МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ «МЭИ»

А.Н. АРХАНГЕЛЬСКИЙ, А.А. БОБОДЖАНОВ,
С.Ф. КУДИН

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

для студентов, обучающихся по направлениям подготовки «Радиоэлектроника», «Электроэнергетика», «Теплоэнергетика», «Атомная энергетика», «Энергомашиностроение»

МОСКВА Издательство «МЭИ» 2020

УДК 519
ББК 22.172
А 723

Утверждено учебным управлением НИУ «МЭИ» в качестве учебного издания

Рецензенты: док. физ-мат. наук, проф. Сакбаев В.Ж., доц. Ю.С. Федоров

Авторы: А.Н. Архангельский, А.А. Бободжанов, С. Ф. Кудин

А 723 Математическая статистика: учебное пособие / А.Н. Архангельский, А.А. Бободжанов, С. Ф. Кудин — М.: Издательство «МЭИ» 2020. 133 с.
ISBN 978-5-7046-2272-7

Настоящее учебное пособие содержит расширенный курс лекций по математической статистике, читаемый авторами на протяжении ряда лет на нескольких факультетах Московского энергетического института студентам второго курса. В результате изучения предмета студенты должны овладеть математическими методами исследования, используя различные модели математической статистики. Особенностью настоящего пособия является его теоретическая направленность.
УДК 519
ББК 22.172

Учебное пособие

Архангельский Алексей Николаевич, **Бободжанов** Абдухафиз Абдурасулович, **Кудин** Сергей Федорович

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Компьютерная верстка А.Н. Архангельский
Редактор издательства О.А. Панова

Подписано в печать xx.xx.2019

Печать офсетная	Физ. печ. л. xxx.xx
Тираж 150 экз.	Заказ № xxx

СОДЕРЖАНИЕ

Предисловие

Настоящее учебное пособие содержит расширенный курс лекций по математической статистике, читаемый авторами на протяжении ряда лет на нескольких факультетах Московского энергетического института студентам второго курса. В результате изучения предмета студенты должны овладеть математическими методами исследования, используя различные модели математической статистики. Особенностью настоящего пособия является его теоретическая направленность. От читателя требуется хорошее знание теории вероятностей в рамках, например, курса [4]. Для будущих инженеров необходимо формирование математического и логического мышления, поэтому изучение курса математической статистики трудно переоценить.

Авторы глубоко признательны своим коллегам по кафедре высшей математики МЭИ и студентам разных лет радиотехнического факультета и отделения атомной энергетики, которые оказали весомую помощь в улучшении курса. Особую признательность авторы выражают профессору Валерию Федоровичу Сафонову за неоценимую помощь в подготовке рукописи.

1. Статистическая структура

Во многих задачах математической статистики приходится иметь дело с наблюдениями, относительно которых нет полной информации. Формально это можно представить следующим образом. Предполагаем, что мы наблюдаем за некоторой случайной величиной (с.в.) ξ, функция распределения которой:

i) либо неизвестна вообще и неизвестен класс функций распределений, которому может принадлежать рассматриваемая случайная величина;

*ii)*либо известен класс функций распределения, зависящий от некоторого неизвестного параметра (скалярного или векторного).

Обычно, конкретное *статистическое наблюдение* представляет из себя число или числовой вектор в зависимости от того над какой случайной величиной производится наблюдение — скалярной или векторной. Для простоты изложения, будем рассматривать исключительно скалярные случайные величины.

Пусть $\Theta \subset \boldsymbol{R}^k$ — некоторое фиксированное подмножество k-мерного евклидова пространства. Определим $\mathfrak{L}(\Theta) = \{\boldsymbol{F}(x, \vec{\boldsymbol{\theta}}),\ \vec{\boldsymbol{\theta}} \in \Theta\}$ или $\mathfrak{L}(\Theta) = \{\boldsymbol{f}(x, \vec{\boldsymbol{\theta}}),\ \vec{\boldsymbol{\theta}} \in \Theta\}$ класс распределений, где $\boldsymbol{F}$ — функция распределения с.в. ξ и $\boldsymbol{f}$ — плотность распределения (в случае абсолютно-непрерывного распределения с.в. ξ) или дискретное распределение ($\boldsymbol{f}(x, \vec{\boldsymbol{\theta}}) = \boldsymbol{P}_{\vec{\boldsymbol{\theta}}}(\xi = x)$ когда ξ — дискретная с.в.) Класс $\mathfrak{L}(\Theta)$ в случае *ii)* будем называть *статистической структурой*.

Замечание. Обычно, $\{(\Omega, \mathfrak{A}, \boldsymbol{P}_{\vec{\boldsymbol{\theta}}}),\ \vec{\boldsymbol{\theta}} \in \Theta \subset \subset \boldsymbol{R}^k\}$ — совокупность вероятностных пространств называют статистической структурой и на этом множестве пространств определяются случайные величины, распределение которых было введено выше.

Примеры

а) Пусть $\Theta = (-\infty, \infty)$, $\boldsymbol{\theta} \in \Theta$,

$$\boldsymbol{F(x, \theta)} = \begin{cases} \mathbf{0}, & \text{если} \quad \boldsymbol{x \le \theta}, \\ \mathbf{1}, & \text{если} \quad \boldsymbol{x > \theta}, \end{cases}$$

$\mathfrak{L}(\Theta)$ – класс вырожденных в точке $\boldsymbol{\theta}$ распределений. Ясно, что тот же самый класс можно задать с помощью дискретного распределения $\boldsymbol{f(\theta, \theta) = 1}$.

б) Пусть $\Theta = (0; 1)$, $\boldsymbol{Bi}(\Theta, 1) = \{\boldsymbol{f(x, \theta)},\ \boldsymbol{\theta} \in \Theta\}$, $\boldsymbol{f(x, \theta)} = \begin{cases} \mathbf{1} - \boldsymbol{\theta}, & \text{если} \quad \boldsymbol{x = 0} \\ \boldsymbol{\theta}, & \text{если} \quad \boldsymbol{x = 1} \end{cases}$ – класс бернуллиевских распределений с вероятностью «успеха» равной $\boldsymbol{\theta}$.

в) Пусть $\Theta = (0, +\infty)$. Класс $\Pi(\Theta) = \{\boldsymbol{f(x, \theta)},\ \boldsymbol{\theta} \in \Theta\}$ пуассоновских распределений, где $\boldsymbol{f(x, \theta) = e^{-\theta}\theta^x / x!}$, $\boldsymbol{x} = \mathbf{0, 1, 2, \ldots}$ – пуассоновское распределение с параметром $\boldsymbol{\theta}$.

г) Пусть $\Theta = \{(\boldsymbol{\theta_1, \theta_2^2})\}$, $\boldsymbol{\theta_1} \in (-\infty, \infty)$, $\boldsymbol{\theta_2^2} \in (0, \infty)$. Класс $\boldsymbol{N}(\Theta) = \{\boldsymbol{f(x, \vec{\theta})},\ \boldsymbol{\vec{\theta}} \in \Theta\}$ нормальных распределений, где $\boldsymbol{f(x, \vec{\theta})} = \frac{1}{\sqrt{2\pi\theta_2^2}} \boldsymbol{exp}\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)$ – плотность нормального распределения.

2. Понятие статистической выборки

Предположим, что задана некоторая статистическая структура $\mathfrak{L}(\Theta)$. Это, как мы знаем, означает что в нашем распоряжении имеется некоторое фиксированное параметрическое вероятностное распределение (т.е. мы знаем тип случайной величины и аналитический вид ее распределения).

Дадим основное определение.

Определение 1. Вектор $\vec{X} = (X_1, X_2, ..., X_n)$, *где* $X_1, X_2, \ldots, X_n$ *— независимые одинаково распределенные случайные величины, определенные статистической структурой* $\mathfrak{L}(\Theta)$, *называется* ***выборкой объема*** n.

Определение 2. Множество $\mathcal{X} \subset R^n$ *всех числовых значений выборки* $\vec{X}$ *называется* ***выборочным пространством*** *или* ***генеральной совокупностью***.

Можно считать, что, в зависимости от задачи, наблюдение за некоторым объектом (или экспериментом) является реализацией случайной величины с заданной статистической структурой. Одно наблюдение — это какое - то число, которое может принимать случайная величина. Если производится серия из n наблюдений, то мы получаем реализацию случайного вектора, т.е. выборки.

