Московский государственный технический университет
им. Н. Э. Баумана

Г. Д. Карташов, И. В. Павлов, В. И. Тимонин

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ ТИПОВОГО РАСЧЕТА ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Издательство МГТУ им. Н. Э. Баумана
1995

Московский государственный технический университет
имени Н.Э. Баумана

Г.Д. Карташов, И.В. Павлов, В.И. Тимонин

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ ТИПОВОГО РАСЧЕТА ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Издательство МГТУ им. Н.Э.Баумана
1995

ББК 22.172
К27

Рецензент О.И.Тескин

К27 Г.Д.Карташов, И.В.Павлов, В.И.Тимонин. Методические указания к выполнению типового расчета по математической статистике. М.: Изд-во МГТУ, 1995. - 54 с., ил.

Дан необходимый теоретический материал и методические указания к решению задач по точечному оцениванию неизвестных параметров, а также задач на построение доверительных интервалов. Подробно рассмотрены соответствующие типовые примеры. Приведены варианты задач для самостоятельного выполнения типового расчета. Для студентов технических вузов.

Ил. 1. Табл. 4. Библиогр. 9 назв.

ББК 22.172

Редакция заказной литературы

Геннадий Дмитриевич Карташов
Игорь Валерианович Павлов
Владимир Иванович Тимонин

Методические указания к выполнению типового расчета по математической статистике

Заведующая редакцией Н.Г.Ковалевская
Редактор Л.М.Элькинд
Корректор Л.И.Малютина

Подписано в печать 29.03.95.Формат 60х84/16. Бумага тип. № 2.
Печ.л. 3,5. Усл.печ.л. 3,26. Уч.-изд.л. 3,12. Тираж 1000 экз.
Заказ №265 Изд. № 10.
Издательство МГТУ, типография МГТУ.
107005, Москва, 2-я Бауманская, 5.

ВВЕДЕНИЕ

1. Основные задачи математической статистики

Математическая статистика, как и теория вероятностей, занимается изучением математических моделей случайных явлений или экспериментов. В то же время задачи математической статистики являются как бы обратными к задачам теории вероятностей. В теории вероятностей после задания математической модели того или иного случайного явления требуется рассчитать различные вероятностные характеристики этого явления в рамках данной модели. В математической статистике, исходя из уже имеющихся результатов эксперимента, называемых статистическими данными, требуется определить или уточнить математическую модель данного явления или найти те или иные параметры, ее определяющие. Перечислим кратко основные задачи математической статистики на простом примере стандартной классической схемы независимых испытаний Бернулли.

Далее рассмотрим задачи математической статистики на примере схемы Бернулли. Предположим, проводится серия из n независимых одинаковых экспериментов, в каждом из которых возможны два исхода - "успех" с вероятностью p и "неуспех" с вероятностью $1-p$. Данная модель полностью задается значением параметра - вероятности "успеха" p в одном испытании. Если параметр p известен, то различные вероятностные характеристики, такие как вероятность наблюдать m "успехов" в серии из n испытаний, математическое ожидание числа "успехов" и др., легко могут быть найдены по известным формулам теории вероятностей.

Задачи математической статистики возникают в тех случаях, когда истинное значение параметра p неизвестно, а известны результаты эксперимента - число m наблюдаемых "успехов" в серии из n испытаний. Требуется, исходя из результатов эксперимента, произвести те или иные выводы относительно неизвестного истинного значения параметра p и, следовательно, относительно исходной математической модели, описывающей данный эксперимент. При этом в математической статистике рассматриваются следующие основные задачи.

1. Задачи оценки неизвестных параметров, т.е. в данном случае параметра p, по результатам эксперимента. При этом требуется найти такую функцию от результатов эксперимента, в данном случае от величины m, которая могла бы служить "достаточ-

но хорошей" оценкой неизвестного истинного значения параметра p. Такой оценкой в данном случае является стандартная оценка вероятности некоторого события на наблюдаемой частоте этого события $\hat{p} = m/n$.

2. Задачи интервального оценивания или, другими словами, оценивания параметров с помощью доверительных интервалов. При этом требуется построить интервал со случайными границами, зависящими от результатов эксперимента $\underline{p} = \underline{p}(m)$, $\bar{p} = \bar{p}(m)$, таким образом, чтобы построенный случайный интервал накрывал неизвестное истинное значение параметра с заданной достаточно высокой вероятностью:

$$P\{\underline{p} \le p \le \bar{p}\} = \gamma,$$

где коэффициент доверия γ выбирается близким к I.

3. Задачи проверки статистических гипотез. В этом случае требуется на основе результатов эксперимента проверить то или иное предположение или гипотезу, например, гипотезу о том, что неизвестное истинное значение параметра удовлетворяет неравенству $p \le p_0$ или равенству $p = p_0$, где p_0 - некоторое заданное число, и т.п.

§ 2. Выборочный метод. Генеральная совокупность

В математической статистике часто используется "выборочная" терминология, основанная на следующей "урновой" схеме. Предположим, что имеется урна, содержащая N чисел

$$X_1, X_2, \ldots, X_N, \tag{I}$$

которые скрыты от наблюдателя. Набор всех чисел (I) называется генеральной совокупностью.

Пусть далее из указанной генеральной совокупности случайным образом выбирается n чисел

$$x_1, x_2, \ldots, x_n, \tag{2}$$

где $n \le N$. Набор чисел (2) называется выборкой объема из генеральной совокупности (I).

В терминах данной схемы задачи математической статистики могут формулироваться следующим образом. Требуется на основе выборки (2) произвести те или иные выводы относительно всей генеральной совокупности (I).

Выборка может производиться двумя основными способами: без возвращения и с возвращением. Если выборка производится с возвращением, то случайные величины (2) независимы. В этом случае говорят, что набор чисел (2) является независимой повторной случайной выборкой объема n из генеральной совокупности (I).

Указанная терминология сохраняется и в случае $N = \infty$, когда исходная генеральная совокупность содержит бесконечный набор чисел

$$X_1, X_2, \ldots, X_N, X_N, X_{N+1}, \ldots \tag{3}$$

В этом случае говорят о бесконечной генеральной совокупности (3) в отличие от конечной генеральной совокупности (I).

Схема с бесконечной генеральной совокупностью используется в математической статистике либо в тех случаях, когда объем генеральной совокупности конечен, но очень велик, например, при статистических исследованиях в экономике, социологии и т.п., либо в тех случаях, когда числа выборки являются повторными реализациями некоторой случайной величины, которая может наблюдаться сколько угодно раз. Нетрудно видеть, что при $N \to \infty$ разница между выборками с возвращением и без возвращения фактически исчезает.

Числа выборки (2), расположенные в порядке возрастания, обозначаются

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n-1)} \le x_{(n)} \tag{4}$$

и называются вариационным рядом.

П р и м е р I. В результате пяти повторных независимых наблюдений некоторой случайной величины получены следующие ее значения: $x_1 = 104$, $x_2 = 95$, $x_3 = 93$, $x_4 = 107$, $x_5 = 101$. Для данной выборки объема $n = 5$ вариационный ряд имеет вид $x_{(1)} = 93$, $x_{(2)} = 95$, $x_{(3)} = 101$, $x_{(4)} = 104$, $x_{(5)} = 107$.

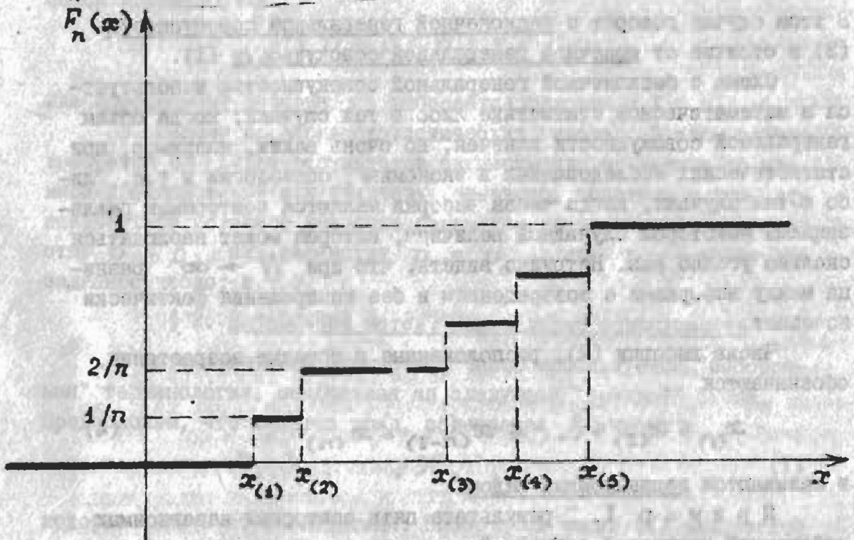
Эмпирической функцией распределения, построенной на основе выборки (2), называется следующая функция:

$$\hat{F}_n(x) = \frac{r(x)}{n},$$

где $r(x)$ - количество таких чисел x_i в выборке, для которых $x_i < x$. Эмпирическая функция распределения может быть также записана с помощью вариационного ряда в следующем виде:

$$\hat{F}_n(x) = \begin{cases} 0, & x \le x_{(1)}, \\ \dfrac{j}{n}, & x_{(j)} < x \le x_{(j+1)},\ j = 1, \ldots, n-1, \\ 1, & x_{(n)} < x. \end{cases}$$

П р и м е р I.2. Построим эмпирическую функцию распределения для выборки объема $n = 5$, приведенной выше в примере I.I. В этом случае график эмпирической функции распределения имеет вид, указанный на рисунке.



Пусть $F(x)$ - истинная или теоретическая функция распределения для данной генеральной совокупности. Тогда в соответствии с известной теоремой Гливенко - Кантелли при $n \to \infty$ с вероятностью I выполняется соотношение

$$\sup_x |\hat{F}_n(x) - F(x)| \to 0,$$

т.е. при увеличении объема выборки эмпирическая функция распределения сходится к теоретической.

Построенные на основе эмпирической функции распределения моменты называются эмпирическими или выборочными моментами.

На основе выборки (2) строятся также следующие характеристики. Величина

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{5}$$

называется эмпирическим или выборочным средним. Величина

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{6}$$

называется эмпирической или выборочной дисперсией. Величина

$$\hat{\sigma} = S = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{7}$$

называется эмпирическим или выборочным среднеквадратичным отклонением.

Величины (6), (7) позволяют оценить разброс выборки относительно ее среднего значения $\bar{x}$. Еще одной часто используемой характеристикой разброса выборки является размах выборки

$$R_n = x_{(n)} - x_{(1)}.$$

Наряду с перечисленными эмпирическими характеристиками используются также эмпирические (выборочные) моменты $\hat{\mu}_r$ и центральные моменты $\hat{\nu}_r$ порядка r, которые вычисляются по формулам

$$\hat{\mu}_r = \frac{1}{n} \sum_{i=1}^{n} x_i^r, \quad \hat{\nu}_r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^r.$$

Выборочные среднее и дисперсия являются частными случаями этих характеристик: $\bar{x} = \hat{\mu}_1$, $S^2 = \hat{\nu}_2$.

