МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное 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

ГУАП
Санкт-Петербург
2013

УДК 519.2(075)
ББК 22.171я73
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.

Ф24 Основы теории вероятностей и математической статистики: учеб. пособие: в 2 ч. Ч. II. Математическая статистика / В. Г. Фарафонов, В. И. Устимов, В. Б. Ильин. – СПб.: ГУАП, 2013. – 80 с.: ил.
ISBN 978-5-8088-0788-4

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(075)
ББК 22.171я73

ISBN 978-5-8088-0788-4

ЧАСТЬ II. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

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, и 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.

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

Очевидно, что

$$\sum_{i=1}^{k} n_i = n; \qquad \sum_{i=1}^{k} \omega_i = 1.$$

Если мы запишем варианты в возрастающем порядке, то получим *вариационный ряд*. Таблица, состоящая из таких упорядоченных вариант 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
ω_1	ω_2	ω_3	...	ω_k

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

Выборочный закон распределения является аналогом закона (ряда) распределения дискретной случайной величины в теории вероятности, который был подробно рассмотрен в первой части нашего пособия.

Если вариационный ряд состоит из очень большого количества чисел или исследуется некоторый непрерывный признак, 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 приравнивают числу вариант, 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:

Номера подинтервалов	Границы подинтервалов	Варианты	Частоты
1	$[a, a+h]$	$a + h/2$	n_1
2	$(a+h, a+2h]$	$a + 3h/2$	n_2
...	...	...	...
s	$(b-h, b]$	$b - h/2$	n_s

где $h = (b - a)/s$, а n_i равно сумме частот вариант, 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 *полигоном частот*. Если вместо абсолютных значений n_i на оси ординат отложить относительные частоты ω_i, то получим полигон относительных частот (рис. 1.1).

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

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

где суммирование 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 ξ 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.

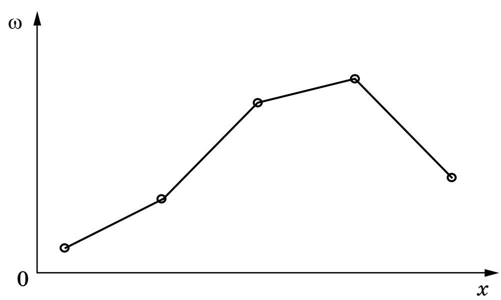

Рис. 1.1. Пример полигона относительных частот ($k = 5$)

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 $F_\xi(x)$:

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

Примечания:

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 ξ_1, значение второго – значением ξ_2 и т.д. Тогда при выполнении условий теоремы о том, что величины ξ_1, $\xi_2, \dots, \xi_n$ 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) $0 \le F_n^*(x) \le 1$;

2) $F_n^*(x)$ – неубывающая функция, т.е. $F_n^*(x_2) \ge F_n^*(x_1)$ при $x_2 > x_1$;

3) Если x_1 – наименьшая варианта, то $F_n^*(x) = 0$ при $x \le x_1$; 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):

Еще одним графическим представлением интересующей нас выборки является *гистограмма* – ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат подинтервалы шириной h, а высотами – отрезки длиной n_i/h (гистограмма частот)

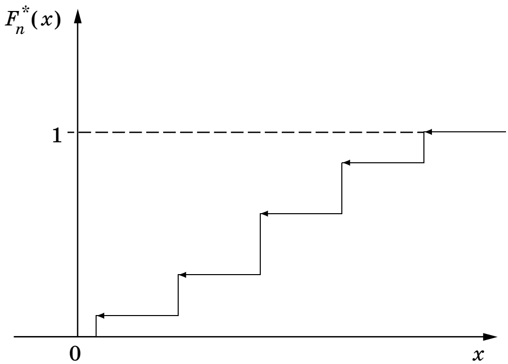

Рис. 1.2. Пример эмпирической функции распределения

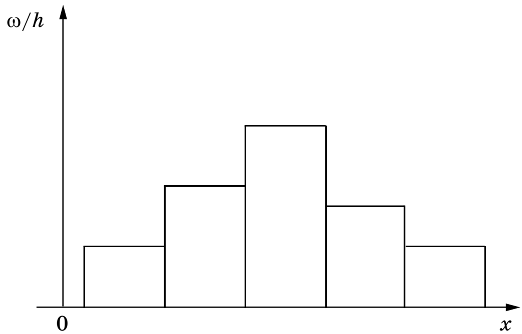

Рис. 1.3. Пример гистограммы относительных частот

или ω_i/h (гистограмма относительных частот). В первом случае площадь гистограммы равна объему выборки n, во втором – единице (рис. 1.3).

1.5. Решение типового примера

Пример 1.1. Проведено 30 серий по 24 броска кубика, и в каждой серии отмечалось число выпадений грани с шестью 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, надо построить выборочный закон распределения для результатов данного опыта.

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

x_i	n_i	ω_i	x_i	n_i	ω_i
1	2	2/30	5	5	5/30
2	4	4/30	6	3	3/30
3	5	5/30	7	1	1/30
4	9	9/30	8	1	1/30

ГЛАВА 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

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i = \sum_{i=1}^{k} \omega_i x_i, \qquad (2.1)$$

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

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

а *выборочным средним квадратическим отклонением* –

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

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

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

где

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

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 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 α_k и β_k теории вероятностей (см. п. 7 части I) *начальным эмпирическим моментом* порядка m называется величина

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

центральным эмпирическим моментом порядка m –

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

Заметим, что первый начальный момент α_1^* – это выборочное среднее $\overline{x}$, первый центральный момент β_1^* равен нулю, а второй центральный момент β_2^* – это выборочная дисперсия D^*.

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

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

Пусть A^* – статистическая оценка неизвестного параметра A теоретического распределения некоторой случайной величины ξ, например, A – дисперсия $D[\xi]$ некоторой случайной величины ξ, 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.

$$M[A^*] = A. \tag{2.8}$$

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 $n \to \infty$ она стремится по вероятности к оцениваемому параметру. Заметим, что несмещенная оценка будет состоятельной, если при $n \to \infty$ 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 $\xi_1, \xi_2, \ldots, \xi_n$, имеющих математическое ожидание $M[\xi_i] = a$ и дисперсию $D[\xi_i] = \sigma^2$. Тогда выборочное среднее можно рассматривать как случайную величину

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

принимающую 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

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

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

Для проверки эффективности этой оценки воспользуемся неравенством Крамера – Рао. Пусть θ_n^* – несмещенная оценка параметра θ случайной величины ξ, сделанная по выборке объемом n. Тогда для непрерывной случайной величины ξ с плотностью распределения вероятностей $f_\xi(x, \theta)$, зависящей от параметра θ, имеем

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

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

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

Несмещенная оценка параметра θ_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 ξ. Возвращаясь к выборочному среднему, в качестве примера рассмотрим случай, когда ξ имеет нормальный закон распределения.