3. Первичная обработка экспериментальных данных

Все методы статистического анализа условно можно разделить на первичные и вторичные.

К первичным методам статистической обработки относят, например, определение выборочной *средней величины*, выборочной *дисперсии*, выборочной *моды* и выборочной *медианы* и ряда других. Вторичные методы это, например, методы сравнения первичных статистик у двух или нескольких выборок, корреляционный и регрессионный анализ и так далее, о чем пойдет речь в других разделах ниже.

а) Выборочная мода — числовая характеристика реализации выборки. *Модой называют количественное значение, наиболее часто встречающееся в наблюдениях.* Для симметричных распределений, в том числе и для нормального распределения, значение моды совпадает со значениями среднего и медианы (см. пункты *б)* и *в)* ниже). Для других типов распределении это не так. К примеру, в наблюдениях 1, 2, 5, 2, 4, 2, 6, 7, 2 модой будет значение 2, так как оно встречается чаще других значений — четыре раза. Моду находят согласно следующим правилам.

i) В том случае, когда все значения в реализации выборки встречаются одинаково часто, принято считать, что эта реализация выборки не имеет моды. Например: 5, 5, 6, 6, 7, 7 — в этой выборке моды нет.

ii) Когда два соседних значения имеют одинаковую частоту и их частота больше частот любых других значений, мода вычисляется как среднее арифметическое этих двух значений. Например, в выборке 1, 2, 2, 2, 5, 5, 5, 6 частоты рядом расположенных значений 2 и 5 совпадают и равняются 3. Эта частота больше, чем частота других значений 1 и 6 (у которых она равна 1). Следовательно, модой этого ряда будет величина 3.5 .

iii) Если два несмежных (не соседних) значения в выборке имеют равные частоты, которые больше частот любого другого значения, то выделяют две моды. Например, в ряду 10, 11, 11, 11, 12, 13, 14, 14, 14, 17 модами являются значения 11 и 14. В таком случае говорят, что выборка является *бимодальной*.

б) Выборочная медиана – такое значение, которое делит выборку, упорядоченную по возрастанию, пополам. Справа и слева от медианы в упорядоченном ряду остается по одинаковому количеству наблюдений. Например, для реализации выборки 2, 3, 4, 4, 5, 6, 8, 7, 9 медианой будет значение 5, так как слева и справа от него остается по четыре числа. Если ряд включает в себя четное число наблюдений, то медианой будет среднее, взятое как полусумма величин двух центральных значений ряда. Для следующего ряда 0, 1, 1, 2, 3, 4, 5, 5, 6, 7 медиана будет равна 3,5.

Медиану полезно использовать для того, чтобы установить, является ли распределение заданной выборки симметричным.

в) Выборочное среднее – статистический показатель, представляющий собой среднее арифметическое наблюдаемых данных. Выборочное среднее определяется при помощи следующей формулы: $\bar{x} = \sum\limits_{i=1}^{n} x_i/n$, где $\bar{x}$ – выборочная средняя величина по выборке; n – количество испытаний; x_i – значения случайной величины, полученные в результате наблюдений.

Часто считают, что если средняя и медиана на практике совпали или достаточно близки, то само распределение наблюдаемой случайной величины либо является нормальным, либо мало отличается от нормального распределения.

г) Разброс реализации выборки и вариационный ряд. Разброс реализации выборки (иногда эту величину называют размахом реализации выборки) обозначается буквой r. Это самый простой показатель, который можно получить для реализации выборки $x_1, x_2, \ldots, x_n$ – разность между максимальной и минимальной величинами данной реализации выборки, т.е. $r = x_{max} - x_{min}$. Вообще, для каждой реализации случайной выборки $x_1, x_2, \ldots, x_n$ можно расставить эти числа по неубыванию: $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. Если теперь при каждом k $(k = 1, \ldots, n)$ ввести случайную величину $X_{(k)}$, принимающую значение $x_{(k)}$ при каждой реализации выборки, то можно составить ряд $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$, который принято называть *вариационным рядом*. Каждую компоненту $X_{(k)}$ вариационного ряда называют *порядковой статистикой* [2], [3]. Величину $R = X_{(n)} - X_{(1)}$ называют *разбросом выборки*.

д) Выборочная дисперсия — это среднее арифметическое квадратов отклонений полученных значений случайной величины от её среднего значения:

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

– выборочная дисперсия; здесь n – количество наблюдений. Чем больше дисперсия, тем больше отклонения или разброс данных. Однако сама дисперсия, как характеристика отклонения от среднего, часто неудобна для интерпретации результатов. Для того, чтобы согласовать размерность отклонения с размерностью измеряемого признака используют среднеквадратичное отклонение – это квадратный корень из дисперсии. Таким образом, *средне-*

квадратичное отклонение:

$$\boldsymbol{S = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n}.}$$

4. Статистики и оценки. Задача оценки параметрических функций

Пусть задана выборка $\vec{X} = (X_1, X_2, ..., X_n)$ объема n, определенная статистической структурой $\mathfrak{L}(\Theta)$. Пусть T– некоторая скалярную функцию от векторного аргумента $T : R^n \to R^1$, т.е. $T = T(x_1, x_2, \ldots, x_n)$.

Определение. Случайная величина $T = T(\vec{X})$ *называется* ***статистикой***.

Отметим, что функция T не должна зависеть от неизвестного параметра $\vec{\theta}$. Таким образом, любая функция от выборки есть статистика.

Рассмотрим одну из важных задач математической статистики. Пусть $\Theta \subset R^1$ – некоторый фиксированный промежуток числовой прямой и имеется выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ объема n. Предположим, что задана некоторая фиксированная функция параметра θ: $\tau = \tau(\theta)$, $\theta \in \Theta$ (далее будем говорить *параметрическая функция*).

Задача оценки функции $\tau(\theta)$ состоит в том, чтобы найти такую статистику, которая приближала бы эту функцию по ряду критериев. Это – задача нахождения *статистической оценки*.

Формально, можно считать, что любая статистика может быть статистической оценкой. Однако, такой подход, вообще говоря, не приводит ни к чему полезному на практике. Необходимо ввести некоторые «характеристики качества» статистик.

5. Несмещенность и состоятельность оценок

Пусть, имеется статистическая структура $\mathfrak{L}(\Theta)$, выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ объема n из случайной величины ξ и заданная функция $\tau = \tau(\theta),\ \theta \in \Theta$.

Определение 1. Статистика $T = T(\vec{X})$ называется ***несмещённой*** *оценкой для параметрической функции $\tau(\theta)$, если $E_\theta T(\vec{X}) = \tau(\theta)$.*

Здесь

$$E_\theta T(\vec{X}) = \int\limits_{\mathcal{X}} T(x_1, \ldots, x_n) f(x_1, \theta) \cdot \ldots \cdot f(x_n, \theta) dx_1 \ldots dx_n$$

обозначает математическое ожидание случайной величины $T(\vec{X})$ в случае абсолютно непрерывного распределения; $f(x, \theta)$ – плотность распределения каждой компоненты выборки $\vec{X}$. Если случайная величины ξ имеет дискретное распределение, то

$$E_\theta T(\vec{X}) = \sum_{(x_1, \ldots, x_n) \in \mathcal{X}} T(x_1, \ldots, x_n) \cdot P_\theta(\xi = x_1) \cdot \ldots \cdot P_\theta(\xi = x_n).$$

Несмещенность, фактически, означает, что среднее значение от статистики по всем возможным наблюдениям из выборочного пространства $\mathcal{X}$ равно некоторому фиксированному значению заданной функции.

Пример. Пусть $Bi(\Theta, 1),\ \Theta = (0, 1)$ – статистическая структура выборки $\vec{X} = (X_1, X_2, \ldots, X_n)$ и пусть $\tau(\theta) = \theta$. Из условия, очевидно, имеем: $P(X_i = 0) = 1 - \theta,\ P(X_i = 1) = \theta, E_\theta X_i = \theta, i = 1, 2, \ldots, n$. Возьмем любой числовой вектор $\vec{a} = (a_1, a_2, \ldots, a_n)$,такой, что $a_1 + a_2 + \cdots + a_n = 1$ и определим статистику $T(\vec{X}) = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$. Очевидно, из свойства линейности математического ожидания получаем, $E_\theta T(\vec{X}) = \theta$, т.е. статистика $T(\vec{X})$, по определению, является несмещенной оценкой для θ.