П р и м е р I.3. Найдем выборочные среднее, дисперсию, среднеквадратичное отклонение и размах для выборки объема $n = 5$ из примера I.I. В этом случае указанные выборочные характеристики равны соответственно:

$$\bar{x} = \frac{1}{5}(93 + 95 + 101 + 104 + 107) = 100;$$

$$S^2 = \frac{1}{5}(7^2 + 5^2 + 1^2 + 4^2 + 7^2) = 28;$$

$$\hat{\sigma} = S = \sqrt{28} \cong 5{,}3;$$

$$R_n = 107 - 93 = 14.$$

П р и м е ч а н и е I.I. Если в выборке имеются совпадающие числа, другими словами, число x_i встречается в выборке n_i раз, $i = 1, \dots, m$; $n_1 + n_2 + \dots + n_m = n$, где m – число различных между собой членов выборки, то формулы для выборочных среднего и дисперсии могут быть записаны в следующем виде:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{m} n_i x_i ; \tag{8}$$

$$S^2 = \frac{1}{n} \sum_{i=1}^{m} n_i (x_i - \bar{x})^2 . \tag{9}$$

П р и м е р I.4. Найдем выборочные среднее и дисперсию и среднеквадратичное отклонение для следующей выборки объема $n = 100$:

x_i	I250	I270	I280	I290
n_i	20	25	50	5

В этом случае число различных между собой членов выборки $m = 4$, и указанные выборочные характеристики определяются как

$$\bar{x} = \frac{1}{100}(20 \cdot 1250 + 25 \cdot 1270 + 50 \cdot 1280 + 5 \cdot 1290) = 1272;$$

$$S^2 = \frac{1}{100}(20 \cdot (22^2) + 25 \cdot 2^2 + 50 \cdot 8^2 + 5 \cdot 18^2) = 125;$$

$$\hat{\sigma} = S = \sqrt{125} = 11{,}2.$$

П р и м е ч а н и е I.2. Если числа выборки велики, то при вычислении выборочных характеристик удобно сначала записать их в виде $x_i = \alpha + y_i$, где α – константа, близкая к среднему (например, для предыдущей выборки полагаем $\alpha = 1270$), после чего формула (8) для выборочного среднего упрощается

$$\bar{x} = \alpha + \frac{1}{n} \sum_{i=1}^{m} n_i y_i . \tag{10}$$

При вычислении выборочной дисперсии часто вместо (9) удобно использовать следующие формулы:

$$S^2 = \frac{1}{n} \sum_{i=1}^{m} n_i x_i^2 - (\bar{x})^2 \tag{11}$$

или, если числа x_i велики,

$$S^2 = \frac{1}{n} \sum_{i=1}^{m} n_i y_i^2 - (\bar{x} - \alpha)^2, \tag{12}$$

где α – указанная выше константа.

Задачи

I. Для следующей выборки объема $n = 9$:

343, 347, 284, 347, 354, 334, 295, 343, 296

построить график эмпирической функции распределения, найти выборочные среднее, дисперсию, среднеквадратичное отклонение.

2. Доказать формулы (I0)...(I2).

Глава I. ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ

Предположим, что имеется некоторая случайная величина ξ с функцией распределения $F(x, \theta)$ и плотностью распределения $f(x, \theta)$, зависящими от некоторого многомерного параметра θ, истинное значение которого неизвестно. Требуется оценить параметр θ по результатам n повторных независимых наблюдений величины ξ или, другими словами, на основе повторной независимой случайной выборки

$$x_1, x_2, \dots, x_n \tag{13}$$

объема n из генеральной совокупности с функцией распределения $F(x, \theta)$.

Совокупность всех распределений $F(x, \theta)$ при различных возможных значениях параметра θ называется параметрическим семейством распределений. (Основные параметрические семейства распределений, используемые в математической статистике, подробнее рассматриваются в приложении.)

Задача построения точечной оценки параметра θ сводится к нахождению такой функции от результатов наблюдений

$$\hat{\theta} = \hat{\theta}(x_1, \dots, x_n), \tag{14}$$

что случайная величина (или случайная точка) $\hat{\theta}$ в некотором смысле близка к неизвестному истинному значению параметра θ, если выборка взята из распределения $F(x, \theta)$ с этим значением параметра.

Заметим, что любая функция от результатов наблюдений $\varphi = \varphi(x_1, \ldots, x_n)$ называется статистикой. Таким образом, точечная оценка параметра (I4) является некоторой статистикой, обладающей теми или иными свойствами, гарантирующими хорошее качество оценки, или близость ее к неизвестному истинному значению параметра. Рассмотрим подробнее эти свойства, определяющие качество оценки.

§ 3. Несмещенные оценки

Оценка $\hat{\theta}$ называется несмещенной оценкой параметра θ, если ее математическое ожидание совпадает со значением параметра $M\hat{\theta} = \theta$ при любом возможном значении θ.

Пример 2.I. Рассмотрим классическую схему независимых испытаний Бернулли. Производится серия из n независимых одинаковых экспериментов, в каждом из которых с некоторой неизвестной вероятностью p наблюдается событие, называемое "успехом" и с вероятностью $1-p$ – противоположное событие, называемое "неуспехом". Пусть в серии из n экспериментов наблюдалось m "успехов". Тогда стандартной оценкой параметра p является частота появления указанного события $\hat{p} = m/n$. Эта оценка является несмещенной, так как

$$M\hat{p} = M\left(\frac{m}{n}\right) = \frac{1}{n} Mm = \frac{1}{n} \cdot np = p.$$

Пример 2.2. Рассмотрим оценку параметра $\mu = M\xi$ математического ожидания некоторой случайной величины ξ по результатам n независимых повторных ее наблюдений $x_1, \ldots, x_n$. Выборочное среднее $\bar{x}$ является несмещенной оценкой параметра μ, так как

$$M\bar{x} = M\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} Mx_i = \frac{1}{n}(n\mu) = \mu.$$

Пример 2.3. В условиях предыдущего примера рассмотрим вопрос о том, является ли выборочная дисперсия S^2 в формуле (6) несмещенной оценкой дисперсии $\sigma^2 = M(\xi - \mu)^2$ случайной величины ξ. Для этого удобно записать выборочную дисперсию в виде

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

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2,$$

после чего, обозначая через D_η дисперсию произвольной случайной величины η и применяя известную формулу $M\eta^2 = (M\eta)^2 + D\eta$, получаем

$$MS^2 = \frac{1}{n}\sum_{i=1}^{n} Mx_i^2 - M(\bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(\mu^2 + \sigma^2) -$$

$$- [(M\bar{x})^2 + D\bar{x}],$$

откуда, учитывая, что математическое ожидание выборочного среднего $M\bar{x} = \mu$, а его дисперсия

$$D\bar{x} = D\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n},$$

следует

$$MS^2 = \left(\frac{n-1}{n}\right)\sigma^2. \qquad \text{(I5)}$$

Таким образом, выборочная дисперсия S^2 является смещенной оценкой дисперсии. В связи с этим наряду с S^2 часто используют также несколько видоизмененную оценку:

$$S_1^2 = \left(\frac{n}{n-1}\right) S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Как следует из равенства (I5), эта величина дает несмещенную оценку дисперсии.

Оценка $\hat{\theta}_n$ называется асимптотически несмещенной оценкой параметра θ, если

$$\lim_{n \to \infty} M\hat{\theta}_n = \theta$$

при всех возможных значениях θ.

Пример 2.4. В рассмотренной выше схеме Бернулли асимптотически несмещенной оценкой такого параметра, как вероятность "успеха", является, например, оценка $\hat{p} = (m+1)/(n+1)$ или любая оценка вида $\hat{p} = (m+c)/(n+c)$, где константа $c > 0$.

Пример 2.5. В условиях рассмотренного примера 2.2 выборочная дисперсия S^2 является асимптотически несмещенной оценкой дисперсии σ^2.

§ 4. Состоятельные оценки

Рассмотрим зависимость оценки от объема выборки, обозначая $\hat{\theta}_n = \hat{\theta}(x_1, \ldots, x_n)$.

Оценка $\hat{\theta}_n$ называется <u>состоятельной</u> оценкой параметра θ, если при $n \to \infty$ она сходится по вероятности к истинному значению параметра, т.е.

$$\lim_{n \to \infty} P\{|\hat{\theta}_n - \theta| > \varepsilon\} = 0$$

для любого $\varepsilon > 0$ и при всех возможных значениях θ.

Пример 2.6. Покажем, что рассмотренная в примере 2.1 оценка $\hat{p} = m/n$ для вероятности p в схеме Бернулли является состоятельной. Для произвольной случайной величины η с конечным математическим ожиданием M_η и дисперсией D_η справедливо известное неравенство Чебышева

$$P\{|\eta - M\eta| > \varepsilon\} \leq \frac{D\eta}{\varepsilon^2}. \qquad (16)$$

Применяя это неравенство и учитывая, что

$$M\hat{p} = p, \quad D\hat{p} = D\left(\frac{m}{n}\right) = \frac{1}{n^2} Dm = \frac{p(1-p)}{n},$$

получаем, что при $n \to \infty$

$$P\{|\hat{p} - p| > \varepsilon\} \leq \frac{D\hat{p}}{\varepsilon^2} = \frac{p(1-p)}{n \cdot \varepsilon^2} \to 0,$$

откуда следует состоятельность оценки $\hat{p}$.

Пример 2.7. Аналогичным образом показывается состоятельность выборочного среднего $\bar{x}$ как оценки для математического ожидания $\mu = M\xi$. Применяя снова неравенство Чебышева (16) и формулу для дисперсии $D\bar{x}$, получаем

$$P\{|\bar{x} - \mu| > \varepsilon\} \leq \frac{D\bar{x}}{\varepsilon^2} = \frac{\sigma^2}{n \cdot \varepsilon^2} \to 0.$$

С помощью неравенства Чебышева легко доказывается и следующий результат о состоятельности асимптотически несмещенных оценок.

<u>Теорема</u>. Пусть $\hat{\theta}_n$ - асимптотически несмещенная оценка параметра θ и, кроме того,

$$\lim_{n \to \infty} D\hat{\theta}_n = 0.$$

Тогда $\hat{\theta}_n$ - состоятельная оценка.

Нетрудно убедиться, что в рассмотренной выше схеме Бернулли состоятельной оценкой параметра p - вероятности "успеха" является не только $\hat{p} = m/n$, но и, например, оценка вида $\hat{p}' = (m+1)/(n+1)$ и т.п. Для дисперсии σ^2 состоятельной является как оценка S^2, так и S_1^2.

§ 5. Эффективные оценки. Неравенство Рао - Крамера

Предположим, что имеются две несмещенные оценки $\hat{\theta}$, $\hat{\theta}'$ для параметра θ. Естественно считать оценку $\hat{\theta}$ более эффективной по сравнению с оценкой $\hat{\theta}'$, если дисперсия первой оценки меньше дисперсии второй:

$$D\hat{\theta} = M(\hat{\theta} - \theta)^2 \leq M(\hat{\theta}' - \theta)^2 = D\hat{\theta}' \qquad (17)$$

при всех возможных значениях θ. Если в некотором классе оценок существует такая оценка $\hat{\theta}$, что неравенство (17) выполняется для всех остальных оценок $\hat{\theta}'$ из этого класса, то говорят, что оценка $\hat{\theta}$ является <u>эффективной</u> (в данном классе оценок).

Известным подходом, позволяющим определять эффективные оценки, является подход, основанный на неравенстве Рао - Крамера.

Введем функцию параметра

$$I(\theta) = M\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2 = \int_{-\infty}^{\infty} \left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2 f(x,\theta)\, dx,$$

которая называется информацией Фишера. Тогда при некоторых общих условиях регулярности для любой несмещенной оценки $\hat{\theta}_n$ параметра θ справедливо следующее <u>неравенство Рао - Крамера</u> для дисперсии оценки:

$$D\hat{\theta}_n \geqslant \frac{1}{n \cdot I(\theta)} \tag{18}$$

при всех возможных значениях θ.