Пусть в соотношениях (2.11), (2.12) рассматривается оценка математического ожидания (т.е. $\theta_n^* = \bar{x}$), а величина ξ 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):

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

$$= M\left[\left(\frac{\partial}{\partial a}\left(\ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(x-a)^2}{2\sigma^2}\right)\right)^2\right] = M\left[\left(\frac{x-a}{\sigma^2}\right)^2\right] =$$

$$= \frac{M\left[(x-a)^2\right]}{\sigma^4} = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}.$$

Таким образом, правая часть неравенства (2.11) равна σ^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 σ^2/n.

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

Заметим, что при доказательстве эффективности оценки мы воспользовались неравенством Крамера – Рао для непрерывной случайной величины. Аналогично может быть рассмотрена и дискретная случайная величина.

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

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

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

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

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

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

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

Таким образом, можно сделать вывод, что 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) $D^* = \overline{X^2} - \left(\bar{X}\right)^2$, поэтому

$$M\left[D^*\right] = M\left[\overline{X^2} - \left(\bar{X}\right)^2\right] = M\left[\frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)^2\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}M\left[\xi_i^2\right] - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}M\left[\xi_i\xi_j\right].$$

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

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

Поскольку ξ_i и ξ_j – независимые случайные величины, то имеем $M[\xi_i\xi_j] = M[\xi_i]\, M[\xi_j]$ (см. п. 8.6 части I). Дисперсия случайной величины ξ равна $\sigma^2 = \alpha_2 - (\alpha_1)^2$, 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

$$M\left[D^*\right] = \frac{1}{n}\sum_{i=1}^{n}\alpha_2 - \frac{1}{n^2}\sum_{i=1}^{n}\alpha_2 - \frac{1}{n^2}\sum_{i\neq j}\alpha_1^2 = \left(\frac{n-1}{n}\right)\alpha_2 - \left(\frac{n-1}{n}\right)\alpha_1^2 =$$

$$= \left(\frac{n-1}{n}\right)\left(\alpha_2 - \alpha_1^2\right) = \left(\frac{n-1}{n}\right)\sigma^2. \qquad \textbf{(2.14)}$$

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

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

что означает асимптотическую несмещенность этой оценки.

Примечания:

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

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

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

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

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

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

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	n_i	ω_i	x_i	n_i	ω_i
1	2	2/30	1	5	5/30
2	4	4/30	2	3	3/30
3	5	5/30	3	1	1/30
4	9	9/30	4	1	1/30

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

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

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

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 $s^2 = n/(n-1)\, D^* = 2{,}76$.

ГЛАВА 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 ξ описывается плотностью распределения вероятностей $f_\xi(x, \boldsymbol{\theta})$, а дискретной величины – таблицей, связывающей значения x_i с их вероятностями $p(\xi = x_i, \boldsymbol{\theta})$. В обоих случаях $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_r)$ – 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 $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_r)$, то и ее моменты зависят от этих же параметров, т.е. $\alpha_k = \alpha_k(\boldsymbol{\theta})$ и $\beta_k = \beta_k(\boldsymbol{\theta})$.

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

$$\alpha_k(\boldsymbol{\theta}) \approx \alpha_k^*; \quad \beta_k(\boldsymbol{\theta}) \approx \beta_k^*. \tag{3.1}$$

Метод моментов, развитый Карлом Пирсоном в 1894 г., основан на использовании этих приближенных равенств: моменты $\alpha_k(\boldsymbol{\theta})$, $\beta_k(\boldsymbol{\theta})$ рассчитываются теоретически по известному закону распределения с параметрами $\boldsymbol{\theta}$, а выборочные моменты α_k^*, β_k^* вычисляются по имеющейся выборке. Неизвестные параметры $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_r)$ определяются в результате решения системы из r

уравнений, связывающих соответствующие теоретический и эмпирический моменты, например, $\alpha_k(\boldsymbol{\theta}) = \alpha_k^*$ или $\beta_k(\boldsymbol{\theta}) = \beta_k^*$.

Можно показать [2], что оценки параметров $\boldsymbol{\theta}$, 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 гг.

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

$$L(x_1, x_2, \ldots, x_n, \boldsymbol{\theta}) = \begin{cases} \prod\limits_{i=1}^{n} f_{\xi_i}(x_i, \boldsymbol{\theta}), & \text{если } \xi_i \text{ непрерывны;} \\ \prod\limits_{i=1}^{n} p_{\xi i}(x_i, \boldsymbol{\theta}), & \text{если } \xi_i \text{ дискретны.} \end{cases} \tag{3.2}$$

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

$$L(x_1, x_2, \ldots, x_n, \tilde{\boldsymbol{\theta}}) = \max_{\boldsymbol{\theta}} \{L(x_1, x_2, \ldots, x_n, \boldsymbol{\theta})\}. \tag{3.3}$$

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

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

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

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 $n \to \infty$ 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.

Примечания:

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

2. Оценки параметров, рассмотренные нами, можно назвать *точечными оценками*, так как для неизвестного параметра θ 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 ξ объектов генеральной совокупности, являясь случайной величиной, имеет равномерное распределение, зависящее от параметров a и b:

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

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

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

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

Первый начальный момент α_1, равный математическому ожиданию, и второй центральный момент β_2, равный дисперсии, в случае равномерного распределения имеют вид (см. п. 7.5 части I):

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

Подставляя соотношения в систему (3.6), получим

$$\begin{cases} \dfrac{a+b}{2}=\overline{x}; \\ \dfrac{(b-a)^2}{12}=D^*=\left(\sigma^*\right)^2. \end{cases}$$

Решая систему, найдем

$$\begin{cases} a=\overline{x}-\sqrt{3}\sigma^*; \\ b=\overline{x}+\sqrt{3}\sigma^*. \end{cases} \tag{3.7}$$

Пример 3.2. Пусть характеристика объектов генеральной совокупности является случайной величиной ξ и имеет нормальное распределение с параметрами a и σ, 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 σ^2.

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

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

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

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

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

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

решая которое, получим:

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

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

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

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 нормального распределения a совпала с выборочным средним $\bar{x}$.