Из этого примера видно, что условие несмещенности оценки может давать широкий класс статистик. Понятно, что ограничиваться только несмещенностью оценки, как правило, не достаточно для заключения о качестве оценки.

На практике могут встретится такие параметрические функции, для которых не существует несмещенных оценок. Поясним это на примере.

Пример. Пусть X_1 - выборка объема один из пуассоновского распределения с параметром $\theta > 0$, функция $\tau(\theta) = 1/\theta$. Предположим, что статистика $T(X_1)$ есть несмещенная оценка для этой функции, т.е. $e^{-\theta} \sum\limits_{i=0}^{\infty} T(i)\frac{\theta^i}{i!} = 1/\theta$. Перенося экспоненту из левой части в правую, а θ из правой части – влево и воспользовавшись разложением экспоненты, получаем: $\sum\limits_{i=0}^{\infty} T(i)\frac{\theta^{i+1}}{i!} = \sum\limits_{i=0}^{\infty} \frac{\theta^i}{i!}$. Отсюда заключаем, что не существует функции T, не зависящей от θ, такой, что будет справедливо последнее равенство (теорема о единственности разложения в степенной ряд).

Предположим теперь, что объем выборки может изменяться. В этом случае в обозначении выборки добавим индекс: $\vec{X}_n = (X_1, X_2, \ldots, X_n)$.

Определение 2. Статистика $T_n = T_n(\vec{X}_n)$ называется ***асимптотически несмещенной*** *оценкой для параметрической функции $\tau(\theta)$, если $\lim\limits_{n\to\infty} E_\theta T_n(\vec{X}_n) = \tau(\theta)$.*

Рассмотрим пример. Пусть имеется выборка $\vec{X}_n = (X_1, \ldots, X_n)$ независимых, одинаково распределенных случайных величин. Пусть математическое ожидание a этих величин известно. Рассмотрим статистику $T_n(\vec{X}_n) = \sum\limits_{i=1}^{n} (X_i - a)^2/n$. Математическое ожидание $ET_n(\vec{X}_n) =$

$= n\theta^2/n = \theta^2$, где $\theta^2 = E(X_i - a)^2$ — дисперсия случайной величины X_i $(i = 1, \ldots, n)$. Следовательно, $T_n(\vec{X}_n)$ является *несмещенной* оценкой дисперсии случайной величины X_i. Теперь допустим, что математическое ожидание не известно. Возьмем в качестве оценки математического ожидания известную статистику $\bar{X} = \sum\limits_{i=1}^{n} X_i/n$. Можно показать, что для статистики $T_n(\vec{X}_n) = \sum\limits_{i=1}^{n}(X_i - \bar{X})^2/n$ справедливо равенство $ET_n(\vec{X}_n) = (n-1)\theta^2/n$, где, как и выше, θ^2 — дисперсия X_i. Таким образом, эта статистика является смещенной оценкой дисперсии. Однако при $n \to \infty$ величина $(n-1)\theta^2/n \to \theta^2$. То есть, эта оценка является *асимптотически несмещенной* для дисперсии.

Заметим, что любая несмещенная оценка также будет и асимптотически несмещенной для той же параметрической функции. В практических задачах желательно выбирать такие статистики, чтобы они были хотя бы асимптотически несмещенными для оцениваемой параметрической функции.

Определение 3. Статистика $T_n = T_n(\vec{X}_n)$ называется ***состоятельной*** *оценкой для параметрической функции $\tau(\theta)$, если*

$$\lim_{n\to\infty} P_\theta(|T_n(\vec{X}_n) - \tau(\theta)| > \varepsilon) = 0$$

для произвольно заданного $\varepsilon > 0$.

Для состоятельности статистики $T_n(\vec{X}_n)$ достаточно, например, асимптотической несмещенности этой статистики для параметрической функции $\tau(\theta)$ и стремление к нулю дисперсии статистики T_n с ростом числа наблюдений. Действительно, из неравенства Чебышева, получаем

$$\lim_{n\to\infty} P_\theta(|T_n(\vec{X}_n) - \tau(\theta)| > \varepsilon) \le \lim_{n\to\infty} P_\theta(|T_n(\vec{X}_n) - ET_n(\vec{X}_n)| +$$

$$+|ET_n(\vec{X}_n) - \tau(\theta)| > \varepsilon) \le \lim_{n\to\infty} P_\theta(|T_n(\vec{X}_n) - ET_n(\vec{X}_n)| > \varepsilon) +$$

$$+ \lim_{n\to\infty} P_\theta(|ET_n(\vec{X}_n) - \tau(\theta)| > \varepsilon) \le \lim_{n\to\infty} \frac{D_\theta T_n(\vec{X}_n)}{\varepsilon^2} = 0.$$

Здесь использовано известное неравенство

$$|a-b| \le |a-c| + |c-b|,$$

справедливое для любых трех чисел a, b, c, а также то, что событие $A = \{|T_n(\vec{X}_n) - \tau(\theta)| > \varepsilon\}$ влечет событие $B = \{|T_n(\vec{X}_n) - ET_n(\vec{X}_n)| + |ET_n(\vec{X}_n) - \tau(\theta)| > \varepsilon\}$, то есть $A \subseteq B$ и, значит, $P(A) \le P(B)$.

6. Выборочная функция распределения и выборочные моменты

а) Выборочная функция распределения. Рассмотрим одну интересную статистику, которая очень часто используется в практических и теоретических вопросах.

Пусть $A \in R^1$ – некоторое множество, функция

$$I_A(x) = \begin{cases} 1, & \text{если } x \in A \\ 0, & \text{если } x \notin A \end{cases}$$

называется индикатором множества A.

Предположим, что выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ из невырожденного распределения, т.е. не являющегося постоянной, с неизвестной функцией распределения $F(x)$. Нашей задачей будет изучение такой статистики (в данном случае функции не только от выборки $\vec{X}$, но и от независимой переменной x), чтобы эта статистика была наилучшей оценкой (несмещенной и состоятельной) для неизвестной функции распределения $F(x)$.

Статистика $F_n(x) = \sum\limits_{i=1}^{n} I_{(-\infty,x]}(X_{(i)})/n = \sum\limits_{i=1}^{n} I_{(-\infty,x]}(X_i)/n$ называется ***выборочной функцией распределения***. Случайная величина $I_{(-\infty,x]}(X_i)$ имеет бернуллиевское распределение с вероятностью «успеха» равной $F(x)$ для всех $i = 1, \ldots, n$. В самом деле, по определению

$$I_{(-\infty,x]}(X_i) = \begin{cases} 1, & \text{если } X_i \leq x; \\ 0, & \text{если } X_i > x, \end{cases}$$

но событие $A = \{X_i \leq x\}$ при каждом фиксированном x имеет вероятность $P(X_i \leq x) = F(x)$, а противоположное событие имеет вероятность, равную $1 - F(x)$. Поэтому, математическое ожидание $EI_{(-\infty,x]}(X_i) = F(x)$ и дисперсия $DI_{(-\infty,x]}(X_i) = F(x)(1-F(x))$. Следовательно, из свойства линейности математического ожидания получаем $EF_n(x) = F(x)$, что означает несмещённость выборочной функции распределения. Используя свойства дисперсии, получаем, $DF_n(x) = \sum\limits_{i=1}^{n} DI_{(-\infty,x]}(X_i)/n^2 = F(x)(1-F(x))/n$. Далее, из неравенства Чебышева, для любого $\varepsilon > 0$, не зависящего от n, следует

$$P(|F_n(x)-EF_n(x)| > \varepsilon) = P(|F_n(x)-F(x)| > \varepsilon) \leq \frac{DF_n(x)}{\varepsilon^2} =$$

$$= \frac{F(x)(1-F(x))}{n\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0,$$

т.е. получаем, что выборочная функция распределения является состоятельной оценкой теоретической функции распределения.

Допустим теперь, что в реальной ситуации необходимо понять, какой же вид имеет функция распределения случайной величины, которую мы изучаем. В нашем распоряжении есть только числовой вектор наблюдений за случайной величиной ξ: $\vec{x} = (x_1, x_2, \ldots, x_n)$, иначе говоря, *реализация выборки*. Из несмещенности и состоятельности выборочной функции распределения можно сделать вывод, что функция $G(x) = \sum\limits_{i=1}^{n} I_{(-\infty,x]}(x_i)/n$, представляет собой наилучшую функцию распределения в качестве «эскиза» реальной функции распределения.

