,99202	0,99224	0,99245	0,99266	0,99286	0,99305	0,99324	0,99343	0,99361
2,5	0,99379	0,99396	0,99413	0,99430	0,99446	0,99461	0,99477	0,99492	0,99506	0,99520
2,6	0,99534	0,99547	0,99560	0,99573	0,99585	0,99598	0,99609	0,99621	0,99632	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,99702	0,99711	0,99720	0,99728	0,99736
2,8	0,99744	0,99752	0,99760	0,99767	0,99774	0,99781	0,99788	0,99795	0,99801	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,99841	0,99846	0,99851	0,99856	0,99861

Критические точки распределения χ^2

Число степеней свободы	Уровень значимости α									
	0,01	0,025	0,05	0,075	0,1	0,9	0,925	0,95	0,975	0,99
1	6,63	5,02	3,84	3,17	2,71	0,02	0,01	0,00	0,00	0,00
2	9,21	7,38	5,99	5,18	4,61	0,21	0,16	0,10	0,05	0,02
3	11,34	9,35	7,81	6,90	6,25	0,58	0,47	0,35	0,22	0,11
4	13,28	11,14	9,49	8,50	7,78	1,06	0,90	0,71	0,48	0,30
5	15,09	12,83	11,07	10,01	9,24	1,61	1,39	1,15	0,83	0,55
6	16,81	14,45	12,59	11,47	10,64	2,20	1,94	1,64	1,24	0,87
7	18,48	16,01	14,07	12,88	12,02	2,83	2,53	2,17	1,69	1,24
8	20,09	17,53	15,51	14,27	13,36	3,49	3,14	2,73	2,18	1,65
9	21,67	19,02	16,92	15,63	14,68	4,17	3,78	3,33	2,70	2,09
10	23,21	20,48	18,31	16,97	15,99	4,87	4,45	3,94	3,25	2,56
11	24,73	21,92	19,68	18,29	17,28	5,58	5,12	4,57	3,82	3,05
12	26,22	23,34	21,03	19,60	18,55	6,30	5,82	5,23	4,40	3,57
13	27,69	24,74	22,36	20,90	19,81	7,04	6,52	5,89	5,01	4,11
14	29,14	26,12	23,68	22,18	21,06	7,79	7,24	6,57	5,63	4,66
15	30,58	27,49	25,00	23,45	22,31	8,55	7,97	7,26	6,26	5,23
16	32,00	28,85	26,30	24,72	23,54	9,31	8,71	7,96	6,91	5,81
17	33,41	30,19	27,59	25,97	24,77	10,09	9,45	8,67	7,56	6,41
18	34,81	31,53	28,87	27,22	25,99	10,86	10,21	9,39	8,23	7,01
19	36,19	32,85	30,14	28,46	27,20	11,65	10,97	10,12	8,91	7,63
20	37,57	34,17	31,41	29,69	28,41	12,44	11,73	10,85	9,59	8,26
21	38,93	35,48	32,67	30,92	29,62	13,24	12,50	11,59	10,28	8,90
22	40,29	36,78	33,92	32,14	30,81	14,04	13,28	12,34	10,98	9,54
23	41,64	38,08	35,17	33,36	32,01	14,85	14,06	13,09	11,69	10,20
24	42,98	39,36	36,42	34,57	33,20	15,66	14,85	13,85	12,40	10,86
25	44,31	40,65	37,65	35,78	34,38	16,47	15,64	14,61	13,12	11,52
26	45,64	41,92	38,89	36,98	35,56	17,29	16,44	15,38	13,84	12,20
27	46,96	43,19	40,11	38,18	36,74	18,11	17,24	16,15	14,57	12,88
28	48,28	44,46	41,34	39,38	37,92	18,94	18,05	16,93	15,31	13,56
29	49,59	45,72	42,56	40,57	39,09	19,77	18,85	17,71	16,05	14,26
30	50,89	46,98	43,77	41,76	40,26	20,60	19,66	18,49	16,79	14,95
31	52,19	48,23	44,99	42,95	41,42	21,43	20,48	19,28	17,54	15,66
32	53,49	49,48	46,19	44,13	42,58	22,27	21,30	20,07	18,29	16,36
33	54,78	50,73	47,40	45,31	43,75	23,11	22,12	20,87	19,05	17,07
34	56,06	51,97	48,60	46,49	44,90	23,95	22,94	21,66	19,81	17,79
35	57,34	53,20	49,80	47,66	46,06	24,80	23,76	22,47	20,57	18,51
36	58,62	54,44	51,00	48,84	47,21	25,64	24,59	23,27	21,34	19,23
37	59,89	55,67	52,19	50,01	48,36	26,49	25,42	24,07	22,11	19,96
38	61,16	56,90	53,38	51,17	49,51	27,34	26,25	24,88	22,88	20,69
39	62,43	58,12	54,57	52,34	50,66	28,20	27,09	25,70	23,65	21,43
40	63,69	59,34	55,76	53,50	51,81	29,05	27,93	26,51	24,43	22,16
99	134,64	128,42	123,23	119,91	117,41	81,45	79,51	77,05	73,36	69,23
100	135,81	129,56	124,34	121,02	118,50	82,36	80,41	77,93	74,22	70,06