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

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

Логарифмическая функция правдоподобия имеет вид

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

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

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

Решая систему (3.12), получим

$$\begin{cases} \tilde{a} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i; \\ \tilde{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \tilde{a})^2. \end{cases} \qquad (3.13)$$

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

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

Получаем

$$\begin{cases} \dfrac{\partial^2}{\partial a^2} l(\mathbf{x}, a, \sigma^2) = -\dfrac{n}{\sigma^2}; \\ \dfrac{\partial^2}{\partial\left(\sigma^2\right)^2} l(\mathbf{x}, a, \sigma^2) = \dfrac{n}{2\left(\sigma^2\right)^2} - \dfrac{1}{\left(\sigma^2\right)^3}\displaystyle\sum_{i=1}^{n}(x_i - a)^2; \\ \dfrac{\partial^2}{\partial a \partial \sigma^2} l(\mathbf{x}, a, \sigma^2) = -\dfrac{1}{\left(\sigma^2\right)^2}\displaystyle\sum_{i=1}^{n}(x_i - a). \end{cases} \qquad (3.14)$$

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

$$\begin{cases} \dfrac{\partial^2}{\partial a^2} l(\mathbf{x},\tilde{a},\tilde{\sigma}^2) = -\dfrac{n}{\tilde{\sigma}^2}; \\ \dfrac{\partial^2}{\partial \left(\sigma^2\right)^2} l(\mathbf{x},\tilde{a},\tilde{\sigma}^2) = \dfrac{n}{2\left(\tilde{\sigma}^2\right)^2} - \dfrac{n}{\left(\tilde{\sigma}^2\right)^2} = -\dfrac{n}{2\left(\tilde{\sigma}^2\right)^2}; \\ \dfrac{\partial^2}{\partial a \partial \sigma^2} l(\mathbf{x},\tilde{a},\tilde{\sigma}^2) = 0. \end{cases} \tag{3.15}$$

Определитель равен

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

Находим, что

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

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 $\overline{x}$, а параметр σ^2 – выборочной дисперсией D^*. Заметим, что полученная оценка σ^2 является смещенной.

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

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

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

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:

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

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 λ в точке $\lambda=\tilde{\lambda}$ равна $-n\bar{x}\,/\,\tilde{\lambda}^2$ и меньше 0, так как $\bar{x}>0$, то экстремум является максимумом. Следовательно, в методе максимального правдоподобия параметр распределения Пуассона λ оценивается выборочным средним $\tilde{\lambda}=\bar{x}$. Этот результат был предсказуем, поскольку математическое ожидание случайной величины, распределенной подобным образом, равно λ (см. п. 7.2 части I).

ГЛАВА 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 θ строится неравенство

$$\left|\tilde{\theta}-\theta\right|<\delta, \qquad (4.1)$$

где число δ характеризует точность оценки: чем меньше δ, 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 δ), в который параметр θ попадает с достаточно большой вероятностью γ:

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

С соотношением (4.2) связаны следующие термины:

1) γ – вероятность, с которой выполняется неравенство, называется *надежностью* (*доверительной вероятностью*) оценки $\tilde{\theta}$ параметра θ;

2) $\alpha = (1 - \gamma)$ – вероятность противоположенного события, которую называют *уровнем значимости*;

3) интервал $\left(\tilde{\theta}-\delta<\theta<\tilde{\theta}+\delta\right)$, в который попадает неизвестный параметр θ с заданной надежностью γ, является *доверительным интервалом*.

4.1. Оценивание математического ожидания нормально распределенной величины при известной дисперсии

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

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

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}\xi_i;\quad M\left[\bar{X}\right]=\frac{1}{n}\sum_{i=1}^{n}M[\xi_i]=a;$$
$$\sigma\left[\bar{X}\right]=\sqrt{\frac{1}{n^2}\sum_{i=1}^{n}D[\xi_i]}=\frac{\sigma}{\sqrt{n}}. \tag{4.3}$$

При достаточно большом n согласно центральной предельной теореме случайная величина $\left(n\bar{X}-nM[\bar{X}]\right)/n\sigma[\bar{X}]$ 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)

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

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

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

Система (4.5) связывает надежность γ и доверительный интервал для математического ожидания, равный

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

Если известна величина γ, 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 γ.

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

4.2. Оценивание математического ожидания нормально распределенной величины при неизвестной дисперсии

Пусть исследуемая случайная величина ξ распределена по нормальному закону с неизвестными математическим ожиданием a и средним квадратическим отклонением σ.

Используя величины $\xi_1, \xi_2, ..., \xi_n$, 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 $\bar{X}$ – случайная величина, определенная в формулах (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

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

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

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

$$p\left(\left|t_n\right| < \tilde{t}_{\gamma,\, n-1}\right) = \int_{-\tilde{t}_{\gamma,\, n-1}}^{\tilde{t}_{\gamma,\, n-1}} f_t(x, n-1)\,\mathrm{d}x = 2\int_{0}^{\tilde{t}_{\gamma,\, n-1}} f_t(x, n-1)\,\mathrm{d}x = \gamma. \qquad (4.8)$$

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

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

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 σ здесь стоит исправленное выборочное среднее квадратическое отклонение s, а параметр $\tilde{t}_{\gamma,\, n-1}$ находится по известным γ 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 $x = 0$ (и $t_n = 0$). Тогда ширина интервала является наименьшей для данного значения γ. Площадь заштрихованной фигуры на рис. 4.1 равна вероятности γ, и в силу симметрии площади криволинейных трапеций слева и справа от заштрихованной фигуры совпадают друг с другом и равны $(1 - \gamma)/2$.

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)

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

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);

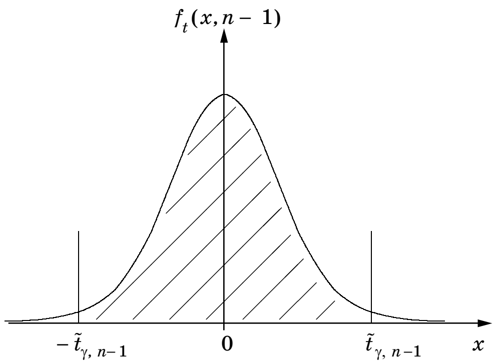

Рис. 4.1. Плотность распределения Стьюдента с n – 1 степенями свободы

2) как критическую точку распределения Стьюдента для уровня значимости $(1-\gamma)/2$, т.е. как число, для которого

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

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, определяемая по таблицам с подобными данными;

3) как критическую точку распределения Стьюдента для уровня значимости $1-\gamma$:

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

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 двусторонней области, определяемая по таблицам с соответствующими данными.

Примечание. При большом числе степеней свободы k распределение Стьюдента (4.7) стремится к нормальному распределению с нулевым математическим ожиданием и единичной дисперсией. Поэтому при $k \geq 30$ доверительный интервал можно на практике находить по формулам

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

4.3. Оценивание среднего квадратического отклонения нормально распределенной величины

Пусть исследуемая случайная величина ξ распределена по нормальному закону с математическим ожиданием a и неизвестным средним квадратическим отклонением σ. Рассмотрим два случая.

4.3.1. Частный случай известного математического ожидания

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

Используя величины $\xi_1, \xi_2, ..., \xi_n$, 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

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

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* χ^2 с n степенями свободы как сумма квадратов n независимых стандартных ($a = 0$, $\sigma = 1$) нормальных случайных величин. Свойства распределения χ^2 рассмотрены в п. 6.3.

