

**ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ**  
Государственное автономное образовательное учреждение  
высшего профессионального образования  
**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**  
**АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ**

---

**В. Г. Фарафонов, В. И. Устимов, В. Б. Ильин**

**ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И  
МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ**

Учебное пособие

**Часть II**

Санкт-Петербург

2012

УДК 519.2  
ББК 22.171  
Ф24

Рецензенты:

доктор физико-математических наук, профессор А.П.Киселев  
доктор физико-математических наук, профессор В.А.Гаген-Торн

Утверждено

редакционно-издательским советом университета  
в качестве учебного пособия

**Фарафонов В. Г., Устимов В. И., Ильин В. Б.**

Ф24 Основы теории вероятностей и математической статистики: учеб. пособие / В.Г. Фарафонов, В.И.Устимов, В.Б. Ильин. – СПб.: ГУАП, 2012. Ч.2 – 75 с.: ил.  
ISBN

Учебное пособие составлено в соответствии с программой по высшей математике для студентов, обучающихся по направлениям «Приборостроение», «Радиофизика» и «Информатика», в том числе на заочном факультете.

Во второй части пособия рассмотрены разделы курса математической статистики, включающие определение основных понятий, оценивание параметров известного распределения и статистическую проверку различных гипотез. Каждая глава содержит сведения и формулы, проиллюстрированные подробно разобранными примерами. Завершают пособие 10 вариантов контрольных работ, каждый из которых содержит по 5 задач из различных разделов математической статистики. Вопросы теории вероятностей рассмотрены в первой части учебного пособия.

УДК 519.2  
ББК 22.171

ISBN

© В.Г.Фарафонов, В.И.Устимов,  
В.Б.Ильин, 2012  
© ГУАП, 2012

## Часть II. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

### ВВЕДЕНИЕ

Мы живем в мире информации. Информация становится частью действительности и нашего сознания. Без адекватных технологий анализа информации (данных) невозможно ориентироваться в информационной среде.

Математическая статистика позволяет компактно описать данные, понять их структуру, провести классификацию, установить закономерности в хаосе случайных явлений.

Математическая статистика решает две основные задачи: первая – сбор и группировка статистических данных, вторая – разработка методов анализа полученных данных в зависимости от целей исследования.

К методам статистического анализа данных относятся:

- оценка неизвестной вероятности события;
- оценка неизвестной функции распределения;
- оценка параметров известного распределения;
- проверка статистических гипотез о виде неизвестного распределения или о значениях параметров известного распределения.

Методы математической статистики применяются практически везде, где используется понятие информации и требуется ее анализ. В частности, статистический анализ является составной частью *инноватики* – науки, охватывающей широкий круг вопросов от создания новых знаний до трансформации их в новшества и распространение новшеств.

В качестве примера рассмотрим довольно часто встречающуюся задачу. Предположим, что продуктом инновационной деятельности является важное нововведение (инновация): изменение системы оплаты труда, переход на выпуск новой продукции, использование новой технологии, методики. Требуется оценить положительный эффект, убедиться, что положительный эффект обусловлен именно инновацией, а не носит случайный характер.

Для решения этой задачи надо сформировать два набора данных (выборок), каждый из которых содержит значения интересующего показателя эффективности до и после инновации. Статистические критерии сравнения этих двух выборок покажут насколько случаен или закономерен рост показателей эффективности и является ли он результатом инновации.

# 1. ОСНОВНЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

## 1.1. Генеральная совокупность и выборка

Дадим ряд определений, связанных со статистическими исследованиями.

*Генеральной совокупностью* будем называть все множество исследуемых объектов, *выборкой* – набор объектов, случайно отобранных из генеральной совокупности для исследования. *Объем* генеральной совокупности и *объем* выборки, т.е. число объектов в генеральной совокупности и выборке, будем обозначать соответственно как  $N$  и  $n$ .

Выборка бывает *повторной*, когда каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность, и *бесповторной*, если отобранный объект в генеральную совокупность не возвращается.

Для того чтобы по выборке можно было сделать выводы о поведении интересующего нас признака (параметра или характеристики объектов) генеральной совокупности, нужно, чтобы выборка правильно представляла особенности генеральной совокупности, т.е. была *репрезентативной* (представительной). Учитывая закон больших чисел, можно утверждать, что это условие выполняется, если:

- 1) объем выборки  $n$  достаточно большой;
- 2) каждый объект выборки выбран случайно;
- 3) для каждого объекта вероятность попасть в выборку одинакова.

Если, например, нас интересует усвоение основ математики школьниками выпускного класса в одном из федеральных округов, то генеральной совокупностью будет множество всех школьников округа. Школьники, отобранные случайным образом для проверки знаний по математике, образуют выборку.

В данном случае генеральная совокупность и выборка являются *одномерными* (*однофакторными*). Если мы будем интересоваться усвоением школьниками материалов сразу по нескольким предметам, то генеральная совокупность и соответствующая выборка будут *многомерными* (*многофакторными*).

## 1.2. Выборочный закон распределения (статистический ряд)

Пусть в выборке объемом  $n$  интересующая нас случайная величина  $\xi$  (какой-либо параметр объектов генеральной совокупности) принимает  $n_1$  раз значение  $x_1$ ,  $n_2$  раз значение  $x_2$ , ... и  $n_k$  раз значение  $x_k$ . Тогда наблюдаемые значения  $x_1, x_2, \dots, x_k$  случайной величины  $\xi$  называются *вариантами*, а  $n_1, n_2, \dots, n_k$  – их *частотами*. Разность  $x_{max} - x_{min}$  есть *размах* выборки, отношение  $\omega_i = n_i/n$  – *относительная частота* варианты  $x_i$ .

Очевидно, что

$$\sum_{i=1}^k n_i = n, \quad \sum_{i=1}^k \omega_i = 1.$$

Если мы запишем варианты в возрастающем порядке, то получим *вариационный ряд*. Таблица, состоящая из таких упорядоченных вариантов и их частот (и/или относительных частот) называется *статистическим рядом* или *выборочным законом распределения*, например

$x_1$	$x_2$	$x_3$	...	$x_k$
$n_1$	$n_2$	$n_3$	...	$n_k$
$\omega_1$	$\omega_2$	$\omega_3$	...	$\omega_k$

где  $x_i \leq x_{i+1}$  при  $i = 1, 2, \dots, k - 1$ .

Выборочный закон распределения является аналогом закона (ряда) распределения дискретной случайной величины в теории вероятности, который был подробно рассмотрен в первой части нашего пособия [1] (далее сокращенно часть I).

Если вариационный ряд состоит из очень большого количества чисел (или исследуется некоторый непрерывный признак), то используют *группированную* выборку. Для ее получения интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько обычно равных частей (подинтервалов) длиной  $h$ . При составлении статистического ряда в качестве  $x_i$  обычно выбирают середины подинтервалов, а  $n_i$  будет равно числу вариантов, попавших в  $i$ -й подинтервал.

Пусть число подинтервалов равно  $s$ ,  $a = \min\{x_i\}$ ,  $b = \max\{x_i\}$ . Тогда для группированной выборки получим следующий статистический ряд:

<i>Номера подинтервалов</i>	1	2	...	$s$
<i>Границы подинтервалов</i>	$[a, a+h]$	$(a+h, a+2h]$	...	$(b-h, b]$
<i>Варианты</i>	$a + h/2$	$a + 3h/2$	...	$b - h/2$
<i>Частоты</i>	$v_1$	$v_2$	...	$v_s$

где  $h = (b - a)/s$ , а  $v_i$  равна сумме частот вариантов, попавших в  $i$ -й подинтервал.

### 1.3. Полигон частот, выборочная функция распределения

Для лучшего представления о поведении исследуемой случайной величины в рамках данной выборки можно строить различные графики.

Отложим значения случайной величины  $x_i$  по оси абсцисс, а значения  $n_i$  – по оси ординат. Ломаная линия, отрезки которой соединяют точки с координатами  $(x_1, n_1)$ ,  $(x_2, n_2)$ , ...,  $(x_k, n_k)$ , называется *полигон частот*. Если вместо абсолютных значений  $n_i$  на оси ординат отложить относительные частоты  $\omega_i$ , то получим *полигон относительных частот* (см. рис. 1.1).

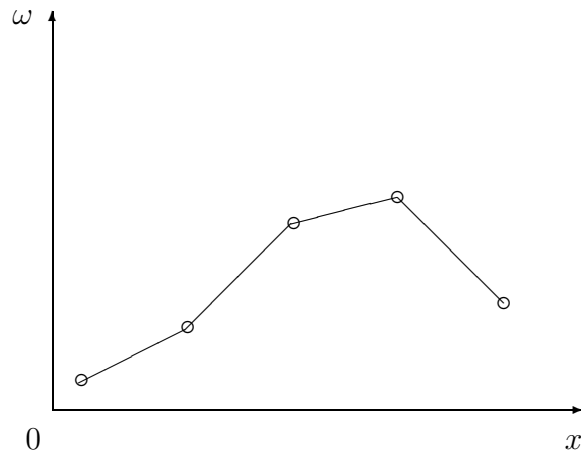


Рис. 1.1. Пример полигона относительных частот ( $k = 5$ )

По аналогии с функцией распределения дискретной случайной величины (см. п. 7.2 части I) по выборочному закону распределения можно построить *выборочную (эмпирическую) функцию распределения*

$$F_n^*(x) = \sum_{x_i < x} \omega_i = \frac{1}{n} \sum_{x_i < x} n_i. \quad (1.1)$$

Суммирование в (1.1) выполняется по всем частотам, которым соответствуют значения вариант, меньшие  $x$ . Заметим, что эмпирическая функция распределения зависит от объема выборки  $n$ .

В отличие от функции  $F_n^*(x)$ , найденной для случайной величины  $\xi$  опытным путем в результате обработки статистических данных, истинную функцию распределения  $F_\xi(x)$ , связанную с генеральной совокупностью, называют *теоретической*. Обычно генеральная совокупность настолько велика, что обработать ее всю невозможно, т.е. исследовать ее можно только теоретически.

Функция  $F_\xi(x)$  по определению есть вероятность события ( $\xi < x$ ):

$$F_\xi(x) = p(\xi < x),$$

а  $F_n^*(x)$  – его относительная частота. При достаточно больших  $n$ , как следует из теоремы Бернулли (см. п. 9.3 части I), относительная частота события ( $\xi < x$ ), т.е.  $F_n^*(x)$ , должна стремиться (по вероятности) к его вероятности, т.е. к  $F_\xi(x)$ :

$$F_n^*(x) \xrightarrow[n \rightarrow \infty]{\text{по вер.}} F_\xi(x).$$

### Замечания:

1. В дальнейшем все эмпирические (найденные по выборке) величины будут иметь индекс \*, чтобы отличать их от соответствующих теоретических (относящихся к генеральной совокупности) величин.
2. В законе больших чисел рассматривается последовательность случайных величин  $\xi_1, \xi_2, \dots, \xi_n$ . Значение первого объекта выборки можно считать значением случайной величины  $\xi_1$ , значение второго – значением  $\xi_2$  и т.д. Тогда при выполнении условий теоремы о том, что величины  $\xi_1, \xi_2,$

...,  $\xi_n$  попарно независимые и число  $n$  достаточно велико, можно показать, что относительная частота события ( $\xi < x$ ) должна быть (по вероятности) сколь угодно близка к его вероятности, и, следовательно,  $F_n^*(x)$  к  $F_\xi(x)$ . Требования случайности и независимости  $\xi_1, \xi_2, \dots, \xi_n$  порождают второе и третье условия репрезентативности выборки (см. п. 1.1).

#### 1.4. Свойства эмпирической функции распределения

Из определения (1.1) функции  $F_n^*(x)$  очевидно, что ее свойства совпадают со свойствами  $F_\xi(x)$ , а именно:

- 1)  $0 \leq F_n^*(x) \leq 1$ ;
- 2)  $F_n^*(x)$  – неубывающая функция, т.е.  $F_n^*(x_2) \geq F_n^*(x_1)$  при  $x_2 > x_1$ ;
- 3) Если  $x_1$  – наименьшая варианта, то  $F_n^*(x) = 0$  при  $x \leq x_1$ ;  
если  $x_k$  – наибольшая варианта, то  $F_n^*(x) = 1$  при  $x > x_k$ .

Соответственно  $F_n^*(-\infty) = 0$  и  $F_n^*(\infty) = 1$ .

График эмпирической функции распределения имеет характерный ступенчатый вид:

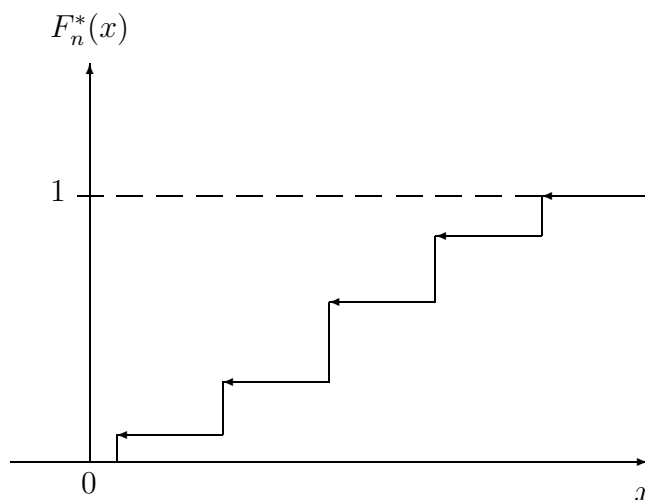


Рис. 1.2. Пример эмпирической функции распределения

Еще одним графическим представлением интересующей нас выборки является *гистограмма* – ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат подынтервалы шириной  $h$ , а высотами – отрезки длиной  $n_i/h$  (гистограмма частот) или  $\omega_i/h$  (гистограмма относительных частот). В первом случае площадь гистограммы равна объему выборки  $n$ , во втором – единице (см. рис. 1.3).

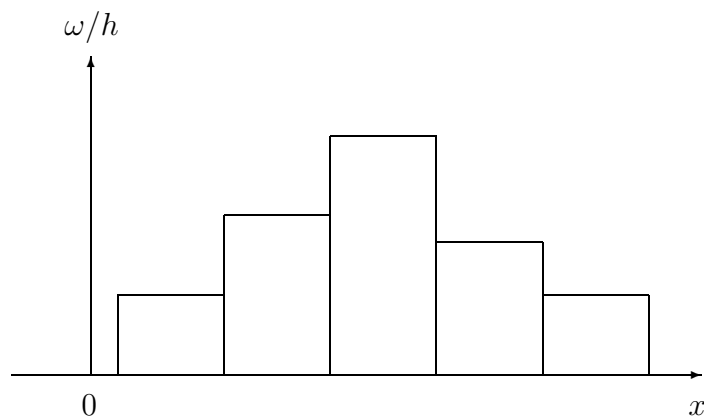


Рис. 1.3. Пример гистограммы относительных частот

### 1.5. Решение типового примера

**Пример 1.1.** Проведено 30 серий по 24 броска кубика, и в каждой серии отмечалось число выпадений грани с 6 очками. Получены следующие результаты: 4, 2, 5, 6, 1, 4, 3, 7, 4, 4, 2, 3, 3, 5, 6, 4, 5, 3, 2, 4, 5, 1, 8, 3, 4, 4, 5, 2, 4, 6. Рассматривая полученные данные как выборку, построить выборочный закон распределения для результатов данного опыта.

**Решение.** Случайная величина  $\xi$  – число выпадений 6 очков при 24 бросках кубика. В принципе, значениями  $\xi$  могли быть числа от 0 до 24. Однако, в данном случае вариационный ряд включает лишь 8 элементов: 1, 2, 3, 4, 5, 6, 7 и 8. Выборочный закон распределения (статистический ряд) имеет вид:

$x_i$	1	2	3	4	5	6	7	8
$n_i$	2	4	5	9	5	3	1	1
$\omega_i$	2/30	4/30	5/30	9/30	5/30	3/30	1/30	1/30



## 2. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ВЫБОРКИ

Задача математической статистики – по имеющейся выборке получить информацию о генеральной совокупности. Числовые характеристики репрезентативной выборки обычно являются хорошей оценкой соответствующих характеристик исследуемой случайной величины, связанной с генеральной совокупностью.

### 2.1. Выборочное среднее и выборочная дисперсия, эмпирические моменты

*Выборочным средним* называется среднее арифметическое значений вариант в выборке:

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \sum_{i=1}^k \omega_i x_i, \quad (2.1)$$

где  $x_i$  – варианты, а  $n_i$  – их частоты.

Выборочное среднее используется для статистической оценки математического ожидания исследуемой случайной величины. Ниже мы рассмотрим вопрос, насколько такая оценка является надежной и точной.

*Выборочной дисперсией* называется величина, равная

$$D^* = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \sum_{i=1}^k \omega_i (x_i - \bar{x})^2, \quad (2.2)$$

а *выборочным средним квадратическим отклонением* –

$$\sigma^* = \sqrt{D^*}. \quad (2.3)$$

Как и в случае теории случайных величин, легко показать, что выполняется следующее соотношение, удобное для вычисления дисперсии:

$$D^* = \overline{x^2} - \bar{x}^2, \quad (2.4)$$

где

$$\overline{x^2} = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 = \sum_{i=1}^k \omega_i x_i^2. \quad (2.5)$$

Напомним, что дисперсия не может быть отрицательной. Получение отрицательной дисперсии по формуле (2.4) свидетельствует об ошибке в вычислениях.

Выборочная дисперсия характеризует разброс вариант относительно выборочного среднего: чем больше разброс, тем больше значение дисперсии. Величина  $D^*$  участвует в статистической оценке дисперсии исследуемой величины.

Другими характеристиками вариационного ряда являются: *мода*  $M_0$  – варианта, имеющая наибольшую частоту и *медиана*  $m_e$  – варианта, которая делит вариационный ряд на две части, равные по числу вариант.

По аналогии с соответствующими теоретическими выражениями можно построить *эмпирические моменты*, применяемые для статистической оценки начальных и центральных моментов исследуемой случайной величины.

По аналогии с моментами  $\alpha_k$  и  $\beta_k$  теории вероятностей (см. п. 7 части I) *начальным эмпирическим моментом* порядка  $m$  называется величина

$$\alpha_m^* = \frac{1}{n} \sum_{i=1}^k n_i x_i^m = \sum_{i=1}^k \omega_i x_i^m, \quad (2.6)$$

*центральным эмпирическим моментом* порядка  $m$  –

$$\beta_m^* = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^m = \sum_{i=1}^k \omega_i (x_i - \bar{x})^m. \quad (2.7)$$

Заметим, что первый начальный момент  $\alpha_1^*$  – это выборочное среднее  $\bar{x}$ , первый центральный момент  $\beta_1^*$  равен нулю, а второй центральный момент  $\beta_2^*$  – это выборочная дисперсия  $D^*$ .

## **2.2. Свойства статистических оценок параметров распределения: несмещенность, эффективность, состоятельность**

Недостаточно получить статистические оценки параметров распределения случайной величины  $\xi$ : выборочное среднее, выборочную дисперсию и т. д., необходимо убедиться, что они являются хорошим приближением для соответствующих параметров теоретического распределения  $\xi$ . Найдем условия, которые должны для этого выполняться.

Пусть  $A^*$  – статистическая оценка неизвестного параметра  $A$  теоретического распределения некоторой случайной величины  $\xi$ , например,  $A$  – дисперсия  $D[\xi]$  некоторой случайной величины  $\xi$ , а  $A^*$  – ее выборочная дисперсия  $D^*$ . Выделим из генеральной совокупности  $l$  выборок одного и того же объема  $n$  и найдем по каждой из них выборочную оценку  $A_j^*$  ( $j = 1, 2, \dots, l$ ) параметра  $A$ . В принципе,  $A^*$  можно рассматривать как некоторую случайную величину, принявшую значения  $A_1^*, A_2^*$  и т.д. Тогда, если математическое ожидание  $A^*$  не равно оцениваемому параметру  $A$ , мы будем получать при вычислении оценок  $A$  систематические ошибки: если  $M[A^*] > A$ , оценка  $A^*$  будет в среднем больше  $A$ , если  $M[A^*] < A$  – меньше. Необходимым условием отсутствия систематических ошибок является требование  $M[A^*] = A$ .

Статистическая оценка  $A^*$  называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру генеральной совокупности  $A$  при любом объеме выборки, т.е.

$$M[A^*] = A. \quad (2.8)$$

Если это условие не выполняется, оценка называется *смещенной*.

Несмещенность оценки не является достаточным условием хорошего приближения статистической оценки  $A^*$  к истинному (теоретическому) значению оцениваемого параметра  $A$ . Разброс отдельных значений  $A_j^*$  относительно среднего значения  $M[A^*]$  зависит от величины дисперсии  $D[A^*]$ . Если дисперсия велика, то значение  $A_j^*$ , найденное по данным одной выборки, может значительно отличаться от оцениваемого параметра.

Следовательно, для надежного оценивания дисперсия  $D[A^*]$  должна быть мала. Статистическая оценка называется *эффективной*, если при заданном объеме выборки  $n$  она имеет наименьшую возможную дисперсию.

К статистическим оценкам предъявляется еще требование состоятельности. Оценка называется *состоятельной*, если при  $n \rightarrow \infty$  она стремится по вероятности к оцениваемому параметру. Заметим, что несмещенная оценка будет состоятельной, если при  $n \rightarrow \infty$  ее дисперсия стремится к 0.

### 2.3. Свойства выборочного среднего

Будем полагать, что варианты  $x_1, x_2, \dots, x_n$  являются значениями соответствующих независимых одинаково распределенных случайных величин  $\xi_1, \xi_2, \dots, \xi_n$ , имеющих математическое ожидание  $M[\xi_i] = a$  и дисперсию  $D[\xi_i] = \sigma^2$ . Тогда выборочное среднее можно рассматривать как случайную величину

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad (2.9)$$

принимаящую значение согласно формуле (2.1). Используя случайную величину  $\bar{X}$ , исследуем свойства выборочного среднего как статистической оценки.

*Несмещенность.* Из свойств математического ожидания (см. п. 7.2 части I) следует, что

$$M[\bar{X}] = M\left[\frac{1}{n} \sum_{i=1}^n \xi_i\right] = \frac{1}{n} M\left[\sum_{i=1}^n \xi_i\right] = \frac{1}{n} \sum_{i=1}^n M[\xi_i] = \frac{1}{n} \sum_{i=1}^n a = a,$$

т.е. выборочное среднее является несмещенной оценкой математического ожидания случайной величины.