б) Выборочные моменты. Пусть $\vec{X} = (X_1, \ldots, X_n)$ – выборка случайной величины ξ с функцией распределения $F(x)$ и $F_n(x)$– выборочная функция распределения. Наряду с теоретической характеристикой распределения $Eg(\xi) = \int\limits_{-\infty}^{+\infty} g(x)dF(x)$ для некоторой функции $g(\xi)$, можно определить *«практическую»* или *выборочную* характеристику $\int\limits_{-\infty}^{+\infty} g(x)dF_n(x) = \frac{1}{n}\sum\limits_{i=1}^{n} g(X_i)$.

Замечание. Здесь используется обозначение

$$\int\limits_{-\infty}^{+\infty} g(x)dF(x),$$

которое надо понимать следующим образом:

1) если функция $F(x)$ дифференцируемая, то

$$\int\limits_{-\infty}^{+\infty} g(x)dF(x) = \int\limits_{-\infty}^{+\infty} g(x)F'(x)dx$$

и, по существу, имеем обычную формулу для математического ожидания

$$\int\limits_{-\infty}^{+\infty} g(x)f(x)dx,$$

где $f(x) = F'(x)$ – плотность распределения случайной величины ξ;

2) если $-\infty = x_0 < x_1 \le \dots \le x_n < +\infty$ – точки разрыва *первого* рода функции $F(x)$, между которыми функция $F(x)$ – дифференцируема, то

$$\int\limits_{-\infty}^{+\infty} g(x)dF(x) = \sum_{k=1}^{n}\left\{\int\limits_{x_{k-1}}^{x_k} g(x)F'(x)dx + g(x_k)p_k\right\} + \\ + \int\limits_{x_n}^{+\infty} g(x)F'(x)dx,$$

где $p_k,\ k = 1, \dots, n$ – скачки функции $F(x)$ в точках разрыва (следует обратить внимание на то, что члены в последовательности $x_1, \dots, x_n$ могут повторяться);

3) в частности, если между точками разрыва функция

$F(x)$ сохраняет постоянное значение, то в этих промежутках $F'(x) = 0$ и в формуле выше все интегралы равны нулю. Поэтому,

$$\int_{-\infty}^{+\infty} g(x)dF(x) = \sum_{k=1}^{n} g(x_k)p_k.$$

Отсюда вытекает, что в случае выборочной функции распределения, когда скачки в точках разрывов $X_{(i)}$ одинаковы и равны $1/n$, имеем

$$\int_{-\infty}^{+\infty} g(x)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(X_{(i)}),$$

поскольку для порядковых статистик справедливы неравенства $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Окончательная же формула получается согласно правилу независимости суммы чисел от порядка следования слагаемых.

Если взять функцию $g(x) = x^k,\ k > 0$, то величина $\alpha_k = E\xi^k$, как известно из курса теории вероятностей, называется *моментом k-го порядка*, а статистику $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ будем называть ***выборочным моментом k−го порядка***. Обычно статистику $A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$ принято обозначать $\bar{X}$ и называть ***выборочным средним***.

Если в качестве функции выбрать $g(x) = (x-\alpha_1)^k$, то получаем $\beta_k = E(\xi-\alpha_1)^k$−*центральный момент k-го порядка*. Для того, чтобы задать статистический аналог последней теоретической величины β_k, берем функцию $g(x) = (x - \bar{X})^k$ и получаем $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$−***выборочный центральный момент k−го порядка***.

В частности, при $k = 2$ $\beta_2 = D\xi = \alpha_2 - \alpha_1^2$ − *дисперсия* *с.в.* ξ и $B_2 = S^2(\vec{X})-$ ***выборочная дисперсия***.

Аналогично, если выбрать $g(x) = |x|^k$, то получаем $\lambda_k = E|\xi|^k$ *– абсолютный момент k-го порядка* и $C_k = \frac{1}{n}\sum\limits_{i=1}^{n}|X_i|^k$ **– выборочный абсолютный момент k-го порядка**. Наконец, для $g(x) = |x - \alpha_1|^k$, получаем $\mu_k = E|\xi - \alpha_1|^k$ *– абсолютный центральный момент k-го порядка* и для $g(x) = |x - \bar{X}|^k$, получаем $D_k = \frac{1}{n}\sum\limits_{i=1}^{n}|X_i - \bar{X}|^k$**–выборочный абсолютный центральный момент k-го порядка**.

Отметим, что введенные в этом разделе выборочные моменты являются случайными величинами, в отличие от введенных ранее моментов, построенных по *реализации* выборки.

в) Связь между выборочными моментами и выборочными центральными моментами. Используя бином Ньютона и определение центральных выборочных моментов, введенных выше, получим

$$B_k = \sum_{l=0}^{k}(-1)^l C_k^l A_l \bar{X}^{k-l} = (1-k)\bar{X}^k + \sum_{l=2}^{k}(-1)^l C_k^l A_l \bar{X}^{k-l}. \quad (1)$$

В частности, при $k = 2$, получаем для выборочной дисперсии $S^2(\vec{X}) = A_2 - \bar{X}^2$. ***г) Математическое ожидание и дисперсия для выборочного среднего и выборочной дисперсии.*** Воспользовавшись свойством линейности математического ожидания и свойствами дисперсии, очевидно, получим

$$E\bar{X} = \frac{1}{n}\sum_{j=1}^{n} EX_i = E\xi = \alpha_1, \quad D\bar{X} = \frac{1}{n^2}\sum_{j=1}^{n} DX_i =$$

$$= \frac{\beta_2}{n} \to 0 \quad \text{при} \quad n \to \infty,$$

откуда следует, что выборочное среднее является *несмещенной* и *состоятельной* оценкой математического ожидания. Кстати, в силу равенства

$$EA_k = \frac{1}{n}\sum_{j=1}^{n} EX_i^k = E\xi^k = \alpha_k$$

заключаем, что выборочный момент любого порядка k будет несмещенной оценкой для теоретического момента того же порядка k.

Из установленных соотношений, в частности, получаем

$$ES^2 = EA_2 - E\bar{X}^2 = \alpha_2 - \alpha_1^2 - \frac{\beta_2}{n} = \beta_2 - \frac{\beta_2}{n} = (1 - \frac{1}{n})\beta_2.$$

Можно показать, что дисперсия статистики S^2 равна

$$DS^2 = \frac{(n-1)^2}{n^3}\left(\beta_4 - \frac{n-3}{n-1}\beta_2^2\right).$$

Из последних соотношений вытекает, что статистика S^2 является *асимптотически несмещенной* оценкой для теоретической дисперсии $\beta_2 = D\xi$ случайной величины ξ. Более того, в случае, когда для той же случайной величины теоретический центральный момент четвертого порядка $\beta_4 = E(\xi - E\xi)^4$ существует и конечен, то статистика S^2 будет также и состоятельной оценкой той же дисперсии. И, наконец, статистика

$$s^2(\vec{X}) = \frac{n}{n-1}S^2(\vec{X}) = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2$$

будет *несмещенной* оценкой дисперсии.

7. Оптимальные оценки неизвестных параметров

Снова предполагаем, что задана статистическая структура $\mathfrak{L}(\Theta)$, выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ объема n и параметрическая функция $\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta$.

Если рассмотреть все несмещенные оценки для одной и той же параметрической функции (этот класс оценок обозначим $\mathbb{T}_{\boldsymbol{\tau}(\Theta)}(\vec{X})$), то понятно, что предпочтительной будет та оценка, у которой дисперсия будет минимальной для всех значений параметра $\boldsymbol{\theta} \in \Theta$, если, конечно, такая оценка существует.