Определим доверительный интервал из условия

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

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

Плотность распределения χ^2 является несимметричной функцией относительно начала координат, что приводит к особенностям в

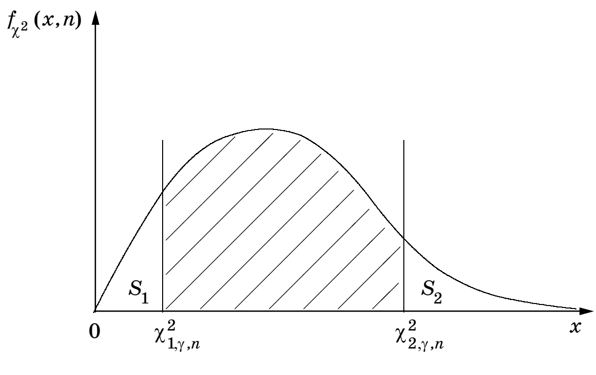

Рис. 4.2. Плотность распределения χ^2 с n степенями свободы

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 $\tilde{\chi}^2_{1,\gamma,n}$ и $\tilde{\chi}^2_{2,\gamma,n}$ используются таблицы квантилей распределения χ^2 (например, табл. 3 приложения), то значения $\tilde{\chi}^2_{1,\gamma,n}$ и $\tilde{\chi}^2_{2,\gamma,n}$ определяются по формулам:

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

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

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

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

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

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

т.е. доверительный интервал для дисперсии равен

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

Для среднего квадратического отклонения σ имеем

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

и доверительный интервал соответственно равен

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

4.3.2. Частный случай неизвестного математического ожидания

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

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 χ^2 с $n-1$ степенями свободы.

Рассмотрим равенство, аналогичное (4.15):

$$p\left(\tilde{\chi}^2_{1,\gamma,n-1} < H_n < \tilde{\chi}^2_{2,\gamma,n-1}\right) = \gamma. \tag{4.20}$$

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

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

Значения $\tilde{\chi}^2_{1,\gamma,n-1}$ и $\tilde{\chi}^2_{2,\gamma,n-1}$ определяются по таблицам квантилей или критических точек распределения χ^2 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 ξ (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 имеет место при выборке большого объема ($n >> 1$).

Как и выше, будем рассматривать варианты $x_1, x_2, \ldots, x_n$ как значения независимых, одинаково распределенных случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, 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_ξ и дисперсией σ_ξ^2 / n.

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

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

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

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

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

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

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 равенства (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

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

$2{,}872 < a < 3{,}528$, т. е. $2{,}872...3{,}528$ – доверительный интервал, в который попадает математическое ожидание a с надежностью 0,9.

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

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

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

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 выражению (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 σ^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 $\tilde{\chi}^2_{1,\gamma,n-1}$, $\tilde{\chi}^2_{2,\gamma,n-1}$ и исправленная выборочная дисперсия s^2.

Поскольку в приложении имеется только таблица квантилей распределения χ^2 (табл. 3), то значения параметров $\tilde{\chi}^2_{1,\gamma,n-1}$, $\tilde{\chi}^2_{2,\gamma,n-1}$ будем искать, используя соотношения (4.16): $\tilde{\chi}^2_{1,0{,}9;20} = \chi^2_{0{,}05;20} = 10{,}85$, $\tilde{\chi}^2_{2,0{,}9;20} = \chi^2_{0{,}95;20} = 31{,}41$.

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 σ^2 при надежности $\gamma = 0{,}9$ равен (5,35, 15,38).

ГЛАВА 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 α.

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 α и по известному закону распределения T вычисляют его критическое значение T_α. Оно делит область возможных значений T на две части: критическую область и область принятия гипотезы (например, области $T > T_\alpha$ и $T \le T_\alpha$);

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 β, 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 она больше, тем меньше вероятность совершить ошибку второго рода, поэтому после выбора уровня значимости следует строить критическую область так, чтобы мощность критерия была максимальной.

Примечания:

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 гл. 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 ξ имеет нормальное распределение. Требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу a_0. Рассмотрим отдельно случаи, когда дисперсия ξ 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

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

Для проверки этой гипотезы выберем следующий статистический критерий:

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

где учтено, что $\sigma\left[\bar{X}\right] = \sqrt{D[\bar{X}]}$ и в нашем случае $D\left[\bar{X}\right] = \sigma^2 / n$ (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, которая может состоять в том, что $M\left[\bar{X}\right] > a_0$, $M\left[\bar{X}\right] < a_0$ или $M\left[\bar{X}\right] \neq a_0$.

1. Если $M\left[\bar{X}\right] > a_0$, критическая область является *правосторонней*. На рис. 5.1 подобная критическая область заштрихована.

Поскольку можно считать, что случайная величина U распределена нормально, то критическое значение U_α 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

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

т.е. U_α 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_α. Если $U^* < U_\alpha$, то нулевую гипотезу H_0 можно принять, а если $U^* > U_\alpha$, то ее следует отвергнуть.

2. Если $M\left[\bar{X}\right] < a_0$, критическая область является *левосторонней* (рис. 5.2) и U_α определяется из условия $p(U < -U_\alpha) = \alpha$.

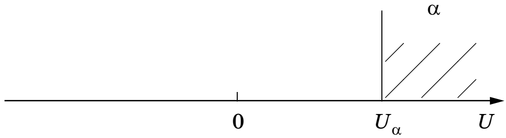

Рис. 5.1. Правосторонняя критическая область

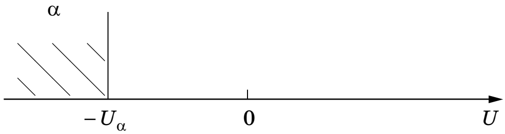

Рис. 5.2. Левосторонняя критическая область

Поскольку $p(U < -U_\alpha) = F_U(-U_\alpha) = \Phi(-U_\alpha) + 1/2 = 1/2 - \Phi(U_\alpha)$, то опять получаем

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

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

3. Если $M\left[\bar{X}\right] \neq a_0$, критическая область удовлетворяет условию $p(|U| > U_\alpha) = \alpha$ и является *двусторонней*, поскольку состоит из двух частей: $U < -U_\alpha$ и $U > U_\alpha$ (рис. 5.3).

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

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

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

В случае неизвестной дисперсии величины ξ в качестве статистического критерия выберем

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

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

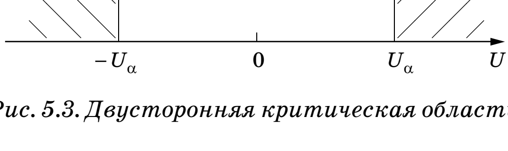

Рис. 5.3. Двусторонняя критическая область

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

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

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 ξ_1 и ξ_2. Для них по независимым выборкам объемом n_1 и n_2 соответственно получены исправленные выборочные дисперсии $s^2_{\xi_1}$ и $s^2_{\xi_2}$. Будем считать, что $s^2_{\xi_1} > s^2_{\xi_2}$. Требуется при заданном уровне значимости α проверить нулевую гипотезу H_0 о равенстве дисперсий рассматриваемых случайных величин.

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

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

где случайная величина

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

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

В качестве статистического критерия выберем случайную величину

$$F = \frac{Z_{\xi_1}}{Z_{\xi_2}}, \tag{5.8}$$

принимающую значение отношения бóльшей 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 бóльшая 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 $\left(s_{\xi_1}^2 > s_{\xi_2}^2\right.$ и $\left.s_{\xi_1}^2 \neq s_{\xi_2}^2\right)$.