*Эффективность.* Найдем сначала дисперсию выборочного среднего. Из свойств дисперсии (см. п. 7.2 части I) вытекает, что

$$D[\bar{X}] = D\left[\frac{1}{n} \sum_{i=1}^n \xi_i\right] = \frac{1}{n^2} D\left[\sum_{i=1}^n \xi_i\right] = \frac{1}{n^2} \sum_{i=1}^n D[\xi_i] = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}. \quad (2.10)$$

Из формулы видно, что  $D[\bar{X}]$  стремится к нулю при  $n \rightarrow \infty$ .

Для проверки эффективности этой оценки воспользуемся неравенством Крамера–Рао. Пусть  $\theta_n^*$  – несмещенная оценка параметра  $\theta$  случайной величины  $\xi$ , сделанная по выборке объемом  $n$ . Тогда для непрерывной случайной ве-

личины  $\xi$  с плотностью распределения вероятностей  $f_\xi(x, \theta)$ , зависящей от параметра  $\theta$ , имеем

$$D[\theta_n^*] \geq \frac{1}{nI(\theta)}, \quad (2.11)$$

где  $I(\theta)$  – информация Фишера, т.е.

$$I(\theta) = M \left[ \left( \frac{\partial \ln f_\xi(x, \theta)}{\partial \theta} \right)^2 \right]. \quad (2.12)$$

Несмещенная оценка параметра  $\theta_n^*$  будет эффективной, если для нее неравенство Крамера–Рао (2.11) обращается в равенство. В этом случае дисперсия данной оценки наименьшая из возможных и в этом смысле лучше всех остальных.

Очевидно, что дальнейшее рассмотрение предполагает знание распределения случайной величины  $\xi$ . Возвращаясь к выборочному среднему в качестве примера, мы рассмотрим случай, когда  $\xi$  имеет нормальный закон распределения.

Пусть в соотношениях (2.11), (2.12) рассматривается оценка математического ожидания (т.е.  $\theta_n^* = \bar{x}$ ), а величина  $\xi$  имеет нормальный закон распределения (т.е. параметр  $\theta = M[\xi] = a$ ):

$$f_\xi(x, a) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x-a)^2}{2\sigma^2} \right\}.$$

Вычислим информацию Фишера (2.12)

$$\begin{aligned} I(a) &= M \left[ \left( \frac{\partial}{\partial a} \ln \left( \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x-a)^2}{2\sigma^2} \right\} \right) \right)^2 \right] = \\ &= M \left[ \left( \frac{\partial}{\partial a} \left( \ln \left( \frac{1}{\sigma\sqrt{2\pi}} \right) - \frac{(x-a)^2}{2\sigma^2} \right) \right)^2 \right] = M \left[ \left( \frac{x-a}{\sigma^2} \right)^2 \right] = \frac{M[(x-a)^2]}{\sigma^4} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}. \end{aligned}$$

Таким образом, правая часть неравенства (2.11) равна  $\sigma^2 / n$ . Левую часть неравенства (2.11) с учетом  $D[\theta_n^*] = D[\bar{X}]$  мы уже вычислили – согласно соотношению (2.10) она также равна  $\sigma^2 / n$ .

Неравенство (2.11) превращается в равенство. Это означает эффективность выборочного среднего как оценки математического ожидания нормально распределенной случайной величины.

Заметим, что мы воспользовались при доказательстве эффективности оценки неравенством Крамера–Рао для непрерывной случайной величины. Аналогично может быть рассмотрен и дискретный случай.

*Состоятельность.* Пусть  $a$  – оцениваемый параметр, а именно математическое ожидание генеральной совокупности  $M[\xi]$ ,  $\sigma^2$  – дисперсия генеральной совокупности  $D[\xi]$ . Рассмотрим неравенство Чебышева (см. п. 9.1 части I)

$$p(|\xi - M[\xi]| \geq \varepsilon) \leq \frac{D[\xi]}{\varepsilon^2}.$$

В нашем случае  $\xi = \bar{X}$ ,  $M[\xi] = M[\bar{X}] = a$ ,  $D[\xi] = D[\bar{X}] = \sigma^2/n$  и тогда

$$p(|\bar{X} - a| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

При  $n \rightarrow \infty$  правая часть неравенства стремится к нулю для любого  $\varepsilon > 0$ , т.е.

$$\lim_{n \rightarrow \infty} (p(|\bar{X} - a| \geq \varepsilon)) = 0,$$

и, следовательно, величина  $\bar{X}$ , представляющая выборочную оценку, стремится к оцениваемому параметру  $a$  по вероятности.

Таким образом, выборочное среднее является несмещенной, эффективной (по крайней мере, для нормального распределения) и состоятельной оценкой математического ожидания случайной величины, связанной с генеральной совокупностью.

## 2.4. Свойства выборочной дисперсии

Исследуем несмещенность выборочной дисперсии  $D^*$  как оценки дисперсии случайной величины  $D[\xi] = \sigma^2$ .

По аналогии с п. 2.3 представим  $\bar{x}$  и  $\overline{x^2}$  как суммы случайных величин:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad \overline{X^2} = \frac{1}{n} \sum_{i=1}^n \xi_i^2.$$

Согласно (2.4)  $D^* = \overline{X^2} - (\bar{X})^2$ . Поэтому

$$\begin{aligned} M[D^*] &= M[\overline{X^2} - (\bar{X})^2] = M\left[\frac{1}{n} \sum_{i=1}^n \xi_i^2 - \left(\frac{1}{n} \sum_{i=1}^n \xi_i\right)^2\right] = \\ &= \frac{1}{n} \sum_{i=1}^n M[\xi_i^2] - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n M[\xi_i \xi_j]. \end{aligned}$$

Разобьем двойную сумму на две, выделив отдельно суммирование по  $i = j$ ,

$$M[D^*] = \frac{1}{n} \sum_{i=1}^n M[\xi_i^2] - \frac{1}{n^2} \sum_{i=1}^n M[\xi_i^2] - \frac{1}{n^2} \sum_{i \neq j} M[\xi_i \xi_j]. \quad (2.13)$$

Поскольку  $\xi_i$  и  $\xi_j$  – независимые случайные величины, то имеем  $M[\xi_i \xi_j] = M[\xi_i] M[\xi_j]$  (см., п. 8.6 части I). Дисперсия генеральной совокупности  $\sigma^2 = \alpha_2 - (\alpha_1)^2$ , где  $\alpha_1 = M[\xi_i]$  и  $\alpha_2 = M[\xi_i^2]$  – соответственно первый и второй начальные моменты. Тогда из соотношения (2.13) получим

$$\begin{aligned} M[D^*] &= \frac{1}{n} \sum_{i=1}^n \alpha_2 - \frac{1}{n^2} \sum_{i=1}^n \alpha_2 - \frac{1}{n^2} \sum_{i \neq j} \alpha_1^2 = \left(\frac{n-1}{n}\right) \alpha_2 - \left(\frac{n-1}{n}\right) \alpha_1^2 = \\ &= \left(\frac{n-1}{n}\right) (\alpha_2 - \alpha_1^2) = \left(\frac{n-1}{n}\right) \sigma^2. \end{aligned} \quad (2.14)$$

Итак,  $M[D^*] \neq \sigma^2$ , т.е.  $D^*$  – смещенная оценка дисперсии случайной величины  $\xi$ . Однако,

$$\lim_{n \rightarrow \infty} M[D^*] = \lim_{n \rightarrow \infty} \left(\frac{n-1}{n}\right) \sigma^2 = \sigma^2,$$

что означает *асимптотическую несмещенность* этой оценки.

### Замечания:

1. Можно предложить другую оценку дисперсии – *исправленную выборочную дисперсию*  $s^2$ , вычисляемую по формуле

$$s^2 = \frac{n}{n-1} D^* = \frac{1}{n-1} \sum_{i=1}^k n_i (x_i - \bar{x})^2. \quad (2.15)$$

Такая оценка будет несмещенной. Ей соответствует *исправленное выборочное среднее квадратическое отклонение*

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^k n_i (x_i - \bar{x})^2}. \quad (2.16)$$

2. Исправленная выборочная дисперсия  $s^2$  есть несмещенная оценка дисперсии случайной величины при неизвестном математическом ожидании. Если же математическое ожидание известно ( $a = a_0$ ), то несмещенной оценкой дисперсии будет выборочная дисперсия

$$s_0^2 = \frac{1}{n} \sum_{i=1}^k n_i (x_i - a_0)^2. \quad (2.17)$$

## 2.5. Решение типового примера

**Пример 2.1.** Найти выборочное среднее, выборочную дисперсию и среднее квадратическое отклонение, моду и исправленную выборочную дисперсию для выборки из примера 1.1, имеющей следующий закон распределения:

$x_i$	1	2	3	4	5	6	7	8
$n_i$	2	4	5	9	5	3	1	1
$\omega_i$	2/30	4/30	5/30	9/30	5/30	3/30	1/30	1/30

**Решение.** В данном примере  $n = 30$  и  $k = 8$ . Получаем

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{1}{30} (2 \cdot 1 + 4 \cdot 2 + 5 \cdot 3 + 9 \cdot 4 + 5 \cdot 5 + 3 \cdot 6 + 1 \cdot 7 + 1 \cdot 8) = 3,97;$$

$$\overline{x^2} = \frac{1}{30} (2 \cdot 1^2 + 4 \cdot 2^2 + 5 \cdot 3^2 + 9 \cdot 4^2 + 5 \cdot 5^2 + 3 \cdot 6^2 + 1 \cdot 7^2 + 1 \cdot 8^2) = 18,43$$

и по формуле (2.4) найдем  $D^* = 18,43 - 15,76 = 2,67$  и  $\sigma^* = \sqrt{2,67} = 1,63$ .

Мода  $M_0$ , т.е. варианта, имеющая наибольшую частоту, равна 4.

Исправленная выборочная дисперсия  $s^2 = n / (n - 1) D^* = 2,76$ .

### 3. ТОЧЕЧНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ ИЗВЕСТНОГО РАСПРЕДЕЛЕНИЯ

Получив по выборке оценки характеристик случайной величины, связанной с генеральной совокупностью, можно перейти к дальнейшему ее изучению. На первом этапе из каких-либо соображений надо сделать предположения о законе распределения данной случайной величины. В дальнейшем мы подробно рассмотрим способы проверки справедливости подобных гипотез.

В настоящем же разделе будем считать, что общий вид закона распределения нам известен и остается уточнить только детали – параметры, определяющие его окончательную форму. Существует несколько методов решения этой задачи, два из которых мы рассмотрим ниже.

#### 3.1. Метод моментов

Напомним, что закон распределения непрерывной случайной величины  $\xi$  описывается плотностью распределения вероятностей  $f_\xi(x, \theta)$ , а дискретной величины – таблицей, связывающей значения  $x_i$  с их вероятностями  $p(\xi = x_i, \theta)$ . В обоих случаях  $\theta = (\theta_1, \theta_2, \dots, \theta_r)$  – вектор (набор) параметров. Например, для нормального распределения имеем два ( $r = 2$ ) параметра:  $\theta_1 = a$  и  $\theta_2 = \sigma$ , для распределения Пуассона – только один ( $r = 1$ ):  $\theta_1 = \lambda$ .

Очевидно, что если закон распределения случайной величины зависит от параметров  $\theta = (\theta_1, \theta_2, \dots, \theta_r)$ , то и ее моменты зависят от этих же параметров, т.е.  $\alpha_k = \alpha_k(\theta)$  и  $\beta_k = \beta_k(\theta)$ .

Можно показать, что при объеме выборки  $n \rightarrow \infty$  выборочные моменты  $\alpha_k^*$  и  $\beta_k^*$  сходятся по вероятности соответственно к  $\alpha_k(\theta)$  и  $\beta_k(\theta)$ . Поэтому для достаточно больших выборок можно с большой вероятностью полагать, что для всех  $k \geq 1$

$$\alpha_k(\theta) \approx \alpha_k^*, \quad \beta_k(\theta) \approx \beta_k^*. \quad (3.1)$$

*Метод моментов*, развитый Карлом Пирсоном в 1894 году, основан на использовании этих приближенных равенств: моменты  $\alpha_k(\theta)$ ,  $\beta_k(\theta)$  рассчитываются теоретически по известному закону распределения с параметрами  $\theta$ , а выборочные моменты  $\alpha_k^*$ ,  $\beta_k^*$  вычисляются по имеющейся выборке. Неизвестные параметры  $\theta = (\theta_1, \theta_2, \dots, \theta_r)$  определяется в результате решения системы из  $r$  уравнений, связывающих соответствующие теоретический и эмпирический моменты, например,  $\alpha_k(\theta) = \alpha_k^*$  или  $\beta_k(\theta) = \beta_k^*$ .

Можно показать [2], что при довольно общих условиях оценки параметров  $\theta$ , полученные методом моментов, состоятельны, их математические ожидания отличаются от истинных значений параметров на величину порядка  $n^{-1}$ , а средние квадратические отклонения являются величинами порядка  $n^{-0,5}$ .



### 3.2. Метод наибольшего правдоподобия

Этот метод в той или иной форме использовался еще Гауссом, Лапласом и др. Однако популярность метод наибольшего правдоподобия приобрел после анализа, выполненного Рональдом Фишером в 1910-х годах.

В основе метода лежит *функция правдоподобия*  $L(x_1, x_2, \dots, x_n, \theta)$ , являющаяся законом распределения вектора  $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ , где случайные величины  $\xi_i$  принимают значения вариант выборки, т.е. имеют одинаковое распределение (см. подробнее п. 2.3). Поскольку случайные величины  $\xi_i$  независимы, функция правдоподобия имеет вид:

$$L(x_1, x_2, \dots, x_n, \theta) = \begin{cases} \prod_{i=1}^n f_{\xi_i}(x_i, \theta), & \text{если } \xi_i \text{ непрерывны;} \\ \prod_{i=1}^n p_{\xi_i}(x_i, \theta), & \text{если } \xi_i \text{ дискретны.} \end{cases} \quad (3.2)$$

Идея *метода максимального правдоподобия* состоит в том, что мы ищем такие значения параметров  $\theta$ , при которых вероятность появления в выборке значений вариант  $x_1, x_2, \dots, x_n$  является наибольшей. Иными словами, в качестве оценки параметров  $\theta$  берется вектор  $\tilde{\theta}$ , при котором функция правдоподобия имеем локальный максимум при заданных  $x_1, x_2, \dots, x_n$ :

$$L(x_1, x_2, \dots, x_n, \tilde{\theta}) = \max_{\theta} \{L(x_1, x_2, \dots, x_n, \theta)\}. \quad (3.3)$$

В результате каждый из элементов вектора  $\tilde{\theta}$  является функцией  $x_1, x_2, \dots, x_n$ .

Как правило, оценки по методу максимального правдоподобия получаются из необходимого условия экстремума функции  $L(x_1, x_2, \dots, x_n, \theta)$  в точке  $\theta = \tilde{\theta}$ :

$$\begin{cases} \left. \frac{\partial}{\partial \theta_1} L(x_1, x_2, \dots, x_n, \theta) \right|_{\theta = \tilde{\theta}} = 0, \\ \left. \frac{\partial}{\partial \theta_2} L(x_1, x_2, \dots, x_n, \theta) \right|_{\theta = \tilde{\theta}} = 0, \\ \dots \\ \left. \frac{\partial}{\partial \theta_r} L(x_1, x_2, \dots, x_n, \theta) \right|_{\theta = \tilde{\theta}} = 0. \end{cases} \quad (3.4)$$

Обычно эту систему называют *уравнениями правдоподобия*. Решая уравнения (3.4), находим значения  $\tilde{\theta}$ . После этого с помощью достаточного условия экстремума необходимо убедиться, что полученные значения  $\tilde{\theta}$  действительно дают максимум функции правдоподобия.

Достоинства метода наибольшего правдоподобия состоят в том, что:

- 1) полученная оценка состоятельна (хотя может быть смещенной);
- 2) при  $n \rightarrow \infty$  оценка  $\tilde{\theta}$  приближается к эффективной оценке;

3) метод наиболее полно использует данные выборки и поэтому особенно полезен в случае выборок небольшого объема.

Недостатком данного метода является относительная сложность вычислений.

#### Замечания:

1. При поиске максимума функции правдоподобия для упрощения расчетов можно выполнить действия, не изменяющие результата: во-первых, использовать вместо  $L(x_1, x_2, \dots, x_n, \theta)$  логарифмическую функцию правдоподобия  $l(x_1, x_2, \dots, x_n, \theta) = \ln L(x_1, x_2, \dots, x_n, \theta)$ ; во-вторых, отбросить в выражении для функции правдоподобия не зависящие от  $\theta$  слагаемые (для  $l$ ) или положительные сомножители (для  $L$ ).

2. Оценки параметров, рассмотренные нами, можно назвать *точечными оценками*, так как для неизвестного параметра  $\theta$  определяется одна единственная точка  $\tilde{\theta}$ , являющаяся его приближенным значением. Однако такой подход может приводить к грубым ошибкам, и точечная оценка может значительно отличаться от истинного значения оцениваемого параметра (особенно в случае выборки малого объема).

### 3.3. Решение типовых примеров

**Пример 3.1.** Известно, что характеристика  $\xi$  объектов генеральной совокупности, являясь случайной величиной, имеет равномерное распределение, зависящее от параметров  $a$  и  $b$ :

$$f_{\xi}(x, a, b) = \begin{cases} 0, & \text{если } x \in (-\infty, a); \\ \frac{1}{b-a}, & \text{если } x \in [a, b]; \\ 0, & \text{если } x \in (b, \infty). \end{cases} \quad (3.5)$$

Требуется определить методом моментов параметры  $a$  и  $b$  по известному выборочному среднему  $\bar{x}$  и выборочной дисперсии  $D^* = (\sigma^*)^2$ .

**Решение.** Неизвестных параметров два:  $a, b$ . Выборочное среднее равно первому выборочному моменту  $\bar{x} = \alpha_1^*$ , а выборочная дисперсия – второму центральному моменту  $D^* = \beta_2^*$ . Поэтому в методе моментов (здесь  $\theta = (a, b)$ ) будем использовать два уравнения:

$$\begin{cases} \alpha_1(a, b) = \alpha_1^*, \\ \beta_2(a, b) = \beta_2^*. \end{cases} \quad (3.6)$$

Первый начальный момент  $\alpha_1$ , равный математическому ожиданию, и второй центральный момент  $\beta_2$ , равный дисперсии, в случае равномерного распределения имеют вид (см. п. 7.5 части I):

$$\alpha_1(a, b) = \frac{a+b}{2}, \quad \beta_2(a, b) = \frac{(b-a)^2}{12}.$$

Подставляя соотношения в систему (3.6), получим

$$\begin{cases} \frac{a+b}{2} = \bar{x}, \\ \frac{(b-a)^2}{12} = D^* = (\sigma^*)^2. \end{cases}$$

Решая систему, найдем

$$\begin{cases} a = \bar{x} - \sqrt{3}\sigma^*, \\ b = \bar{x} + \sqrt{3}\sigma^*. \end{cases} \quad (3.7)$$

**Пример 3.2.** Пусть характеристика  $\xi$  объектов генеральной совокупности, являясь случайной величиной, имеет нормальное распределение с параметрами  $a$  и  $\sigma$ , т.е.

$$f_{\xi}(x) = f_N(x, a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Методом максимального правдоподобия найти параметр  $a$ , если известна дисперсия  $\sigma^2$ .

**Решение.** Используя определение (3.2), построим функцию правдоподобия для данной непрерывной случайной величины ( $n$  – объем выборки)

$$L(x_1, x_2, \dots, x_n, a) = \prod_{i=1}^n f_{\xi_i}(x_i, a) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^n \exp\left[-\frac{(x_i - a)^2}{2\sigma^2}\right].$$

Отбросим множитель  $(\sigma\sqrt{2\pi})^{-n}$ , не зависящий от параметра  $a$ , и рассмотрим логарифмическую функцию правдоподобия  $l(x_1, x_2, \dots, x_n, a) = l(\mathbf{x}, a)$

$$l(\mathbf{x}, a) = \ln\left\{\prod_{i=1}^n \exp\left[-\frac{(x_i - a)^2}{2\sigma^2}\right]\right\} = -\sum_{i=1}^n \frac{(x_i - a)^2}{2\sigma^2}. \quad (3.8)$$

Составим уравнение правдоподобия для  $\tilde{a}$  – оценки параметра  $a$

$$\frac{\partial}{\partial a} l(\mathbf{x}, a) \Big|_{a=\tilde{a}} = \sum_{i=1}^n \frac{(x_i - \tilde{a})}{\sigma^2} = 0, \quad (3.9)$$

решая которое, получим:

$$\sum_{i=1}^n x_i = \sum_{i=1}^n \tilde{a} = n\tilde{a}, \quad \tilde{a} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (3.10)$$

Осталось проверить, что при  $a = \tilde{a}$  функция  $l(\mathbf{x}, a)$  достигает своего максимума. Для этого найдем значение второй производной при  $a = \tilde{a}$

$$\frac{\partial^2}{\partial a^2} l(\mathbf{x}, a) \Big|_{a=\tilde{a}} = -\frac{n}{\sigma^2}.$$

Поскольку эта производная всегда меньше 0, то  $\tilde{a}$  действительно дает максимум логарифмической функции правдоподобия.

Таким образом, полученная методом наибольшего правдоподобия выборочная оценка  $\tilde{a}$  параметра ожидания нормального распределения  $a$  совпала с выборочным средним  $\bar{x}$ .

**Пример 3.3.** Для характеристики  $\xi$  объектов генеральной совокупности, имеющей нормальную функцию распределения  $F_\xi(x) = F_N(x, a, \sigma)$ , методом максимального правдоподобия найти математическое ожидание  $a$  и дисперсию  $\sigma^2$ .

**Решение.** В отличие от предыдущего примера в данной задаче следует оценить сразу два неизвестных параметра:  $a$  и  $\sigma^2$ .