<u>Эффективностью</u> несмещенной оценки $\hat{\theta}_n$ назовем величину

$$e(\theta) = \frac{1}{n \cdot I(\theta) \cdot D\hat{\theta}_n}.$$

В соответствии с неравенством Рао – Крамера эффективность любой оценки удовлетворяет неравенству $0 \leqslant e(\theta) \leqslant 1$. Несмещенную оценку $\hat{\theta}_n$ назовем <u>эффективной</u>, если для нее $e(\theta) = 1$, или, другими словами, ее дисперсия $D\hat{\theta}_n$ достигает нижней границы, указанной в неравенстве Рао – Крамера (18) при всех возможных значениях θ.

Пример 2.8. Для рассмотренной в примере 2.1 схемы Бернулли неизвестным параметром θ является вероятность "успеха" $\theta = p$, $f(x,p) = p$ при $x = 1$ и $f(x,p) = (1-p)$ при $x = 0$. Тогда информация Фишера

$$I(p) = M\left[\frac{\partial \ln f(x,p)}{\partial p}\right]^2 = P(x=0)\left[\frac{\partial \ln f(0,p)}{\partial p}\right]^2 +$$

$$+ P(x=1)\left[\frac{\partial \ln f(1;p)}{\partial p}\right]^2 = (1-p)\left(\frac{1}{1-p}\right)^2 + p\left(\frac{1}{p}\right)^2 =$$

$$= \frac{1}{p(1-p)}.$$

С другой стороны, дисперсия оценки $\hat{p} = m/n$ равна $p(1-p)/n$ (см. пример 2.6). Отсюда следует, что оценка $\hat{p}$ является эффективной.

Пример 2.9. Рассмотрим выборку объема n из нормального распределения с известной дисперсией σ^2 и с плотностью

$$f(x,\mu) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \exp - \frac{(x-\mu)^2}{2\sigma^2}.$$

В данном случае $\theta = \mu$, информация Фишера

$$I(\mu) = M\left[\frac{\partial \ln f(x,\mu)}{\partial \mu}\right]^2 = M\left[\frac{x-\mu}{\sigma^2}\right]^2 = \frac{1}{\sigma^2}.$$

С другой стороны, дисперсия оценки $\hat{\mu} = (x_1 + \ldots + x_n)/n$ равна σ^2/n, откуда с учетом неравенства Рао – Крамера следует, что указанная оценка параметра μ нормального закона распределения является эффективной.

Пример 2.10. Рассмотрим выборку объема n из экспоненциального распределения плотностью $f(x\,\theta) = \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right)$, $x > 0$. В этом случае информация Фишера

$$I(\theta) = M\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2 = M\left[\frac{(x-\theta)^2}{\theta^4}\right] = \frac{1}{\theta^2}.$$

Дисперсия оценки $\hat{\theta} = (x_1 + \ldots + x_n)/n$ равна

$$D\hat{\theta} = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i = \frac{\theta^2}{n},$$

т.е. данная оценка параметра экспоненциального распределения является эффективной.

Пример 2.11. В случае пуассоновского распределения

$$f(x,\theta) = \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, 2, \ldots$$

информация Фишера

$$I(\theta) = M\left[\frac{\partial}{\partial\theta}(x\ln\theta - \theta - \ln x!)\right]^2 = M\left(\frac{x}{\theta} - 1\right)^2 = \frac{1}{\theta}.$$

При этом дисперсия оценки параметра $\hat{\theta} = (x_1 + \ldots + x_n)/n$ равна $D\hat{\theta} = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i = \frac{1}{n^2}(n\theta) = \theta/n$, т.е. оценка $\hat{\theta}$ является эффективной.

§ 6. <u>Методы построения точечных оценок</u>

Наиболее часто используемыми для построения оценок являются рассматриваемые ниже <u>метод моментов</u> и <u>метод максимального правдоподобия</u>.

Метод моментов

Пусть $x_1, \ldots, x_n$ – независимая случайная повторная выборка из распределения с плотностью $f(x,\theta)$, зависящей от неизвестного параметра θ. Найдем первый момент (математическое ожидание) для этого распределения как функцию от параметра θ

$$\mu_1 = \mu_1(\theta) = \int_{-\infty}^{\infty} x f(x,\theta)\,dx .$$

После этого в качестве оценки $\hat{\theta}$ возьмем решение относительно θ следующего уравнения:

$$\mu_1(\theta) = \bar{x}, \qquad (19)$$

где $\bar{x}$ – эмпирическое среднее, или оценка первого момента, найденная по выборке.

Другими словами, в качестве оценки параметра берется то его значение, при котором точное или "теоретическое" значение первого момента совпадает с его эмпирическим значением, найденным по результатам выборки.

Аналогичным образом метод моментов применяется и в случае нескольких неизвестных параметров. А именно, если $\theta = (\theta_1, \ldots, \theta_l)$ – параметр размерности $l \geqslant 1$, то оценки $\hat{\theta}_1, \ldots, \hat{\theta}_l$ находятся как решение системы l уравнений

$$\begin{cases} \mu_1(\theta_1, \ldots, \theta_l) = \hat{\mu}_l, \\ \ldots\ldots\ldots\ldots \\ \mu_l(\theta_1, \ldots, \theta_l) = \hat{\mu}_l, \end{cases}$$

где $\mu_r(\theta_1, \ldots, \theta_l) = \int_{-\infty}^{\infty} x^r f(x, \theta_1, \ldots, \theta_l)\,dx$ –

– "теоретическое" значение момента порядка r, а

$$\hat{\mu}_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r –$$

его эмпирическое значение, найденное по выборке.

В случае $l = 2$ указанная система уравнений может быть также записана в следующем эквивалентном виде:

$$\begin{cases} \mu_1(\theta_1, \theta_2) = \bar{x}, \\ D(\theta_1, \theta_2) = S^2, \end{cases} \qquad (20)$$

где $D(\theta_1, \theta_2)$ – точное, или теоретическое значение дисперсии, а S^2 – эмпирическая дисперсия.

Пример 2.12. Рассмотрим оценку неизвестного параметра – вероятности "успеха" p в схеме независимых испытаний Бернулли. Пусть m – число "успехов", наблюдаемых в серии из n испытаний. Первый момент, или математическое ожидание числа "успехов"

$$\mu_1 = \mu_1(p) = Mm = \sum_{m=0}^{n} m \cdot C_n^m p^m (1-p)^{n-m} = np .$$

В соответствии с методом моментов оценка $\hat{p}$ параметра p находится из уравнения $np = m$, откуда $\hat{p} = m/n$.

Пример 2.13. Рассмотрим выборку из экспоненциального распределения с плотностью $f(x,\lambda) = \lambda \exp(-\lambda x)$, $x > 0$ с одним неизвестным параметром λ. В этом случае первый момент

$$\mu_1 = \mu_1(\lambda) = \int_{-\infty}^{\infty} x \cdot \lambda \cdot \exp(-\lambda x)\,dx = \frac{1}{\lambda} .$$

Уравнение (19) имеет вид $1/\lambda = \bar{x}$, откуда $\hat{\lambda} = 1/\bar{x}$.

Пример 2.14. В случае распределения Эрланга порядка r с плотностью вида

$$f(x,\lambda) = \frac{\lambda^r x^{r-1}}{(r-1)!} \exp(-\lambda x),\ x > 0$$

первый момент определяется выражением

$$\mu_1(\lambda) = \int_{-\infty}^{\infty} x f(x,\lambda)\,dx = r/\lambda .$$

Соответственно оценка параметра λ методом моментов находится из уравнения $r\lambda = \bar{x}$, откуда $\hat{\lambda} = r/\bar{x}$.

Пример 2.15. Рассмотрим ситуацию, когда выборка производится из нормального распределения с двумя неизвестными параметрами μ, σ. В этом случае математическое ожидание и дисперсия равны соответственно μ и σ^2. Система уравнений (20) имеет простой вид

$$\begin{cases} \mu = \bar{x}, \\ \sigma^2 = S^2, \end{cases}$$

откуда $\hat{\mu} = \bar{x}$, $\hat{\sigma} = S$.

Пример 2.16. Рассмотрим случай гамма-распределения с плотностью вида $f(x, \lambda, \alpha) = \dfrac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} \cdot \exp(-\lambda x)$, $x > 0$ с двумя неизвестными параметрами λ, α. В этом случае, используя известное выражение для гамма-функции

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t)\,dt,$$

а также рекуррентное соотношение $\Gamma(\alpha+1) = \alpha\,\Gamma(\alpha)$, получаем следующие выражения для первого, второго моментов и дисперсии:

$$\mu_1(\lambda, \alpha) = \int_{-\infty}^{\infty} \frac{\lambda^\alpha x^\alpha}{\Gamma(\alpha)} e^{-\lambda x} dx = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\cdot\lambda} = \frac{\alpha}{\lambda},$$

$$\mu_2(\lambda, \alpha) = \int_{-\infty}^{\infty} \frac{\lambda^\alpha x^{\alpha+1}}{\Gamma(\alpha)} e^{-\lambda x} dx = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)\cdot\lambda^2} = \frac{\alpha\cdot(\alpha+1)}{\lambda^2},$$

$$D(\lambda, \alpha) = \mu_2(\lambda, \alpha) - \mu_1^2(\lambda, \alpha) = \frac{\alpha}{\lambda^2}.$$

Система уравнений для нахождения оценок параметров в данном случае имеет вид

$$\begin{cases} \dfrac{\alpha}{\lambda} = \bar{x}, \\ \dfrac{\alpha}{\lambda^2} = S^2, \end{cases}$$

откуда $\hat{\lambda} = \bar{x}/S^2$, $\hat{\alpha} = \bar{x}^2/S^2$.

Метод моментов дает состоятельные оценки параметров, но не всегда достаточно эффективен.

Метод максимального правдоподобия

Пусть $x_1, \ldots, x_n$ – по-прежнему независимая повторная случайная выборка объема n из распределения с плотностью $f(x, \theta)$. Совместная плотность распределения всех результатов выборки

$$L(x_1, \ldots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \ldots \cdot f(x_n, \theta) \quad (21)$$

называется функцией правдоподобия.

В качестве оценки максимального правдоподобия $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$ берется то значение параметра θ, при котором функция правдоподобия (21), рассматриваемая как функция от θ при фиксированных результатах наблюдений $x_1, \ldots, x_n$, обращается в максимум. Другими словами, оценка максимального правдоподобия находится из условия

$$L(x_1, \ldots, x_n, \hat{\theta}) = \max_\theta L(x_1, \ldots, x_n, \theta).$$

Во многих случаях удобно искать максимум не из самой функции правдоподобия, а из ее логарифма. Таким образом, предполагая, что функция $f(x, \theta)$ дифференцируема по θ, получаем, что оценка максимального правдоподобия $\hat{\theta}$ находится из уравнения относительно θ

$$\frac{\partial}{\partial\theta} \ln L(x_1, \ldots, x_n, \theta) = 0. \quad (22)$$

Уравнение (22) называется уравнением правдоподобия.

Примечание. Уравнение (22) является необходимым, но недостаточным условием максимума. Тем не менее, для многих часто используемых семейств распределений оказывается, что это уравнение имеет единственное решение, которое и дает искомую оценку максимального правдоподобия.

Аналогично в случае многомерного параметра $\theta = (\theta_1, \ldots, \theta_l)$ для нахождения методом максимального правдоподобия оценок параметров $\hat{\theta}_1, \ldots, \hat{\theta}_l$ необходимо решить систему уравнений относительно $\theta_1, \ldots, \theta_l$ при фиксированных $x_1, \ldots, x_n$:

$$\begin{cases} \dfrac{\partial}{\partial\theta_1} \ln L(x_1, \ldots, x_n, \theta_1, \ldots, \theta_l) = 0, \\ \ldots\ldots\ldots\ldots \\ \dfrac{\partial}{\partial\theta_l} \ln L(x_1, \ldots, x_n, \theta_1, \ldots, \theta_l) = 0. \end{cases} \quad (23)$$

Важным свойством оценок максимального правдоподобия является следующее. В тех случаях, когда существует эффективная оценка, метод максимального правдоподобия дает эту оценку. В общем случае при довольно общих условиях метод максимального правдоподобия дает асимптотически несмещенные и асимптотически эффективные оценки при $n \to \infty$.