Критические точки распределения Стьюдента

Число степеней свободы	Уровень значимости α (двусторонняя критическая область)									
	0,01	0,025	0,05	0,075	0,1	0,2	0,25	0,3	0,4	0,5
1	63,66	25,45	12,71	8,45	6,31	3,08	2,41	1,96	1,38	1,00
2	9,92	6,21	4,30	3,44	2,92	1,89	1,60	1,39	1,83	0,82
3	5,84	4,18	3,18	2,68	2,35	1,64	1,42	1,25	2,95	0,76
4	4,60	3,50	2,78	2,39	2,13	1,53	1,34	1,19	4,04	0,74
5	4,03	3,16	2,57	2,24	2,02	1,48	1,30	1,16	5,13	0,73
6	3,71	2,97	2,45	2,15	1,94	1,44	1,27	1,13	6,21	0,72
7	3,50	2,84	2,36	2,09	1,89	1,41	1,25	1,12	7,28	0,71
8	3,36	2,75	2,31	2,05	1,86	1,40	1,24	1,11	8,35	0,71
9	3,25	2,69	2,26	2,01	1,83	1,38	1,23	1,10	9,41	0,70
10	3,17	2,63	2,23	1,99	1,81	1,37	1,22	1,09	10,47	0,70
11	3,11	2,59	2,20	1,97	1,80	1,36	1,21	1,09	11,53	0,70
12	3,05	2,56	2,18	1,95	1,78	1,36	1,21	1,08	12,58	0,70
13	3,01	2,53	2,16	1,94	1,77	1,35	1,20	1,08	13,64	0,69
14	2,98	2,51	2,14	1,92	1,76	1,35	1,20	1,08	14,69	0,69
15	2,95	2,49	2,13	1,91	1,75	1,34	1,20	1,07	15,73	0,69
16	2,92	2,47	2,12	1,90	1,75	1,34	1,19	1,07	16,78	0,69
17	2,90	2,46	2,11	1,90	1,74	1,33	1,19	1,07	17,82	0,69
18	2,88	2,45	2,10	1,89	1,73	1,33	1,19	1,07	18,87	0,69
19	2,86	2,43	2,09	1,88	1,73	1,33	1,19	1,07	19,91	0,69
20	2,85	2,42	2,09	1,88	1,72	1,33	1,18	1,06	20,95	0,69
21	2,83	2,41	2,08	1,87	1,72	1,32	1,18	1,06	21,99	0,69
22	2,82	2,41	2,07	1,87	1,72	1,32	1,18	1,06	23,03	0,69
23	2,81	2,40	2,07	1,86	1,71	1,32	1,18	1,06	24,07	0,69
24	2,80	2,39	2,06	1,86	1,71	1,32	1,18	1,06	25,11	0,68
25	2,79	2,38	2,06	1,86	1,71	1,32	1,18	1,06	26,14	0,68
26	2,78	2,38	2,06	1,85	1,71	1,31	1,18	1,06	27,18	0,68
27	2,77	2,37	2,05	1,85	1,70	1,31	1,18	1,06	28,21	0,68
28	2,76	2,37	2,05	1,85	1,70	1,31	1,17	1,06	29,25	0,68
29	2,76	2,36	2,05	1,85	1,70	1,31	1,17	1,06	30,28	0,68
30	2,75	2,36	2,04	1,84	1,70	1,31	1,17	1,05	31,32	0,68
40	2,70	2,33	2,02	1,83	1,68	1,30	1,17	1,05	41,62	0,68
50	2,68	2,31	2,01	1,82	1,68	1,30	1,16	1,05	51,89	0,68
60	2,66	2,30	2,00	1,81	1,67	1,30	1,16	1,05	62,13	0,68
70	2,65	2,29	1,99	1,81	1,67	1,29	1,16	1,04	72,36	0,68
80	2,64	2,28	1,99	1,80	1,66	1,29	1,16	1,04	82,57	0,68
90	2,63	2,28	1,99	1,80	1,66	1,29	1,16	1,04	92,76	0,68
100	2,63	2,28	1,98	1,80	1,66	1,29	1,16	1,04	102,95	0,68