Логарифмическая функция правдоподобия имеет вид

$$l(\mathbf{x}, a, \sigma^2) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi - \sum_{i=1}^n \frac{(x_i - a)^2}{2\sigma^2}. \quad (3.11)$$

Отбросив в этой формуле слагаемое, которое не зависит от  $a$  и  $\sigma^2$ , составим систему уравнений правдоподобия (3.4)

$$\begin{cases} \left. \frac{\partial}{\partial a} l(\mathbf{x}, a, \sigma^2) \right|_{a=\tilde{a}, \sigma^2=\tilde{\sigma}^2} = \sum_{i=1}^n \frac{x_i - \tilde{a}}{\tilde{\sigma}^2} = 0, \\ \left. \frac{\partial}{\partial \sigma^2} l(\mathbf{x}, a, \sigma^2) \right|_{a=\tilde{a}, \sigma^2=\tilde{\sigma}^2} = \sum_{i=1}^n \frac{(x_i - \tilde{a})^2}{2(\tilde{\sigma}^2)^2} - \frac{n}{2\tilde{\sigma}^2} = 0. \end{cases} \quad (3.12)$$

Решая систему (3.12), получим

$$\begin{cases} \tilde{a} = \frac{1}{n} \sum_{i=1}^n x_i, \\ \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \tilde{a})^2. \end{cases} \quad (3.13)$$

Чтобы убедиться, что при  $\tilde{a}$  и  $\tilde{\sigma}^2$  достигается максимум функции правдоподобия, следует найти вторые производные

$$\frac{\partial^2}{\partial a^2} l(\mathbf{x}, a, \sigma^2), \quad \frac{\partial^2}{\partial (\sigma^2)^2} l(\mathbf{x}, a, \sigma^2), \quad \frac{\partial^2}{\partial a \partial \sigma^2} l(\mathbf{x}, a, \sigma^2).$$

Получаем

$$\begin{cases} \frac{\partial^2}{\partial a^2} l(\mathbf{x}, a, \sigma^2) = -\frac{n}{\sigma^2}, \\ \frac{\partial^2}{\partial (\sigma^2)^2} l(\mathbf{x}, a, \sigma^2) = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^n (x_i - a)^2, \\ \frac{\partial^2}{\partial a \partial \sigma^2} l(\mathbf{x}, a, \sigma^2) = -\frac{1}{(\sigma^2)^2} \sum_{i=1}^n (x_i - a). \end{cases} \quad (3.14)$$

При  $a = \tilde{a}$  и  $\sigma^2 = \tilde{\sigma}^2$  система (3.14) переписывается как

$$\begin{cases} \frac{\partial^2}{\partial a^2} l(\mathbf{x}, \tilde{a}, \tilde{\sigma}^2) = -\frac{n}{\tilde{\sigma}^2}, \\ \frac{\partial^2}{\partial (\sigma^2)^2} l(\mathbf{x}, \tilde{a}, \tilde{\sigma}^2) = \frac{n}{2(\tilde{\sigma}^2)^2} - \frac{n}{(\tilde{\sigma}^2)^2} = -\frac{n}{2(\tilde{\sigma}^2)^2}, \\ \frac{\partial^2}{\partial a \partial \sigma^2} l(\mathbf{x}, \tilde{a}, \tilde{\sigma}^2) = 0. \end{cases} \quad (3.15)$$

Определитель равен

$$\Delta = \begin{vmatrix} \frac{\partial^2}{\partial a^2} l(\mathbf{x}, \tilde{a}, \tilde{\sigma}^2) & \frac{\partial^2}{\partial a \partial \sigma^2} l(\mathbf{x}, \tilde{a}, \tilde{\sigma}^2) \\ \frac{\partial^2}{\partial \sigma^2 \partial a} l(\mathbf{x}, \tilde{a}, \tilde{\sigma}^2) & \frac{\partial^2}{\partial (\sigma^2)^2} l(\mathbf{x}, \tilde{a}, \tilde{\sigma}^2) \end{vmatrix} = \begin{vmatrix} -\frac{n}{\tilde{\sigma}^2} & 0 \\ 0 & -\frac{n}{2(\tilde{\sigma}^2)^2} \end{vmatrix}.$$

Находим, что

$$\Delta = \frac{n^2}{2(\tilde{\sigma}^2)^3} > 0, \quad \frac{\partial^2}{\partial a^2} l(\mathbf{x}, \tilde{a}, \tilde{\sigma}^2) = -\frac{n}{\tilde{\sigma}^2} < 0,$$

т.е. достаточные условия существования максимума функции двух переменных выполнены и мы можем утверждать, что задача решена.

Таким образом, при использовании метода максимального правдоподобия параметр  $a$  оценивается выборочным средним  $\bar{x}$ , а параметр  $\sigma^2$  – выборочной дисперсией  $D^*$ . Заметим, что полученная оценка  $\sigma^2$  является смещенной.

**Пример 3.4.** Пусть  $x_1, x_2, \dots, x_n$  – выборка значений дискретной случайной величины  $\xi$ , связанной с некоторой генеральной совокупностью и имеющей распределение Пуассона:

$$p(\xi = x, \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda),$$

где  $\lambda > 0$  и значения  $x = x_1, x_2, \dots$  – целые неотрицательные числа. Найти методом максимального правдоподобия оценку  $\tilde{\lambda}$  параметра  $\lambda$ .

**Решение.** Построим функцию правдоподобия (3.2)

$$L(\mathbf{x}, \lambda) = \prod_{i=1}^n p(\xi_i = x_i, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda) = \frac{\lambda^{\sum x_i}}{\prod x_i!} \exp(-n\lambda) = \frac{\lambda^{n\bar{x}}}{\prod x_i!} \exp(-n\lambda),$$

где  $\bar{x} = \left( \sum_{i=1}^n x_i \right) / n$  – выборочное среднее.

Как и в предыдущих примерах, в дальнейших расчетах удобнее использовать логарифмическую функцию правдоподобия

$$l(\mathbf{x}, \lambda) = \ln \left[ \frac{\lambda^{n\bar{x}}}{\prod x_i!} \exp(-n\lambda) \right] = n\bar{x} \ln \lambda - \ln \prod_{i=1}^n x_i! - n\lambda. \quad (3.16)$$

Тогда уравнение правдоподобия принимает вид:

$$\left. \frac{\partial}{\partial \lambda} l(\mathbf{x}, \lambda) \right|_{\lambda=\tilde{\lambda}} = \frac{n\bar{x}}{\tilde{\lambda}} - n = 0.$$

Решение данного уравнения дает точку экстремума  $\tilde{\lambda} = \bar{x}$ . Поскольку вторая производная  $l(\mathbf{x}, \lambda)$  по  $\lambda$  в точке  $\lambda = \tilde{\lambda}$  равна  $-n\bar{x}/\tilde{\lambda}^2$  и меньше 0, т.к.  $\bar{x} > 0$ , то экстремум является максимумом.

Таким образом, в методе максимального правдоподобия параметр распределения Пуассона  $\lambda$  оценивается выборочным средним  $\tilde{\lambda} = \bar{x}$ . Этот результат был предсказуем, поскольку математическое ожидание случайной величины, распределенной подобным образом, равно  $\lambda$  (см. п. 7.2 части I).

## 4. ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ ИЗВЕСТНОГО РАСПРЕДЕЛЕНИЯ

Задачу оценивания параметра известного распределения можно также решать путем построения интервала, в который с заданной вероятностью попадает истинное значение параметра. Такой метод оценивания называется *интервальной оценкой*.

Обычно в математике для оценки  $\tilde{\theta}$  параметра  $\theta$  строится неравенство

$$|\tilde{\theta} - \theta| < \delta, \quad (4.1)$$

где число  $\delta$  характеризует точность оценки: чем меньше  $\delta$ , тем лучше оценка.

Статистические методы, однако, позволяют говорить только о том, что неравенство (4.1) выполняется с некоторой вероятностью. Поэтому “хорошей” оценкой здесь нужно считать построение достаточно узкого интервала (т.е. с достаточно малым  $\delta$ ), в который параметр  $\theta$  попадает с достаточно большой вероятностью  $\gamma$ :

$$p(|\tilde{\theta} - \theta| < \delta) = p(\tilde{\theta} - \delta < \theta < \tilde{\theta} + \delta) = \gamma. \quad (4.2)$$

С соотношением (4.2) связаны следующие термины:

- 1)  $\gamma$  – вероятность, с которой выполняется неравенство, называется *надежностью* (*доверительной вероятностью*) оценки  $\tilde{\theta}$  параметра  $\theta$ ;
- 2)  $\alpha = (1 - \gamma)$  – вероятность противоположенного события, которую называют *уровнем значимости*;
- 3) интервал  $(\tilde{\theta} - \delta < \theta < \tilde{\theta} + \delta)$ , в который попадает неизвестный параметр  $\theta$  с заданной надежностью  $\gamma$ , – *доверительный интервал*.

### 4.1. Оценивание математического ожидания нормально распределенной величины при известной дисперсии

Пусть исследуемая случайная величина  $\xi$  распределена по нормальному закону с известным средним квадратическим отклонением  $\sigma$  и неизвестным математическим ожиданием  $a$ . Требуется по значению выборочного среднего  $\bar{x}$  оценить математическое ожидание  $\xi$ .

Как и в п. 2.3 будем рассматривать получаемое выборочное среднее  $\bar{x}$  как значение случайной величины  $\bar{X}$ , а значения вариант выборки  $x_1, x_2, \dots, x_n$  соответственно как значения одинаково распределенных независимых случайных величин  $\xi_1, \xi_2, \dots, \xi_n$ , каждая из которых имеет математическое ожидание  $a$  и среднее квадратическое отклонение  $\sigma$ . Тогда имеем

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad M[\bar{X}] = \frac{1}{n} \sum_{i=1}^n M[\xi_i] = a, \quad \sigma[\bar{X}] = \sqrt{\frac{1}{n^2} \sum_{i=1}^n D[\xi_i]} = \frac{\sigma}{\sqrt{n}}. \quad (4.3)$$

При достаточно большом  $n$  согласно центральной предельной теореме случайная величина  $(n\bar{X} - nM[\bar{X}])/n\sigma[\bar{X}]$  имеет практически нормальное распределение с нулевым математическим ожиданием и единичной дисперсией. Тогда вероятность попадания этой величины в интервал  $(-x, x)$  равна (см. п. 7.5.5 части I)

$$p\left(\left|\frac{\bar{X} - M[\bar{X}]}{\sigma[\bar{X}]}\right| < x\right) = 2\Phi(x), \quad (4.4)$$

где  $\Phi(x)$  – функция Лапласа.

Пусть  $\tilde{x}_\gamma$  – число, при котором  $\Phi(\tilde{x}_\gamma) = \gamma/2$ . Тогда для данной выборки, где величина  $\bar{X}$  принимает значение  $\bar{x}$ , из формулы (4.4) с учетом (4.3) после несложных преобразований получим

$$\begin{cases} p\left(\bar{x} - \frac{\sigma \tilde{x}_\gamma}{\sqrt{n}} < a < \bar{x} + \frac{\sigma \tilde{x}_\gamma}{\sqrt{n}}\right) = 2\Phi(\tilde{x}_\gamma), \\ 2\Phi(\tilde{x}_\gamma) = \gamma. \end{cases} \quad (4.5)$$

Система (4.5) связывает надежность  $\gamma$  и доверительный интервал для математического ожидания, равный

$$\left(\bar{x} - \frac{\sigma \tilde{x}_\gamma}{\sqrt{n}}, \bar{x} + \frac{\sigma \tilde{x}_\gamma}{\sqrt{n}}\right). \quad (4.6)$$

Если известна величина  $\gamma$ , то из второго уравнения системы (4.5), пользуясь таблицами значений функций Лапласа (см. Приложение), находим параметр  $\tilde{x}_\gamma$  и затем доверительный интервал из первого уравнения этой системы. И, наоборот, если известен доверительный интервал, то сначала, используя интервал (4.6), определяют  $\tilde{x}_\gamma$ , а затем находят  $\Phi(\tilde{x}_\gamma)$  и  $\gamma$ .

**Замечание.** Величина  $\tilde{x}_\gamma$  равна квантили порядка  $(\gamma + 1)/2$  нормального распределения  $x_{(\gamma+1)/2}$  (см. подробнее п. 6.1).

## 4.2. Оценивание математического ожидания нормально распределенной величины при неизвестной дисперсии

Пусть исследуемая случайная величина  $\xi$  распределена по нормальному закону с неизвестными математическим ожиданием  $a$  и средним квадратическим отклонением  $\sigma$ .

Используя величины  $\xi_1, \xi_2, \dots, \xi_n$ , определенные в п. 4.1, введем случайную величину  $Z$ , принимающую значения исправленной выборочной дисперсии  $s^2$ :

$$Z = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - a)^2.$$

Рассмотрим случайную величину



$$t_n = \frac{\bar{X} - a}{\sqrt{Z/n}},$$

где случайная величина  $\bar{X}$  определена в соотношениях (4.3),  $a$  – неизвестное математическое ожидание,  $n$  – объем выборки.

Известно, что случайная величина  $t_n$ , заданная таким образом, имеет *распределение Стьюдента* с  $k = n - 1$  степенями свободы (см. подробнее п. 6.2). Плотность распределения вероятностей такой величины есть

$$f_t(x, n-1) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)} \Gamma\left(\frac{n-1}{2}\right)} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}, \quad (4.7)$$

где  $\Gamma(x)$  – гамма-функция – известная специальная функция (см. п. 6.2).

Пусть число  $\tilde{t}_{\gamma, n-1}$  определяется следующим соотношением, где учтена четность функции  $f_t(x, n-1)$ :

$$p(|t_n| < \tilde{t}_{\gamma, n-1}) = \int_{-\tilde{t}_{\gamma, n-1}}^{\tilde{t}_{\gamma, n-1}} f_t(x, n-1) dx = 2 \int_0^{\tilde{t}_{\gamma, n-1}} f_t(x, n-1) dx = \gamma. \quad (4.8)$$

Отсюда для данной выборки, где случайная величина  $\bar{X}$  принимает значение выборочного среднего  $\bar{x}$ , а  $Z$  – исправленной выборочной дисперсии  $s^2$ , получим

$$p\left(\bar{x} - \frac{s\tilde{t}_{\gamma, n-1}}{\sqrt{n}} < a < \bar{x} + \frac{s\tilde{t}_{\gamma, n-1}}{\sqrt{n}}\right) = \gamma. \quad (4.9)$$

Это равенство похоже на уравнения (4.5), позволяющие найти доверительный интервал для математического ожидания  $a$  в случае известной дисперсии. Различие заключается в том, что вместо  $\sigma$  здесь стоит исправленной (несмещенное) выборочное среднее квадратическое отклонение  $s$ , а параметр  $\tilde{t}_{\gamma, n-1}$  находится по известным  $\gamma$  и  $n$  из таблиц квантилей или критических точек распределения Стьюдента.

Способы определения  $\tilde{t}_{\gamma, n-1}$  становятся более понятными после рассмотрения плотности распределения Стьюдента на рис. 4.1. В силу четности функции  $f_t(x, n)$  целесообразно выбрать доверительный интервал симметрично относительно  $a = \bar{x}$ , т.е. относительно  $x = 0$  (т.е.  $t_n = 0$ ) – тогда ширина интервала является наименьшей для данного значения  $\gamma$ . Площадь заштрихованной фигуры на рис. 4.1 равна вероятности  $\gamma$ , и в силу симметрии площади криволинейных трапеций слева и справа от заштрихованной фигуры совпадают друг с другом и равны  $(1 - \gamma) / 2$ .

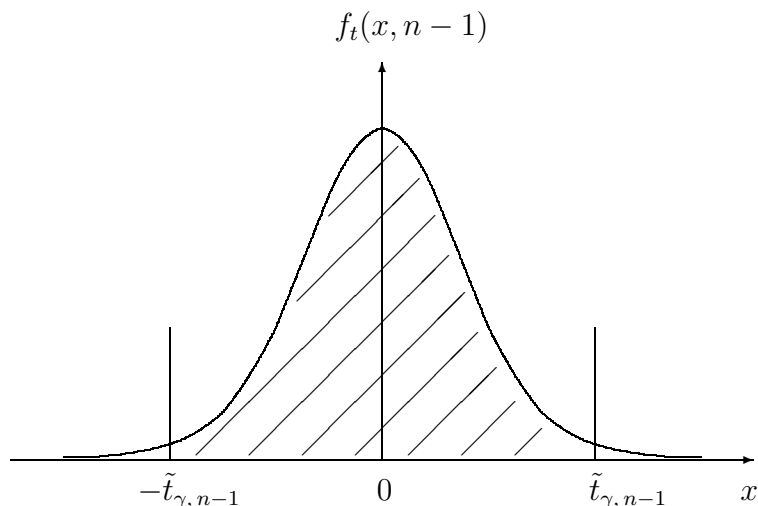


Рис. 4.1. Плотность распределения Стьюдента с  $n - 1$  степенями свободы

В зависимости от имеющихся таблиц параметр  $\tilde{t}_{\gamma, n-1}$  находят следующими способами:

- 1) как квантиль распределения Стьюдента с  $n - 1$  степенями свободы, т.е. как число, для которого (см. п. 6.2)

$$p(t_n < \tilde{t}_{\gamma, n-1}) = \gamma + \frac{1 - \gamma}{2} = \frac{1 + \gamma}{2}, \quad (4.10)$$

и тогда  $\tilde{t}_{\gamma, n-1} = t_{(1+\gamma)/2, n-1}$ , где  $t_{(1+\gamma)/2, n-1}$  – квантиль порядка  $(1 + \gamma) / 2$ , определяемый по таблицам квантилей данного распределения (см., например, табл. 2 Приложения);

- 2) как критическую точку распределения Стьюдента для уровня значимости  $(1 - \gamma) / 2$  (в случае таблиц с данными для односторонней критической области), т.е. как число, для которого

$$p(t_n > \tilde{t}_{\gamma, n-1}) = 1 - \frac{1 + \gamma}{2} = \frac{1 - \gamma}{2}, \quad (4.11)$$

и тогда  $\tilde{t}_{\gamma, n-1} = t_1((1 - \gamma)/2, n - 1)$ , где  $t_1((1 - \gamma)/2, n - 1)$  – критическая точка с уровнем значимости  $(1 - \gamma)/2$  для односторонней области;

- 3) как критическую точку распределения Стьюдента для уровня значимости  $1 - \gamma$  (в случае таблиц с данными для двухсторонней критической области)

$$p(|t_n| > \tilde{t}_{\gamma, n-1}) = 1 - \gamma, \quad (4.12)$$

и тогда  $\tilde{t}_{\gamma, n-1} = t_2(1 - \gamma, n - 1)$ , где  $t_2(1 - \gamma, n - 1)$  – критическая точка с уровнем значимости  $(1 - \gamma)$  для двухсторонней области.

**Замечание.** При больших  $k$  распределение Стьюдента стремится к нормальному распределению с нулевым математическим ожиданием и единичной дисперсией, и доверительный интервал (при  $k \geq 30$ ) можно на практике находить по формулам

$$\begin{cases} p\left(\bar{x} - \frac{s\tilde{x}_\gamma}{\sqrt{n}} < a < \bar{x} + \frac{s\tilde{x}_\gamma}{\sqrt{n}}\right) = 2\Phi(\tilde{x}_\gamma), \\ 2\Phi(\tilde{x}_\gamma) = \gamma. \end{cases} \quad (4.13)$$

### 4.3. Оценивание среднего квадратического отклонения нормально распределенной величины

Пусть исследуемая случайная величина  $\xi$  распределена по нормальному закону с математическим ожиданием  $a$  и неизвестным средним квадратическим отклонением  $\sigma$ . Рассмотрим два случая.

#### 4.3.1. Случай известного математического ожидания

Пусть известно значение  $M[\xi] = a$  и требуется оценить только  $\sigma$  или дисперсию  $D[\xi] = \sigma^2$ . Напомним, что при известном математическом ожидании несмещенной оценкой дисперсии является выборочная дисперсия  $D^* = (\sigma^*)^2$  (см. соотношение (2.4)).

Используя величины  $\xi_1, \xi_2, \dots, \xi_n$ , определенные выше, введем случайную величину  $Y$ , принимающую значения выборочной дисперсии  $D^*$ :

$$Y = \frac{1}{n} \sum_{i=1}^n (\xi_i - a)^2.$$

Рассмотрим случайную величину

$$H_n = \frac{nY}{\sigma^2} = \sum_{i=1}^n \left( \frac{\xi_i - a}{\sigma} \right)^2. \quad (4.14)$$

Стоящие под знаком суммы (4.14) случайные величины  $(\xi_i - a) / \sigma$  имеют нормальное распределение с плотностью  $f_N(x, 0, 1)$ . Тогда  $H_n$  имеет *распределение  $\chi^2$  (хи-квадрат)* с  $n$  степенями свободы как сумма квадратов  $n$  независимых стандартных ( $a = 0, \sigma = 1$ ) нормальных случайных величин. Свойства распределения хи-квадрат рассмотрены в п. 6.3.

Определим доверительный интервал из условия

$$p\left(\tilde{\chi}_{1,\gamma,n}^2 < H_n < \tilde{\chi}_{2,\gamma,n}^2\right) = \int_{\tilde{\chi}_{1,\gamma,n}^2}^{\tilde{\chi}_{2,\gamma,n}^2} f_{\chi^2}(x, n) dx = \gamma, \quad (4.15)$$

где  $f_{\chi^2}(x, n)$  – плотность распределения хи-квадрат и  $\gamma$  – надежность (доверительная вероятность). Величина  $\gamma$  численно равна площади заштрихованной фигуры на рис. 4.2.

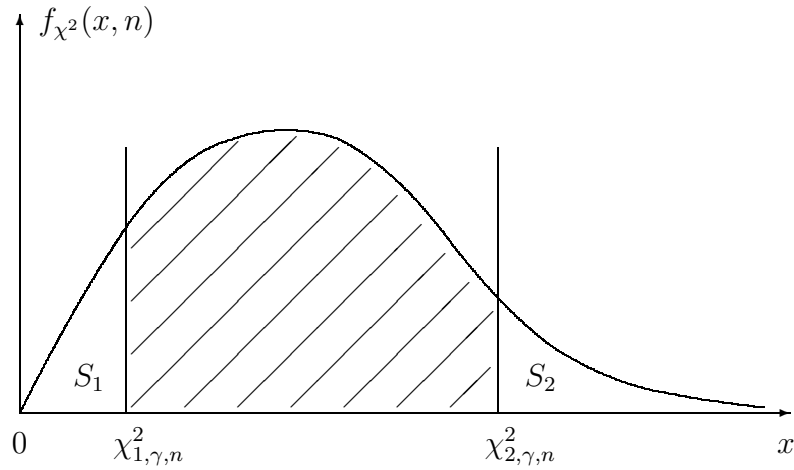


Рис. 4.2. Плотность распределения хи-квадрат с  $n$  степенями свободы

Плотность распределения хи-квадрат является несимметричной функцией относительно начала координат, что приводит к особенностям в построении доверительного интервала. Например, можно потребовать, чтобы были одинаковы и равны  $(1 - \gamma) / 2$  вероятности попадания случайной величины в области слева и справа от доверительного интервала. На рис. 4.2 это условие соответствует равенству площадей  $S_1$  и  $S_2$ .