П р и м е р 2.17. Применим метод максимального правдоподобия для нахождения оценки для неизвестного параметра - вероятности "успеха" в схеме Бернулли. В этом случае функция правдоподобия - вероятность наблюдать m "успехов" в серии из n опытов равна $L(m,p) = C_n^m p^m (1-p)^{n-m}$.

Уравнение правдоподобия (22) имеет вид

$$\frac{\partial}{\partial p}[m \ln p + (n-m)\ln(1-p)] = \frac{m}{p} - \frac{n-m}{1-p} = 0,$$

откуда получаем $\hat{p} = m/n$.

П р и м е р 2.18. В условиях примера 2.13 найдем оценку параметра λ экспоненциального закона распределения методом максимального правдоподобия. Функция правдоподобия в этом случае

$$L(x_1,\dots,x_n,\lambda) = \lambda \cdot e^{-\lambda x_1} \cdot \dots \cdot \lambda e^{-\lambda x_n} = \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i).$$

Уравнение правдоподобия имеет вид

$$\frac{\partial}{\partial \lambda}(n \ln \lambda - \lambda \sum_{i=1}^{n} x_i) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0,$$

откуда $\hat{\lambda} = n / \sum_{i=1}^{n} x_i$.

П р и м е р 2.19. Найдем методом максимального правдоподобия точечные оценки параметров μ, σ нормального закона распределения с плотностью

$$f(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

В этом случае функция правдоподобия и ее логарифм равны соответственно

$$L = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right],$$

$$\ln L = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

В данном случае число неизвестных параметров $l = 2$. Система уравнений (23) для нахождения оценок максимального правдоподобия имеет вид

$$\begin{cases} \dfrac{\partial}{\partial \mu} \ln L = \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu) = 0, \\ \dfrac{\partial}{\partial \sigma} \ln L = -\dfrac{n}{\sigma} + \dfrac{1}{\sigma^3}\sum\limits_{i=1}^{n}(x_i-\mu)^2 = 0, \end{cases}$$

откуда получаем

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

т.е. оценки максимального правдоподобия для математического ожидания μ и дисперсии σ^2 нормального распределения совпадают с выборочным средним $\bar{x}$ и выборочной дисперсией S^2.

Задачи

1. Доказать теорему 2.1.
2. Найти дисперсии оценок S^2, S_1^2 для величины σ^2.
3. Доказать состоятельность оценок S_1^2, S^2 для σ^2.
4. Доказать несмещенность оценки - выборочного момента порядка r $\hat{\mu}_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$ для момента $\mu_r = M\xi^r$.
5. Для случая экспоненциального распределения с плотностью вида $f(x,\theta) = \frac{1}{\theta}\exp(-x/\theta)$, $x > 0$ показать, что оценка параметра $\hat{\theta} = (x_1 + \dots + x_n)/n$ является эффективной на основании неравенства Рао - Крамера.
6. Для случая дискретного пуассоновского распределения с $f(x,\theta) = \frac{\theta^x}{x!}e^{-\theta}$, $x = 0, 1, \dots$ показать, что оценка параметра $\hat{\theta} = (x_1 + \dots + x_n)/n$ является эффективной на основании неравенства Рао - Крамера.

Найти систему уравнений правдоподобия для нахождения оценок максимального правдоподобия параметров гамма-распределения с плотностью вида

$$f(x,\lambda,\alpha)=\frac{\lambda^{\alpha}x^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda x},\quad x>0.$$

8. Найти систему уравнений правдоподобия для нахождения оценок максимального правдоподобия параметров распределения Вейбулла с плотностью

$$f(x,\lambda,\alpha)=\lambda\alpha x^{\alpha-1}\exp(-\lambda x^{\alpha}),\quad x>0.$$

9. Найти оценки максимального правдоподобия для параметров a, b равномерного распределения на интервале (a,b).

Глава 2. ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

При оценивании неизвестных параметров наряду с рассмотренными выше точечными оценками часто используются также доверительные интервалы. Пусть $x_1, x_2, \ldots, x_n$ – выборка объема n из генеральной совокупности с функцией распределения $F(x,\theta)$ и плотностью распределения $f(x,\theta)$, зависящими от параметра θ, значение которого неизвестно. Предположим, что для параметра θ на основе выборки построен интервал $[\underline{\theta}, \bar{\theta}]$, нижняя и верхняя границы которого являются некоторыми функциями от выборки

$$\underline{\theta}=\underline{\theta}(x_1,\ldots,x_n),\quad \bar{\theta}=\bar{\theta}(x_1,\ldots,x_n).$$

Интервал $[\underline{\theta}, \bar{\theta}]$ называется доверительным интервалом для параметра θ с коэффициентом доверия γ, если

$$P(\underline{\theta}\le\theta\le\bar{\theta})=\gamma \tag{24}$$

при всех возможных значениях θ. При этом значение коэффициента доверия обычно выбирается достаточно близким к единице, например, $\gamma = 0{,}9$; $0{,}95$ и т.п.

Таким образом, доверительный интервал $[\underline{\theta}, \bar{\theta}]$ представляет собой интервал со случайными границами, построенный на основе выборки таким образом, что он накрывает неизвестное истинное значение параметра θ с заданной достаточно высокой вероятностью γ. В отличие от точечных оценок доверительный интервал дает информацию не только о том, где находится неизвестное

истинное значение параметра, но и о точности или достоверности оценивания, которая характеризуется коэффициентом доверия γ.

§ 7. Методы построения доверительных интервалов

Одним из часто используемых для построения доверительных интервалов является метод, основанный на предварительном построении некоторой центральной статистики.

Центральной статистикой называется любая функция, зависящая от параметра θ и от выборки

$$T=T(x_1,\ldots,x_n,\theta), \tag{25}$$

т.е. такая, что функция распределения статистики $F(t)=P(T\le t)$ не зависит от параметра θ.

Предположим, что центральная статистика (25) является монотонной функцией параметра θ, например, монотонно убывающей.

Квантилем уровня α функции распределения $F(t)$ случайной величины T называется величина $K=K(\alpha)$, определяемая из условия

$$P(T\le K)=F(K)=\alpha.$$

Зададимся двумя малыми числами ε_1, ε_2 и определим величины t_1, t_2 из условий

$$P(T\le t_1)=F(t_1)=\varepsilon_1;$$

$$P(T>t_2)=1-F(t_2)=\varepsilon_2.$$

Для этого, очевидно, нужно положить $t_1=K(\varepsilon_1)$, $t_2=K(1-\varepsilon_2)$. Тогда для центральной статистики (25) будет выполняться равенство

$$P\{t_1\le T(x_1,\ldots,x_n,\theta)\le t_2\}=\gamma, \tag{26}$$

где $\gamma=1-\varepsilon_1-\varepsilon_2$.

Далее нижняя и верхняя границы $\underline{\theta}$, $\bar{\theta}$ доверительного интервала для параметра θ определяются как соответственно минимальное и максимальное значение среди всех θ, удовлетворяющих неравенствам

$$t_1\le T(x_1,\ldots,x_n,\theta)\le t_2.$$

При этом из равенства (26) следует равенство

$$P(\underline{\theta} \le \theta \le \bar{\theta}) = \gamma,$$

т.е. определенный таким образом интервал $[\underline{\theta}, \bar{\theta}]$ является доверительным интервалом с коэффициентом доверия $\gamma = 1 - \varepsilon_1 - \varepsilon_2$.

В большинстве случаев центральная статистика (25) монотонно зависит от параметра θ. Пусть, например, она монотонно убывает по θ. Тогда нижняя и верхняя границы доверительного интервала определяются из уравнений

$$\begin{cases} T(x_1, \ldots, x_n, \underline{\theta}) = t_2, \\ T(x_1, \ldots, x_n, \bar{\theta}) = t_1. \end{cases} \qquad (27)$$

Если центральная статистика монотонно возрастает по θ, то границы доверительного интервала находятся из уравнений

$$\begin{cases} T(x_1, \ldots, x_n, \underline{\theta}) = t_1, \\ T(x_1, \ldots, x_n, \bar{\theta}) = t_2. \end{cases}$$

При этом на практике чаще всего доверительный интервал строится симметричным образом, т.е. полагается $\varepsilon_1 = \varepsilon_2 = \varepsilon = (1-\gamma)/2$. Рассмотрим далее построение доверительных интервалов для параметров основных, наиболее часто используемых законов распределения.

§ 8. Доверительный интервал для параметра экспоненциального распределения

Рассмотрим случай экспоненциального закона распределения с плотностью $f(x,\theta) = \frac{1}{\theta} \exp(-\frac{x}{\theta})$, $x > 0$. В качестве исходной центральной статистики возьмем

$$T = 2 \sum_{i=1}^{n} \frac{x_i}{\theta}.$$

Эта статистика имеет стандартное χ^2-распределение с $2n$ степенями свободы (см. приложение). Система уравнений (27) в данном случае имеет вид

$$2 \sum_{i=1}^{n} \frac{x_i}{\underline{\theta}} = t_2 = \chi^2_{1-\varepsilon}(2n),$$

$$2 \sum_{i=1}^{n} x_i / \bar{\theta} = t_1 = \chi^2_{\varepsilon}(2n),$$

где $\chi^2_{\alpha}(2n)$ - квантиль уровня α для χ^2-распределения с $2n$ степенями свободы, $\varepsilon = (1-\gamma)/2$.

Отсюда получаем, что нижняя и верхняя границы доверительного интервала с коэффициентом доверия $\gamma = 1 - 2\varepsilon$ для параметра экспоненциального закона распределения имеют вид

$$\underline{\theta} = 2 \sum_{i=1}^{n} x_i / \chi^2_{1-\varepsilon}(2n),$$

$$\bar{\theta} = 2 \sum_{i=1}^{n} x_i / \chi^2_{\varepsilon}(2n).$$

§ 9. Доверительные интервалы для параметров нормального распределения

Рассмотрим построение доверительных интервалов для параметров μ, σ нормального закона распределения, которые имеют смысл соответственно среднего значения (математического ожидания) и среднеквадратичного отклонения для этого закона.

Доверительный интервал для среднего μ при известной дисперсии σ^2

В этом случае в качестве исходной центральной статистики возьмем следующую:

$$T = \left(\frac{\bar{x} - \mu}{\sigma}\right)\sqrt{n},$$

которая имеет стандартное нормальное распределение с параметрами 0, 1. Система уравнений (27) соответственно принимает вид

$$\begin{cases} \sqrt{n}\,(\bar{x} - \underline{\mu})/\sigma = t_2 = u_{1-\varepsilon}, \\ \sqrt{n}\,(\bar{x} - \bar{\mu})/\sigma = t_1 = u_{\varepsilon}, \end{cases}$$

где u_{α} - квантиль уровня α стандартного нормального закона; $\varepsilon = (1-\gamma)/2$.