Определение 1. Статистика $T^* = T^*(\vec{X})$ называется ***оптимальной*** *оценкой для параметрической функции* $\boldsymbol{\tau} = \boldsymbol{\tau}(\boldsymbol{\theta})$*, если для любой несмещенной оценки $T(\vec{X})$ выполняется неравенство $D_{\boldsymbol{\theta}} T^*(\vec{X}) = \inf\limits_{T(\vec{X}) \in \mathbb{T}_{\boldsymbol{\tau}(\Theta)}(\vec{X})} D_{\boldsymbol{\theta}} T(\vec{X})$ для всех $\boldsymbol{\theta} \in \Theta$.*

Замечание. В определении 1 предполагается, что оптимальная оценка также принадлежит классу несмещенных. Далее, напомним, что называется инфинумом (**inf**). Пусть дано числовое множество B. Оно называется ограниченным снизу, если найдется такое число $\boldsymbol{\alpha}$, что для всякого $b \in B$ выполняется неравенство $\boldsymbol{\alpha} \leq b$. Наибольшее из таких чисел $\boldsymbol{\alpha}$ называется точной нижней границей, или нижней гранью, или инфинумом. В математическом анализе доказывается, что любое, ограниченное снизу множество, обладает инфинумом. Однако инфинум может принадлежать, а может и не принадлежать самому множеству. Если в множестве есть наименьший элемент, то инфинум принадлежит множеству и совпадает с этим элементом. Иначе, если в множестве наименьшего элемента нет, то инфинум есть все равно, но он уже не принадлежит самому множеству. Например, на отрезке [2,5] имеется наименьшее число 2, поэтому, $2 = \mathbf{inf}[2, 5]$. С другой стороны, скажем, в множестве положительных чисел $(0, +\infty)$ наименьшего числа, очевидно, нет, а инфинум есть и равен нулю.

Утверждение 1. Оптимальная оценка, если она существует, единственна п.н.

Здесь *п.н.* означает *почти наверное*, т.е. если T_1^* и T_2^*—оптимальные оценки, то событие $T_1^* = T_2^*$ выполняется с вероятностью единица.

Доказательство. Предположим, что имеются две разные п.н. оптимальные оценки T_1 и T_2 для одной и той же параметрической функции $\tau(\theta)$, $\theta \in \Theta$. Из этого предположения имеем: $E_\theta T_1 = E_\theta T_2 = \tau(\theta)$ и $D_\theta T_1 = D_\theta T_2 = \sigma^2$, где $\sigma^2 > 0$. Определим новую статистику $T_3 = (T_1 + T_2)/2$. Очевидно,

$$E_\theta T_3 = \frac{1}{2}(E_\theta T_1 + E_\theta T_2) = \tau(\theta), \tag{2}$$

$$D_\theta T_3 = \frac{1}{4}(D_\theta T_1 + D_\theta T_2) + \frac{1}{2} cov_\theta(T_1, T_2). \tag{3}$$

Из неравенства Коши—Буняковского имеем

$$cov_\theta(T_1, T_2) \leq \sqrt{D_\theta T_1 D_\theta T_2} = \sigma^2 \tag{4}$$

и равенство справедливо тогда и только тогда, когда

$$T_2 = aT_1 + b \quad \text{п.н.} \tag{5}$$

Поэтому, из (3) и (4) получаем $D_\theta T_3 \leq \sigma^2$. С другой стороны, вспоминаем, что T_1—оптимальная оценка и, поэтому, $D_\theta T_3 \geqslant D_\theta T_1 = \sigma^2$. Таким образом, $D_\theta T_3 = \sigma^2$ и из (2) следует, что T_3—оптимальная оценка. Следовательно, из (3) получаем, что в (4) выполняется равенство, откуда заключаем, что выполняется равенство (5). Поэтому, $\sigma^2 =$

$= cov_{\theta}(T_1, aT_1 + b) = aD_{\theta}T_1 = a\sigma^2,$ откуда следует, что $a = 1$. На основании несмещенности оценок T_1 и T_2 из (5) следует, что $b = 0$ и поэтому получаем $T_1 = T_2$ п.н., что и доказывает утверждение 1.

Утверждение 2. Пусть T_i^* – оптимальные оценки для параметрических функций $\tau_i = \tau_i(\theta)$ для всех $\theta \in \Theta$ и a_i – произвольные постоянные $i = 1, \ldots, k$. Тогда статистика $T^* = a_1T_1^* + \cdots + a_kT_k^*$ является оптимальной оценкой для параметрической функции $\tau(\theta) = a_1\tau_1(\theta) + \ldots + a_k\tau_k(\theta)$ для всех $\theta \in \Theta$.

Последние утверждение оставим без доказательства.

Иногда оптимальную оценку называют *несмещённой оценкой с равномерно (относительно параметра $\theta \in \Theta$) минимальной дисперсией* или, сокращенно, ***нормд** оценкой.*

8. Эффективные оценки неизвестных параметров для регулярных распределений. Неравенство Рао—Крамера

Предположим, что дифференцируемая по $\theta \in \Theta$ функция $f(x, \theta)$ обозначает или распределение некоторой дискретной случайной величины ($P_\theta(\xi = x) = f(x, \theta)$, например, для ξ), или плотность некоторой абсолютно-непрерывной случайной величины. Пусть имеется выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ объема n. По определению, $L(\vec{x}, \theta) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta) -$ совместное распределение выборки $\vec{X}$, $L(\vec{x}, \theta) > 0$, если $\vec{x} \in \mathcal{X}$, где $\mathcal{X}-$выборочное пространство выборки $\vec{X}$. Заменим, теперь, числовой вектор $\vec{x}$ в $L(\vec{x}, \theta)$ выборкой $\vec{X}$. Очевидно, в результате этого получаем случайную величину, зависящую от параметра θ.

Определение 1. Случайная величина $L(\vec{X}, \theta)$, $\theta \in \Theta$ называется ***функцией правдоподобия****.*

Определение 2. Распределение выборки $\vec{X}$ называется ***регулярным****, если для любой непрерывной и дифференцируемой по θ функции Ψ выполняется равенство*

$$\frac{\partial}{\partial \theta} E_\theta \Psi(\vec{X}, \theta) = E_\theta \Psi'_\theta(\vec{X}, \theta).$$

Замечание. Из мат. анализа известно, что для того, чтобы распределение было регулярным, необходимо, чтобы выборочное пространство $\mathcal{X}$ не зависело от параметра θ.

Определение 3. Случайная величина $Q_n(\vec{X}, \theta) = \frac{\partial ln L(\vec{X}, \theta)}{\partial \theta}$ называется ***вкладом выборки*** *$\vec{X}$ объёма n.*

Определение 4. Величина $i_n(\theta) = D_\theta Q_n(\vec{X}, \theta)$ *называется* ***информацией Фишера*** *для выборки* $\vec{X}$ *объёма* n.

Легко проверить, что $i_n(\theta) = n i_1(\theta)$. Это следует из свойства дисперсии суммы независимых случайных величин, свойства логарифма и определения выборки.

Теорема 1. Пусть задана выборка $\vec{X} = (X_1, X_2, \dots, X_n)$ *объема* n *из некоторого регулярного распределения* $\mathfrak{L}(\Theta)$*, и пусть* $T = T(\vec{X})$ *несмещенная статистика для некоторой гладкой (непрерывно дифференцируемой) функции* $\tau = \tau(\theta)$*,* $\theta \in \Theta$*. Тогда справедливо* ***неравенство Рао—Крамера***

$$D_\theta T(\vec{X}) \geqslant \frac{(\tau'(\theta))^2}{n i_1(\theta)} \quad \text{для любого} \quad \theta \in \Theta. \tag{6}$$

Доказательство. Покажем сначала, что для регулярного распределения функция вклада $Q_n(\theta)$ имеет нулевое математическое ожидание. Действительно, из определения распределения выборки $\vec{X}$ имеем

$$\int\limits_{\mathcal{X}} L(\vec{x}, \theta) d\vec{x} = 1 \quad \text{для любого} \quad \theta \in \Theta.$$

Взяв производные по θ левой и правой частей последнего равенства и воспользовавшись условием регулярности, получаем

$$\frac{\partial}{\partial\theta} \int\limits_{\mathcal{X}} L(\vec{x}, \theta) d\vec{x} = 0 \quad \Rightarrow \int\limits_{\mathcal{X}} \frac{\partial L(\vec{x}, \theta)}{\partial\theta} d\vec{x} = 0 \quad \Rightarrow$$

$$\Rightarrow \int\limits_{\mathcal{X}} \frac{L'_\theta(\vec{x}, \theta)}{L(\vec{x}, \theta)} L(\vec{x}, \theta) d\vec{x} = 0 \quad \Rightarrow \int\limits_{\mathcal{X}} \frac{\partial ln L(\vec{x}, \theta)}{\partial\theta} L(\vec{x}, \theta) d\vec{x} = 0 \quad \Rightarrow$$

$$\Rightarrow E_\theta\left(\frac{\partial lnL(\vec{X},\theta)}{\partial\theta}\right)=0 \quad \Rightarrow E_\theta Q_n(\vec{X},\theta)=0.$$