Если для нахождения  $\tilde{\chi}_{1,\gamma,n}^2$  и  $\tilde{\chi}_{2,\gamma,n}^2$  используются таблицы квантилей распределения хи-квадрат (см., например, табл. 3 Приложения), то значения  $\tilde{\chi}_{1,\gamma,n}^2$  и  $\tilde{\chi}_{2,\gamma,n}^2$  определяются из соотношений:

$$p(H_n < \tilde{\chi}_{1,\gamma,n}^2) = \frac{1-\gamma}{2}, \quad p(H_n < \tilde{\chi}_{2,\gamma,n}^2) = \gamma + \frac{1-\gamma}{2} = \frac{1+\gamma}{2}, \quad (4.16)$$

т.е.  $\tilde{\chi}_{1,\gamma,n}^2 = \chi_{(1-\gamma)/2,n}^2$  и  $\tilde{\chi}_{2,\gamma,n}^2 = \chi_{(1+\gamma)/2,n}^2$ , где  $\chi_{\alpha,n}^2$  есть квантиль порядка  $\alpha$  распределения хи-квадрат с  $n$  степенями свободы.

В случае, когда  $\tilde{\chi}_{1,\gamma,n}^2$  и  $\tilde{\chi}_{2,\gamma,n}^2$  находятся по таблицам критических точек, следует пользоваться соотношениями:

$$p(H_n > \tilde{\chi}_{1,\gamma,n}^2) = \frac{1+\gamma}{2}, \quad p(H_n > \tilde{\chi}_{2,\gamma,n}^2) = \frac{1-\gamma}{2}. \quad (4.17)$$

По известным  $\tilde{\chi}_{1,\gamma,n}^2$  и  $\tilde{\chi}_{2,\gamma,n}^2$  легко вычислить интервал, вероятность попадания дисперсии  $D[\xi] = \sigma^2$  в который равна  $\gamma$ . Для данной выборки, где  $Y$  принимает значение выборочной дисперсии  $D^* = (\sigma^*)^2$ , из соотношения (4.15) следует, что

$$p\left(\tilde{\chi}_{1,\gamma,n}^2 < \frac{n(\sigma^*)^2}{\sigma^2} < \tilde{\chi}_{2,\gamma,n}^2\right) = p\left(\frac{n(\sigma^*)^2}{\tilde{\chi}_{2,\gamma,n}^2} < D[\xi] < \frac{n(\sigma^*)^2}{\tilde{\chi}_{1,\gamma,n}^2}\right) = \gamma, \quad (4.18)$$

т.е. доверительный интервал для дисперсии равен

$$\left( \frac{n(\sigma^*)^2}{\tilde{\chi}_{2,\gamma,n}^2}, \frac{n(\sigma^*)^2}{\tilde{\chi}_{1,\gamma,n}^2} \right).$$

Для среднего квадратического отклонения  $\sigma$  имеем

$$P\left( \sqrt{\frac{n}{\tilde{\chi}_{2,\gamma,n}^2}} \sigma^* < \sigma < \sqrt{\frac{n}{\tilde{\chi}_{1,\gamma,n}^2}} \sigma^* \right) = \gamma, \quad (4.19)$$

и доверительный интервал соответственно равен

$$\left( \sqrt{\frac{n}{\tilde{\chi}_{2,\gamma,n}^2}} \sigma^*, \sqrt{\frac{n}{\tilde{\chi}_{1,\gamma,n}^2}} \sigma^* \right).$$

### 4.3.2. Случай неизвестного математического ожидания

Наиболее распространенной является ситуация, когда неизвестны оба параметра нормального распределения: математическое ожидание  $a$  и среднее квадратическое отклонение  $\sigma$ .

В этом случае построение доверительного интервала основывается на теореме Фишера, из которой следует, что случайная величина

$$H_n = \frac{(n-1)Z}{\sigma^2},$$

где  $Z$  – случайная величина, определенная в п. 4.2 и принимающая значения несмещенной выборочной дисперсии  $s^2$ , имеет распределение хи-квадрат с  $n - 1$  степенями свободы.

Рассмотрим равенство, аналогичное (4.15):

$$P(\tilde{\chi}_{1,\gamma,n-1}^2 < H_n < \tilde{\chi}_{2,\gamma,n-1}^2) = \gamma. \quad (4.20)$$

Для данной выборки, когда величина  $Z$  принимает значение  $s^2$ , после несложных алгебраических преобразований получим

$$P\left( \frac{(n-1)s^2}{\tilde{\chi}_{2,\gamma,n-1}^2} < \sigma^2 < \frac{(n-1)s^2}{\tilde{\chi}_{1,\gamma,n-1}^2} \right) = \gamma. \quad (4.21)$$

Значения  $\tilde{\chi}_{1,\gamma,n-1}^2$  и  $\tilde{\chi}_{2,\gamma,n-1}^2$  определяются по таблицам квантилей или критических точек распределения хи-квадрат с  $n - 1$  степенью свободы, как описано в п. 4.3.1.

## 4.4. Оценивание математического ожидания случайной величины для произвольной выборки

Интервальные оценки математического ожидания  $M[\xi]$ , полученные для нормально распределенной случайной величины  $\xi$  (см. формулы (4.6) и (4.9)), являются, вообще говоря, непригодными для случайных величин, имеющих иной вид распределения. Однако есть ситуация, когда для любых случайных величин можно пользоваться соотношениями, аналогичными (4.6) и (4.9), – это случай выборок большого объема ( $n \gg 1$ ).

Как и выше, будем рассматривать варианты  $x_1, x_2, \dots, x_n$ , как значения независимых, одинаково распределенных случайных величин  $\xi_1, \xi_2, \dots, \xi_n$ , имеющих математическое ожидание  $M[\xi_i] = m_\xi$  и дисперсию  $D[\xi_i] = \sigma_\xi^2$ , а полученное выборочное среднее  $\bar{x}$  как значение случайной величины

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n \xi_i.$$

Согласно центральной предельной теореме величина  $\bar{X}$  имеет асимптотически нормальный закон распределения с математическим ожиданием  $m_\xi$  и дисперсией  $\sigma_\xi^2/n$ .

Поэтому, если известно значение дисперсии случайной величины  $\xi$ , то можно пользоваться приближенными формулами, аналогичными соотношениям (4.5):

$$\begin{cases} p \left( \bar{x} - \frac{\sigma_\xi \tilde{x}_\gamma}{\sqrt{n}} < m_\xi < \bar{x} + \frac{\sigma_\xi \tilde{x}_\gamma}{\sqrt{n}} \right) \approx 2\Phi(\tilde{x}_\gamma), \\ 2\Phi(\tilde{x}_\gamma) = \gamma, \end{cases} \quad (4.22)$$

которые будут точны только в пределе  $n \rightarrow \infty$ .

Если же значение дисперсии величины  $\xi$  неизвестно, то при больших  $n$  можно использовать формулу, аналогичную (4.13):

$$\begin{cases} p \left( \bar{x} - \frac{s \tilde{x}_\gamma}{\sqrt{n}} < m_\xi < \bar{x} + \frac{s \tilde{x}_\gamma}{\sqrt{n}} \right) \approx 2\Phi(\tilde{x}_\gamma), \\ 2\Phi(\tilde{x}_\gamma) = \gamma, \end{cases} \quad (4.23)$$

где  $s$  – исправленное среднее квадратическое отклонение (2.16). Точное равенство в (4.23) возможно, как и в (4.22), лишь в пределе  $n \rightarrow \infty$ .

#### 4.5. Решение типовых примеров

**Пример 4.1.** Найти доверительный интервал для математического ожидания  $a$  нормально распределенной случайной величины с известной дисперсией  $\sigma^2 = 4$ , если объем выборки  $n = 100$ , выборочное среднее  $\bar{x} = 3,2$ , а требуемая доверительная вероятность  $\gamma = 0,9$ .

**Решение.** Будем использовать соотношения (4.5). Из условия  $\Phi(\tilde{x}_{0,9}) = 0,9/2$

определим  $\tilde{x}_{0,9} = 1,64$ . Тогда интервал, определяемый соотношениями (4.5), равен

$$3,2 - \frac{2,0 \cdot 1,64}{\sqrt{100}} < a < 3,2 + \frac{2,0 \cdot 1,64}{\sqrt{100}}$$

или  $2,872 < a < 3,528$ .

Таким образом,  $(2,872, 3,528)$  – доверительный интервал, в который попадает математическое ожидание  $a$  с надежностью  $0,9$ .

**Пример 4.2.** Найти  $\gamma$  – надежность (доверительную вероятность) попадания математического ожидания  $a$  нормально распределенной случайной величины с известной дисперсией  $\sigma^2 = 9$  в интервал длиной  $2$  при объеме выборки  $n = 49$ .

**Решение.** Опять используем соотношения (4.5). Ширина доверительного интервала, определяемого этими соотношениями, равна  $2\sigma \tilde{x}_\gamma / \sqrt{n}$ . Поэтому имеем

$$\frac{2 \cdot 3 \cdot \tilde{x}_\gamma}{\sqrt{49}} = \frac{6}{7} \tilde{x}_\gamma = 2.$$

Следовательно, параметр  $\tilde{x}_\gamma = 2,33$  и  $\Phi(2,33) = 0,4901$ . Таким образом,  $\gamma = 2 \cdot 0,4901 = 0,98$ .

**Пример 4.3.** Пусть при выборке объемом  $n = 25$  для нормально распределенной величины получено среднее выборочное  $\bar{x} = 7$  и исправленная выборочная дисперсия  $s^2 = 9$ . Найти доверительный интервал для параметра  $a$  при надежности  $\gamma = 0,99$ , используя таблицу квантилей распределения Стьюдента, приведенную в *Приложении*.

**Решение.** Исходя из имеющейся таблицы, используем подход (4.10). Тогда надо найти квантиль порядка  $(1 + \gamma) / 2 = 0,995$  распределения Стьюдента с числом степеней свободы, равным  $k = n - 1 = 24$ . Из таблицы находим значение квантили и получаем  $\tilde{t}_{0,99,24} = 2,797$ .

Тогда доверительный интервал согласно (4.9) равен

$$7 - \frac{3 \cdot 2,797}{\sqrt{25}} < a < 7 + \frac{3 \cdot 2,797}{\sqrt{25}}.$$

Таким образом, искомый интервал:  $5,32 < a < 8,68$ .

**Пример 4.4.** Найти доверительный интервал для дисперсии  $\sigma^2$  случайной величины с надежностью  $\gamma = 0,9$ , если при объеме выборки  $n = 21$  выборочная дисперсия  $(\sigma^*)^2 = 8$ .

**Решение.** Поскольку математическое ожидание случайной величины неизвестно, для интервальной оценки дисперсии следует применять формулу

(4.21), в которую входят параметры  $\tilde{\chi}_{1,\gamma,n-1}^2$ ,  $\tilde{\chi}_{2,\gamma,n-1}^2$  и исправленная выборочная дисперсия  $s^2$ .

В табл. 3 *Приложения* приведены квантили распределения хи-квадрат  $\chi_{\alpha,n}^2$ , поэтому значения параметров  $\tilde{\chi}_{1,\gamma,n-1}^2$ ,  $\tilde{\chi}_{2,\gamma,n-1}^2$  будем искать, используя соотношения (4.16) и данную таблицу:  $\tilde{\chi}_{1,0,9,20}^2 = \chi_{0,05,20}^2 = 10,85$ ,  $\tilde{\chi}_{2,0,9,20}^2 = \chi_{0,95,20}^2 = 31,41$ .

Исправленная выборочная дисперсия равна  $s^2 = D^* n / (n - 1) = 8,4$ . Подставляем полученные результаты в формулу (4.21) и находим, что  $p(5,35 < \sigma^2 < 15,38) = 0,9$ . Таким образом, доверительный интервал для дисперсии  $\sigma^2$  при надежности  $\gamma = 0,9$  равен (5,35, 15,38).



## 5. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Изучение подходов к статистической проверке гипотез начнем с ряда определений. При этом под случайной величиной будем понимать некую числовую характеристику объектов данной генеральной совокупности.

*Статистической гипотезой* называют гипотезу о виде неизвестного распределения или о параметрах известного распределения случайной величины.

Проверяемая гипотеза, обозначаемая обычно как  $H_0$ , получила название *нулевой* или *основной* гипотезы. Дополнительно используемая гипотеза  $H_1$ , противоречащая гипотезе  $H_0$ , называется *конкурирующей* или *альтернативной*.

*Статистическая проверка* выдвинутой нулевой гипотезы  $H_0$  состоит в ее сопоставлении с выборочными данными. При такой проверке возможно появление ошибок двух видов:

- а) *ошибки первого рода* – случаи, когда отвергается правильная гипотеза  $H_0$ ;
- б) *ошибки второго рода* – случаи, когда принимается неверная гипотеза  $H_0$ .

Вероятность ошибки первого рода будем называть *уровнем значимости* и обозначать как  $\alpha$ .

Основной прием проверки статистических гипотез заключается в том, что по имеющейся выборке вычисляется значение *статистического критерия* – некоторой случайной величины  $T$ , имеющей известный закон распределения. Область значений  $T$ , при которых основная гипотеза  $H_0$  должна быть отвергнута, называют *критической*, а область значений  $T$ , при которых эту гипотезу можно принять, – *областью принятия гипотезы*.

Процесс проверки гипотезы  $H_0$  состоит в следующем:

- 1) Определяют статистический критерий  $T$  и по имеющейся выборке вычисляют его эмпирическое (наблюдаемое) значение  $T^*$ .
- 2) Выбирают уровень значимости  $\alpha$  и по известному закону распределения  $T$  вычисляют его критическое значение  $T_\alpha$ . Оно делит область возможных значений  $T$  на две части: критическую область и область принятия гипотезы (например, области  $T > T_\alpha$  и  $T \leq T_\alpha$ ).
- 3) Если значение  $T^*$  попадает в область принятия гипотезы, то гипотезу  $H_0$  можно принять, если в критическую область, то гипотезу  $H_0$  следует отвергнуть.

В некоторых ситуациях непринятие правильной гипотезы  $H_0$  (ошибка первого рода) может быть менее важным, чем принятие неправильной нулевой гипотезы (ошибка второго рода). Пусть вероятность такой ошибки равна  $\beta$ , тогда  $(1 - \beta)$  – вероятность попадания критерия  $T$  в критическую область при условии, что верна конкурирующая гипотеза  $H_1$ . Вероятность  $1 - \beta$  называется *мощностью критерия*. Чем больше мощность критерия, тем меньше вероятность совершить ошибку второго рода. Поэтому после выбора уровня значимости следует строить критическую область так, чтобы мощность критерия была максимальной.

симальной.

### Замечания:

1. Статистические методы позволяют лишь опровергнуть или не опровергнуть статистическую гипотезу, но не доказать ее.
2. Проверка статистических гипотез имеет много общего с построением доверительного интервала, обсуждавшимся в главе 4. Для статистического критерия  $T$  мы строим критическую область и ее дополнение – область принятия решения, являющуюся, по сути, доверительным интервалом для  $T$ . Если значение критерия  $T^*$ , полученное для выборки, должно было с большой вероятностью попадать в область принятия решения, но не попало в нее (т.е. попало в критическую область), то с большой вероятностью (надежностью) можно отвергнуть данную гипотезу.

## 5.1. Проверка гипотез о параметрах известного распределения

Приведем несколько примеров построения статистических критериев для проверки подобных гипотез.

### 5.1.1. Проверка гипотезы о математическом ожидании нормально распределенной случайной величины

Проверка такой гипотезы сходна с процедурой построения доверительного интервала для математического ожидания нормально распределенной случайной величины (см. п. 4.1).

Пусть случайная величина  $\xi$  имеет нормальное распределение. Требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу  $a_0$ . Рассмотрим отдельно случаи, когда дисперсия  $\xi$  известна и когда она неизвестна.

В случае известной дисперсии  $D[\xi] = \sigma^2$ , как и в п. 4.1, определим случайную величину  $\bar{X}$ , принимающую значения выборочного среднего  $\bar{x}$ . Гипотеза  $H_0$  изначально формулируется как  $M[\xi] = a_0$ . Поскольку выборочное среднее является несмещенной оценкой  $M[\xi]$ , то гипотезу  $H_0$  можно представить как

$$M[\bar{X}] = a_0.$$

Для проверки этой гипотезы выберем следующий статистический критерий:

$$U = \frac{\bar{X} - a_0}{\sigma[\bar{X}]} = \frac{(\bar{X} - a_0)\sqrt{n}}{\sigma}, \quad (5.1)$$

где учтено, что  $\sigma[\bar{X}] = \sqrt{D[\bar{X}]}$  и в нашем случае  $D[\bar{X}] = \sigma^2/n$  (см. п. 2.3).

Отметим, что если нулевая гипотеза справедлива, то при достаточно большом объеме выборки  $n$  согласно центральной предельной теореме случайная

величина  $U$  должна иметь практически нормальное распределение с  $M[U] = 0$  и  $D[U] = 1$  (см. п. 4.1).

Конфигурация критической области определяется видом конкурирующей гипотезы  $H_1$ :

- 1) Если гипотеза  $H_1$  заключается в том, что  $M[\bar{X}] > a_0$ , то критическая область является *правосторонней*. На рис. 5.1 подобная критическая область заштрихована.

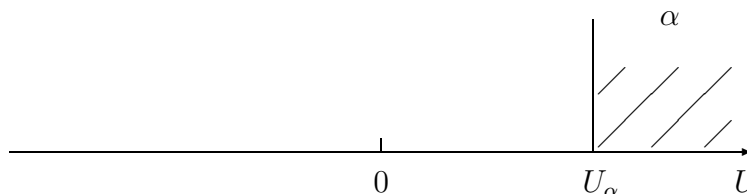


Рис. 5.1. Правосторонняя критическая область

Поскольку можно считать, что случайная величина  $U$  распределена нормально, то критическое значение  $U_\alpha$  можно найти из условия  $p(U > U_\alpha) = \alpha$ . Имеем  $p(U > U_\alpha) = 1 - F_U(U_\alpha) = 1 - (\Phi(U_\alpha) + 1/2) = 1/2 - \Phi(U_\alpha)$ , где  $\Phi(x)$  – функция Лапласа, и получаем

$$\Phi(U_\alpha) = \frac{1}{2} - \alpha, \quad (5.2)$$

т.е.  $U_\alpha$  равно квантили нормального распределения  $x_{1-\alpha}$ .

Остается по выборке найти значение

$$U^* = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$$

и сравнить его с  $U_\alpha$ : если  $U^* < U_\alpha$ , то нулевую гипотезу  $H_0$  можно принять, а если  $U^* > U_\alpha$ , то ее следует отвергнуть.

- 2) Если гипотеза  $H_1$  состоит в том, что  $M[\bar{X}] < a_0$ , то критическая область является *левосторонней* (см. рис. 5.2) и  $U_\alpha$  определяется из условия  $p(U < -U_\alpha) = \alpha$ . Поскольку  $p(U < -U_\alpha) = F_U(-U_\alpha) = \Phi(-U_\alpha) + 1/2 = 1/2 - \Phi(U_\alpha)$ , то опять получаем

$$\Phi(U_\alpha) = \frac{1}{2} - \alpha. \quad (5.3)$$

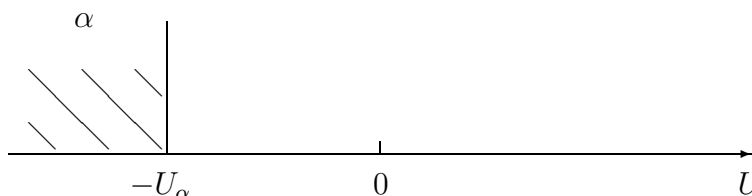


Рис. 5.2. Левосторонняя критическая область

Здесь нулевую гипотезу  $H_0$  можно принять, если  $U^* > -U_\alpha$ , и следует отвергнуть, если  $U^* < -U_\alpha$ .

- 3) Если гипотеза  $H_1$  состоит в том, что  $M[\bar{X}] \neq a_0$ , то критическая область удовлетворяет условию  $p(|U| > U_\alpha) = \alpha$  и является *двухсторонней*, поскольку состоит из двух частей:  $U < -U_\alpha$  и  $U > U_\alpha$  (см. рис. 5.3).

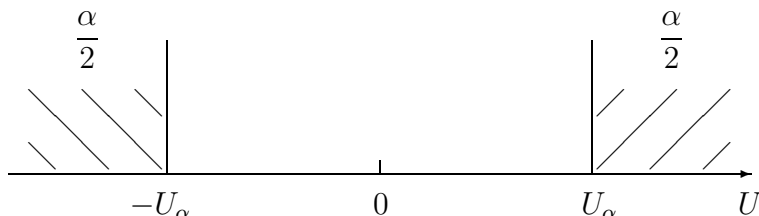


Рис. 5.3. Двухсторонняя критическая область

Вероятность попадания критерия  $U$  в каждую из половин критической области равна  $\alpha/2$ . Поэтому  $U_\alpha$  определяется из соотношения  $p(U > U_\alpha) = \alpha/2$ . Поскольку  $p(U > U_\alpha) = 1/2 - \Phi(U_\alpha)$ , то получаем

$$\Phi(U_\alpha) = \frac{1 - \alpha}{2}. \quad (5.4)$$

Здесь нулевую гипотезу  $H_0$  можно принять, если  $|U^*| < U_\alpha$ , и следует отвергнуть, если  $|U^*| > U_\alpha$ .

В случае неизвестной дисперсии величины  $\xi$  в качестве статистического критерия выберем

$$T = \frac{(\bar{X} - a_0)}{\sqrt{Z/n}}, \quad (5.5)$$

где случайные величины  $\bar{X}$  и  $Z$ , определенные в п. 4.2, принимают данной выборки значения соответственно выборочного среднего  $\bar{x}$  и исправленной (несмещенной) выборочной дисперсии  $s^2$ .

Случайная величина  $T$  имеет распределение Стьюдента с  $k = n - 1$  степенями свободы (см. подробнее п. 6.2).

Если конкурирующая (альтернативная) гипотеза  $H_1$  базируется на неравенстве  $M[X] > a_0$ , то критическая область является правосторонней (см. п. 4.2) и согласно формуле (4.10) критическое значение равно  $T_\alpha = \tilde{t}_{1-\alpha, n-1} = t_{1-\alpha/2, n-1}$ , где  $t_{1-\alpha/2, n-1}$  – квантиль порядка  $(1 - \alpha/2)$  распределения Стьюдента с  $n - 1$  степенями свободы, которую можно найти по соответствующим таблицам.