Отсюда, учитывая, что для нормального закона $u_\varepsilon = -u_{1-\varepsilon}$, получаем следующие нижнюю и верхнюю границы доверительного интервала для параметра μ :

$$\underline{\mu} = \bar{x} - u_{1-\varepsilon}\cdot\left(\frac{\sigma}{\sqrt{n}}\right),$$

$$\bar{\mu} = \bar{x} + u_{1-\varepsilon}\cdot\left(\frac{\sigma}{\sqrt{n}}\right).$$

Доверительный интервал для среднего μ при неизвестной дисперсии σ^2

В этом случае исходной центральной статистикой является

$$T = \left(\frac{\bar{x}-\mu}{S}\right)\sqrt{n-1},$$

где $\bar{x}$ – выборочное среднее; S^2 – выборочная дисперсия (см. § 2). Эта статистика имеет распределение Стьюдента с $(n-1)$ степенями свободы (см. приложение). Система уравнений (27) в данном случае принимает вид

$$\begin{cases} \sqrt{n-1}\,(\bar{x}-\underline{\mu})/S = t_{1-\varepsilon}(n-1); \\ \sqrt{n-1}\,(\bar{x}-\bar{\mu})/S = t_{\varepsilon}(n-1), \end{cases}$$

где $t_\alpha(n-1)$ – квантиль уровня α распределения Стьюдента с $(n-1)$ степенями свободы; $\varepsilon = (1-\gamma)/2$.

Поскольку распределение Стьюдента симметрично, то для него $t_\varepsilon(n-1) = -t_{1-\varepsilon}(n-1)$. Таким образом, нижняя и верхняя границы доверительного интервала с коэффициентом доверия $\gamma = 1-2\varepsilon$ для параметра μ в случае с неизвестной дисперсией определяются по формулам

$$\underline{\mu} = \bar{x} - t_{1-\varepsilon}(n-1)\cdot\left(\frac{S}{\sqrt{n-1}}\right),$$

$$\bar{\mu} = \bar{x} + t_{1-\varepsilon}(n-1)\cdot\left(\frac{S}{\sqrt{n-1}}\right).$$

Доверительный интервал для σ при известном среднем μ

В этом случае центральная статистика вида

$$T = \sum_{i=1}^{n}(x_i-\mu)^2/\sigma^2$$

имеет χ^2-распределение с n степенями свободы. Система уравнений (27) принимает вид

$$\begin{cases} \sum_{i=1}^{n}(x_i-\mu)^2/\underline{\sigma}^2 = \chi^2_{1-\varepsilon}(n), \\ \sum_{i=1}^{n}(x_i-\mu)^2/\bar{\sigma}^2 = \chi^2_{\varepsilon}(n), \end{cases}$$

откуда получаем нижнюю и верхнюю границы доверительного интервала с коэффициентом доверия $\gamma = 1-2\varepsilon$ для σ

$$\underline{\sigma} = \left[\sum_{i=1}^{n}(x_i-\mu)^2/\chi^2_{1-\varepsilon}(n)\right]^{1/2};$$

$$\bar{\sigma} = \left[\sum_{i=1}^{n}(x_i-\mu)^2/\chi^2_{\varepsilon}(n)\right]^{1/2}.$$

Доверительный интервал для σ при неизвестном среднем μ

В этом случае в качестве исходной центральной статистики возьмем

$$T = (nS^2)/\sigma^2,$$

которая имеет χ^2-распределение с $(n-1)$ степенями свободы.

Система уравнений (27) в этом случае имеет вид

$$\begin{cases} nS^2/\underline{\sigma}^2 = \chi^2_{1-\varepsilon}(n-1), \\ nS^2/\bar{\sigma}^2 = \chi^2_{\varepsilon}(n-1), \end{cases}$$

откуда таким же образом получаем нижнюю и верхнюю границы доверительного интервала с коэффициентом доверия $\gamma = 1-2\varepsilon$ для σ в случае неизвестного среднего

$$\underline{\sigma} = \sqrt{n}\cdot S/\sqrt{\chi^2_{1-\varepsilon}(n-1)}\ ;$$

$$\bar{\sigma} = \sqrt{n}\cdot S/\sqrt{\chi^2_{\varepsilon}(n-1)}\ .$$

§ 10. Приближенный доверительный интервал для параметра биномиального распределения

Пусть m – число наблюдаемых "успехов" в серии из n испытаний по схеме Бернулли с неизвестным параметром – вероятностью p в одном испытании. Случайная величина m имеет биномиальное распределение с параметром p. Для построения доверительного интервала для p возьмем в качестве исходной центральной статистики величину

$$T = \frac{m-np}{\sqrt{np(1-p)}}\ .$$

В соответствии с предельной теоремой Муавра – Лапласа эта статистика при достаточно большом объеме наблюдений n имеет приближенно стандартное нормальное распределение. Таким образом, неравенство

$$-u_{1-\varepsilon} \leq \frac{m-np}{\sqrt{np(1-p)}} \leq u_{1-\varepsilon} \qquad (28)$$

выполняется с вероятностью, которую при большом n можно считать приближенно равной $\gamma = 1-2\varepsilon$. Последнее неравенство может быть записано в виде

$$\frac{m}{n} - \frac{u_{1-\varepsilon}}{\sqrt{n}}\sqrt{p(1-p)} \leq p \leq \frac{m}{n} + \frac{u_{1-\varepsilon}}{\sqrt{n}}\sqrt{p(1-p)}\ .$$

Эти неравенства еще не дают доверительного интервала для p, так как их левая и правая части также содержат p. Поэтому на практике часто подставляют в указанные части неравенств вместо неизвестного значения параметра p его оценку $\hat{p} = m/n$.

В результате получают следующие нижнюю и верхнюю границы доверительного интервала с коэффициентом доверия $\gamma \cong 1-2\varepsilon$ для параметра p:

$$\underline{p} = \frac{m}{n} - \frac{u_{1-\varepsilon}}{\sqrt{n}}\cdot\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}\ ;$$

$$\bar{p} = \frac{m}{n} + \frac{u_{1-\varepsilon}}{\sqrt{n}}\sqrt{\frac{m}{n}\left(1-\frac{m}{n}\right)}\ .$$

Эти доверительные границы являются приближенными и могут использоваться в случае достаточно большого объема наблюдений n. При небольших значениях n обычно используются точные доверительные границы Клоппера – Пирсона, определяемые при данном коэффициенте доверия $\gamma = 1-2\varepsilon$ из уравнений

$$\sum_{j=m}^{n} C_n^j\, \underline{p}^j (1-\underline{p})^{n-j} = \varepsilon\ ;$$

$$\sum_{j=0}^{m} C_n^j\, \bar{p}^j (1-\bar{p})^{n-j} = \varepsilon\ .$$

§ 11. Приближенный доверительный интервал для математического ожидания

Пусть $x_1, \ldots, x_n$ – повторная независимая случайная выборка или, другими словами, n повторных независимых наблюдений, некоторой случайной величины ξ с конечными (неизвестными) математическим ожиданием $\mu = M\xi$ и дисперсией σ^2. При этом закон распределения наблюдаемой случайной величины здесь (в отличие от предыдущих параграфов) неизвестен.

Рассмотрим статистику

$$T = \left(\frac{\bar{x}-\mu}{\sigma}\right)\sqrt{n}\ .$$

В соответствии с центральной предельной теоремой эта статистика при больших объемах наблюдений n имеет приближенно нормальное распределение. Отсюда, аналогично определениям в предыдущем параграфе, получаем, что неравенства

$$-u_{1-\varepsilon} \leq \left(\frac{\bar{x}-\mu}{\sigma}\right)\sqrt{n} \leq u_{1-\varepsilon}$$

выполняются с веростностью, близкой при больших n к величине $\gamma = 1-2\varepsilon$. Неравенства (II) эквивалентны следующим:

$$\bar{x} - u_{1-\varepsilon}\left(\frac{\sigma}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + u_{1-\varepsilon}\cdot\left(\frac{\sigma}{\sqrt{n}}\right).$$

Эти неравенства не дают еще доверительного интервала для параметра μ, так как их левая и правая части содержат неизвестный параметр σ. Применяя еще одно приближение, а именно, подставляя в указанные неравенства вместо неизвестного точного значения σ его оценку $\hat{\sigma} = S$, получаем нижнюю и верхнюю границы доверительного интервала с коэффициентом доверия $\gamma \approx 1-2\varepsilon$ для μ

$$\underline{\mu} = \bar{x} - u_{1-\varepsilon}\cdot\left(\frac{S}{\sqrt{n}}\right),$$

$$\bar{\mu} = \bar{x} + u_{1-\varepsilon}\left(\frac{S}{\sqrt{n}}\right).$$

Задачи

1. Построить доверительные интервалы для параметров равномерного закона распределения на отрезке $[\alpha, b]$ по выборке объема n из этого распределения в предположении, что параметры α, b неизвестны.

2. Построить доверительный интервал для разности $\mu_1 - \mu_2$ параметров – средних двух нормальных распределений с параметрами (μ_1, σ_1) и (μ_2, σ_2) по результатам двух соответствующих независимых выборок объемов n_1 и n_2 в предположении, что дисперсии σ_1, σ_2 известны.

3. Предыдущую задачу решить в предположении, что дисперсии неизвестны, но равны между собой: $\sigma_1 = \sigma_2 = \sigma$.

СПИСОК ЛИТЕРАТУРЫ

Гнеденко Б.В. Курс теории вероятностей. – М.: Высшая школа, 1972. – М.: Физматгиз, 1961. – 448 с.

Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 1972. – 614 с.

Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 1979. – 400 с.

Сборник задач по теории вероятностей, математической статистике и теории случайных функций / Под ред. Свешникова А.А. – М.: Наука, 1970. – 656 с.

Севастьянов Б.А. Курс теории вероятностей и математической статистики. – М.: Наука, 1982. – 256 с.

Крамер Г. Математические методы статистики. Пер. с англ. – М.: Мир, 1975. – 618 с.

Уилкс С. Математическая статистика. Пер. с англ. – М.: Наука, 1967. – 554 с.

Боровков А.А. Математическая статистика. – М.: Наука, 1984. – 472 с.

Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 441 с.

ПРИЛОЖЕНИЯ

Достаточные статистики

Пусть $x_1, x_2, \ldots, x_n$ - повторная независимая случайная выборка объема n из генеральной совокупности с плотностью распределения $f(x, \theta)$ или, другими словами, n повторных независимых наблюдений некоторой случайной величины ξ с указанной плотностью распределения. Ограничимся для простоты дискретным случаем, когда наблюдаемая случайная величина принимает конечный или счетный набор значений. В этом случае $f(x, \theta)$ представляет собой непосредственно вероятность того, что наблюдаемая случайная величина принимает значение x.

Пусть $T = T(x_1, \ldots, x_n)$ - некоторая статистика или, другими словами, некоторая функция результатов наблюдений. Предположим, что после проведения эксперимента нам известна не вся выборка $(x_1, x_2, \ldots, x_n)$ в целом, а только значение статистики

$$T(x_1, x_2, \ldots, x_n) = t. \tag{П.1}$$

Статистика $T(x_1, x_2, \ldots, x_n)$ называется достаточной для параметра θ (или для параметрического семейства распределений $f(x, \theta)$), если для любого события A условная вероятность этого события при условии, что произошло событие (П.1), т.е. величина

$$P\{A \mid T(x_1, x_2, \ldots, x_n) = t\}$$

не зависит от значения параметра θ (при любом возможном значении статистики t). В дискретном случае указанная условная вероятность вычисляется по формуле

$$P\{A \mid T(x_1, \ldots, x_n) = t\} = \frac{\sum\limits_{\substack{(x_1,\ldots,x_n)\in A \\ T(x_1,\ldots,x_n)=t}} L(x_1, \ldots, x_n, \theta)}{\sum\limits_{T(x_1,\ldots,x_n)=t} L(x_1, \ldots, x_n, \theta)}.$$

Смысл приведенного выше определения заключается в том, что при фиксированном значении достаточной статистики изменение параметра θ не влияет на вероятности тех или иных событий или, более точно, на условное распределение выборки при условии (П.1). Это означает, что знание статистики T дает полную информацию о параметре θ. Следующая теорема дает простой критерий для определения достаточной статистики.