Из полученного равенства, в частности, получаем, что информацию Фишера можно считать по формуле

$$i_n(\theta)=E_\theta Q_n^2(\vec{X},\theta).$$

Из курса теории вероятностей вспомним, что если имеются две случайные величины ξ и η, у одной из которых нулевое математическое ожидание, то $cov(\xi,\eta)=E\xi\eta$. Итак, пусть статистика $T(\vec{X})$ есть несмещенная оценка для $\tau(\theta)$, т.е.

$$\tau(\theta)=\int\limits_{\mathcal{X}} T(\vec{x})L(\vec{x},\theta)d\vec{x} \quad \text{для любого} \quad \theta\in\Theta.$$

Как и выше, возьмем производные левой и правой частей по θ последнего равенства и воспользуемся регулярностью распределения, получим

$$\tau'(\theta)=\int\limits_{\mathcal{X}} T(\vec{x})\frac{\partial L(\vec{x},\theta)}{\partial\theta}d\vec{x}=\int\limits_{\mathcal{X}} T(\vec{x})\frac{\partial lnL(\vec{x},\theta)}{\partial\theta}L(\vec{x},\theta)d\vec{x}=$$

$$=E_\theta\left(T(\vec{X})Q_n(\vec{X},\theta)\right)=cov_\theta\left(T(\vec{X}),Q_n(\vec{X},\theta)\right)\leqslant$$

$$\leqslant\sqrt{D_\theta T(\vec{X})i_n(\theta)}, \tag{7}$$

откуда и следует неравенство (6).

Определение 5. Статистика $T(\vec{X})$ называется ***эффективной оценкой*** *для параметрической функции $\tau(\theta)$, если в неравенстве (6)* достигается равенство для всех $\theta\in\Theta$.

Знак неравенства в доказательстве теоремы появляется в (7) из неравенства Коши-Буняковского. Если оценка будет эффективной, то в (7) вместо знака неравенства должен быть знак равенства, что сводится к условию равенства в неравенстве Коши-Буняковского. Вспомним, что для этого необходимо, чтобы случайные величины, стоящие под знаком ковариации были линейно связаны. В нашем случае должно быть $T(\vec{X}) = \alpha(\theta)Q_n(\vec{X}, \theta) + \beta(\theta)$, т.е. статистика T должна быть линейной функцией от функции вклада выборки. Отметим, что $\alpha(\theta)$ и $\beta(\theta)$ не должны зависеть от выборки $\vec{X}$. Взяв математическое ожидание левой и правой частей последнего равенства и вспомнив, что функция вклада выборки для регулярных распределений имеет нулевое математическое ожидание, а статистика $T-$несмещенная оценка для τ, получим следующее

Утверждение. Для того, чтобы статистика $T(\vec{X})$ *была эффективной оценкой для параметрической функции* $\tau(\theta)$ *необходимо, чтобы выполнялось равенство*

$$T(\vec{X}) = \alpha(\theta)Q_n(\vec{X}, \theta) + \tau(\theta). \tag{8}$$

Пример. Пусть имеется выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ из бернуллиевской статистической структуры с параметром $\theta \in (0, 1)$. Покажем, что статистика $\overline{X} = \sum\limits_{i=1}^{n} X_i$ является несмещенной эффективной оценкой для параметрической функции $\tau(\theta) = \theta$. Сначала проверим несмещенность. Поскольку, для любого $i = 1, \ldots, n$ имеем $E_\theta X_i = 1 \cdot \theta + 0 \cdot (1 - \theta) = \theta$, то, из свойства линейности математического ожидания, получаем $E_\theta \overline{X} = \theta$, т.е., статистика $\overline{X}$ является несмещенной оценкой для θ. Дальше, поскольку распределение случайной величины Бернулли равно $f(x) = \theta^x(1 - \theta)^{(1-x)}, \quad x = 0, 1$, то функция правдоподобия в нашем случае равна $L(\vec{X}, \theta) = \theta^{n\overline{X}} \cdot (1 - \theta)^{n(1-\overline{X})} = exp\left(n\overline{X}ln\theta + nln(1 - \theta) - n\overline{X}ln(1 - \theta)\right)$. Поэтому, функция вклада $Q_n(\vec{X}, \theta) = n\overline{X}/(\theta(1-\theta)) - n/(1-\theta)$

и информация Фишера будет равна $i_n(\theta) = D_\theta Q_n(\vec{X},\theta) = n/(\theta(1-\theta))$ (воспользовались свойством дисперсии $D(\xi+c) = D\xi$). Наконец, в силу того, что $D_\theta\overline{X} = \theta(1-\theta)/n$ (из свойства дисперсии $Dc(\xi_1+\xi_2) = c^2(D\xi_1 + D\xi_2)$, где ξ_1 и ξ_2 – независимые случайные величины) и в нашем случае $\tau'(\theta) = 1$, получаем, что в (6) достигается равенство для всех $0 < \theta < 1$, что и доказывает наше утверждение.

Дополнение. Рассмотрим аналогичные результаты в случае векторного неизвестного параметра $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta \subset R^k$. Пусть задана параметрическая функция $\tau = \tau(\vec{\theta})$, а статистика $T = T(\vec{X})$ есть несмещенная оценка для этой параметрической функции $\tau(\vec{\theta})$, т.е. $E_{\vec{\theta}}T(\vec{X}) = \tau(\vec{\theta})$.

Пусть выполняются условия регулярности

$$\frac{\partial}{\partial\theta_i}E_{\vec{\theta}}\,\Psi(\vec{X},\vec{\theta}) = E_{\vec{\theta}}\,\frac{\partial}{\partial\theta_i}\Psi(\vec{X},\vec{\theta}),\quad i = 1, 2, \ldots, k.$$

Определим вектор, зависящий от векторного параметра $\vec{\theta}$

$$\overrightarrow{J(\vec{\theta})} = \left(\frac{\partial\tau(\vec{\theta})}{\partial\theta_1},\ \frac{\partial\tau(\vec{\theta})}{\partial\theta_2}, \ldots,\ \frac{\partial\tau(\vec{\theta})}{\partial\theta_k}\right).$$

Теорема 2. Пусть информационная матрица

$$I(\vec{\theta}) = \left\| E_{\vec{\theta}}\left(\frac{\partial lnL(\vec{X},\vec{\theta})}{\partial\theta_i}\cdot\frac{\partial lnL(\vec{X},\vec{\theta})}{\partial\theta_j}\right)\right\|,\ i, j = 1, \ldots, k$$

положительно определена и ее определитель $det\, I(\vec{\theta}) < \infty$. *Тогда справедливо обобщённое неравенство Рао—Крамера*

$$D_{\vec{\theta}}\,T(\vec{X}) \ \geqslant\ \overrightarrow{J(\vec{\theta})}I^{-1}(\vec{\theta})\left(\overrightarrow{J(\vec{\theta})}\right)^T.$$

Это неравенство превращается в равенство тогда и *только тогда, когда*

$$\boldsymbol{T(\vec{X}) - \tau(\vec{\theta}) = \sum_{i=1}^{k} a_i(\vec{\theta}) \frac{\partial ln L(\vec{X}, \vec{\theta})}{\partial \theta_i}},$$

где $a_i(\vec{\boldsymbol{\theta}})$ *– некоторые функции* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*.*

Эту теорему оставляем без доказательства.

8.1. Неравенство Бхаттачария. Оптимальные оценки неизвестных параметров по Бхаттачария

Если дисперсия $D_\theta T(\vec{X})$ несмещённой оценки для параметрической функции $\tau(\theta)$ в неравенстве Рао—Крамера не достигает равенства, то в некоторых случаях удается найти лучшую нижнюю границу для $D_\theta T(\vec{X})$. В этом разделе, для простоты, будем рассматривать лишь скалярный параметр, то есть $\theta \in \Theta \subset R^1$. Пусть $L(\theta) = L(\vec{X}, \theta)$ – функция правдоподобия, $L^{(i)}(\theta) = \partial^i L(\vec{X}, \theta)/\partial\theta^i$, $\tau^{(i)}(\theta) = \partial^i \tau(\theta)/\partial\theta^i$ и пусть $\overrightarrow{J(\theta)} = \left(\tau^{(1)}(\theta), \tau^{(2)}(\theta), \ldots, \tau^{(r)}(\theta)\right)$, $r \geqslant 2$ (вектор понимается как вектор-строка). Следующее утверждение приведем без доказательства.