Эмпирическое значение критерия для данной выборки равно

$$T^* = \frac{(\bar{x} - a_0)\sqrt{n}}{s}. \quad (5.6)$$

Нулевую гипотезу  $H_0$  можно принять, если  $|T^*| < T_\alpha$ , и следует отвергнуть, если  $|T^*| > T_\alpha$ .

Аналогичным образом рассматриваются и остальные варианты выбора гипотезы  $H_1$  (см. п. 4.2).

### 5.1.2. Сравнение дисперсий нормально распределенных случайных величин

Пусть имеются две нормально распределенные случайные величины  $\xi_1$  и  $\xi_2$ . Для них по независимым выборкам объемом  $n_1$  и  $n_2$  соответственно получены исправленные выборочные дисперсии  $s_{\xi_1}^2$  и  $s_{\xi_2}^2$ . Будем считать, что  $s_{\xi_1}^2 > s_{\xi_2}^2$ . Требуется при заданном уровне значимости  $\alpha$  проверить нулевую гипотезу  $H_0$  о равенстве дисперсий рассматриваемых случайных величин.

Учитывая несмещенность исправленных выборочных дисперсий, нулевую гипотезу можно записать следующим образом:

$$M[Z_{\xi_1}] = M[Z_{\xi_2}], \quad (5.7)$$

где случайная величина

$$Z_{\xi} = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - M[\xi])^2$$

принимает значения исправленной выборочной дисперсии величины  $\xi$  и аналогична случайной величине  $Z$ , рассмотренной в п. 4.2.

В качестве статистического критерия примем случайную величину

$$F = \frac{Z_{\xi_1}}{Z_{\xi_2}}, \quad (5.8)$$

принимавшую значение отношения большей выборочной дисперсии к меньшей.

Случайная величина  $F$  имеет распределение Фишера–Снедекора (см. п. 6.4) с числом степеней свободы  $k_1 = n_1 - 1$  и  $k_2 = n_2 - 1$ , где  $n_1$  – объем выборки, по которой вычислена большая исправленная дисперсия  $s_{\xi_1}^2$ , а  $n_2$  – объем второй выборки, по которой найдена меньшая дисперсия  $s_{\xi_2}^2$ .

Рассмотрим два вида конкурирующих гипотез  $H_1$ :

1) гипотеза  $H_1$ :  $s_{\xi_1}^2 > s_{\xi_2}^2$

В этом случае строят одностороннюю (а именно правостороннюю) критическую область, исходя из условия

$$p(F > F(\alpha, k_1, k_2)) = \alpha.$$

Критическую точку  $F(\alpha, k_1, k_2)$  находят по таблице значений таких точек для распределения Фишера–Снедекора с числом степеней свободы  $k_1$  и  $k_2$  при уровне значимости  $\alpha$ .

Затем для данных выборок вычисляют эмпирическое значение критерия

$$F^* = \frac{s_{\xi_1}^2}{s_{\xi_2}^2}$$

и сравнивают его с  $F(\alpha, k_1, k_2)$ . Если  $F^* < F(\alpha, k_1, k_2)$ , то нулевая гипотеза принимается, если  $F^* > F(\alpha, k_1, k_2)$ , то отвергается.

2) гипотеза  $H_1: s_{\xi_1}^2 \neq s_{\xi_2}^2$

В этом случае строят двустороннюю критическую область из условия, что

$$p((F < F_1) \cup (F > F_2)) = \alpha.$$

Оказывается, что наибольшая мощность (вероятность попадания критерия в критическую область при справедливости конкурирующей гипотезы) достигается, когда вероятность попадания критерия в каждый из двух интервалов критической области равна  $\alpha/2$ :

$$p(F < F_1) = \frac{\alpha}{2}; \quad p(F > F_2) = \frac{\alpha}{2}.$$

Критическая точка  $F_2 = F(\alpha/2, k_1, k_2)$  находится по таблице критических точек для распределения Фишера–Снедекора с числом степеней свободы  $k_1$  и  $k_2$  при уровне значимости  $\alpha/2$ . Критическую точку  $F_1$  можно и не отыскивать. Действительно, если вероятность попадания критерия в “правую часть” критической области равна  $\alpha/2$ , то и вероятность попадания в “левую часть” также равна  $\alpha/2$ . Так как эти события несовместны, то вероятность попадания рассматриваемого критерия во всю двустороннюю критическую область будет равна

$$\frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.$$

Таким образом, если  $F^* < F(\alpha/2, k_1, k_2)$ , то нулевая гипотеза принимается, а при  $F^* > F(\alpha/2, k_1, k_2)$  – отвергается.

**?Замечание.** В случае  $\xi_1 = \xi_2 = \xi$  имеет место сравнение выборочных дисперсий одной случайной величины  $\xi$ . Такая задача возникает, когда требуется сопоставить различные методы сбора статистической информации. Очевидно, что предпочтительнее тот метод, который дает наименьший разброс данных, т.е. имеет наименьшую дисперсию.

Кроме того, при статистической обработке нескольких выборок часто бывает необходимо проверить насколько значимо (существенно) различаются для них выборочные дисперсии.

### 5.1.3. Сравнение математических ожиданий независимых случайных величин

Сначала рассмотрим случай нормально распределенных случайных величин с известными дисперсиями, а затем на его основе более общий случай произвольно распределенных величин при достаточно больших независимых выборках.

Пусть случайные величины  $\xi_1$  и  $\xi_2$  независимы и распределены нормально, и пусть их дисперсии  $D[\xi_1]$  и  $D[\xi_2]$  известны. Например, они могут быть найдены из какого-то другого опыта или рассчитаны теоретически. Извлечены выборки объемом  $n_1$  и  $n_2$  соответственно. Пусть  $\bar{x}_1$  и  $\bar{x}_2$  – выборочные средние для этих выборок.

Требуется по выборочным средним при заданном уровне значимости  $\alpha$  проверить гипотезу о равенстве математических ожиданий рассматриваемых случайных величин

$$M[\xi_1] = M[\xi_2].$$

Введем случайные величины  $\bar{X}_1$  и  $\bar{X}_2$ , принимающие значения выборочных средних  $\bar{x}_1$  и  $\bar{x}_2$  соответственно. Поскольку выборочные средние – это несмещенные оценки математических ожиданий, то нулевую гипотезу  $H_0$  можно записать в следующем виде:

$$M[\bar{X}_1] = M[\bar{X}_2]. \quad (5.9)$$

В качестве статистического критерия для проверки  $H_0$  возьмем случайную величину

$$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{D[\xi_1]}{n_1} + \frac{D[\xi_2]}{n_2}}}. \quad (5.10)$$

Можно показать, что при  $n_1, n_2 \rightarrow \infty$  сумма  $(D[\xi_1] / n_1 + D[\xi_2] / n_2)$  стремится к  $D[\bar{X}_1 - \bar{X}_2]$ . Тогда, если нулевая гипотеза (5.9) справедлива, то согласно центральной предельной теореме с ростом объемов выборок распределение величины  $U$  стремится к нормальному с нулевым математическим ожиданием и единичной дисперсией.

Критическую область строят в зависимости от конкурирующей гипотезы:

1) гипотеза  $H_1: M[\xi_1] \neq M[\xi_2]$

В этом случае строят двустороннюю критическую область. Как и в разделе 5.1.1, критическая точка  $U_\alpha$  находится из условия

$$\Phi(U_\alpha) = \frac{1 - \alpha}{2}. \quad (5.11)$$

Затем по имеющимся выборкам вычисляют эмпирическое значение критерия

$$U^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{D[\xi_1]}{n_1} + \frac{D[\xi_2]}{n_2}}}$$

и сравнивают его с теоретическим значением  $U_\alpha$ :

- а) если  $|U^*| < U_\alpha$ , то нет оснований отвергать нулевую гипотезу;
- б) если  $|U^*| > U_\alpha$ , то нулевую гипотезу отвергают.

2) гипотеза  $H_1: M[\xi_1] > M[\xi_2]$

На практике такая ситуация возникает тогда, когда из априорных соображений можно ожидать, что математическое ожидание одной случайной величины должно быть больше, чем математическое ожидание другой.

В этом случае строят правостороннюю критическую область. Критическая точка  $U_\alpha$  определяется из условия, аналогичного соотношению (5.2),

$$\Phi(U_\alpha) = \frac{1}{2} - \alpha. \quad (5.12)$$

Вывод делается по результатам сравнения  $U^*$  и  $U_\alpha$ :

- а) если  $U^* < U_\alpha$ , то нулевую гипотезу можно принять;
- б) если  $U^* > U_\alpha$ , то ее следует отвергнуть.

3) гипотеза  $H_1: M[\xi_1] < M[\xi_2]$

В этом случае строят левостороннюю критическую область. Критическая точка  $U_\alpha$  определяется, как и в п. 5.1.1, из условия

$$\Phi(U_\alpha) = \frac{1}{2} - \alpha. \quad (5.13)$$

Нулевую гипотезу  $H_0$  можно принять, если  $U^* > -U_\alpha$ , и следует отвергнуть, если  $U^* < -U_\alpha$ .

Рассмотрим теперь сравнение математических ожиданий двух произвольно распределенных случайных величин  $\xi_1$  и  $\xi_2$  при больших ( $n > 30$ ) независимых выборках.

В этом случае в силу центральной предельной теоремы выборочные средние распределены приближенно по нормальному закону вне зависимости от того, как распределены сами случайные величины. Поскольку выборочные дисперсии являются при этом достаточно хорошими оценками дисперсий случайных величин, последние можно считать приближенно известными, т.е.

$$D[\xi_1] \approx D^*[\xi_1], \quad D[\xi_2] \approx D^*[\xi_2].$$

Поэтому можно полагать, что статистический критерий

$$\tilde{U} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{D^*[\xi_1]}{n_1} + \frac{D^*[\xi_2]}{n_2}}} \quad (5.14)$$

имеет распределение близкое к нормальному, причем:



- а) если справедлива нулевая гипотеза  $H_0: M[\xi_1] = M[\xi_2]$ , то  $M[\tilde{U}] \approx 0$ ;  
 б) если рассматриваемые выборки независимы, то  $D[\tilde{U}] \approx 1$ .

Дальнейшая проверка статистических гипотез проводится как выше для критерия  $U$ .

## 5.2. Проверка гипотез о виде закона распределения случайной величины. Критерий Пирсона

Надежное предположение о распределении случайной величины, связанной с генеральной совокупностью, можно иногда сделать из априорных соображений, основываясь на условиях эксперимента, и тогда предположения о параметрах распределения исследуются, как показано в п. 5.1. Однако весьма часто возникает необходимость проверить выдвинутую гипотезу о законе распределения. Статистические критерии, предназначенные для таких проверок, обычно называются *критериями согласия*.

Известно несколько критериев согласия. Достоинством критерия Пирсона является его универсальность. С его помощью можно проверять гипотезы о различных законах распределения.

*Критерий Пирсона* основан на сравнении частот, найденных по выборке (будем называть их эмпирическими частотами), с частотами, рассчитанными с помощью проверяемого закона распределения (теоретическими частотами).

Обычно эмпирические и теоретические частоты различаются. Следует выяснить, случайно ли расхождение частот или оно значимо и объясняется тем, что теоретические частоты вычислены, исходя из неверной гипотезы о распределении генеральной совокупности.

Критерий Пирсона, как и любой другой критерий, отвечает на вопрос, есть ли согласие выдвинутой гипотезы с эмпирическими данными при заданном уровне значимости.

### 5.2.1. Проверка гипотезы о нормальном распределении

Пусть имеется случайная величина  $\xi$  и сделана выборка достаточно большого объема  $n$  с большим количеством различных значений вариантов. Требуется при уровне значимости  $\alpha$  проверить нулевую гипотезу  $H_0$  о том, что случайная величина  $\xi$  распределена нормально.

Для удобства обработки выборки возьмем два числа  $\alpha, \beta: \alpha < x_1, \beta > x_n$ , и разделим интервал  $[\alpha, \beta]$  на  $s$  подинтервалов. Будем считать, что значения вариантов, попавших в каждый подинтервал, приближенно равны числу, задающему середину подинтервала. Подсчитав число вариантов, попавших в каждый интервал, составим группированную выборку:

варианты	$x_1$	$x_2$	...	$x_s$
----------	-------	-------	-----	-------

частоты	$n_1$	$n_2$	...	$n_s$
---------	-------	-------	-----	-------

где  $x_j = (b_j + a_j)/2$  – середина  $j$ -го подинтервала  $(a_j, b_j]$ ,  $n_j$  – количество вариантов, попавших в этот подинтервал, т.е. эмпирическая частота.

По полученным данным можно вычислить выборочное среднее  $\bar{x}$

$$\bar{x} = \frac{1}{n} \sum_{j=1}^s n_j x_j$$

и выборочное среднее квадратическое отклонение  $\sigma^*$

$$\sigma^* = \sqrt{D^*} = \sqrt{\frac{1}{n} \sum_{j=1}^s n_j (x_j - \bar{x})^2}.$$

Найдем количество вариантов, которое должно оказаться в каждом подинтервале при выборке объемом  $n$  в предположении, что случайная величина  $\xi$  распределена по нормальному закону с параметрами  $M[\xi] = \bar{x}$ ,  $D[\xi] = (\sigma^*)^2$ .

Вероятность того, что значение такой величины попадет в  $j$ -й подинтервал, равна

$$p_j = \Phi\left(\frac{b_j - \bar{x}}{\sigma^*}\right) - \Phi\left(\frac{a_j - \bar{x}}{\sigma^*}\right), \quad (5.15)$$

где  $a_j$  и  $b_j$  – границы подинтервала,  $\Phi(x)$  – функция Лапласа. Умножив вероятности  $p_j$  на объем выборки  $n$ , найдем теоретические частоты  $n'_j = n \cdot p_j$ .

Выберем статистический критерий, равный

$$H = \sum_{j=1}^s \frac{(\eta_j - n'_j)^2}{n'_j}, \quad (5.16)$$

где  $\eta_j$  – случайная величина, принимающая для данной выборки значение частоты  $n_j$ .

Можно показать, что вне зависимости от реального закона распределения случайной величины  $\xi$ , распределение  $H$  при  $n \rightarrow \infty$  стремится к распределению хи-квадрат с числом степеней свободы  $k = s - 1 - r$ , где  $r$  – число параметров предполагаемого распределения, оцениваемых по выборке. В случае нормального распределения, характеризуемого двумя параметрами,  $k = s - 3$ .

Для выбранного критерия построим правостороннюю критическую область, определяемую условием

$$p(H > \chi^2(\alpha, k)) = \alpha,$$

где  $\alpha$  – уровень значимости. Критическая точка  $\chi^2(\alpha, k) = \chi_{1-\alpha, k}^2$ , где  $\chi_{1-\alpha, k}^2$  – квантиль порядка  $(1 - \alpha)$  распределения хи-квадрат с числом степеней свободы  $k = s - 3$ , которая определяется из соответствующих таблиц.

Значение критерия  $H^*$  для данной выборки равно

$$H^* = \sum_{j=1}^s \frac{(n_j - n'_j)^2}{n'_j}. \quad (5.17)$$

Если  $H^* < \chi^2(\alpha, k)$ , то нулевую гипотезу  $H_0$  о том, что случайная величина  $\xi$  распределена нормально, можно принять с уровнем значимости  $\alpha$ . Если  $H^* > \chi^2(\alpha, k)$ , то гипотезу  $H_0$  необходимо отвергнуть.

**Замечание.** Объем выборки  $n$  должен быть достаточно большим, во всяком случае, не менее 50. Каждый подинтервал должен содержать не менее 5–8 вариант. Если в подинтервале слишком мало точек, то его следует объединить с соседним.

### 5.2.2. Проверка гипотезы о равномерном распределении

Пусть имеется случайная величина  $\xi$  и сделана выборка достаточно большого объема  $n$ . Требуется при уровне значимости  $\alpha$  проверить нулевую гипотезу  $H_0$  о том, что случайная величина  $\xi$  имеет равномерное распределение, т.е.

$$f_\xi(x) = \begin{cases} \frac{1}{b-a}, & \text{если } x \in [a, b]; \\ 0, & \text{если } x \in (-\infty, a) \cup (b, \infty), \end{cases} \quad (5.18)$$

где  $a, b$  – некоторые параметры.

Для проверки гипотезы используем статистический критерий  $H$ , определенный формулой (5.16).

Оценим параметры распределения  $a$  и  $b$ , применяя метод моментов (см. п. 3.1). Получим (см. пример 3.1)

$$a^* = \bar{x} - \sqrt{3}\sigma^*, \quad b^* = \bar{x} + \sqrt{3}\sigma^*, \quad (5.19)$$

где  $\bar{x}$  – выборочное среднее, а  $\sigma^*$  – выборочное среднее квадратическое отклонение. Полагая  $a = a^*, b = b^*$ , найдем теоретические частоты:

$$\begin{aligned} n'_1 &= np_1 = n \frac{b_1 - a^*}{b^* - a^*}; \\ n'_j &= np_j = n \frac{b_j - b_{j-1}}{b^* - a^*}, \quad j = 2, 3, \dots, s-1; \\ n'_s &= np_s = n \frac{b^* - b_{s-1}}{b^* - a^*}, \end{aligned} \quad (5.20)$$

где  $b_j$  – правая граница  $j$ -го подинтервала. Используя эти частоты, получим эмпирическое значение критерия  $H^*$  по формуле (5.17).

Критическая область значений критерия  $H$  ограничена критической точкой  $\chi^2(\alpha, k) = \chi^2_{1-\alpha, k}$ , где  $\chi^2_{1-\alpha, k}$  – квантиль порядка  $(1 - \alpha)$  распределения хи-квадрат

с числом степеней свободы  $k = s - 3$ , которая определяется из соответствующих таблиц.

Если  $H^* < \chi^2(\alpha, k)$ , то нулевую гипотезу  $H_0$  о том, что случайная величина  $\xi$  имеет равномерное распределение, можно принять с уровнем значимости  $\alpha$ . Если  $H^* > \chi^2(\alpha, k)$ , то гипотезу  $H_0$  следует отвергнуть.

**Замечание.** На практике может оказаться, что  $b_1 < a^*$  или  $b_{s-1} > b^*$ . Такая ситуация свидетельствует не в пользу гипотезы о равномерном распределении исследуемой случайной величины.

### 5.3. Решение типовых примеров

**Пример 5.1.** Пусть  $\xi$  – нормально распределенная случайная величина с дисперсией  $\sigma^2 = 9$ . По выборке объемом  $n = 25$  определено выборочное среднее  $\bar{x} = 4,7$ . Проверить при уровне значимости  $\alpha = 0,1$  гипотезу о том, что математическое ожидание  $M[\xi]$  равно  $a_0 = 5$ , используя разные виды конкурирующих гипотез.

**Решение.** Нулевая гипотеза  $H_0$  состоит в том, что  $M[\xi] = a_0$ . Для ее проверки, как и п. 5.1.1, выберем критерий  $U$ , заданный соотношением (5.1). При определении критической области рассмотрим все возможные альтернативные гипотезы  $H_1$ :

1) Конкурирующая гипотеза  $H_1: M[\xi] \neq a_0$ . В таком случае согласно п. 5.1.1 границы критической области задаются величиной  $U_\alpha$ , определяемой для  $\alpha = 0,1$  по формуле (5.4):

$$\Phi(U_{0,1}) = \frac{1 - 0,1}{2} = 0,45,$$

что дает  $U_{0,1} = 1,65$ .

Находим значение критерия для данной выборки

$$U^* = \frac{(\bar{x} - a_0)\sqrt{n}}{\sigma} = \frac{(4,7 - 5)\sqrt{25}}{3} = -0,5.$$

Поскольку  $|U^*| < U_{0,1}$ , нулевую гипотезу можно принять.

2) Конкурирующая гипотеза  $H_1: M[\xi] > a_0$ . Тогда граница критической области задается величиной  $U_\alpha$ , определяемой для  $\alpha = 0,1$  по формуле (5.2):

$$\Phi(U_{0,1}) = \frac{1}{2} - 0,1 = 0,40,$$

что дает  $U_{0,1} = 1,28$ .

Поскольку  $U^* < U_{0,1}$ , то нет оснований отвергать нулевую гипотезу.

3) Конкурирующая гипотеза  $H_1: M[\xi] < a_0$ . Здесь граница критической области  $U_\alpha$  определяется для  $\alpha = 0,1$  по формуле (5.3):

$$\Phi(U_{0,1}) = \frac{1}{2} - 0,1 = 0,40,$$

которая дает  $U_{0,1} = 1,28$ .

Нулевую гипотезу можно принять, т.к.  $U^* > -U_{0,1}$ .

**Пример 5.2.** Две независимые выборки объемом  $n_1 = 13$  и  $n_2 = 18$  дали для нормально распределенных случайных величин  $\xi_1$  и  $\xi_2$  соответственно следующие исправленные выборочные дисперсии:  $s_{\xi_1}^2 = 1,52$  и  $s_{\xi_2}^2 = 0,60$ .

Проверить при уровне значимости  $\alpha = 0,1$  нулевую гипотезу о равенстве дисперсий величин  $\xi_1$  и  $\xi_2$  при конкурирующей гипотезе  $D[\xi_1] \neq D[\xi_2]$ .

**Решение.** Следуя п. 5.1.2, выберем в данном случае критерий  $F$ , заданный формулой (5.8) и равный отношению большей выборочной дисперсии к меньшей. Поскольку исправленная выборочная дисперсия для  $\xi_1$  больше, чем для  $\xi_2$ , то порядок величин менять не надо.

Критическая область при конкурирующей гипотезе  $H_1: D[\xi_1] \neq D[\xi_2]$  определяется критической точкой  $F(\alpha/2, k_1, k_2)$ , где число степеней свободы равно  $k_1 = n_1 - 1$  и  $k_2 = n_2 - 1$ . В нашем случае  $\alpha/2 = 0,05$ ,  $k_1 = 12$  и  $k_2 = 17$ , и требуемую критическую точку распределения Фишера–Снедекора можно найти по табл. 4а *Приложения*:  $F(0,05, 12, 17) = 2,38$ .

Эмпирическое значение критерия  $F^*$  есть отношение большей исправленной выборочной дисперсии к меньшей:

$$F^* = \frac{s_{\xi_1}^2}{s_{\xi_2}^2} = \frac{1,52}{0,60} = 2,53.$$

Поскольку получаем  $F^* > F(0,05, 12, 17)$ , то нулевую гипотезу о равенстве дисперсий данных случайных величин необходимо отвергнуть, т.е. исправленные выборочные дисперсии различаются значимо.