Теорема. (Критерий факторизации): статистика $T = T(x_1, \ldots, x_n)$ является достаточной тогда и только тогда, когда функция правдоподобия имеет вид

$$L(x_1, \ldots, x_n, \theta) = h(x_1, \ldots, x_n) \cdot g\,[T(x_1, \ldots, x_n), \theta].$$

Другими словами, функция правдоподобия записывается в виде произведения двух сомножителей, из которых первый не зависит от θ, а второй (зависящий от θ) зависит от результатов наблюдений $(x_1, \ldots, x_n)$ только через статистику T.

Приведенные результаты остаются аналогичными по смыслу и в непрерывном случае, а также для многомерных параметров θ и многомерных достаточных статистик T. Непосредственно из критерия факторизации следует, например, что рассмотренный выше метод максимального правдоподобия всегда приводит к оценкам параметров через достаточные статистики. Еще одно важное свойство достаточных статистик состоит в том, что если существует эффективная оценка параметра, то она выражается через достаточную статистику.

П р и м е р П.1. Рассмотрим случай экспоненциального распределения с плотностью $f(x,\theta) = \frac{1}{\theta}\exp(-\frac{x}{\theta})$, $x > 0$. В этом случае функция правдоподобия

$$L(x_1, \ldots, x_n, \theta) = \frac{1}{\theta}\exp(-\frac{x_1}{\theta}) \cdot \ldots \cdot \frac{1}{\theta}\exp(-\frac{x_n}{\theta}) = \frac{1}{\theta^n}\exp(-\frac{1}{\theta}\sum_{i=1}^{n} x_i),$$

откуда в соответствии с критерием факторизации следует, что статистика $T = x_1 + x_2 + \ldots + x_n$ является достаточной. В данном случае существует эффективная оценка параметра θ, которая выражается через достаточную статистику (см. § 5).

П р и м е р П.2. В случае нормального распределения с двумя неизвестными параметрами μ, σ функция правдоподобия имеет вид

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

$$=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right]=$$

$$=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\exp\left(-\frac{n\mu^2}{2\sigma^2}\right)\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2+\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i\right),$$

откуда следует, что достаточной для двумерного параметра (μ, σ) является двумерная достаточная статистика (T_1, T_2), где

$$T_1=\sum_{i=1}^{n}x_i\,,\quad T_2=\sum_{i=1}^{n}x_i^2\,.$$

Основные распределения, используемые в математической статистике

Числовые характеристики перечисляемых ниже основных стандартных распределений подробно затабулированы и приводятся в соответствующих таблицах, например, в Таблицах математической статистики Л.В.Большева и Н.В.Смирнова, и др. Таблицы для нахождения квантилей нормального распределения, χ^2-распределения и распределения Стьюдента приводятся также ниже в приложении.

Нормальное распределение

Стандартное нормальное распределение имеет плотность

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

и функцию распределения (Лапласа)

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{t^2}{2}\right)dt\,.$$

Математическое ожидание и дисперсия для случайной величины ξ, распределенной по указанному стандартному нормальному закону, равны соответственно $\mu=M\xi=0$, $\sigma^2=D\xi=1$. Поскольку данное распределение симметрично относительно точки $x=0$, то квантиль u_α уровня α стандартного нормального закона удовлетворяет равенству $u_{1-\alpha}=-u_\alpha$, а функция распределения $\Phi(x)=1-\Phi(-x)$.

Общее нормальное распределение с математическим ожиданием μ и дисперсией σ^2 задается плотностью

$$f(x)=\frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

и функцией распределения, которая выражается через функцию Лапласа

$$F(x)=\Phi\left(\frac{x-\mu}{\sigma}\right).$$

Случайная величина η, распределенная по нормальному закону с параметрами μ, σ, может быть получена из случайной величины ξ, распределенной по стандартному нормальному закону с параметрами $\mu=0$, $\sigma=1$ путем преобразования $\eta=\mu+\sigma\xi$. Соответственно квантиль уровня ξ нормального закона с параметрами μ, σ выражается через квантиль стандартного нормального закона u_α следующим образом: $K_\alpha=\mu+\sigma u_\alpha$.

Сумма независимых случайных величин, распределенных по нормальному закону

Пусть имеется m независимых случайных величин $\xi_1,\dots,\xi_m$, где ξ_i имеет нормальное распределение с параметрами (μ_i, σ_i), $i=1,\dots,m$. Тогда их сумма $\xi_1+\dots+\xi_m$ также имеет нормальное распределение с математическим ожиданием $\mu=\mu_1+\dots+\mu_m$ и дисперсией $\sigma^2=\sigma_1^2+\dots+\sigma_m^2$.

Гамма-распределение

Это распределение задается плотностью

$$g(x)=\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x},\quad x>0,$$

где $\lambda > 0$, $\alpha > 0$ - параметры распределения, $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt$ - гамма-функция.

Математическое ожидание и дисперсия для случайной величины ξ, имеющей гамма-распределение, равны соответственно

$$M\xi = \frac{\alpha}{\lambda}, \quad D\xi = \frac{\alpha}{\lambda^2}.$$

Сумма независимых случайных величин $\xi_1 + \ldots + \xi_m$, где ξ_i имеет гамма-распределение с параметрами (λ, α_i), $i = 1, \ldots, m$ имеет также гамма-распределение с параметрами $(\lambda, \alpha_1 + \ldots + \alpha_m)$.

Частный случай гамма-распределения при целом $\alpha = n$ с плотностью

$$f(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad x > 0$$

называется распределением Эрланга порядка n.

Экспоненциальное распределение

Частным случаем гамма-распределения при $\alpha = 1$ является экспоненциальное распределение с плотностью

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

и функцией распределения $F(x) = 1 - e^{-\lambda x}$, $x > 0$. Математическое ожидание и дисперсия для экспоненциального распределения равны соответственно $1/\lambda$ и $1/\lambda^2$. Сумма m независимых экспоненциальных случайных величин с одинаковым параметром λ имеет гамма-распределение с параметрами (λ, m) и с плотностью

$$f(x) = \frac{\lambda^m}{(m-1)!} x^{m-1} e^{-\lambda x}, \quad x > 0.$$

Распределение Релея

Если случайная величина ξ имеет нормальное распределение с математическим ожиданием 0 и дисперсией σ^2, то случайная величина $\eta = \xi^2$ имеет гамма-распределение с параметрами $\lambda = 1/2\sigma^2$, $\alpha = 1/2$ и плотностью

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi} \cdot \sqrt{x}} \exp\left(-\frac{x}{2\sigma^2}\right), \quad x > 0,$$

которое называется распределением Релея.

χ^2-распределение

Пусть $\xi_1, \ldots, \xi_m$ - независимые одинаково распределенные случайные величины, каждая из которых имеет стандартное нормальное распределение с параметрами $\mu = 0$, $\sigma = 1$. Тогда сумма квадратов этих случайных величин $\xi_1^2 + \ldots + \xi_m^2$ имеет χ^2-распределение с m степенями свободы, плотность которого имеет вид

$$f(x) = \frac{1}{2^{m/2} \Gamma\left(\frac{m}{2}\right)} x^{(m/2)-1} e^{-x/2}.$$

Это распределение является частным случаем гамма-распределения с параметрами $\lambda = 1/2$, $\alpha = m/2$. Тем самым, сумма независимых случайных величин, имеющих χ^2-распределение, также имеет χ^2-распределение с числом степеней свободы, равным сумме степеней свободы для отдельных слагаемых.

Распределение Стьюдента

Пусть случайные величины ξ, η независимы, причем ξ имеет стандартное нормальное распределение с параметрами 0, 1, а η имеет χ^2-распределение с m степенями свободы. Тогда статистика

$$t = \frac{\xi}{\sqrt{\eta}} \cdot \sqrt{m}$$

имеет распределение Стьюдента с m степенями свободы, плотность которого имеет вид

$$f(x) = C\left(1 + \frac{x^2}{m}\right)^{-(m+1)/2},$$

где C - нормировочная константа; $C = 1/B(1/2, m/2)\sqrt{m}$; $B(y, z) = \int_0^1 t^{y-1}(1-t)^{z-1} dt$ - бета-функция.

П р и м е ч а н и е. В случае нормального распределения статистики

$$\xi = \sqrt{n}\,(\bar{x}-\mu)/\sigma, \quad \eta = nS^2/\sigma^2$$

независимы и имеют соответственно стандартное нормальное распределение и χ^2-распределение с ($n-1$) степенями свободы. Следовательно, статистика

$$T = \frac{\xi}{\sqrt{\eta}}\sqrt{n-1} = \left(\frac{\bar{x}-\mu}{S}\right)\sqrt{n-1}$$

имеет распределение Стьюдента с ($n-1$) степенями свободы, на основе чего строится доверительный интервал для параметра μ при неизвестной дисперсии σ^2.

Распределение дисперсионного отношения Фишера

Если случайные величины ξ, η независимы и имеют χ^2-распределение соответственно с n и m степенями свободы, то статистика

$$\varphi = \frac{\xi m}{\eta n}$$

имеет распределение Фишера с плотностью

$$f(x) = C\,x^{(n-2)-1} \Big/ [1+(nx/m)]^{-(n+m)/2},$$

где C – нормировочная константа; $C = \left(\frac{n}{m}\right)^{n/2} / B\left(\frac{n}{2}, \frac{m}{2}\right)$.

Бета-распределение

Плотность бета-распределения имеет вид

$$f(x) = C\,x^{a-1}(1-x)^{b-1}, \quad 0 \le x \le 1,$$

где нормировочная константа $C = 1/B(a, b)$; a, b – параметры распределения.

Математическое ожидание и дисперсия для бета-распределения равны соответственно

$$a/(a+b), \qquad ab/[(a+b)^2(a+b+1)].$$

Частным случаем бета-распределения при $a = b = 1$ является равномерное распределение на отрезке $[0, 1]$.

Логарифмически нормальное распределение

Неотрицательная случайная величина имеет логарифмически нормальное (логнормальное) распределение, если ее логарифм распределен по нормальному закону. Плотность логнормального распределения имеет следующий вид:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma x}\exp\left[-\frac{(\ln x-\mu)^2}{2\sigma^2}\right].$$

Математическое ожидание и дисперсия для него равны соответственно

$$\exp[a+(\sigma^2/2)], \quad \exp(2a+\sigma^2)[\exp(\sigma^2)-1].$$

Распределение Вейбулла

Функция распределения Вейбулла (неотрицательной случайной величины) и его плотность имеют вид

$$F(x) = 1-\exp(-\lambda x^{\alpha}),$$

$$f(x) = \lambda\alpha x^{\alpha-1}\cdot\exp(-\lambda x^{\alpha}), \quad x>0,$$

где $\lambda > 0$, $\alpha > 0$ – параметры распределения.

Математическое ожидание и дисперсия для распределения Вейбулла равны соответственно

$$\frac{\Gamma\left(1+\frac{1}{\alpha}\right)}{\lambda^{1/\alpha}}, \quad \frac{\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)}{\lambda^{2/\alpha}}.$$

Частным случаем распределения Вейбулла при $\alpha = 1$ является экспоненциальное распределение.

Распределение Колмогорова

Данное распределение характеризует асимптотическое при $n\to\infty$ распределение статистики

$$T_n = \sqrt{n}\,\sup_x |\hat{F}_n(x) - F(x)|,$$

где $\hat{F}_n(x)$ – эмпирическая функция распределения, построенная по выборке объема n ; $F(x)$ – истинная (теоретическая) функция распределения.