Теорема. Пусть выполняются более строгие условия регулярности

$$E_\theta \left(T(\vec{X}) \frac{\partial^r L(\vec{X}, \theta)}{\partial\theta^r} \right) = \tau^{(r)}(\theta),$$

выборочное множество $\mathcal{X} \subset R^n$ не зависит от θ. Пусть квадратная матрица размерности $r \times r$

$$A(\theta) = \left\| E_\theta \frac{L^{(i)}(\theta)\ L^{(j)}(\theta)}{L^2(\theta)} \right\|, \quad i, j = 1, 2, \ldots, r.$$

является невырожденной, то есть существует обратная матрица $A^{-1}(\theta)$. Тогда справедливо неравенство Бхаттачария

$$D_\theta T(\vec{X}) \geqslant \overrightarrow{J(\theta)}\ A^{-1}(\theta)\ \left(\overrightarrow{J(\theta)}\right)^T$$

для любого $\theta \in \Theta$.

Это неравенство превращается в строгое равенство тогда и только тогда, когда

$$T(\vec{X}) - \tau(\theta) = \sum_{i=1}^{r} a_i(\theta) \frac{L^{(i)}(\theta)}{L(\theta)},$$

где $a_i(\theta)$, $i = 1, \ldots, r$*—некоторые функции.*

Замечание. Если выполняется последнее равенство, то статистику $T(\vec{X})$ называют *оптимальной по Бхаттачария* оценкой для параметрической функции $\tau(\theta)$.

Пример. Пусть имеется выборка $\vec{X} = (X_1, \ldots, X_n) \sim \sim N(\theta, \sigma^2)$ из нормального распределения с неизвестным математическим ожиданием θ и известной дисперсией σ^2. Функция правдоподобия равна

$$L(\vec{X}, \theta) = \frac{1}{(\sqrt{2\pi}\sigma)^n} exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \theta)^2\right).$$

Легко проверить, что

$$\frac{1}{L}\left(\frac{2\theta\sigma^2}{n} \frac{\partial L}{\partial \theta} + \frac{\sigma^4}{n^2} \frac{\partial^2 L}{\partial \theta^2}\right) = \bar{X}^2 - \frac{\sigma^2}{n} - \theta^2.$$

Отсюда, из неравенства Бхаттачария следует, что статистика $T(\vec{X}) = \bar{X}^2 - \sigma^2/n$ является оптимальной (несмещенной с минимальной дисперсией) оценкой для параметрической функции $\tau(\theta) = \theta^2$.

9. Класс регулярных экспоненциальных распределений

Пусть, как и в предыдущем разделе, $f(x, \theta)$ обозначает распределение в случае дискретной случайной величины, или — плотность распределения в случае абсолютно-непрерывной случайной величины (для простоты изложения, будем предполагать скалярным параметр θ).

Определение. Класс распределений $\mathfrak{L}(\Theta)$ *называется* ***регулярным экспоненциальным****, если*

$$f(x, \theta) = exp\left\{a(\theta)b(x) + c(\theta) + d(x)\right\}, \qquad (9)$$

где $a(\theta)$, $b(x)$, $c(\theta)$, $d(x)$*—некоторые фиксированные функции, а* $a(\theta)$ *и* $c(\theta)$ *— дифференцируемые по параметру* $\theta \in \Theta$ *функции и множество* $D = \{x \in R : f(x, \theta) \neq 0\}$ *не зависит от* θ*.*

Замечание. Последнее условие, по сути, есть необходимое условие регулярности распределения и требуется для справедливости неравенства Рао—Крамера (6).

Теорема. Эффективная оценка $T(\vec{X})$ *для некоторой параметрической функции* $\tau(\theta)$ *существует тогда и только тогда, когда статистическая структура является экспоненциальным регулярным классом. При этом статистика*

$$T(\vec{X}) = \sum_{i=1}^{n} b(X_i)/n$$

есть эффективная оценка для

$$\tau(\theta) = -c'(\theta)/a'(\theta).$$

Доказательство. Предположим, что имеется выборка $\vec{X}$ объема n экспоненциальной статистической структуры (9). Поэтому, функция правдоподобия будет равна

$$L(\vec{X}, \theta) = exp\left\{a(\theta)\sum_{i=1}^{n} b(X_i) + nc(\theta) + \sum_{i=1}^{n} d(X_i)\right\},$$

а функция вклада в этом случае будет иметь вид

$$Q_n(\vec{X}, \theta) = a'(\theta)\sum_{i=1}^{n} b(X_i) + nc'(\theta).$$

Для того, чтобы статистика $T(\vec{X})$ была бы эффективной оценкой для $\tau(\theta)$ требуется выполнение равенства (8). В нашем случае должно быть

$$T(\vec{X}) = \alpha(\theta)\left(a'(\theta)\sum_{i=1}^{n} b(X_i) + nc'(\theta)\right) + \tau(\theta),$$

или

$$T(\vec{X}) - \tau(\theta) = n\alpha(\theta)a'(\theta)\left(\frac{1}{n}\sum_{i=1}^{n} b(X_i) - \left(-\frac{c'(\theta)}{a'(\theta)}\right)\right),$$

откуда получаем, что

$$T(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n} b(X_i)$$

и

$$\tau(\theta) = -\frac{c'(\theta)}{a'(\theta)},$$

что и утверждалось.

Обратно, пусть выполняется (8), т.е. статистика $T(\vec{X})$−эффективная оценка для параметрической функции $\tau(\theta)$. Вся информация о статистической структуре в этом равенстве находится в функции вклада $Q_n(\vec{X}, \theta)$. Из (8) получаем

$$Q_n(\vec{X}, \theta) = \frac{\partial ln L(\vec{X}, \theta)}{\partial \theta} = \alpha^{-1}(\theta) T(\vec{X}) - \alpha^{-1}(\theta)\tau(\theta).$$

Взяв неопределенный интеграл по θ левой и правой части последнего равенства для всех $\theta \in \Theta$, очевидно, получаем

$$L(\vec{X}, \theta) = exp\left\{a(\theta)T(\vec{X}) + C(\theta) + D(\vec{X})\right\},$$

где $a'(\theta) = \alpha^{-1}(\theta)$, $C'(\theta) = \alpha^{-1}(\theta)\tau(\theta)$ и $D(\vec{X})$—величина, не зависящая от θ, появляющаяся после неопределенного интегрирования. Далее, из определения функции правдоподобия и выборки следует, что

$$L(\vec{X}, \theta) = \prod_{i=1}^{n} f(X_i, \theta) = exp\left\{n(a(\theta)\frac{1}{n}\sum_{i=1}^{n} b(X_i) + c(\theta) + \right.$$

$$\left. + \frac{1}{n}\sum_{i=1}^{n} d(X_i))\right\},$$

где $\sum\limits_{i=1}^{n} b(X_i) = nT(\vec{X})$, $nc(\theta) = C(\theta)$ и $\sum\limits_{i=1}^{n} d(X_i) = D(\vec{X})$, откуда следует вид (9). Теорема полностью доказана.