**Пример 5.3.** Для двух нормально распределенных случайных величин  $\xi_1$  и  $\xi_2$  с известными дисперсиями ( $D[\xi_1] = 20$  и  $D[\xi_2] = 18$ ) независимые выборки объемом  $n_1 = 15$  и  $n_2 = 20$  дали соответственно следующие выборочные средние:  $\bar{x}_1 = 15,0$  и  $\bar{x}_2 = 13,8$ . Проверить при уровне значимости  $\alpha = 0,05$  нулевую гипотезу о равенстве математических ожиданий данных случайных величин при конкурирующей гипотезе  $M[\xi_1] > M[\xi_2]$ .

**Решение.** Используем описанный в п. 5.1.3 критерий  $U$ , определенный формулой (5.10). Заданная конкурирующая гипотеза требует рассмотрения правосторонней критической области, ограниченной значением  $U_\alpha$ , получаемым из соотношения (5.12). При  $\alpha = 0,05$  это соотношение приобретает вид:

$$\Phi(U_{0,05}) = \frac{1}{2} - 0,05 = 0,45,$$

где  $\Phi(x)$  – функция Лапласа. Используя табл. 1 *Приложения*, находим  $U_{0,05} = 1,64$ .

Эмпирическое значение критерия для данных выборок равно

$$U^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{D[\xi_1]}{n_1} + \frac{D[\xi_2]}{n_2}}} = \frac{15,0 - 13,8}{\sqrt{\frac{20}{15} + \frac{18}{20}}} = \frac{1,2}{\sqrt{2,23}} = \frac{1,2}{1,49} = 0,80.$$

Поскольку  $U^* < U_\alpha$ , то нет оснований отвергать нулевую гипотезу, т.е. выборочные средние  $\bar{x}_1 = 15,0$  и  $\bar{x}_2 = 13,8$  различаются незначимо.

**Пример 5.4.** Для двух случайных величин  $\xi_1$  и  $\xi_2$  с известными дисперсиями ( $D[\xi_1] = 140$  и  $D[\xi_2] = 100$ ) независимые выборки объемом  $n_1 = 70$  и  $n_2 = 50$  дали соответственно следующие выборочные средние:  $\bar{x}_1 = 150$  и  $\bar{x}_2 = 175$ . Проверить при уровне значимости  $\alpha = 0,01$  нулевую гипотезу о равенстве математических ожиданий данных случайных величин при конкурирующей гипотезе  $M[\xi_1] \neq M[\xi_2]$ .

**Решение.** Выборки независимы и достаточно велики, поэтому можно использовать критерий  $\tilde{U}$ , определенный формулой (5.14). Критическая область для этого критерия строится, как и для критерия  $U$ , рассмотренного в п. 5.1.3. Заданная конкурирующая гипотеза требует рассмотрения двухсторонней критической области, связанной со значением  $U_\alpha$ , определяемым из соотношения (5.11). При  $\alpha = 0,01$  это соотношение приобретает вид:

$$\Phi(\tilde{U}_{0,01}) = \frac{1 - 0,01}{2} = 0,495,$$

где  $\Phi(x)$  – функция Лапласа. Используя табл. 1 *Приложения*, находим  $\tilde{U}_{0,01} = 2,58$ .

Эмпирическое значение критерия для данных выборок равно

$$\tilde{U}^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{D[\xi_1]}{n_1} + \frac{D[\xi_2]}{n_2}}} = \frac{150 - 175}{\sqrt{\frac{140}{70} + \frac{100}{50}}} = -\frac{25}{2} = -12,5.$$

Поскольку  $|\tilde{U}^*| > U_\alpha$ , то нулевую гипотезу отвергаем, т.е. найденные выборочные средние  $\bar{x}_1 = 150$  и  $\bar{x}_2 = 175$  различаются значимо.

**Замечание.** Примеры применения критерия Пирсона для проверки гипотез о нормальном и равномерном распределении случайной величины рассмотрены в главе 7 (см. соответственно примеры 7.4 и 7.5).

## 6. ВАЖНЕЙШИЕ РАСПРЕДЕЛЕНИЯ И ИХ КВАНТИЛИ

### 6.1. Нормальное распределение

По определению нормально распределенная случайная величина  $\xi$  имеет плотность распределения вероятностей

$$f_{\xi}(x) = f_N(x, a, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad (6.1)$$

где  $a$  и  $\sigma$  являются параметрами. Свойства таких случайных величин обсуждались в п. 7.5.5 части I.

*Квантилью* порядка  $\alpha$  ( $0 < \alpha < 1$ ) непрерывной случайной величины  $\xi$  называется такое число  $x_{\alpha}$ , для которого выполняется равенство

$$P(\xi < x_{\alpha}) = \alpha.$$

Квантиль  $x_{1/2}$  называется *медианой* случайной величины  $\xi$ , квантили  $x_{1/4}$  и  $x_{3/4}$  – ее *квартилями*, а  $x_{0,1}, x_{0,2}, \dots, x_{0,9}$  – *децилями*.

Для стандартного нормального распределения  $a = 0$ ,  $\sigma = 1$  и, следовательно,

$$P(\xi < x_{\alpha}) = F_N(x_{\alpha}, 0, 1) = \frac{1}{2} + \Phi(x_{\alpha}),$$

где  $F_N(x, a, \sigma)$  – функция распределения нормально распределенной случайной величины (см. п. 7.5.5 части I), а  $\Phi(x)$  – функция Лапласа. Последняя функция не выражается через элементарные функции, и ее значения табулируются. Поскольку  $\Phi(-x) = -\Phi(x)$  и при  $x > 3$  имеем  $\Phi(x) \approx 0,5$ , то таблицы обычно составляются только для  $0 \leq x \leq 3$ .

Квантиль стандартного нормального распределения  $x_{\alpha}$  для заданного  $\alpha$  можно найти из соотношения

$$\Phi(x_{\alpha}) = \alpha - \frac{1}{2}, \quad (6.2)$$

используя таблицы значений функции Лапласа (например, табл. 1 *Приложения*).

### 6.2. Распределение Стьюдента

Если  $\xi_0, \xi_1, \xi_2, \dots, \xi_n$  – независимые случайные величины, имеющие нормальное распределение с нулевым математическим ожиданием и единичной дисперсией, то распределение случайной величины

$$t_n = \frac{\xi_0}{\sqrt{\frac{(\xi_1^2 + \dots + \xi_n^2)}{n}}} \quad (6.3)$$

называют *распределением Стьюдента* с  $n$  степенями свободы. Впервые это распределение применил английский математик Уильям Госсет (W.S. Gosset), использовавший псевдонимом Student.

Плотность распределения вероятностей  $t_n$  равна (см. рис. 6.1)

$$f_t(x, n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad (6.4)$$

где  $\Gamma(x)$  – гамма-функция, имеющая следующее определение и значения:

$$\Gamma(x) = \int_0^{\infty} z^{x-1} \exp(-z) dz; \quad \Gamma(n+1) = 1 \cdot 2 \cdot 3 \cdots n = n! \quad (6.5)$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}; \quad \Gamma\left(n + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \Gamma\left(\frac{1}{2}\right).$$

Однопараметрическая функция  $f_t(x, n)$  является четной:  $f_t(-x, n) = f_t(x, n)$ .

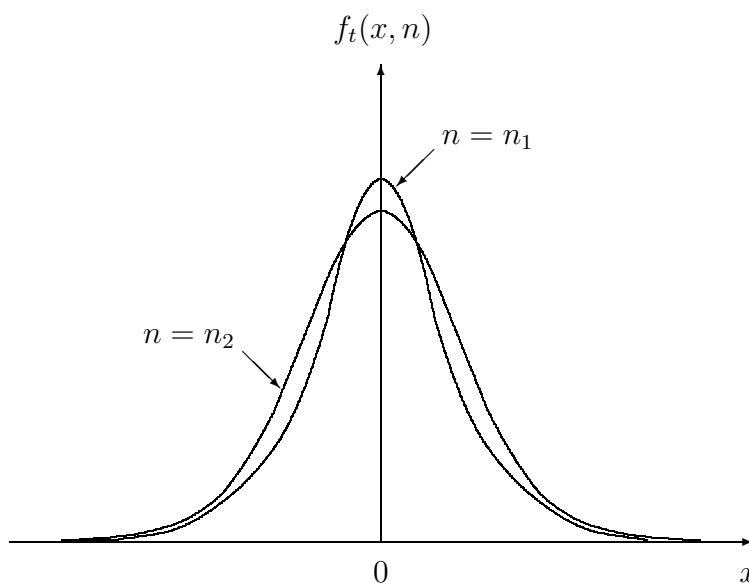


Рис. 6.1. Плотности распределения Стьюдента при  $n_1 > n_2$

Случайная величина  $t_n$  имеет моменты только при  $k < n$ , причем начальные моменты равны

$$\alpha_k[t_n] = M[t_n^k] = \begin{cases} 0, & \text{если } k \text{ нечетно;} \\ \frac{\Gamma\left(\frac{k+1}{2}\right) \Gamma\left(\frac{n-k}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{n}{2}\right)} \sqrt{n^k}, & \text{если } k \text{ четно.} \end{cases}$$



Математическое ожидание и дисперсия равны соответственно:  $M[t_n] = 0$  и  $D[t_n] = \frac{n}{n-2}$ , если  $n > 2$ .

При  $n \rightarrow \infty$  распределение Стьюдента стремится к нормальному с  $f_N(x, 0, 1)$ , т.е.

$$\lim_{n \rightarrow \infty} f_t(x, n) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

По определению квантиль порядка  $\alpha$  распределения Стьюдента с  $n$  степенями свободы есть такое число  $t_{\alpha, n}$ , что вероятность  $p(t_n < t_{\alpha, n}) = \alpha$  или эквивалентно

$$\int_{-\infty}^{t_{\alpha, n}} f_t(x, n) dx = \int_{-\infty}^{t_{\alpha, n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = \alpha.$$

Очевидно, что вследствие симметрии распределения Стьюдента  $t_{1-\alpha, n} = -t_{\alpha, n}$ .

При больших  $n$  ( $n \geq 30$ ) выполняется приближенное равенство  $t_{1-\alpha, n} \approx x_\alpha$ , где  $x_\alpha$  – квантиль порядка  $\alpha$  нормального распределения (см. п. 6.1). Более точное выражение для  $t_{\alpha, n}$  при больших  $n$  дает формула

$$t_{\alpha, n} = \frac{x_\alpha}{\sqrt{\left(1 - \frac{1}{4n}\right)^2 - \frac{x_\alpha^2}{2n}}}. \quad (6.6)$$

Критическая точка с уровнем значимости  $\alpha$  для односторонней критической области равна квантили  $t_{1-\alpha, n}$ , а для двусторонней критической области – квантили  $t_{1-\alpha/2, n}$ .

Квантили  $t_{\alpha, n}$  для распределения Стьюдента с  $n$  степенями свободы приведены в табл. 2 Приложения.

### 6.3. Распределение хи-квадрат

Если  $\xi_1, \xi_2, \dots, \xi_n$  – независимые случайные величины, имеющие нормальное распределение с нулевым математическим ожиданием и единичной дисперсией, то распределение случайной величины

$$H_n = \sum_{j=1}^n \xi_j^2 \quad (6.7)$$

называют *распределением “хи-квадрат”* с  $n$  степенями свободы и обозначают символом  $\chi^2$ . Обычно и для самой случайной величины  $H_n$  используется тот же символ, т.е. вместо  $H_n$  пишут  $\chi^2$ .

Плотность распределения вероятностей  $H_n$  (или  $\chi^2$ ) является однопараметрической функцией, равной (см. рис. 6.2)

$$f_{\chi^2}(x, n) = \frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{2\Gamma\left(\frac{n}{2}\right)} \exp\left(-\frac{x}{2}\right), \quad (6.8)$$

где  $\Gamma(x)$  – гамма-функция (6.5).

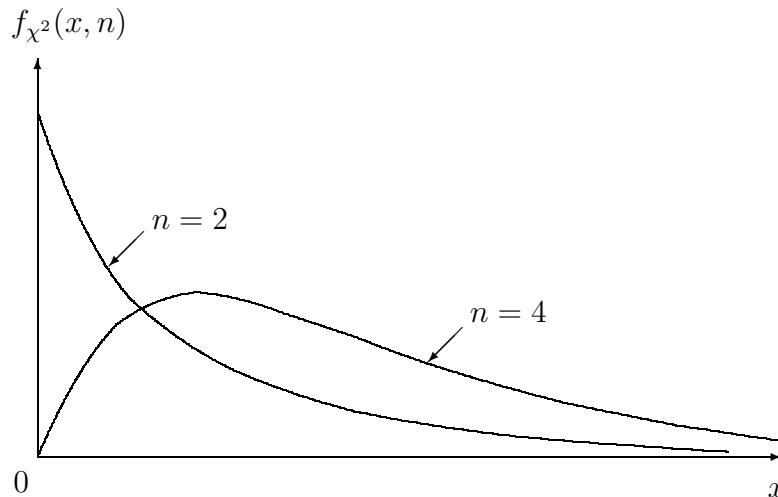


Рис. 6.2. Плотности распределения хи-квадрат с разным числом степеней свободы  $n = 2$  и  $4$

Заметим, что распределение хи-квадрат устойчиво относительно суммирования. Например, если независимые случайные величины  $H_{n_1}$  и  $H_{n_2}$  имеют распределение хи-квадрат с  $n_1$  и  $n_2$  степенями свободы соответственно, то их сумма  $H_{n_1} + H_{n_2}$  имеет распределение хи-квадрат с  $n_1 + n_2$  степенями свободы.

Математическое ожидание и дисперсия  $H_n$  равны соответственно:

$$M[H_n] = n, \quad D[H_n] = 2n.$$

В силу центральной предельной теоремы при большом числе степеней свободы распределение случайной величины  $H_n$  можно рассматривать как нормальное с параметрами:  $a = n$ ,  $\sigma^2 = 2n$ . Другими словами, функция распределения  $(H_n - n)/\sqrt{2n}$  стремится к  $F_N(x, 0, 1)$  при  $n \rightarrow \infty$ .

Квантилью порядка  $\alpha$  распределения хи-квадрат называется такое число  $\chi_{\alpha, n}^2$ , для которого справедливо равенство

$$P(H_n < \chi_{\alpha, n}^2) = \alpha.$$

Значения квантилей даны в виде таблиц в учебниках и специальных справочниках (см., например, Приложение или [3]).

Для приближенного вычисления квантилей при больших  $n$  ( $n \geq 30$ ) используют асимптотическую нормальность распределения хи-квадрат. Поскольку при  $n \rightarrow \infty$  функция распределения  $(H_n - n) / \sqrt{2n}$  стремится к нормальной, то

$$p\left(\frac{H_n - n}{\sqrt{2n}} < x_\alpha\right) \approx \alpha.$$

где  $x_\alpha$  – квантиль порядка  $\alpha$  нормального распределения. Используя определение  $\chi_{\alpha,n}^2$  и свойства неравенств, получаем

$$p(H_n < \chi_{\alpha,n}^2) = p\left(\frac{H_n - n}{\sqrt{2n}} < \frac{\chi_{\alpha,n}^2 - n}{\sqrt{2n}}\right) = \alpha,$$

что дает

$$\frac{\chi_{\alpha,n}^2 - n}{\sqrt{2n}} \approx x_\alpha.$$

Таким образом, при больших  $n$  справедливо приближенное выражение

$$\chi_{\alpha,n}^2 \approx n + x_\alpha \sqrt{2n}. \quad (6.9)$$

Можно получить и другие приближенные формулы, точность которых возрастает с увеличением  $n$ :

$$\chi_{\alpha,n}^2 \approx \frac{1}{2}(x_\alpha + \sqrt{2n-1})^2, \quad (6.10)$$

$$\chi_{\alpha,n}^2 \approx n \left(1 - \frac{2}{9n} + x_\alpha \sqrt{\frac{2}{9n}}\right)^3. \quad (6.11)$$

Критическая точка  $\chi^2(\alpha, n)$  определяется соотношением  $p(H_n > \chi^2(\alpha, n)) = \alpha$  и связана с квантилью соотношением

$$\chi^2(\alpha, n) = \chi_{1-\alpha,n}^2.$$

Квантили  $\chi_{\alpha,n}^2$  распределения хи-квадрат с  $n$  степенями свободы приведены в табл. 3 *Приложения*.

#### 6.4. Распределение Фишера–Снедекора

В многомерном статистическом анализе, дисперсионном анализе, регрессионном анализе и т.д. широко используется распределение Фишера–Снедекора – непрерывное двухпараметрическое распределение вероятностей с плотностью

$$f_F(x, v_1, v_2) = \frac{x^{-1}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} \left(x \frac{v_1}{v_2}\right)^{\frac{v_1}{2}} \left(1 + x \frac{v_1}{v_2}\right)^{-\frac{v_1+v_2}{2}}, \quad (6.12)$$

где  $x > 0$ ,  $v_1 > 0$ ,  $v_2 > 0$  и  $B(l_1, l_2) = \Gamma(l_1) \Gamma(l_2) / \Gamma(l_1 + l_2)$  – бета-функция.

Математическое ожидание и дисперсия случайной величины  $F(v_1, v_2)$ , имеющей такое распределение, равны

$$M[F(v_1, v_2)] = \frac{v_2}{v_2 + 1} \quad \text{при } v_2 > 2 \text{ и}$$

$$D[F(v_1, v_2)] = \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)} \quad \text{при } v_2 > 4.$$

При целых значениях параметров, когда  $v_1 = k_1$  и  $v_2 = k_2$ , где  $k_1, k_2$  – целые числа, распределение Фишера–Снедекора есть распределение отношения  $\frac{\chi_{k_1}^2}{k_1} \frac{k_2}{\chi_{k_2}^2}$ , где  $\frac{\chi_{k_1}^2}{k_1}$  и  $\frac{\chi_{k_2}^2}{k_2}$  – независимые случайные величины, имеющие распределения хи-квадрат с  $k_1$  и  $k_2$  степенями свободы соответственно.

Квантили  $F_\alpha(k_1, k_2)$  распределения  $F(k_1, k_2)$  содержатся в справочниках и другой литературе (см., например, [3]).

При  $k_1, k_2 \gg 1$  для вычисления квантили  $F_\alpha(k_1, k_2)$  можно использовать приближенную формулу

$$F_\alpha(k_1, k_2) \approx \frac{k_2}{k_2 - 2} \sqrt{\frac{2(k_1 + k_2 - 2)}{k_1(k_2 - 4)}} x_\alpha + \frac{k_2}{k_2 - 2}, \quad (6.13)$$

где  $x_\alpha$  – квантиль нормального распределения.

Для квантили  $F_\alpha(k_1, k_2)$  выполняется следующее соотношение:

$$F_\alpha(k_1, k_2) = \frac{1}{F_{1-\alpha}(k_2, k_1)}, \quad (6.14)$$

что позволяет в таблицах приводить значения только для  $\alpha \geq 0,5$ .

Критическая точка распределения  $F(\alpha, k_1, k_2)$  связана с квантилью соотношением

$$F(\alpha, k_1, k_2) = F_{1-\alpha}(k_1, k_2). \quad (6.15)$$

Критические точки распределения Фишера–Снедекора приведены для  $\alpha = 0,05$  и  $0,01$  соответственно в табл. 4а и 4б *Приложения*.

## 6.5. Решение типовых примеров

**Пример 6.1.** Найти квантиль порядка 0,9 нормального распределения.

**Решение.** Используем соотношение (6.2) при  $\alpha = 0,9$

$$\Phi(x_{0,9}) = 0,9 - \frac{1}{2} = 0,4$$

и из табл.1 *Приложения* получим значение искомой квантили  $x_{0,9} = 1,28$ .

**Пример 6.2.** Найти квантиль порядка 0,95 распределения Стьюдента с 30 степенями свободы, используя таблицу *Приложения* и приближенную формулу из п. 6.2.

**Решение.** Согласно табл. 2 *Приложения* имеем квантиль  $t_{0,95; 30} = 1,6973$ .

Поскольку число степеней  $n = 30$  достаточно велико, можно использовать приближенную формулу (6.6).

Квантиль порядка 0,95 нормального распределения  $x_{0,95}$ , входящая в эту формулу, определяется согласно п. 6.1 из соотношения (6.2):  $\Phi(x_{0,95}) = 0,95 - 0,5 = 0,45$ . Из табл. 1 *Приложения* находим  $x_{0,95} = 1,64$ .

Подставляя  $n = 30$  и  $x_{0,95} = 1,64$  в формулу (6.6), получаем оценку искомой квантили  $t_{0,95; 30} \approx 1,69$ .

**Пример 6.3.** Найти квантиль порядка 0,9 распределения хи-квадрат с 30 степенями свободы по таблицам и оценить ее значение по приближенным формулам (6.9)–(6.11).

**Решение.** Из табл. 3 *Приложения* найдем искомую квантиль  $\chi_{0,9, 30}^2 = 40,256$ .

С учетом того, что квантиль порядка 0,9 нормального распределения  $x_{0,9} = 1,28$ , по формуле (6.9) получаем

$$\chi_{0,9, 30}^2 \approx 30 + 1,28\sqrt{2 \cdot 30} = 30 + 9,91 = 39,91.$$

Согласно соотношению (6.10)

$$\chi_{0,9, 30}^2 \approx \frac{1}{2} (1,28 + \sqrt{2 \cdot 30 - 1})^2 = \frac{1}{2} (8,961)^2 = 40,15.$$

Наконец, согласно формуле (6.11) имеем

$$\chi_{0,9, 30}^2 \approx 30 \left( 1 - \frac{2}{9 \cdot 30} + 1,28 \sqrt{\frac{2}{9 \cdot 30}} \right)^3 = 30 \cdot (1,1028)^3 = 40,236.$$

**Пример 6.4.** Определить квантиль порядка 0,9 распределения Фишера–Снедекора с числами степеней свободы  $k_1 = 40$  и  $k_2 = 60$ , используя таблицы из *Приложения* и приближенную формулу (6.13).

**Решение.** Табл. 4б *Приложения* дает значение критической точки  $F(0,1, 40, 60) = 1,44$ .

Применяя соотношение (6.15), получаем искомую квантиль  $F_{0,9}(40, 60) = 1,44$ .