Функция распределения Колмогорова задается формулой

$$K(t) = 1 - 2\sum_{j=1}^{\infty}(-1)^j \exp(-2j^2t^2).$$

Приведем также основные, наиболее часто используемые дискретные распределения: биномиальное, отрицательное биномиальное, пуассоновское и мультиномиальное распределения.

Биномиальное распределение

Пусть проводится n независимых испытаний в схеме Бернулли с вероятностью "успеха" в одном испытании, равной p . Тогда суммарное число m "успехов" в серии из n испытаний имеет биномиальное распределение

$$P(m=k) = C_n^k p^k (1-p)^{n-k}, \ k = 0, 1, \dots, n.$$

Отрицательное биномиальное распределение

Пусть ν – номер того испытания, на котором впервые наблюдается "успех" в рассмотренной выше схеме испытаний Бернулли. Тогда случайная величина ν имеет "отрицательное биномиальное" распределение

$$P(\nu = k) = (1-p)^{k-1} p, \ k = 1, 2, \dots .$$

Пуассоновское распределение

Пусть время τ между двумя последовательными "вызовами" (например, телефонными звонками) на временной оси является случайной величиной с экспоненциальным распределением с параметром λ . Тогда число вызовов m за фиксированное время t является случайной величиной с пуассоновским распределением с параметром $\Lambda = \lambda t$

$$P(m=k) = \frac{\Lambda^k}{k!} e^{-\Lambda}, \ k = 0, 1, 2, \dots .$$

Данное распределение является также предельным для биномиального распределения при $n \to \infty, p \to 0, np = \Lambda = \mathrm{const}$.

Мультиномиальное распределение

Данное распределение является обобщением биномиального распределения на тот случай, когда в серии из n независимых испытаний в каждом отдельном испытании могут наблюдаться $l \geq 2$ событий, p_j – вероятность j-го события в одном испытании, $p_1 + p_2 + \dots + p_l = 1$. Пусть ν_j – число наблюдений j-го события в серии из n испытаний $\nu_1 + \nu_2 + \dots + \nu_l = n$. Тогда набор случайных величин $\nu_1, \nu_2, \dots, \nu_l$ имеет мультиноминальное распределение следующего вида:

$$P\{\nu_1 = k_1, \dots, \nu_l = k_l\} = \frac{n!}{k_1! \cdot \ldots \cdot k_l!} p_1^{k_1} \cdot \ldots \cdot p_l^{k_l}.$$

Таблицы

Таблица П.I

Значения функции нормального распределения

$$\Phi(x) = \int_{-\infty}^{x} e^{t^2/2} dt \,/\sqrt{2\pi}$$

$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x
0,50	0,000	0,68	0,468	0,86	I,080
0,5I	0,025	0,69	0,496	0,87	I,I26
0,52	0,05I	0,70	0,524	0,88	I,I75
0,53	0,075	0,7I	0,553	0,89	I,227
0,54	0,I00	0,72	0,583	0,90	I,28I
0,55	0,I25	0,73	0,6I3	0,9I	I,34I
0,56	0,I50	0,74	0,643	0,92	I,405
0,57	0,I76	0,75	0,674	0,93	I,476
0,58	0,202	0,76	0,706	0,94	I,555
0,59	0,228	0,77	0,739	0,95	I,645
0,60	0,254	0,78	0,772	0,96	I,75I
0,6I	0,279	0,79	0,806	0,97	I,88I
0,62	0,306	0,80	0,842	0,98	2,054
0,63	0,332	0,8I	0,878	0,99	2,326
0,64	0,358	0,82	0,9I5	0,999	3,090
0,65	0,385	0,83	0,954	0,9999	3,720
0,66	0,4I2	0,84	0,995	–	–
0,67	0,440	0,85	I,036	–	–

Для $x < 0$ значения $\Phi(x)$ вычисляются по формуле $\Phi(x) = 1 - \Phi(-x)$.

Таблица П.2

Квантили $z_{1-\alpha}(n)$ распределения Стьюдента

n	α			
	0,I	0,05	0,025	0,0I
I	3,08	6,3I	I2,7I	3I,82
2	I,89	2,92	4,30	6,97
3	I,64	2,35	3,I8	4,54
4	I,53	2,I3	2,78	3,75
5	I,48	2,02	2,57	3,37
6	I,44	I,94	2,45	3,I4
7	I,42	I,90	2,37	3,00
8	I,40	I,86	2,3I	2,90
9	I,38	I,83	2,26	2,82
I0	I,37	I,8I	2,23	2,76
II	I,36	I,80	2,20	2,72
I2	I,36	I,78	2,I8	2,68
I3	I,35	I,77	2,I6	2,65
I4	I,35	I,76	2,I5	2,62
I5	I,34	I,75	2,I3	2,60
I6	I,34	I,75	2,I2	2,58
I7	I,33	I,74	2,II	2,57
I8	I,33	I,73	2,I0	2,55
I9	I,33	I,73	2,09	2,54
20	I,33	I,73	2,08	2,52
2I	I,32	I,72	2,08	2,52
22	I,32	I,72	2,07	2,5I
23	I,32	I,7I	2,07	2,50
24	I,32	I,7I	2,06	2,49
25	I,32	I,7I	2,06	2,49
26	I,32	I,7I	2,06	2,48
27	I,3I	I,70	2,05	2,47
30	I,3I	I,70	2,04	2,46
60	I,30	I,67	2,00	2,39
I20	I,29	I,66	I,98	2,36

Таблица П.3

Значения квантилей $z_{1-\alpha}(n)$ распределения $\chi^2(n)$

n	α				
	0,I	0,05	0,025	0,0I	0,95
I	2,706	3,84I	5,024	6,635	0,004
2	4,605	5,99I	7,378	9,2I0	0,I03
3	6,25I	7,8I5	9,348	II,34	0,352
4	7,779	9,488	II,I4	I3,28	0,7II
5	9,236	II,07	I2,83	I5,09	I,I45
6	I0,64	I2,59	I4,45	I6,8I	I,635
7	I2,02	I4,07	I6,0I	I8,48	2,I67
8	I3,36	I5,5I	I7,53	20,09	2,733
9	I4,68	I6,92	I9,02	2I,67	3,325
I0	I5,99	I8,3I	20,48	23,2I	3,940
II	I7,28	I9,68	2I,92	24,72	4,575
I2	I8,55	2I,03	23,34	26,22	5,226
I3	I9,8I	22,36	24,74	27,69	5,892
I4	2I,06	23,68	26,I2	29,I4	6,57I
I5	22,3I	25,00	27,49	30,58	7,26I
I6	23,54	26,30	28,85	32,00	7,962
I7	24,77	27,59	30,I9	33,4I	8,672
I8	25,99	28,87	3I,53	34,8I	9,390
I9	27,20	30,I4	32,85	36,I9	I0,I2
20	28,4I	3I,4I	34,I7	37,57	I0.85
22	30,8I	33,92	36,78	40,29	I2,34
24	33.20	36,42	39,36	42,98	I3,85
26	35,56	38,89	4I,92	45,64	I5,38
28	37,92	4I,34	44,46	48,28	I6,93
30	40,26	43,77	46,98	50,89	I8,49
40	5I,8I	55,76	59,34	63,39	26,5I
50	63,I7	67,50	7I,42	76,I5	34,76

Типовой расчет

Ниже приводится 30 вариантов типового расчета, состоящего из 6 задач.

З а д а ч а I. На основе приведенной ниже для каждого варианта выборки x_1, x_2, ..., x_n построить вариационный ряд, эмпирическую функцию распределения $F_n(x)$, найти выборочные среднее $\bar{x}$, дисперсию S^2, среднеквадратичное отклонение S, размах выборки R_n.

I. II,8; I2,4; I6,I; 8,4; I7,5; II,6; I8,8.
2. 7; 4; 3; II; 7; 9; II; 7; 5.
3. 4,8; 5,4; 5,9; 6,4; II,3; 9,5.
4. 25,3; 23,I; 28,4; 26,9; 30,5; 22,8; I9,7.
5. 4I,9; 44,9; 44,7; 5I,8; 43,8; 56,9.
6. 33,9; 35,3; 39,2; 44,4; 4I,7.
7. I5; I2; I6; I7; I5; I3; I8; I6; II; I4.
8. 47; 64; 44; 5I; 75; 82; 9I; II4.
9. I3; 20; 24; 29; 24; I8; II; 24.
I0. I07; II4; II9; 95; I24; II8.
II. 28; 26; I9; 34; I7; 48; 28; 24.
I2. 259; 264; 253; 25I; 296; I87.
I3. I8,I4; 42; [illegible]I; I7; 44; I7.
I4. 5I; 66; 9I; 54; 43; 8I; 86; II8; 7I.
I5. 3,6; 4,8; 8,3; I4,5; 5,6; 7,4.
I6. 24; 54; 4I; I7; 48; 33; 42; 44; 92; 27.
I7. 564; 783; 664; 86I; 775; 496; 335.
I8. 68; 34; 75; 83; 75; 95; 7I; 29.
I9. I7; 24; 8I; I9; 44; 75; 29; 38.
20. 7,8; 8,[illegible]; I4,3; I2,5; 8,9; 9,5.
2I. 77; 95; 7I; I07; I24; I45; 83.
22. I4; 42; 32; 9; 38; 24; 30; 34; 84; I7.
23. 66I; 884; 562; 964; 875; 598; 434.
24. 98; 64; I05; II4; I05; I24; I0I; 64.
25. 37; 53; 34; 4I; 64; 77; 8I; I04; II6.
26. 6,8; 7,4; 7,7; 8,6; I4,3; II,5.
27. 24,9; 25,[illegible]; 29,2; 34,4; 3I,7; 28,I.
28. 47; 54; III; 54; 75; I07; 56; 68.
29. I57; I78; 2II; I44; II8; I94; I24.
30. 54; 6I; 97; 56; 78; 4I; II4; 47.

З а д а ч а 2. Найти методом моментов по выборке x_1, x_2, ..., x_n точечные оценки параметров для плотности распределения:

I. $f(x) = \dfrac{\theta^3}{2!} x^2 e^{-\theta x}$, $x > 0$.

2. $f(x) = \dfrac{\sqrt{\theta}}{\sqrt{\pi x}} e^{-\theta x}$, $x > 0$,

3. $f(x) = \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, $x > 0$.

4. $f(x) = \dfrac{\lambda\sqrt{\lambda x}}{\Gamma(3/2)} e^{-\lambda x}$, $x > 0$.

5. $f(x) = \dfrac{1}{\Gamma(\theta)} x^{\theta-1} e^{-x}$, $x > 0$.

6. $f(x) = \theta x^{-(\theta+1)}$, $x > 1$.

7. $f(x) = 4\theta^3 x^2 e^{-\theta^2 x^2}$, $x > 0$.

8. $f(x) = 3\theta^3 x^4 e^{-\theta x^3}$, $x > 0$.