1. Если $s_{\xi_1}^2 > s_{\xi_2}^2$, строят одностороннюю (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 числами степеней свободы k_1 и k_2 при уровне значимости α. 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. Если $s_{\xi_1}^2 \neq s_{\xi_2}^2$, строят двустороннюю критическую область из условия, что

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

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 $\alpha/2$:

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

Критическая точка $F_2 = F(\alpha/2, k_1, k_2)$ находится по таблице критических точек для распределения Фишера – Снедекора с числами 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.

Примечание. В случае, когда $\xi_1 = \xi_2 = \xi$, имеет место сравнение выборочных дисперсии одной случайной величины ξ. Такая задача возникает при необходимости 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 нормального распределения случайных величин с известными дисперсиями, а затем на его основе – более общий случай произвольного распределения величин при достаточно больших независимых выборках.

Пусть случайные величины ξ_1 и ξ_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 α 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, нулевую гипотезу H_0 можно записать в следующем виде:

$$M[\bar{X}_1] = M[\bar{X}_2]. \tag{5.9}$$

В качестве статистического критерия для проверки H_0 возьмем случайную величину

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

Можно показать, что при n_1, $n_2 \to \infty$ 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 вида конкурирующей гипотезы H_1 $\left(M[\xi_1] \neq M[\xi_2],\ M[\xi_1] > M[\xi_2],\ M[\xi_1] < M[\xi_2]\right)$.

1. Если $M[\xi_1] \neq M[\xi_2]$, строят двустороннюю критическую область. Как и в п. 5.1.1, критическая точка U_α находится из условия

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

Затем по имеющимся выборкам вычисляют эмпирическое значение критерия

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

и сравнивают его с теоретическим значением U_α:

а) если $|U^*| < U_\alpha$, нет оснований отвергать нулевую гипотезу;

б) если $|U^*| > U_\alpha$, нулевую гипотезу отвергают.

2. Если $M[\xi_1] > M[\xi_2]$, строят правостороннюю критическую область. 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.

Критическая точка U_α определяется из условия, аналогичного соотношению (5.2),

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

Вывод делается по результатам сравнения U^* и U_α:

а) если $U^* < U_\alpha$, нулевую гипотезу можно принять;

б) если $U^* > U_\alpha$, ее следует отвергнуть.

3. Если $M[\xi_1] < M[\xi_2]$, строят левостороннюю критическую область. Критическая точка U_α определяется, как и в п. 5.1.1, из условия

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

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

Сравним математические ожидания двух произвольно распределенных случайных величин ξ_1 и ξ_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 распределения самих случайных величин. Поскольку выборочные дисперсии являются при этом достаточно хорошими оценками дисперсий случайных величин, последние можно считать приближенно известными, т.е. $D[\xi_1] \approx D^*[\xi_1]$, $D[\xi_2] \approx D^*[\xi_2]$. Тогда можно полагать, что статистический критерий

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

имеет распределение, близкое к нормальному, причем:

а) если справедлива нулевая гипотеза $M[\xi_1] = M[\xi_2]$, то $M[\tilde{U}] \approx 0$;

б) если рассматриваемые выборки независимы, то $D[\tilde{U}] \approx 1$.

Дальнейшая проверка статистических гипотез проводится так же, как 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, 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 ξ и сделана выборка достаточно большого объема n с большим количеством различных значений вариант. Требуется при уровне значимости α проверить нулевую гипотезу H_0 о том, что случайная величина ξ распределена нормально.

Для удобства обработки выборки возьмем два числа α и β: $\alpha < x_1$, $\beta > x_n$ и разделим интервал $[\alpha, \beta]$ на s подинтервалов. Будем считать, что значения вариант, попавших в каждый подинтервал, приближенно равны числу, задающему середину подинтервала. Подсчитав число вариант, попавших в каждый интервал, составим группированную выборку с вариантами: $x_1, x_2, ..., x_s$ и их частотами $n_1, n_2, ..., n_s$, где $x_j = (b_j + a_j)/2$ – середина j-го подинтервала $(a_j, b_j]$; n_j – количество вариант, попавших в этот подинтервал, т.е. эмпирическая частота.

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

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

и выборочное среднее квадратическое отклонение σ^*:

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

Найдем количество вариант, которое должно оказаться в каждом подинтервале при выборке объемом n в предположении, что случайная величина ξ 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

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

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

Выберем статистический критерий, равный

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

где η_j – случайная величина, принимающая для данной выборки значение частоты n_j. Можно показать, что вне зависимости от реального закона распределения случайной величины ξ распределение H при $n \to \infty$ стремится к распределению χ^2 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 α – уровень значимости. Критическая точка $\chi^2(\alpha, k) = \chi^2_{1-\alpha,k}$, $\chi^2_{1-\alpha,k}$ – квантиль порядка $(1 - \alpha)$ распределения χ^2 с числом степеней свободы $k = s - 3$, которая определяется из соответствующих таблиц.

Значение критерия H^* для данной выборки равно

$$H^* = \sum_{j=1}^{s} \frac{(n_j - n_j')^2}{n_j'}. \tag{5,17}$$

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

Примечание. 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 ξ и сделана выборка достаточно большого объема n. Требуется при уровне значимости α проверить нулевую гипотезу H_0 о том, что случайная величина ξ имеет равномерное распределение, т.е.

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

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 и пример 3.1):

$$a^* = \overline{x} - \sqrt{3}\sigma^*, \quad b^* = \overline{x} + \sqrt{3}\sigma^*, \tag{5.19}$$

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

$$n_1' = n\,p_1 = n\frac{b_1 - a^*}{b^* - a^*};$$

$$n_j' = np_j = n\frac{b_j - b_{j-1}}{b^* - a^*} \text{ при } j = 2, 3, \ldots, s-1; \tag{5.20}$$

$$n_s = np_s = n\frac{b^* - b_{s-1}}{b^* - a^*},$$

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 χ^2 с числом степеней свободы $k = s - 3$, которую находим в соответствующих таблицах.

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

Примечание. 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 ξ – 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 ($M[\xi] \neq a_0$, $M[\xi] > a_0$, $M[\xi] < a_0$).

1. Если $M[\xi] \neq a_0$, то согласно п. 5.1.1 границы критической области задаются величиной U_α, 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. Если $M[\xi] > a_0$, граница критической области задается величиной U_α, 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. Если $M[\xi] < a_0$, граница критической области U_α 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, так как $U^* > -U_{0,1}$.