Подставляя значение квантили нормального распределения  $x_{0,9} = 1,28$  в формулу (6.13), получаем оценку

$$F_{0,9}(40, 60) \approx \frac{60}{60 - 2} \sqrt{\frac{2(40 + 60 - 2)}{40(60 - 4)}} \cdot 1,28 + \frac{60}{60 - 2} =$$

$$= \frac{60}{58} \sqrt{\frac{2 \cdot 98}{4 \cdot 56}} \cdot 1,28 + \frac{60}{58} = 1,43.$$

**Пример 6.5.** Найти квантиль порядка 0,05 распределения Фишера–Снедекора с числами степеней свободы  $k_1 = 10$  и  $k_2 = 9$ , используя таблицы из *Приложения*.

**Решение.** Согласно соотношениям (6.14) и (6.15) искомая квантиль равна (критическую точку  $F(0,05, 9, 10) = 2,35$  находим из табл. 4а *Приложения*)

$$F_{0,05}(10, 9) = \frac{1}{F_{0,95}(9, 10)} = \frac{1}{F(0,05, 9, 10)} = \frac{1}{2,35} = 0,4255.$$

## 7. ПРИМЕР СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ВЫБОРКИ

Будем считать максимальную дневную температуру в Санкт-Петербурге 1 сентября случайной величиной  $\xi$ . Генеральная совокупность – это данные Гидрометеослужбы о такой температуре в разные годы. Сделана следующая выборка из генеральной совокупности (в градусах Цельсия):

14	16	16	18	13	13	16	16	25	10
19	14	15	17	12	16	13	19	16	17
17	12	15	19	14	15	17	19	18	14
26	21	17	15	19	18	18	13	15	18
17	17	16	18	16	21	15	13	20	14

Рассмотрим задачи, на которые разбивается статистическая обработка выборки, направленная на определение свойств данной случайной величины  $\xi$ .

**Пример 7.1.** Для приведенной выборки построить вариационный ряд и выборочный закон распределения  $\xi$ . Найти выборочное среднее  $\bar{x}$ , выборочную дисперсию  $D^*$  и исправленную выборочную дисперсию  $s^2$ .

**Решение.** Вариационный ряд, построенный по данной выборке, извлеченный из генеральной совокупности, включает 13 вариантов: 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 25, 26° С.

Выборочный закон распределения оказывается следующим:

$x_i$	10	12	13	14	15	16	17	18	19	20	21	25	26
$n_i$	1	2	5	5	6	8	7	6	5	1	2	1	1
$\omega_i$	0,02	0,04	0,10	0,10	0,12	0,16	0,14	0,12	0,10	0,02	0,04	0,02	0,02

Выборочное среднее равно

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{13} n_i x_i = \frac{1}{50} (10 + 2 \cdot 12 + \dots + 26) = \frac{822}{50} = 16,44^\circ \text{С}.$$

Найдем

$$\overline{x^2} = \frac{1}{n} \sum_{i=1}^{13} n_i x_i^2 = \frac{1}{50} \{ (10)^2 + 2 \cdot (12)^2 + \dots + (26)^2 \} = \frac{13966}{50} = 279,32 \text{ град}^2$$

и получим выборочные дисперсии:

$$D^* = (\sigma^*)^2 = \overline{x^2} - (\bar{x})^2 = 279,32 - (16,44)^2 = 9,0464 \text{ град}^2,$$

$$s^2 = \frac{n}{n-1} (\sigma^*)^2 = \frac{50}{49} \cdot 9,0464 = 9,231 \text{ град}^2.$$

Заметим, что полученные результаты говорят о том, что 1 сентября в Санкт-Петербурге следует ожидать максимальную дневную температуру, равную примерно 16,5° С с разбросом порядка  $\sigma^* \approx 3^\circ \text{С}$ , т.е. примерно в 70% случаев температура должна попадать в интервал 13,5 – 19,5° С (в выборке в этот интервал попадает 74% значений).

**Пример 7.2.** Построить с надежностью  $\gamma = 0,90$  доверительные интервал для математического ожидания случайной величины  $\xi$ .

**Решение.** Будем считать, что рассматриваемая выборка достаточно велика ( $n \gg 1$ ) и применима формула (4.23)

$$\begin{cases} p\left(\bar{x} - \frac{s \tilde{x}_\gamma}{\sqrt{n}} < M[\xi] < \bar{x} + \frac{s \tilde{x}_\gamma}{\sqrt{n}}\right) \approx 2\Phi(\tilde{x}_\gamma), \\ 2\Phi(\tilde{x}_\gamma) = \gamma, \end{cases}$$

в которой, учитывая полученные выше результаты:

$$\bar{x} = 16,44^\circ\text{C}, \quad s = \sqrt{9,231} \approx 3,038^\circ\text{C}, \quad \text{а } 2\Phi(\tilde{x}_\gamma) = 0,9 \text{ и } \tilde{x}_\gamma = 1,65.$$

Тогда доверительный интервал для математического ожидания  $M[\xi]$  с надежностью (доверительной вероятностью)  $\gamma = 0,9$  определяется следующим образом:

$$\begin{aligned} p\left(16,44 - \frac{3,038 \times 1,65}{\sqrt{50}} < M[\xi] < 16,44 + \frac{3,038 \times 1,65^\circ\text{C}}{\sqrt{50}}\right) = \\ = p(15,73 < M[\xi] < 17,15^\circ\text{C}) = 0,9. \end{aligned}$$

Это означает, что если нет тренда, то средний за все годы максимум температуры в Санкт-Петербурге 1 сентября с вероятностью 90% заключен в интервале  $15,7 - 17,2^\circ\text{C}$ .

**Пример 7.3.** Построить с надежностью  $\gamma = 0,90$  доверительный интервал для дисперсии  $D[\xi]$  случайной величины  $\xi$  в предположении, что она имеет нормальное распределение.

**Решение.** Здесь можно воспользоваться формулой (4.21)

$$p\left(\frac{(n-1)s^2}{\tilde{\chi}_{2,\gamma,n-1}^2} < D[\xi] < \frac{(n-1)s^2}{\tilde{\chi}_{1,\gamma,n-1}^2}\right) = \gamma.$$

Поскольку в *Приложении* дана таблица квантилей распределения хи-квадрат, то для определения значений  $\tilde{\chi}_{1,\gamma,n-1}^2$  и  $\tilde{\chi}_{2,\gamma,n-1}^2$  будем использовать соотношения (4.16), которые дают

$$\tilde{\chi}_{1,\gamma,n-1}^2 = \chi_{(1-\gamma)/2, n-1}^2, \quad \tilde{\chi}_{2,\gamma,n-1}^2 = \chi_{(1+\gamma)/2, n-1}^2,$$

где  $\chi_{\alpha,n}^2$  есть квантиль порядка  $\alpha$  распределения хи-квадрат с  $n$  степенями свободы. В нашем случае требуется найти квантили  $\chi_{0,05,49}^2$  и  $\chi_{0,95,49}^2$ . Поскольку табл. 3 *Приложения* не содержит этих квантилей, но объем выборки достаточно велик ( $n > 30$ ), для вычисления квантилей можно применить асимптотическую формулу (6.11)

$$\chi_{\alpha,n}^2 \approx n \left(1 - \frac{2}{9n} + x_\alpha \sqrt{\frac{2}{9n}}\right)^3.$$



Квантили нормального распределения равны:  $x_{0,05} = -1,65$ ,  $x_{0,95} = 1,65$ , и для  $n = 50$ ,  $\gamma = 0,9$  имеем

$$\tilde{\chi}_{1,\gamma,n-1}^2 = \chi_{0,05,49}^2 \approx 49 \left( 1 - \frac{2}{9 \times 49} - 1,65 \sqrt{\frac{2}{9 \times 49}} \right)^3 = 33,89,$$

$$\tilde{\chi}_{2,\gamma,n-1}^2 = \chi_{0,95,49}^2 \approx 49 \left( 1 - \frac{2}{9 \times 49} + 1,65 \sqrt{\frac{2}{9 \times 49}} \right)^3 = 66,40.$$

Таким образом, доверительный интервал для дисперсии случайной величины  $\xi$  (если предположить, что она имеет нормальное распределение) с надежностью (доверительной вероятностью)  $\gamma = 0,9$  определяется следующим образом:

$$p \left( \frac{49 \times 9,231}{66,40} < D[\xi] < \frac{49 \times 9,231}{33,89} \right) = \\ = p(6,81 < D[\xi] < 13,35 \text{ град}^2) = 0,90.$$

Иными словами среднее квадратическое отклонение величины  $\xi$ , характеризующее ее разброс относительно  $M[\xi] = 16,44^\circ \text{ С}$ , с вероятностью 90% заключено в пределах от  $2,6$  до  $3,7^\circ \text{ С}$ .

**Пример 7.4.** Используя критерий согласия Пирсона, проверить гипотезу о нормальном распределении случайной величины  $\xi$  с уровнем значимости  $\alpha = 0,1$ .

**Решение.** Построим группированную выборку, разбив интервал, в который попали все наши варианты, на 8 подинтервалов ( $j = 1, 2, \dots, 8$ ):

$[a_j, b_j]$	[9,5; 13,5]	[13,5; 14,5]	[14,5; 15,5]	[15,5; 16,5]
$x_j$	11,5	14	15	16
$n_j$	8	5	6	8

$[a_j, b_j]$	[16,5; 17,5]	[17,5; 18,5]	[18,5; 19,5]	[19,5; 26,5]
$x_j$	17	18	19	23
$n_j$	7	6	5	5

Гистограмма частот для группированной выборки представлена на рис. 7.1, где  $h$  – ширина соответствующего интервала (см. п. 1.4).

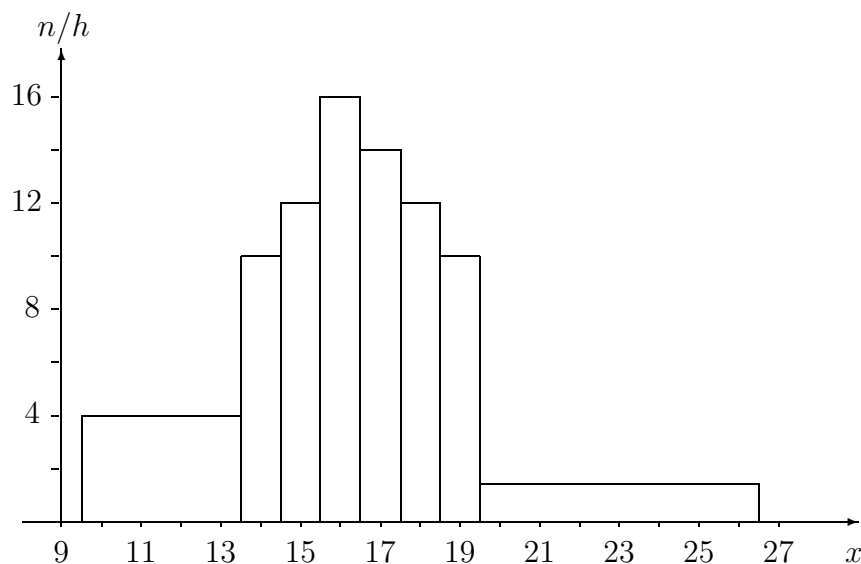


Рис. 7.1. Гистограмма частот группированной выборки

Применим критерий  $H$ , определенный формулой (5.16). Теоретические частоты рассчитаем по формуле  $n'_j = np_j$ , где вероятности  $p_j$  найдем из формулы (5.15)

$$n'_j = np_j = n \left( \Phi \left( \frac{b_j - \bar{x}}{\sigma^*} \right) - \Phi \left( \frac{a_j - \bar{x}}{\sigma^*} \right) \right).$$

Тогда

$$\begin{aligned} n'_1 &= n \left( \Phi \left( \frac{13,5 - 16,44}{\sqrt{9,0464}} \right) - \Phi \left( \frac{9,5 - 16,44}{\sqrt{9,0464}} \right) \right) = \\ &= 50 \left( \Phi \left( \frac{-2,94}{3,01} \right) - \Phi \left( \frac{-6,94}{3,01} \right) \right) = 50(\Phi(-0,98) - \\ &- \Phi(-2,31)) = 50(-0,336 + 0,490) = 50 \cdot 0,154 = 7,7; \\ n'_2 &= 50 \left( \Phi \left( \frac{14,5 - 16,44}{3,01} \right) - \Phi \left( \frac{13,5 - 16,44}{3,01} \right) \right) = 50(\Phi(-0,64) - \\ &- \Phi(-0,98)) = 50(-0,239 + 0,336) = 50 \cdot 0,097 = 4,85; \\ n'_3 &= 50 \left( \Phi \left( \frac{15,5 - 16,44}{3,01} \right) - \Phi \left( \frac{14,5 - 16,44}{3,01} \right) \right) = 50(\Phi(-0,31) - \\ &- \Phi(-0,64)) = 50(-0,122 + 0,239) = 50 \cdot 0,117 = 5,85; \\ n'_4 &= 50 \left( \Phi \left( \frac{16,5 - 16,44}{3,01} \right) - \Phi \left( \frac{15,5 - 16,44}{3,01} \right) \right) = 50(\Phi(0,02) - \\ &- \Phi(-0,31)) = 50(0,008 + 0,122) = 50 \cdot 0,130 = 6,5; \\ n'_5 &= 50 \left( \Phi \left( \frac{17,5 - 16,44}{3,01} \right) - \Phi \left( \frac{16,5 - 16,44}{3,01} \right) \right) = 50(\Phi(0,35) - \end{aligned}$$

$$- \Phi(0,02)) = 50(0,134 - 0,008) = 50 \cdot 0,126 = 6,3;$$

$$n'_6 = 50 \left( \Phi \left( \frac{18,5 - 16,44}{3,01} \right) - \Phi \left( \frac{17,5 - 16,44}{3,01} \right) \right) = 50(\Phi(0,68) - \Phi(0,35)) = 50(0,252 - 0,134) = 50 \cdot 0,118 = 5,9;$$

$$n'_7 = 50 \left( \Phi \left( \frac{19,5 - 16,44}{3,01} \right) - \Phi \left( \frac{18,5 - 16,44}{3,01} \right) \right) = 50(\Phi(1,02) - \Phi(0,68)) = 50(0,346 - 0,252) = 50 \cdot 0,094 = 4,7;$$

$$n'_8 = 50 \left( \Phi \left( \frac{26,5 - 16,44}{3,01} \right) - \Phi \left( \frac{19,5 - 16,44}{3,01} \right) \right) = 50(\Phi(3,34) - \Phi(1,02)) = 50(0,5000 - 0,346) = 50 \cdot 0,154 = 7,7.$$

Результаты вычислений можно проверить, просуммировав теоретические частоты. Очевидно, что сумма всех  $n'_j$ , приведенных в таблице ниже, должна примерно равняться числу вариантов, т.е. 50.

$[a_j, b_j]$	[9,5; 13,5]	[13,5; 14,5]	[14,5; 15,5]	[15,5; 16,5]
$n_j$	8	5	6	8
$n'_j$	7,7	4,85	5,85	6,5
$(n_j - n'_j)^2 / n'_j$	0,012	0,005	0,004	0,346

$[a_j, b_j]$	[16,5; 17,5]	[17,5; 18,5]	[18,5; 19,5]	[19,5; 26,5]
$n_j$	7	6	5	5
$n'_j$	6,3	5,9	4,7	7,7
$(n_j - n'_j)^2 / n'_j$	0,078	0,002	0,019	0,947

В нашем случае сумма теоретических частот равна 49,5, что связано с недостаточным отличием нижней границы первого интервала 9,5 от среднего значения 16,44.

Эмпирическое значение критерия  $H^*$  для данной группированной выборки ( $s = 8$ ) вычисляется по формуле (5.17)

$$H^* = \sum_{i=1}^8 \frac{(n_i - n'_i)^2}{n'_i} = 1,413.$$

Построим правостороннюю критическую область, удовлетворяющую равенству

$$p(H > \chi^2(\alpha, k)) = \alpha.$$

Критическая точка  $\chi^2(\alpha, k)$  равна квантили порядка  $(1 - \alpha)$  распределения хи-квадрат  $\chi^2_{1-\alpha, k}$ . Из табл. 3 *Приложения* для уровня значимости  $\alpha = 0,1$  и  $k = s - 3 = 5$  имеем

$$\chi^2_{0,9,5} = 9,236.$$

В силу того, что наблюдаемое значение критерия не попадает в критическую область ( $H^* < \chi^2(0,1, 5)$ ), можно принять гипотезу о нормальном распределении случайной величины  $\xi$  при уровне значимости  $\alpha = 0,1$ .

**Пример 7.5.** Используя критерий согласия Пирсона, проверить гипотезу о равномерном законе распределения случайной величины  $\xi$  с уровнем значимости  $\alpha = 0,1$ .

**Решение.** При равномерном законе плотность распределения вероятностей представляется выражением (5.18) и имеет параметры  $a$  и  $b$ . Оценим их, применяя метод моментов. Тогда параметры выражаются через характеристики выборки следующим образом (см. п. 5.2.2):

$$a^* = \bar{x} - \sqrt{3}\sigma^*, \quad b^* = \bar{x} + \sqrt{3}\sigma^*.$$

Используя найденные выше значения выборочного среднего и выборочной дисперсии, получим:  $a^* = 16,44 - 1,732 \cdot 3,008 = 11,23$ ;  $b^* = 16,44 + 1,732 \cdot 3,008 = 21,65$ .

Применим критерий  $H$ , определенный формулой (5.16). Используем группированную выборку, сделанную выше, и вычислим теоретические частоты по формуле (5.20). В результате получим:

$[a_j, b_j]$	[9,5; 13,5]	[13,5; 14,5]	[14,5; 15,5]	[15,5; 16,5]
$n_j$	8	5	6	8
$n'_j$	10,89	4,80	4,80	4,80
$(n_j - n'_j)^2/n'_j$	0,767	0,008	0,300	2,133

$[a_j, b_j]$	[16,5; 17,5]	[17,5; 18,5]	[18,5; 19,5]	[19,5; 26,5]
$n_j$	7	6	5	5
$n'_j$	4,80	4,80	4,80	10,32
$(n_j - n'_j)^2/n'_j$	1,008	0,300	0,008	2,742

Наблюдаемое (эмпирическое) значение критерия согласно формуле (5.17) равно  $H^* = 7.266$ .

Как и в п. 5.2.2, будем рассматривать правостороннюю критическую область. Тогда критическая точка равна  $\chi^2_{1-\alpha, k}$  – квантили порядка  $(1 - \alpha)$  распределения хи-квадрат с числом степеней свободы  $k = s - 3$ . В нашем

случае  $k = 8 - 3 = 5$  и  $\alpha = 0,1$ , и из табл. 3 *Приложения* находим  $\chi_{0,9,5}^2 = 9,236$ .

Поскольку эмпирическое значение критерия не попадает в критическую область, определяемую как  $H^* > \chi_{0,9,5}^2$ , то гипотеза о равномерном распределении данной случайной величины может быть принята при уровне значимости  $\alpha = 0,1$ . Однако, заметим, что поскольку эмпирическое значение критерия близко к критическому, гипотеза была близка к тому, чтобы быть отвергнутой.

## ЛИТЕРАТУРА

1. *Фарафонов В.Г., Ильин В.Б.* Теория вероятностей и математическая статистика. Ч.1. СПб: СПбГУАП, 2012. 112 с.
2. *Лексаченко В. А.* Логика. Множества. Вероятность. СПб: СПбГУАП 2005. 135 с.
3. *Гмурман В.Е.* Теория вероятностей и математическая статистика. М.: Высшее образование, 2006. 575 с.
4. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. М.: Высшее образование, 2006. 476 с.