Используемая литература

- Гмурман В.Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 1998.
- Гнеденко Б.В. Курс теории вероятностей. — М.: Наука, 1967.
- Красс М.С., Чупрынов Б.П. Основы математики и ее приложения в экономическом образовании. — М.: Дело, 2002.
- Кузнецов Б.Т. Математика. — М.: ЮНИТИ-ДАНА, 2004.
- Черняк А.А., Новиков В.А., Мельников О.И., Кузнецов А.В. Математика для экономистов на базе Mathcad. — СПб.: БХВ-Петербург, 2003

Оглавление

Введение.....	3
Глава 1. Случайные события и их вероятности.....	4
§1. События. Действия с событиями	4
§2. Общее определение и свойства вероятности.....	6
ГЛАВА 2. Классическая и геометрическая вероятности.....	7
§1. Классическое определение вероятности	7
§2. Применение комбинаторного анализа.....	8
§3. Геометрическое определение вероятности	10
Глава 3. Условная вероятность. Независимость событий. Формулы полной вероятности и Байеса.....	13
§1. Условная вероятность	13
§2. Теоремы сложения и умножения вероятностей	14
§3. Независимость событий.....	17
§4. Формула полной вероятности	18
§5. Формула Байеса	20
Глава 4. Схема независимых испытаний. Схема Бернулли	23
§1. Формула Бернулли	23
§2. Формула Пуассона.....	25
§3. Формулы Муавра – Лапласа.....	26
Глава 5. Случайные величины и их распределения	29
§1. Понятие случайной величины	29
§2. Функция распределения случайной величины	29
§3. Дискретные случайные величины	30
§4. Непрерывные случайные величины	32
§5. Функция от случайных величин	36
Глава 6. Числовые характеристики случайных величин	40
§1. Математическое ожидание случайной величины	40
§2. Математическое ожидание функции от случайной величины. Свойства математического ожидания	42
§3. Дисперсия. Моменты высших порядков	46
Глава 7. Элементы математической статистики	49
§1. Основные понятия и основные задачи математической статистики.....	49
§2. Простейшие статистические преобразования	49
§3. Эмпирическая функция распределения	51
§4. Полигон и гистограмма.....	53
Глава 8. Статистическое оценивание	55
§1. Точечные оценки. Выборочная средняя и выборочная дисперсия	55
§2. Метод моментов	56
§3. Метод максимального правдоподобия.....	57
§4. Интервальные оценки (доверительные интервалы).....	58
Глава 9. Проверка статистических гипотез.....	64
§1. Основные понятия	64
§2. Проверка гипотезы о значении математического ожидания	65
§3. Проверка гипотезы о равенстве математических ожиданий двух генеральных совокупностей.....	66
§4. Проверка гипотезы о значении дисперсии генеральной совокупности.....	68
§5. Проверка гипотезы о равенстве дисперсий двух генеральных совокупностей	69
§6. Проверка гипотезы о распределении. Критерий Пирсона	71
Приложения	77
Используемая литература	82