Пример 5.2. Две независимые выборки объемом $n_1 = 13$ и $n_2 = 18$ дали для нормально распределенных случайных величин ξ_1 и ξ_2 соответственно следующие исправленные выборочные дисперсии: $s_{\xi_1} = 1,52$ и $s_{\xi_2} = 0,60$. Проверить при уровне значимости $\alpha = 0,1$ нулевую гипотезу о равенстве дисперсий величин ξ_1 и ξ_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 бóльшей выборочной дисперсии к меньшей. Поскольку исправленная выборочная дисперсия для ξ_1 больше, чем для ξ_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 бóльшей исправленной выборочной дисперсии к меньшей:

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

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 ξ_1 и ξ_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: $\overline{x}_1 = 15,0$ и $\overline{x}_2 = 13,8$. 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_α, получаемым из соотношения (5.11). 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

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

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 ξ_1 и ξ_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 двусторонней критической области, связанной со значением U_α, определяемым из формулы (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 $\left|\tilde{U}^*\right| > U_\alpha$, то нулевую гипотезу отвергаем, т.е. найденные выборочные средние $\bar{x}_1 = 150$ и $\bar{x}_2 = 175$ различаются значимо.

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

ГЛАВА 6. ВАЖНЕЙШИЕ РАСПРЕДЕЛЕНИЯ И ИХ КВАНТИЛИ

6.1. Нормальное распределение

По определению нормально распределенная случайная величина ξ имеет плотность распределения вероятностей

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

где a и σ являются параметрами. Свойства таких случайных величин обсуждались в п. 7.5.5 ч. I.

Квантилью порядка α $(0 < \alpha < 1)$ непрерывной случайной величины ξ называется такое число x_α, для которого выполняется равенство $p(\xi < x_\alpha) = \alpha$.

Квантиль $x_{1/2}$ называется *медианой* случайной величины ξ, *квантили* $x_{1/4}$ и $x_{3/4}$ – ее *квартилями*, а $x_{0,1}, x_{0,2}, \ldots, x_{0,9}$ – *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 ч. I), а $\Phi(x)$ – функция Лапласа. Последняя 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_α для заданного α можно найти из соотношения

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

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

6.2. Распределение Стьюдента

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

$$t_n = \frac{\xi_0}{\sqrt{\dfrac{(\xi_1^2 + \ldots + \xi_n^2)}{n}}} \qquad (6.3)$$

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 (рис. 6.1)

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

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

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

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

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

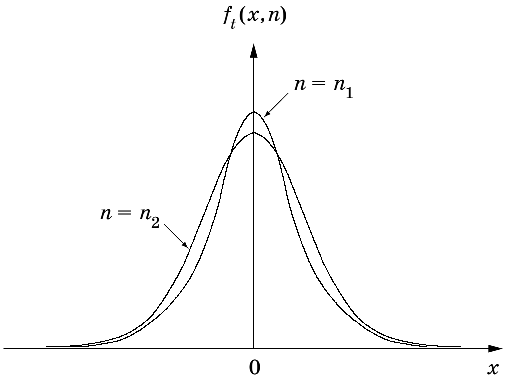

Рис. 6.1. Плотности распределения Стьюдента при $n_1 > n_2$

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

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

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

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

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

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

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

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

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

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

Критическая точка с уровнем значимости α 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 χ^2

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

$$H_n = \sum_{j=1}^{n} \xi_j^2 \tag{6.7}$$

называют *распределением* χ^2 с n степенями свободы. Обычно и для самой случайной величины H_n используется тот же символ, т.е. вместо H_n пишут χ^2.

Плотность распределения вероятностей H_n (или χ^2) является однопараметрической функцией, равной (рис. 6.2)

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

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

Заметим, что распределение χ^2 устойчиво относительно суммирования. Например, если независимые случайные величины H_{n_1}

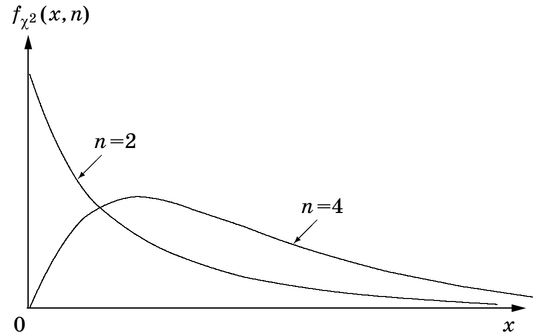

Рис. 6.2. Плотности распределения χ^2 с разным числом степеней свободы $n = 2$ и 4

и H_{n_2} имеют распределение χ^2 с n_1 и n_2 степенями свободы соответственно, то их сумма $H_{n_1} + H_{n_2}$ имеет распределение χ^2 с $n_1 + n_2$ степенями свободы.

Математическое ожидание и дисперсия H_n равны соответственно $M[H_n] = n$ и $D[H_n] = 2n$.

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 $n \to \infty$.

Квантилью порядка α распределения χ^2 называется такое число $\chi^2_{\alpha,n}$, для которого справедливо равенство $p(H_n < \chi^2_{\alpha,n}) = \alpha$.

Значения квантилей даны в виде таблиц в учебниках и специальных справочниках (например, приложение или [2]).

Для приближенного вычисления квантилей при больших n ($n \geq 30$) используют асимптотическую нормальность распределения χ^2. Поскольку при $n \to \infty$ 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_α – квантиль порядка α нормального распределения. Используя определение $\chi^2_{\alpha,n}$ и свойства неравенств, получаем

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

что дает

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

Таким образом, при больших n справедливо приближенное выражение

$$\chi^2_{\alpha,n} \approx n + x_\alpha\sqrt{2n}. \qquad (6.9)$$

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

$$\chi^2_{\alpha,n} \approx \frac{1}{2}(x_\alpha + \sqrt{2n-1})^2; \qquad (6.10)$$

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

Критическая точка $\chi^2(\alpha,\ n)$ определяется соотношением $p(H_n > \chi^2(\alpha,\ n)) = \alpha$ и связана с квантилью соотношением $\chi^2(\alpha, n) = \chi^2_{1-\alpha,\ n}$.

Квантили $\chi^2_{\alpha,n}$ распределения χ^2 с n степенями свободы приведены в табл. 3 приложения.

6.4. Распределение Фишера – Снедекора

В многомерном статистическом, дисперсионном, регрессионном и других видах анализа широко используется распределение Фишера – Снедекора – непрерывное двухпараметрическое распределение вероятностей с плотностью

$$f_F(x,\nu_1,\nu_2) = \frac{x^{-1}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}\left(x\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}\left(1+x\frac{\nu_1}{\nu_2}\right)^{-\frac{\nu_1+\nu_2}{2}}, \qquad (6.12)$$

где $x > 0$, $\nu_1 > 0$, $\nu_2 > 0$ 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 $F(\nu_1,\nu_2)$, имеющей такое распределение, равны

$$M[F(\nu_1,\nu_2)] = \frac{\nu_2}{\nu_2+1} \text{ при } \nu_2 > 2 \text{ и}$$

$$D[F(\nu_1,\nu_2)] = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)} \text{ при } \nu_2 > 4.$$

При целых значениях параметров, когда $\nu_1 = k_1$ и $\nu_2 = k_2$, где k_1, k_2 – целые числа, распределение Фишера – Снедекора есть распределение отношения $\frac{\chi^2_{k_1}}{k_1}\frac{k_2}{\chi^2_{k_2}}$, где $\frac{\chi^2_{k_1}}{k_1}$ и $\frac{\chi^2_{k_2}}{k_2}$ – независимые случайные величины, имеющие распределения χ^2 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 (например, в книге [2]).