## ПРИЛОЖЕНИЯ

Таблица 1

**Значения функции Лапласа  $\Phi(x)$**

<b>x</b>	<b>0,00</b>	<b>0,02</b>	<b>0,04</b>	<b>0,06</b>	<b>0,08</b>
<b>0,0</b>	0,00000	0,00798	0,01595	0,02392	0,03188
<b>0,1</b>	0,03983	0,04776	0,05567	0,06356	0,07142
<b>0,2</b>	0,07926	0,08706	0,09483	0,10257	0,11026
<b>0,3</b>	0,11791	0,12552	0,13307	0,14058	0,14803
<b>0,4</b>	0,15542	0,16276	0,17003	0,17724	0,18439
<b>0,5</b>	0,19146	0,19847	0,20540	0,21226	0,21904
<b>0,6</b>	0,22575	0,23237	0,23891	0,24537	0,25175
<b>0,7</b>	0,25804	0,26424	0,27035	0,27637	0,28230
<b>0,8</b>	0,28814	0,29389	0,29955	0,30511	0,31057
<b>0,9</b>	0,31594	0,32121	0,32639	0,33147	0,33646
<b>1,0</b>	0,34134	0,34614	0,35083	0,35543	0,35993
<b>1,1</b>	0,36433	0,36864	0,37286	0,37698	0,38100
<b>1,2</b>	0,38493	0,38877	0,39251	0,39617	0,39973
<b>1,3</b>	0,40320	0,40658	0,40988	0,41308	0,41621
<b>1,4</b>	0,41924	0,42220	0,42507	0,42786	0,43056
<b>1,5</b>	0,43319	0,43574	0,43822	0,44062	0,44295
<b>1,6</b>	0,44520	0,44738	0,44950	0,45154	0,45352
<b>1,7</b>	0,45543	0,45728	0,45907	0,46080	0,46246
<b>1,8</b>	0,46407	0,46562	0,46712	0,46856	0,46995
<b>1,9</b>	0,47128	0,47257	0,47381	0,47500	0,47615
<b>2,0</b>	0,47725	0,47831	0,47932	0,48030	0,48124
<b>2,1</b>	0,48214	0,48300	0,48382	0,48461	0,48537
<b>2,2</b>	0,48610	0,48679	0,48745	0,48809	0,48870
<b>2,3</b>	0,48928	0,48983	0,49036	0,49086	0,49134
<b>2,4</b>	0,49180	0,49224	0,49266	0,49305	0,49343
<b>2,5</b>	0,49379	0,49413	0,49446	0,49477	0,49506
<b>2,6</b>	0,49534	0,49560	0,49585	0,49609	0,49632
<b>2,7</b>	0,49653	0,49674	0,49693	0,49711	0,49728
<b>2,8</b>	0,49744	0,49760	0,49774	0,49788	0,49801
<b>2,9</b>	0,49813	0,49825	0,49836	0,49846	0,49856
<b>x</b>	<b>3,0</b>	<b>3,2</b>	<b>3,4</b>	<b>3,6</b>	<b>3,8</b>
<b><math>\Phi(x)</math></b>	0,49865	0,49931	0,49966	0,49984	0,49993
<b>x</b>	<b>4,0</b>	<b>4,2</b>			
<b><math>\Phi(x)</math></b>	0,49997	0,49999			

Квантили распределения Стьюдента  $t_{\alpha, n}$ 

$n \backslash \alpha$	<b>0,7500</b>	<b>0,9000</b>	<b>0,9500</b>	<b>0,9750</b>	<b>0,9900</b>	<b>0,9950</b>	<b>0,9995</b>
<b>1</b>	1,0000	3,0777	6,3138	12,706	31,821	63,657	636,62
<b>2</b>	0,8165	1,8856	2,9110	4,3027	6,9646	9,9248	31,599
<b>3</b>	0,7649	1,6377	2,3534	3,1825	4,5407	5,8409	12,924
<b>4</b>	0,7407	1,5332	2,1319	2,7765	3,7470	4,6041	8,6103
<b>5</b>	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321	6,8688
<b>6</b>	0,7176	1,4398	1,9432	2,4469	3,1427	3,7074	5,9588
<b>7</b>	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995	5,4079
<b>8</b>	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554	5,0413
<b>9</b>	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498	4,7809
<b>10</b>	0,6998	1,3722	1,8125	2,2281	2,7638	3,1693	4,5869
<b>11</b>	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	4,4370
<b>12</b>	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545	4,3178
<b>13</b>	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123	4,2208
<b>14</b>	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768	4,1405
<b>15</b>	0,6912	1,3406	1,7531	2,1315	2,6025	2,9467	4,0728
<b>16</b>	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	4,0150
<b>17</b>	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
<b>18</b>	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784	3,9216
<b>19</b>	0,6876	1,3277	1,7291	2,0930	2,5395	2,8609	3,8834
<b>20</b>	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453	3,8495
<b>21</b>	0,6864	1,3232	1,7207	2,0796	2,5177	2,8314	3,8193
<b>22</b>	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188	3,7921
<b>23</b>	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
<b>24</b>	0,6849	1,3179	1,7109	2,0639	2,4922	2,7969	3,7454
<b>25</b>	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
<b>26</b>	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787	3,7066
<b>27</b>	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707	3,6896
<b>28</b>	0,6834	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739
<b>29</b>	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564	3,6594
<b>30</b>	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500	3,6460
$\infty$	0,6745	1,2816	1,6449	1,9600	2,3264	2,5758	3,2905



Квантили распределения хи-квадрат  $\chi_{\alpha,n}^2$ 

$n \backslash \alpha$	0,01	0,025	0,05	0,1	0,2	0,3	0,4	0,5
1	0,0002	0,0010	0,0039	0,0158	0,0642	0,1485	0,2750	0,4549
2	0,0201	0,0506	0,1026	0,2107	0,4463	0,7133	1,0217	1,3863
3	0,1148	0,2158	0,3519	0,5844	1,0052	1,4237	1,8692	2,3660
4	0,2971	0,4844	0,7107	1,0636	1,6488	2,1947	2,7528	3,3567
5	0,5543	0,8312	1,1455	1,6103	2,3425	2,9999	3,6555	4,3515
6	0,8721	1,2373	1,6354	2,2041	3,0701	3,8276	4,5702	5,3481
7	1,2390	1,6899	2,1674	2,8331	3,8223	4,6713	5,4932	6,3458
8	1,6465	2,1797	2,7326	3,4895	4,5936	5,5274	6,4226	7,3441
9	2,0879	2,7004	3,3251	4,1682	5,3801	6,3933	7,3570	8,3428
10	2,5582	3,2470	3,9403	4,8652	6,1791	7,2672	8,2955	9,3418
11	3,0535	3,8158	4,5748	5,5778	6,9887	8,1479	9,2373	10,3410
12	3,5706	4,4038	5,2260	6,3038	7,8073	9,0343	10,1820	11,3403
13	4,1069	5,0088	5,8919	7,0415	8,6339	9,9257	11,1291	12,3398
14	4,6604	5,6287	6,5706	7,7895	9,4673	10,8215	12,0785	13,3393
15	5,2294	6,2621	7,2609	8,5468	10,3070	11,7212	13,0297	14,3389
16	5,8122	6,9077	7,9617	9,3122	11,1521	12,6243	13,9827	15,3385
17	6,4078	7,5642	8,6718	10,0852	12,0023	13,5307	14,9373	16,3382
18	7,0149	8,2308	9,3905	10,8649	12,8570	14,4399	15,8932	17,3379
19	7,6327	8,9065	10,117	11,6509	13,7158	15,3517	16,8504	18,3377
20	8,2604	9,5908	10,851	12,4426	14,5784	16,2659	17,8088	19,3374
21	8,8972	10,283	11,591	13,2396	15,4446	17,1823	18,7683	20,3372
22	9,5425	10,982	12,338	14,0415	16,3140	18,1007	19,7288	21,3370
23	10,196	11,689	13,091	14,8480	17,1865	19,0211	20,6902	22,3369
24	10,856	12,401	13,848	15,6587	18,0618	19,9432	21,6525	23,3367
25	11,524	13,120	14,611	16,4734	18,9398	20,8670	22,6156	24,3366
26	12,198	13,844	15,379	17,2919	19,8202	21,7924	23,5794	25,3365
27	12,879	14,573	16,151	18,1139	20,7030	22,7192	24,5440	26,3363
28	13,565	15,308	16,928	18,9392	21,5880	23,6475	25,5093	27,3362
29	14,256	16,047	17,708	19,7677	22,4751	24,5770	26,4751	28,3361
30	14,953	16,791	18,493	20,5992	23,3641	25,5078	27,4416	29,3360

Квантили распределения хи-квадрат  $\chi^2_{\alpha, n}$ 

$n \backslash \alpha$	0,6	0,7	0,8	0,9	0,95	0,975	0,99
1	0,7083	1,0742	1,6424	2,7055	3,8415	5,0239	6,6349
2	1,8326	2,4079	3,2189	4,6052	5,9915	7,3778	9,2103
3	2,9462	3,6649	4,6416	6,2514	7,8147	9,3484	11,345
4	4,0446	4,8784	5,9886	7,7794	9,4877	11,143	13,277
5	5,1319	6,0644	7,2893	9,2364	11,071	12,833	15,086
6	6,2108	7,2311	8,5581	10,645	12,592	14,449	16,812
7	7,2832	8,3834	9,8032	12,017	14,067	16,013	18,475
8	8,3505	9,5245	11,0301	13,362	15,507	17,535	20,090
9	9,4136	10,6564	12,2421	14,684	16,919	19,023	21,666
10	10,4732	11,7807	13,4420	15,987	18,307	20,483	23,209
11	11,5298	12,8987	14,6314	17,275	19,675	21,920	24,725
12	12,5838	14,0111	15,8120	18,549	21,026	23,337	26,217
13	13,6356	15,1187	16,9848	19,812	22,362	24,736	27,688
14	14,6853	16,2221	18,1508	21,064	23,685	26,119	29,141
15	15,7332	17,3217	19,3107	22,307	24,996	27,488	30,578
16	16,7795	18,4179	20,4651	23,542	26,296	28,845	32,000
17	17,8244	19,5110	21,6146	24,769	27,587	30,191	33,409
18	18,8679	20,6014	22,7595	25,989	28,869	31,526	34,805
19	19,9102	21,6891	23,9004	27,203	30,144	32,852	36,191
20	20,9514	22,7745	25,0375	28,412	31,410	34,170	37,566
21	21,9915	23,8578	26,1711	29,615	32,671	35,479	38,932
22	23,0307	24,9390	27,3015	30,813	33,924	36,781	40,289
23	24,0689	26,0184	28,4288	32,007	35,172	38,076	41,638
24	25,1063	27,0960	29,5533	33,196	36,415	39,364	42,980
25	26,1430	28,1719	30,6752	34,382	37,652	40,646	44,314
26	27,1789	29,2463	31,7946	35,563	38,885	41,923	45,642
27	28,2141	30,3193	32,9117	36,741	40,113	43,195	46,963
28	29,2486	31,3909	34,0266	37,916	41,337	44,461	48,278
29	30,2825	32,4612	35,1394	39,087	42,557	45,722	49,588
30	31,3159	33,5302	36,2502	40,256	43,773	46,979	50,892

**Критические точки распределения Фишера–Снедекора  
 $F(\alpha, k_1, k_2)$  при  $\alpha = 0,05$**

$k_2 \backslash k_1$	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>12</b>	<b>15</b>	<b>20</b>	<b>40</b>
<b>5</b>	5,05	4,95	4,88	4,82	4,77	4,68	4,62	4,56	4,46
<b>6</b>	4,39	4,28	4,21	4,15	4,10	4,00	3,94	3,87	3,77
<b>7</b>	3,97	3,87	3,79	3,73	3,68	3,57	3,51	3,44	3,34
<b>8</b>	3,69	3,58	3,50	3,44	3,39	3,28	3,22	3,15	3,04
<b>9</b>	3,48	3,37	3,29	3,23	3,18	3,07	3,01	2,94	2,83
<b>10</b>	3,33	3,22	3,14	3,07	3,02	2,91	2,85	2,77	2,66
<b>11</b>	3,20	3,09	3,01	2,95	2,90	2,79	2,72	2,65	2,53
<b>12</b>	3,12	3,00	2,91	2,85	2,80	2,69	2,62	2,54	2,43
<b>13</b>	3,03	2,92	2,83	2,77	2,71	2,60	2,53	2,46	2,34
<b>14</b>	2,96	2,85	2,76	2,70	2,65	2,53	2,46	2,39	2,27
<b>15</b>	2,90	2,79	2,71	2,64	2,59	2,48	2,40	2,33	2,20
<b>16</b>	2,85	2,74	2,66	2,59	2,54	2,42	2,35	2,28	2,15
<b>17</b>	2,81	2,70	2,61	2,55	2,49	2,38	2,31	2,23	2,10
<b>18</b>	2,77	2,66	2,58	2,51	2,46	2,34	2,27	2,19	2,06
<b>19</b>	2,74	2,63	2,54	2,48	2,42	2,31	2,23	2,16	2,03
<b>20</b>	2,71	2,60	2,51	2,45	2,39	2,28	2,20	2,12	1,99
<b>21</b>	2,68	2,57	2,49	2,42	2,37	2,25	2,18	2,10	1,96
<b>22</b>	2,66	2,55	2,46	2,40	2,34	2,23	2,15	2,07	1,94
<b>23</b>	2,64	2,53	2,44	2,37	2,32	2,20	2,13	2,05	1,91
<b>24</b>	2,62	2,51	2,42	2,36	2,30	2,18	2,11	2,03	1,89
<b>25</b>	2,60	2,49	2,40	2,34	2,28	2,16	2,09	2,01	1,87
<b>26</b>	2,59	2,47	2,39	2,32	2,27	2,15	2,07	1,99	1,85
<b>27</b>	2,57	2,46	2,37	2,31	2,25	2,13	2,06	1,97	1,84
<b>28</b>	2,56	2,45	2,36	2,29	2,24	2,12	2,04	1,96	1,82
<b>29</b>	2,55	2,43	2,35	2,28	2,22	2,10	2,03	1,94	1,81
<b>30</b>	2,53	2,42	2,33	2,27	2,21	2,09	2,01	1,93	1,79
<b>40</b>	2,45	2,34	2,25	2,18	2,12	2,00	1,92	1,84	1,69
<b>60</b>	2,37	2,25	2,17	2,10	2,04	1,92	1,84	1,75	1,59

**Критические точки распределения Фишера–Снедекора**  
 **$F(\alpha, k_1, k_2)$  при  $\alpha = 0,1$**

$k_2 \backslash k_1$	5	6	7	8	9	12	15	20	40
5	3,45	3,40	3,37	3,34	3,32	3,27	3,24	3,21	3,16
6	3,11	3,05	3,01	2,98	2,96	2,90	2,87	2,84	2,78
7	2,88	2,83	2,78	2,75	2,72	2,67	2,63	2,59	2,54
8	2,73	2,67	2,62	2,59	2,56	2,50	2,46	2,42	2,36
9	2,61	2,55	2,51	2,47	2,44	2,38	2,34	2,30	2,23
10	2,52	2,46	2,41	2,38	2,35	2,28	2,24	2,20	2,13
11	2,45	2,39	2,34	2,30	2,27	2,21	2,17	2,12	2,05
12	2,39	2,33	2,28	2,24	2,21	2,15	2,10	2,06	1,99
13	2,35	2,28	2,23	2,20	2,16	2,10	2,05	2,01	1,93
14	2,31	2,24	2,19	2,15	2,12	2,05	2,01	1,96	1,89
15	2,27	2,21	2,16	2,12	2,09	2,02	1,97	1,92	1,85
16	2,24	2,18	2,13	2,09	2,06	1,99	1,94	1,89	1,81
17	2,22	2,15	2,10	2,06	2,03	1,96	1,91	1,86	1,78
18	2,20	2,13	2,08	2,04	2,00	1,93	1,89	1,84	1,75
19	2,18	2,11	2,06	2,02	1,98	1,91	1,86	1,81	1,73
20	2,16	2,09	2,04	2,00	1,96	1,89	1,84	1,79	1,71
21	2,14	2,08	2,02	1,98	1,95	1,87	1,83	1,78	1,69
22	2,13	2,06	2,01	1,97	1,93	1,86	1,81	1,76	1,67
23	2,11	2,05	2,00	1,95	1,92	1,84	1,80	1,74	1,66
24	2,10	2,04	1,98	1,94	1,91	1,83	1,78	1,73	1,64
25	2,09	2,02	1,97	1,93	1,89	1,82	1,77	1,72	1,63
26	2,08	2,01	1,96	1,92	1,88	1,81	1,76	1,71	1,61
27	2,07	2,00	1,95	1,91	1,87	1,80	1,75	1,70	1,60
28	2,06	2,00	1,94	1,90	1,87	1,79	1,74	1,69	1,59
29	2,06	1,99	1,93	1,89	1,86	1,78	1,73	1,68	1,58
30	2,05	1,98	1,93	1,88	1,85	1,77	1,72	1,67	1,57
40	2,00	1,93	1,87	1,83	1,79	1,71	1,66	1,61	1,51
60	1,95	1,87	1,82	1,77	1,74	1,66	1,60	1,54	1,44

## ВАРИАНТЫ КОНТРОЛЬНЫХ РАБОТ ДЛЯ ЗАОЧНИКОВ

Для приводимых ниже вариантов выборок решить следующие задачи:

**Задача 1.** Для приведенной выборки случайной величины  $\xi$  построить вариационный ряд и выборочный закон распределения  $\xi$ . Найти выборочное среднее  $\bar{x}$ , выборочную дисперсию  $D^*$  и исправленную выборочную дисперсию  $s^2$ .

**Задача 2.** Построить с надежностью  $\gamma = 0,90$  доверительные интервал для математического ожидания случайной величины  $\xi$ .

**Задача 3.** Построить с надежностью  $\gamma = 0,90$  доверительный интервал для дисперсии  $D[\xi]$  случайной величины  $\xi$  в предположении, что она имеет нормальное распределение.

**Задача 4.** Используя критерий согласия Пирсона, проверить гипотезу о нормальном распределении случайной величины  $\xi$  с уровнем значимости  $\alpha = 0,1$ .

**Задача 5.** Используя критерий согласия Пирсона, проверить гипотезу о равномерном законе распределения случайной величины  $\xi$  с уровнем значимости  $\alpha = 0,1$ .

### Вариант 1

Считать максимальную дневную температуру в Санкт-Петербурге 23 февраля случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

+2	-12	-5	-7	+2	-12	-9	-7	-3	-1
-5	-14	-7	-4	+3	-5	-6	-18	-12	-4
-1	-2	-4	-3	-11	-6	+4	+2	0	0
-1	-13	-4	-5	-4	+2	+2	+1	-1	-7
-10	0	-2	0	-5	+3	-5	+1	-3	-7

Решить задачи 1 – 5.

### Вариант 2

Считать максимальную дневную температуру в Санкт-Петербурге 8 марта случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

-3	+1	-5	+3	+1	+4	-7	-4	+1	+3
-3	-4	-3	-2	-4	+1	-3	-3	0	-3

+3	-8	-4	-4	-5	0	-2	+2	-5	-3
+1	+1	-2	0	+3	+5	-2	-8	+7	+1
+3	+5	0	0	-6	-3	0	-1	+2	+1

Решить задачи 1 – 5.

### Вариант 3

Считать максимальную дневную температуру в Санкт-Петербурге 1 апреля случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

0	+5	+13	+3	+14	+6	+8	+6	-1	-3
+6	+3	+3	+5	-4	+7	+3	+6	+12	+2
+12	+4	+2	+5	+9	0	+3	+8	+2	+3
+1	0	+1	-4	+3	-1	-1	+9	-1	+6
+1	+7	+2	-3	+2	0	+3	+4	+3	+4

Решить задачи 1 – 5.

### Вариант 4

Считать максимальную дневную температуру в Санкт-Петербурге 1 мая случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

11	5	10	12	22	4	16	4	14	10
19	12	4	5	20	10	14	6	13	13
7	15	9	9	2	13	24	24	14	11
12	7	6	12	10	8	21	10	24	9
21	24	8	9	10	5	3	7	11	10

Решить задачи 1 – 5.

### Вариант 5

Считать максимальную дневную температуру в Санкт-Петербурге 12 июня случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

14	25	18	19	15	18	19	19	19	16
23	11	12	23	30	16	25	17	11	11
22	10	15	15	16	18	22	15	28	16
17	22	15	10	23	21	22	20	12	13

21	20	8	24	20	21	24	10	18	19
----	----	---	----	----	----	----	----	----	----

Решить задачи 1 – 5.

### Вариант 6

Считать максимальную дневную температуру в Санкт-Петербурге 14 июля случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

20	23	32	19	23	19	23	26	18	28
24	22	20	32	22	17	21	28	24	21
19	23	33	27	22	17	22	24	19	21
22	19	17	29	24	25	22	18	19	22
23	25	26	30	18	24	18	20	23	24

Решить задачи 1 – 5.

### Вариант 7

Считать максимальную дневную температуру в Санкт-Петербурге 1 августа случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

18	17	25	21	23	19	21	21	24	28
28	17	19	23	21	23	23	23	29	22
15	25	25	22	27	20	17	18	19	20
15	22	21	24	32	21	23	22	24	21
20	26	19	26	22	20	20	24	22	23

Решить задачи 1 – 5.

### Вариант 8

Считать максимальную дневную температуру в Санкт-Петербурге 1 октября случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

13	12	12	6	8	18	9	18	13	12
7	10	18	14	6	8	10	12	20	12
13	20	12	18	10	13	10	9	12	9
9	8	11	13	13	7	11	12	14	13
13	11	4	11	10	12	6	7	12	11

Решить задачи 1 – 5.

### Вариант 9

Считать максимальную дневную температуру в Санкт-Петербурге 6 ноября случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

+3	+2	0	+2	-1	-4	+5	+6	+7	+2
+1	+8	0	+1	+1	+9	+8	+1	+6	+7
+5	+7	+6	+4	+1	+6	+10	+8	+4	+6
+5	+5	+1	+11	+2	0	+5	+5	+3	-5
+5	+2	+4	-9	+1	+3	+5	+2	+4	+3

Решить задачи 1 – 5.

### Вариант 10

Считать максимальную дневную температуру в Санкт-Петербурге 31 декабря случайной величиной  $\xi$ . Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (в градусах Цельсия):

0	-10	-8	+1	+5	+2	+1	+2	-2	-24
-8	-2	-5	-1	-3	+5	-9	-4	0	-1
0	-9	+2	+3	-7	-6	-9	-2	+2	-3
+2	-9	-6	-10	-10	+1	0	0	-10	-20
0	-2	-1	-2	+1	-4	-2	-6	-2	-3

Решить задачи 1 – 5.



# ОГЛАВЛЕНИЕ

## Часть II. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

<b>Введение.....</b>	<b>3</b>
<b>1. Основные понятия математической статистики.....</b>	<b>4</b>
1.1. Генеральная совокупность и выборка.....	4
1.2. Выборочный закон распределения (статистический ряд).....	4
1.3. Полигон частот, выборочная функция распределения .....	5
1.4. Свойства эмпирической функции распределения.....	7
1.5. Решение типового примера.....	8
<b>2. Числовые характеристики выборки.....</b>	<b>9</b>
2.1. Выборочное среднее и выборочная дисперсия, эмпирические моменты.....	9
2.2. Свойства статистических оценок параметров распределения: несмещенность, эффективность, состоятельность.....	10
2.3. Свойства выборочного среднего.....	11
2.4. Свойства выборочной дисперсии.....	13
2.5. Решение типового примера.....	14
<b>3. Точечное оценивание параметров известного распределения.....</b>	<b>16</b>
3.1. Метод моментов.....	16
3.2. Метод наибольшего правдоподобия.....	17
3.3. Решение типовых примеров.....	18
<b>4. Интервальное оценивание параметров известного распределения.....</b>	<b>23</b>
4.1. Оценивание математического ожидания нормально распреде- ленной величины при известной дисперсии.....	23
4.2. Оценивание математического ожидания нормально распреде- ленной величины при неизвестной дисперсии.....	24
4.3. Оценивание среднего квадратического отклонения нормально распределенной величины.....	27
4.3.1. Случай известного математического ожидания.....	27
4.3.2. Случай неизвестного математического ожидания.....	29
4.4. Оценивание математического ожидания случайной величины для произвольной выборки.....	29
4.5. Решение типовых примеров.....	30
<b>5. Проверка статистических гипотез.....</b>	<b>33</b>
5.1. Проверка гипотез о параметрах известного распределения.....	34
5.1.1. Проверка гипотезы о математическом ожидании нормально распределенной случайной величины.....	34
5.1.2. Сравнение дисперсий нормально распределенных случайных величин.....	37
5.1.3. Сравнение математических ожиданий независимых	

случайных величин.....	38
5.2. Проверка гипотез о виде закона случайной величины. Критерий Пирсона.....	41
5.2.1. Проверка гипотезы о нормальном распределении.....	41
5.2.2. Проверка гипотезы о равномерном распределении.....	43
5.3. Решение типовых примеров.....	44
<b>6. Важнейшие распределения и их квантили.....</b>	<b>47</b>
6.1. Нормальное распределение.....	47
6.2. Распределение Стьюдента.....	47
6.3. Распределение хи-квадрат.....	49
6.4. Распределение Фишера–Снедекора .....	51
6.5. Решение типовых примеров.....	52
<b>7. Пример статистической обработки выборки.....</b>	<b>55</b>
<b>Литература.....</b>	<b>61</b>
<b>Приложение.....</b>	<b>62</b>
<b>Варианты контрольных работ для заочников.....</b>	<b>68</b>