9. $f(x) = \dfrac{1}{2^{\theta/2}\,\Gamma(\theta/2)}\, x^{\frac{\theta}{2}-1} \cdot e^{-\frac{x}{2}}, \quad x > 0.$

10. $f(x) = 2\theta^2 x e^{-\theta^2 x^2}, \quad x > 0.$

11. $f(x) = \theta^2 x e^{-\theta x^3}, \quad x > 0.$

12. $f(x) = \dfrac{2}{\theta}\, e^{-\frac{(x+c)}{\theta}}, \quad x > 2, \ c = -2.$

13. $f(x) = \dfrac{\lambda^{\alpha+1}}{\Gamma(\alpha+1)}\, x^{\alpha} e^{-\lambda x}, \quad x > 0.$

14. $f(x) = \dfrac{1}{\sigma\,\sqrt{2\pi}\cdot\sqrt{x}}\, e^{-\frac{x}{2\sigma^2}}, \quad x > 0.$

15. $f(x) = \dfrac{\theta^4}{3!}\, x^3 e^{-\theta x}, \quad x > 0.$

16. $f(x) = \dfrac{\theta^5}{4!}\, x^4 e^{-\theta x}, \quad x > 0.$

17. $f(x) = \dfrac{1}{2^{\theta}\,\Gamma(\theta)}\, x^{\theta-1} e^{-\frac{x}{2}}, \quad x > 0.$

18. $f(x) = \dfrac{3\theta^3}{x^4}, \quad x \geqslant 1.$

19. $f(x) = \dfrac{x^{\theta}}{\Gamma(\theta+1)}\, e^{-x}, \quad x > 0.$

20. $f(x) = \dfrac{1}{\theta^4 \cdot 3!}\, x^3 e^{-\frac{x}{\theta}}, \quad x > 0.$

21. $f(x) = \dfrac{1}{\theta^5 \cdot 4!}\, x^4 e^{-\frac{x}{\theta}}, \quad x > 0.$

22. $f(x) = \dfrac{1}{4^{\theta}\,\Gamma(\theta)}\, x^{\theta-1} e^{-\frac{x}{4}}, \quad x > 0.$

23. $f(x) = \theta^2 x e^{-\theta x}, \quad x > 0.$

24. $f(x) = \dfrac{\theta^2}{2}\, |x|\, e^{-\theta|x|}.$

25. $f(x) = 2\theta x e^{-\theta x^2}, \quad x > 0.$

26. $f(x) = \dfrac{\theta}{9}\left(\dfrac{3}{x}\right)^{\theta+1}, \quad x > 3.$

27. $f(x) = \dfrac{\theta^5}{4} x^4 e^{-(\theta x)^2}, \quad x > 0.$

28. $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$

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

30. $p_m = \dfrac{\theta^m}{m!} e^{-\theta}, \quad m = 0, 1, 2, \ldots .$

З а д а ч а 3. Найти методом максимального правдоподобия по выборке $x_1, x_2, \ldots, x_n$ точечную оценку параметра для плотности распределения:

I) закон распределения Пуассона

$$p_x = \frac{\theta^x}{x!} e^{-\theta}, \quad x = 0, 1, 2, \ldots,$$

$x_1 = 1, \; x_2 = 3, \; x_3 = 4, \; x_4 = 5, \; x_5 = 7;$

2) экспоненциальное распределение

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

$x_1 = 2, \; x_2 = 3, \; x_3 = 5, \; x_4 = 8, \; x_5 = 22;$

3) распределение Релея

$$f(x) = \frac{1}{\theta\sqrt{2\pi} \cdot \sqrt{x}} e^{-\frac{x}{2\theta^2}}, \quad x > 0;$$

$x_1 = 4,2; \; x_2 = 7,8, \; x_3 = 16,3, \; x_4 = 11,6, \; x_5 = 8,3;$

4) распределение Вейбулла

$$f(x) = \alpha \theta x^{\alpha-1} e^{-\theta x^\alpha}, \quad x > 0, \; \alpha = 2;$$

$x_1 = 2, \; x_2 = 5, \; x_3 = 6, \; x_4 = 8, \; x_5 = 10;$

5) гамма-распределение

$$f(x) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}, \quad x > 0, \; \alpha = 3;$$

$x_1 = 4, \; x_2 = 1, \; x_3 = 3, \; x_4 = 5, \; x_5 = 7;$

6) $$f(x) = \frac{1}{\theta^{\alpha+1} \cdot \Gamma(\alpha+1)} x^\alpha e^{-(x/\theta)}, \quad x > 0, \; \alpha = 4;$$

$x_1 = 1, \; x_2 = 4, \; x_3 = 5, \; x_4 = 2, \; x_5 = 7;$

7) $$f(x) = 3\theta x^2 e^{-\theta x^3}, \quad x > 0;$$

$x_1 = 4, \; x_2 = 7, \; x_3 = 9, \; x_4 = 5, \; x_5 = 3;$

8) $$f(x) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}, \quad x > 0, \; \alpha = 7;$$

$x_1 = 5, \; x_2 = 10, \; x_3 = 15, \; x_4 = 17, \; x_5 = 4;$

9) $$f(x) = \theta x^{-(\theta+1)}, \quad x > 0;$$

$x_1=4,\ x_2=16,\ x_3=22,\ x_4=7,\ x_5=17;$

10) $f(x)=2\theta x e^{-\theta x^2},\ x>0;$

$x_1=2,\ x_2=4,\ x_3=3,\ x_4=6,\ x_5=1;$

11) $f(x)=\dfrac{\theta\sqrt{\theta}}{\Gamma(3/2)}\sqrt{x}\,e^{-\theta x},\ x>0;$

$x_1=1{,}2,\ x_2=0{,}7,\ x_3=1{,}6,\ x_4=0{,}8,\ x_5=2{,}4;$

12) $f(x)=\dfrac{\theta^3}{2!}x^2e^{-\theta x},\ x>0;$

$x_1=0{,}4;\ x_2=1{,}5;\ x_2=0{,}8;\ x_4=0{,}7;\ x_5=2{,}7;$

13) $f(x)=\dfrac{\theta}{2}\left(\dfrac{2}{x}\right)^{\theta+1},\ x>2;$

$x_1=4,\ x_2=7,\ x_3=5,\ x_4=3,\ x_5=9.$

14) $f(x)=\dfrac{\theta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\theta x},\ x>0,\ \alpha=3{,}5;$

$x_1=0{,}6;\ x_2=1{,}8;\ x_3=1{,}4;\ x_4=0{,}8;\ x_5=0{,}7;$

15) $f(x)=\dfrac{4\theta^3}{\sqrt{\pi}}x^2e^{-\theta^2x^2};$

$x_1=1,\ x_2=4,\ x_3=7,\ x_4=2\ \ x_5=3;$

16) $f(x)=\dfrac{1}{x\sqrt{2\pi}\,\sigma}e^{-\frac{(\ln x-\mu)^2}{2\sigma}};$

$x_1=e,\ x_2=e^2,\ x_3=e^3,\ x_4=e^4,\ x_5=e^5;$

17) $f(x)=\dfrac{\theta}{2}\ e^{-\theta|x-2|};$

$x_1=-2,\ x_2=4,\ x_3=-3,\ x_4=5,\ x_5=1;$

18) $f(x)=\theta^2xe^{-\theta x},\ x>0;$

$x_1=0{,}03;\ x_2=0{,}04;\ x_3=0{,}06;\ x_4=0{,}05;\ x_5=0{,}07;$

19) $f(x)=\dfrac{\theta^4}{3!}x^3e^{-\theta x},\ x>0;$

$x_1=0{,}6;\ x_2=1{,}6;\ x_3=1{,}4;\ x_4=1{,}8;\ x_5=2{,}4;$

20) $f(x)=\dfrac{1}{\theta^4\cdot 3!}x^3e^{-\frac{x}{\theta}},\ x>0;$

$x_1=2,\ x_2=7,\ x_3=4,\ x_4=3,\ x_5=6;$

21) $f(x)=\dfrac{1}{2\cdot\theta^3}x^2e^{-\frac{x}{\theta}},\ x>0;$

$x_1=3,\ x_2=5,\ x_3=6,\ x_4=2,\ x_5=9.$

22) $f(x)=\dfrac{\theta^5}{4!}x^4e^{-\theta x},\ x>0;$

$x_1=7,\ x_2=4,\ x_3=11,\ x_4=5,\ x_5=3;$

23) $f(x)=\dfrac{1}{\theta^2}xe^{-\frac{x}{\theta}},\ x>0;$

$x_1=4,\ x_2=12,\ x_3=6,\ x_4=7,\ x_5=9.$

24) $f(x)=\dfrac{\theta}{3}\left(\dfrac{3}{x}\right)^{\theta+1},\ x>3;$

$x_1=4,\ x_2=9,\ x_3=8,\ x_4=5,\ x_5=7.$

25) $f(x)=\dfrac{\theta}{x^{\theta+1}},\ x>1,$

$x_1=e,\ x_2=e^2,\ x_3=e^3,\ x_4=e^4,\ x_5=e^5.$

26) $f(x)=\frac{1}{\theta}\, e^{-\frac{x}{\theta}}, \; x>0;$

$x_1=5, \; x_2=15, \; x_3=11, \; x_4=25, \; x_5=30;$

27) $f(x)=\frac{\theta^4}{3!}\, x^3 e^{-\theta x}, \; x>0;$

$x_1=3, \; x_2=4, \; x_3=8, \; x_4=7, \; x_5=15.$

28) $f(x)=\frac{1}{4!\cdot\theta^5}\, x^4 e^{-\frac{x}{\theta}}, \; x>0;$

$x_1=2, \; x_2=4, \; x_3=3, \; x_4=5, x_5=7.$

29) $f(x)=\frac{\theta}{2}\, e^{-\theta|x|},$

$x_1=-2, \; x_2=4, \; x_3=-6, \; x_4=5, \; x_5=3.$

30) $f(x)=\frac{\theta^7}{6!}\, x^6 e^{-\theta x}, \; x>0;$

$x_1=8, \; x_2=4, \; x_3=16, \; x_4=3, \; x_5=6.$

З а д а ч а 4. На основе выборки $x_1, x_2, \ldots, x_n$ построить доверительный интервал для параметров нормального распределения:

а) для среднего μ при известной дисперсии $\sigma^2 = \sigma_*^2$ с коэффициентом доверия $\gamma = 0{,}9$;

б) для среднего μ при неизвестной дисперсии σ^2 с коэффициентом доверия $\gamma = 0{,}9$;

в) для среднеквадратичного отклонения σ при известном среднем $\mu = \mu_*$ с коэффициентом доверия $\gamma = 0{,}95$;

г) для среднеквадратического отклонения σ при неизвестном среднем μ с коэффициентом доверия $\gamma = 0{,}95$.

Параметры σ_*, μ_* для каждого варианта приводятся в табл. П.4.

З а д а ч а 5. На основе выборки $x_1, x_2, \ldots, x_n$ построить приближенный доверительный интервал для математического ожидания наблюдаемой случайной величины $\mu = M\xi$ с коэффициентом доверия $\gamma = 0{,}8$ в случае, когда закон распределения ξ неизвестен.

З а д а ч а 6. Построить доверительный интервал с коэффициентом доверия $\gamma = 0{,}97$ для параметра - вероятности "успеха" p в схеме Бернулли при условии, что в серии из n испытаний наблюдалось m "успехов". Величины m, n для каждого варианта приводятся в табл. П.4.

Таблица П.4

Параметры для разных вариантов расчета

Номер варианта	σ_*	μ_*	m	n
I	2	3	4	5
I	3,5	I6,5	5	I0
2	3,3	8	4	I6
3	4	6	80	100
4	5,5	25	7	20
5	7,5	47	I8	50
6	6	38	II	I8
7	3	I5,5	8	24
8	34	74	20	45
9	4	I9	7	28
I0	I0	II0	40	I00
II	8	25	34	80
I2	30	255	I2	30
I3	I8	30	75	I00
I4	20	74	II	40
I5	3	8	I5	50
I6	24	55	35	90
I7	I55	565	45	I20
I8	56	60	I5	20
I9	24	48	9	50
20	2,5	[illegible]	5	35
2I	30	II5	7	20
22	25	45	II	25
23	I65	630	6	30
24	90	25	55	I20
25	70	75	I6	40
26	3	II	7	I5
27	4	25	4	I0
28	30	80	8	34
29	44	I55	64	I00
30	74	79	I4	24

ОГЛАВЛЕНИЕ