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

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

где x_α – квантиль нормального распределения.

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

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

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

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

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 формулу (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 формуле (6.2), а именно: $\Phi(x_{0,95}) = 0{,}95 - 0{,}5 = 0{,}45$. 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 χ^2 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 $\chi^2_{0,9;\,30} = 40{,}256$.

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

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

Согласно соотношению (6.10)

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

Наконец, согласно формуле (6.11) имеем

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

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

Решение. Таблица 4б приложения дает значение критической точки $F(0{,}1;\, 40;\, 60) = 1{,}44$.

Применяя равенство (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

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

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

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)

$$F_{0,05}(10;\, 9) = \frac{1}{F_{0,95}(9, 10)} = \frac{1}{F(0{,}05; 9; 10)} = \frac{1}{3{,}02} = 0{,}331.$$

ГЛАВА 7. ПРИМЕР СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ВЫБОРКИ

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

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 ξ.

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

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

Выборочный закон распределения оказывается следующим:

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

Выборочное среднее равно

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

Найдем

$$\overline{x^2}=\frac{1}{n}\sum_{i=1}^{13}n_i x_i^2=\frac{1}{50}\{(10)^2+2(12)^2+\ldots+(26)^2\}=$$

$$=\frac{13966}{50}=279{,}32\ \text{град}^2$$

и получим выборочные дисперсии:

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

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

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 $\sigma^*\approx 3°$ 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 $\gamma = 0{,}90$ доверительный интервал для математического ожидания случайной величины ξ.

Решение. Будем считать, что рассматриваемая выборка достаточна велика $(n>>1)$ и применима формула (4.23):

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

в которой, учитывая полученные выше результаты, $\overline{x}=16{,}44^\circ\text{C}$, $s=\sqrt{9{,}231}\approx 3{,}038^\circ\text{C}$, а $2\Phi(\tilde{x}_\gamma)=0{,}9$ и $\tilde{x}_\gamma=1{,}65$.

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

$$p\left(16{,}44-\frac{3{,}038\cdot 1{,}65}{\sqrt{50}}<M[\xi]<16{,}44+\frac{3{,}038\cdot 1{,}65}{\sqrt{50}}\right)=$$

$$=p\big(15{,}73<M[\xi]<17{,}15\big)=0{,}9\,.$$

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

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

Решение. Здесь можно воспользоваться формулой (4.21):

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

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

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

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

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

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

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

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

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

$$p\left(\frac{49\times 9{,}231}{66{,}40} < D[\xi] < \frac{49\times 9{,}231}{33{,}89}\right) =$$

$$= p\left(6{,}81 < D[\xi] < 13{,}35\ \text{град}^2\right) = 0{,}90.$$

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

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

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

$[a_j, b_j]$	x_j	n_j	$[a_j, b_j]$	x_j	n_j
[9,5; 13,5]	11,5	8	[16,5; 17,5]	17	7
[13,5; 14,5]	14	5	[17,5; 18,5]	18	6
[14,5; 15,5]	15	6	[18,5; 19,5]	19	5
[15,5; 16,5]	16	8	[19,5; 26,5]	23	5

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

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):

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

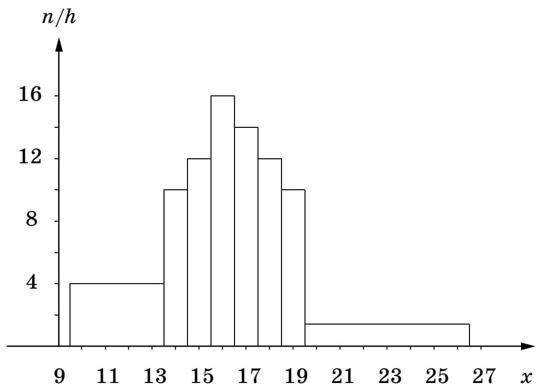

Рис. 7.1. Гистограмма частот группированной выборки

Тогда

$$n_1' = n\left(\Phi\left(\frac{13{,}5-16{,}44}{\sqrt{9{,}0464}}\right)-\Phi\left(\frac{9{,}5-16{,}44}{\sqrt{9{,}0464}}\right)\right)=$$

$$=50\left(\Phi\left(\frac{-2{,}94}{3{,}01}\right)-\Phi\left(\frac{-6{,}94}{3{,}01}\right)\right)=50(\Phi(-0{,}98)-$$

$$-\Phi(-2{,}31))=50(-0{,}336+0{,}490)=50\cdot 0{,}154=7{,}7;$$

$$n_2' = 50\left(\Phi\left(\frac{14{,}5-16{,}44}{3{,}01}\right)-\Phi\left(\frac{13{,}5-16{,}44}{3{,}01}\right)\right)=50(\Phi(-0{,}64)-$$

$$-\Phi(-0{,}98))=50(-0{,}239+0{,}336)=50\cdot 0{,}097=4{,}85;$$

$$n_3' = 50\left(\Phi\left(\frac{15{,}5-16{,}44}{3{,}01}\right)-\Phi\left(\frac{14{,}5-16{,}44}{3{,}01}\right)\right)=50(\Phi(-0{,}31)-$$

$$-\Phi(-0{,}64))=50(-0{,}122+0{,}239)=50\cdot 0{,}117=5{,}85;$$

$$n_4' = 50\left(\Phi\left(\frac{16{,}5-16{,}44}{3{,}01}\right)-\Phi\left(\frac{15{,}5-16{,}44}{3{,}01}\right)\right)=50(\Phi(0{,}02)-$$

$$-\Phi(-0{,}31))=50(0{,}008+0{,}122)=50\cdot 0{,}130=6{,}5;$$

$$n_5' = 50\left(\Phi\left(\frac{17{,}5-16{,}44}{3{,}01}\right)-\Phi\left(\frac{16{,}5-16{,}44}{3{,}01}\right)\right)=50(\Phi(0{,}35)-$$

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

$$n_6' = 50\left(\Phi\left(\frac{18{,}5-16{,}44}{3{,}01}\right)-\Phi\left(\frac{17{,}5-16{,}44}{3{,}01}\right)\right)=50(\Phi(0{,}68)-$$

$$-\Phi(0{,}35))=50(0{,}252-0{,}134)=50\cdot 0{,}118=5{,}9;$$

$$n_7' = 50\left(\Phi\left(\frac{19{,}5-16{,}44}{3{,}01}\right)-\Phi\left(\frac{18{,}5-16{,}44}{3{,}01}\right)\right)=50(\Phi(1{,}02)-$$

$$-\Phi(0{,}68))=50(0{,}346-0{,}252)=50\cdot 0{,}094=4{,}7;$$

$$n_8' = 50\left(\Phi\left(\frac{26{,}5-16{,}44}{3{,}01}\right)-\Phi\left(\frac{19{,}5-16{,}44}{3{,}01}\right)\right)=50(\Phi(3{,}34)-$$

$$-\Phi(1{,}02))=50(0{,}5000-0{,}346)=50\cdot 0{,}154=7{,}7.$$

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

$[a_j, b_j]$	n_j	n'_j	$(n_j - n'_j)^2/n'_j$	$[a_j, b_j]$	n_j	n'_j	$(n_j - n'_j)^2/n'_j$
[9,5; 13,5]	8	7,7	0,012	[16,5; 17,5]	7	6,3	0,078
[13,5; 14,5]	5	4,85	0,005	[17,5; 18,5]	6	5,9	0,002
[14,5; 15,5]	6	5,85	0,004	[18,5; 19,5]	5	4,7	0,019
[15,5; 16,5]	8	6,5	0,346	[19,5; 26,5]	5	7,7	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):

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

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 χ^2, т.е. $\chi^2_{1-\alpha,k}$. 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 ξ при уровне значимости $\alpha = 0{,}1$.

Пример 7.5. Используя критерий согласия Пирсона, проверить гипотезу о равномерном законе распределения случайной величины ξ 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): $a^* = \overline{x} - \sqrt{3}\sigma^*$; $b^* = \overline{x} + \sqrt{3}\sigma^*$.

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:

$[a_j, b_j]$	n_j	n'_j	$(n_j - n'_j)^2 / n'_j$	$[a_j, b_j]$	n_j	n'_j	$(n_j - n'_j)^2 / n'_j$
[9,5; 13,5]	8	10,89	0,767	[16,5; 17,5]	7	4,80	1,008
[13,5; 14,5]	5	4,80	0,008	[17,5; 18,5]	6	4,80	0,300
[14,5; 15,5]	6	4,80	0,300	[18,5; 19,5]	5	4,80	0,008
[15,5; 16,5]	8	4,80	2,133	[19,5; 26,5]	5	10,32	2,742

Наблюдаемое (эмпирическое) значение критерия согласно формуле (5.16), 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 χ^2 с числом степеней свободы $k = s - 3$. В нашем случае $k = 8 - 3 = 5$ и $\alpha = 0{,}1$, и из табл. 3 приложения находим $\chi^2_{0,9;5} = 9{,}236$.

Поскольку эмпирическое значение критерия не попадает в критическую область, определяемую как $H^* > \chi^2_{0,9;5}$, 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.

Таблица 1

Значения функции Лапласа Ф(*x*)

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

Таблица 2

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

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

Таблица 3

Квантили $\chi^2_{\alpha,n}$ распределения χ^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

$n\backslash\alpha$	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

Таблица 4а

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

$k_2 \backslash k_1$	5	6	7	8	9	12	15	20	40
5	5,05	4,95	4,88	4,82	4,77	4,68	4,62	4,56	4,46
6	4,39	4,28	4,21	4,15	4,10	4,00	3,94	3,87	3,77
7	3,97	3,87	3,79	3,73	3,68	3,57	3,51	3,44	3,34
8	3,69	3,58	3,50	3,44	3,39	3,28	3,22	3,15	3,04
9	3,48	3,37	3,29	3,23	3,18	3,07	3,01	2,94	2,83
10	3,33	3,22	3,14	3,07	3,02	2,91	2,85	2,77	2,66
11	3,20	3,09	3,01	2,95	2,90	2,79	2,72	2,65	2,53
12	3,12	3,00	2,91	2,85	2,80	2,69	2,62	2,54	2,43
13	3,03	2,92	2,83	2,77	2,71	2,60	2,53	2,46	2,34
14	2,96	2,85	2,76	2,70	2,65	2,53	2,46	2,39	2,27
15	2,90	2,79	2,71	2,64	2,59	2,48	2,40	2,33	2,20
16	2,85	2,74	2,66	2,59	2,54	2,42	2,35	2,28	2,15
17	2,81	2,70	2,61	2,55	2,49	2,38	2,31	2,23	2,10
18	2,77	2,66	2,58	2,51	2,46	2,34	2,27	2,19	2,06
19	2,74	2,63	2,54	2,48	2,42	2,31	2,23	2,16	2,03
20	2,71	2,60	2,51	2,45	2,39	2,28	2,20	2,12	1,99
21	2,68	2,57	2,49	2,42	2,37	2,25	2,18	2,10	1,96
22	2,66	2,55	2,46	2,40	2,34	2,23	2,15	2,07	1,94
23	2,64	2,53	2,44	2,37	2,32	2,20	2,13	2,05	1,91
24	2,62	2,51	2,42	2,36	2,30	2,18	2,11	2,03	1,89
25	2,60	2,49	2,40	2,34	2,28	2,16	2,09	2,01	1,87
26	2,59	2,47	2,39	2,32	2,27	2,15	2,07	1,99	1,85
27	2,57	2,46	2,37	2,31	2,25	2,13	2,06	1,97	1,84
28	2,56	2,45	2,36	2,29	2,24	2,12	2,04	1,96	1,82
29	2,55	2,43	2,35	2,28	2,22	2,10	2,03	1,94	1,81
30	2,53	2,42	2,33	2,27	2,21	2,09	2,01	1,93	1,79
40	2,45	2,34	2,25	2,18	2,12	2,00	1,92	1,84	1,69
60	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 ξ построить вариационный ряд и выборочный закон распределения ξ. Найти выборочное среднее $\overline{x}$, выборочную дисперсию D^* и исправленную выборочную дисперсию s^2.

Задача 2. Построить с надежностью $\gamma = 0{,}90$ доверительный интервал для математического ожидания случайной величины ξ.

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

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

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

Вариант 1

Считать максимальную дневную температуру в Санкт-Петербурге 23 февраля случайной величиной ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

+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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

−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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

+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 ξ. Из генеральной совокупности – данных Гидрометеослужбы о такой температуре в разные годы сделана следующая выборка (°C):

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.

ЛИТЕРАТУРА

1. *Лексаченко В. А.* Логика. Множества. Вероятность. СПб.: СПбГУАП, 2005. 135 с.

2. *Гмурман В. Е.* Теория вероятностей и математическая статистика. М.: Высшее образование, 2006. 575 с.

3. *Гмурман В. Е.* Руководство к решению задач по теории вероятностей и математической статистике. М.: Высшее образование, 2006. 476 с.

ОГЛАВЛЕНИЕ

Учебное издание

Фарафонов Виктор Георгиевич
Устимов Владимир Иванович
Ильин Владимир Борисович

ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Учебное пособие

Часть II

Редактор *Л. А. Яковлева*
Верстальщик *С. Б. Мацапура*

Сдано в набор 21.01.13. Подписано к печати 23.01.13.
Формат 60×84 1/16. Бумага офсетная. Усл. печ. л. 4,6.
Уч.-изд. л. 5,0. Тираж 100 экз. Заказ № .

Редакционно-издательский центр ГУАП
190000, Санкт-Петербург, Б. Морская ул., 67