Пример 1. Пусть имеется выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ из статистической структуры нормального распределения $N(\theta, \sigma^2)$ (с неизвестным математическим ожиданием θ, $\theta \in (-\infty, \infty)$ и известной дисперсией $\sigma^2 > 0$). Ясно, что плотность распределения с.в. в этом случае имеет вид:

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

$$= exp\left(-\frac{x^2}{2\sigma^2} + \frac{x\theta}{\sigma^2} - \frac{\theta^2}{2\sigma^2} - ln(\sqrt{2\pi\sigma^2})\right).$$

Последнюю функцию можно представить в виде (9), где

$$a(\theta) = \theta/\sigma^2, \quad b(x) = x, \quad c(\theta) = -\theta^2/(2\sigma^2),$$

$$d(x) = -x^2/(2\sigma^2) - ln(\sqrt{2\pi\sigma^2}).$$

Поэтому, из последней теоремы получаем, что статистика

$$T(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

является *эффективной* оценкой для параметрической функции $\tau(\theta) = -c'(\theta)/a'(\theta) = \theta$. Таким образом, *среднеарифметическое наблюдений является эффективной (т.е. самой лучшей) оценкой для неизвестного математического ожидания при известной дисперсии в случае нормально распределенных наблюдений.*

Пример 2. Пусть имеется выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ из статистической структуры нормального распределения $N(a, \theta)$ (с известным математическим ожиданием a, $a \in (-\infty, \infty)$ и неизвестной дисперсией $\theta > 0$). Ясно, что плотность распределения с.в. в этом случае имеет вид:

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

$$= exp\left(-\frac{(x-a)^2}{2\theta} - ln(\sqrt{2\pi}) - ln(\theta)/2\right).$$

Последнюю функцию можно представить в виде (9), где

$$a(\theta) = -1/(2\theta), \quad b(x) = (x-a)^2, \quad c(\theta) = -ln(\theta)/2,$$

$$d(x) = -ln(\sqrt{2\pi}).$$

Поэтому, получаем, что статистика

$$T(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 = S^2(\vec{X})$$

является *эффективной* оценкой для параметрической функции $\tau(\theta) = -c'(\theta)/a'(\theta) = \theta$. Таким образом, статистика $S^2(\vec{X})$, называемая *выборочной дисперсией*, является эффективной оценкой для неизвестной дисперсии при известном математическом ожидании в случае нормально распределенных наблюдений.

Замечание. В случае выборки из статистической структуры нормального распределения с неизвестными математическим ожиданием и дисперсией можно показать, что статистика $\overline{X}$ будет эффективной оценкой для математического ожидания, а статистика

$$s^2(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

будет несмещенной оптимальной, но не эффективной оценкой для дисперсии.

10. Достаточные оценки и оптимальные оценки. Теорема Рао — Блэкуэлла — Колмогорова

В предыдущих разделах было показано, что эффективные оценки существуют только для экспоненциальных статистических структур и только для определенных параметрических функций. Поэтому понятно, что класс эффективных оценок достаточно мал. То же можно сказать и относительно оптимальных по Бхаттачария оценках. В практических задачах зачастую возникают ситуации, когда рассматриваются параметрические функции, для которых не может существовать эффективной или оптимальной по Бхаттачария оценки. Кроме этого, статистические структуры могут быть не регулярны или не экспоненциальны. В таких случаях необходимо выбрать принципиально другой подход для поиска оптимальных оценок.

Как и раньше, для упрощения изложения, будем подробно рассматривать скалярные параметры, а статистические структуры будут исключительно дискретные, хотя все рассматриваемые результаты справедливы и в случае абсолютно-непрерывных статистических структур.

Определение 1. Статистика $T(\vec{X})$ называется ***достаточной*** *для статистической структуры $\mathfrak{L}(\Theta) = \{f(x,\vec{\theta}),\ \vec{\theta} \in \Theta\}$, если условное распределение выборки $\vec{X}$ при фиксированном значении статистики T*

$$L(\vec{x} \mid t;\ \theta) = P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T(\vec{X}) = t)$$

не зависит от неизвестного параметра θ, т.е. в общем случае есть функция только, быть может, от $\vec{x}$ и t, принимающих произвольные значения.

Замечание. В случае абсолютно-непрерывной статистической модели, дискретное условное распределение заменяется на условную плотность

$$L(\vec{x} \mid t;\ \theta) = f_{\vec{X}|T(\vec{X})}(\vec{x} \mid t; \theta) = \frac{f_{\vec{X},\,T(\vec{X})}(\vec{x},\ t; \theta)}{f_{T(\vec{X})}(t; \theta)},$$

где $f_{\vec{X},\,T(\vec{X})}(\vec{x},\ t; \theta)$—совместная плотность распределения $\vec{X}$ и $T(\vec{X})$, а $f_{T(\vec{X})}(t; \theta)$—плотность распределения $T(\vec{X})$.

Это определение, по сути, означает, что все исследования неизвестного параметра статистической структуры используют только статистику T. Отметим, что достаточная статистика не единственна.

Теорема (критерий факторизации). Для того, чтобы статистика $T(\vec{X})$ была достаточной, необходимо и достаточно, чтобы распределение выборки $\vec{X}$ было представимо в виде

$$L(\vec{x};\ \theta) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n;\ \theta) =$$

$$= H(\vec{x}) \cdot G(T(\vec{x}), \theta), \tag{10}$$

где функция $H(\vec{x})$ не зависит от θ, а функция $G(T(\vec{x}), \theta)$ зависит от θ и от $\vec{x}$ через статистику T.

Замечание 1. В случае абсолютно-непрерывной статистической модели, дискретное распределение заменяется на совместную плотность распределения выборки $\vec{X}$

$$L(\vec{x};\ \theta) = f_{\vec{X}}(\vec{x};\ \theta) = \prod_{i=1}^{n} f(x_i; \theta),$$

где $f(x; \theta)$ плотность с.в. $X_i, \quad i = 1, 2, \ldots, n$.

Замечание 2. В общем случае параметр θ и, соответственно, статистика $T(\vec{X})$ в (10) могут быть векторными.

Доказательство. Предположим, что справедливо равенство (10). Тогда для любого числового вектора $\vec{x}$ и t, таких, что $T(\vec{x}) = t$ получаем по определению условной вероятности и независимости с.в. $X_i,\ i = 1, \ldots, n$, что

$$P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T(\vec{X}) = t) =$$

$$= \frac{P_\theta\left(\left\{\bigcap\limits_{i=1}^{n}\{X_i = x_i\}\right\}\bigcap\left\{\{T(\vec{X}) = t\}\right\}\right)}{P_\theta\left(T(\vec{X}) = t\right)} =$$

$$= \frac{P_\theta\left(\vec{X} = \vec{x}\right)}{\sum\limits_{\vec{u}:T(\vec{u})=t} P_\theta\left(\vec{X} = \vec{u}\right)} = \frac{H(\vec{x})G(t,\theta)}{\sum\limits_{\vec{u}:T(\vec{u})=t} H(\vec{u})G(t,\theta)} =$$

$$= \frac{H(\vec{x})}{\sum\limits_{\vec{u}:T(\vec{u})=t} H(\vec{u})}$$

не зависит от θ.

Обратно, пусть статистика T является достаточной. По определению это означает, что для любого $t \in \mathcal{D}(T(\vec{X}))$ из области изменения статистики T, следует, что

$$P_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid T(\vec{X}) = t) = W(\vec{x}, t),$$

где функция W не зависит от θ. Поскольку, очевидно, $\left\{\vec{X} = \vec{x}\right\} \subseteq \left\{T(\vec{X}) = t\right\}$ в случае $T(\vec{x}) = t$, то получим

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n;\ \theta) =$$

$$= P_\theta(\vec{X} = \vec{x},\ T(\vec{X}) = t) =$$

$$= P_\theta(\vec{X} = \vec{x} \mid T(\vec{X}) = t) \cdot P_\theta(T(\vec{X}) = t) =$$

$$= W(\vec{x}, t)\ G(T(\vec{x}), \theta),$$

что представляет из себя вид (10). Теорема полностью доказана.

Легко понять, что если заданная статистика $T(\vec{X})$ является достаточной, то любая взаимно-однозначная функция от нее $U(\vec{X}) = g(T(\vec{X}))$ будет тоже достаточной. Действительно, из взаимной однозначности функции g получаем, что $T(\vec{X}) = g^{-1}(U(\vec{X}))$ и, поэтому, из (10) получаем

$$L(\vec{x};\ \theta) = H(\vec{x}) \cdot G(g^{-1}(U(\vec{X})), \theta) = H(\vec{x}) \cdot G_1(\vec{X}, \theta),$$

т.е. статистика $U(\vec{X})$ из критерия факторизации будет достаточной. В частности, отсюда и представления эффективной оценки в экспоненциальном классе получаем, что любая взаимно-однозначная функция от эффективной оценки будет достаточной статистикой.

Пример 1. Пусть $\Theta = (0, +\infty)$ и статистическая структура является классом

$$R(\Theta) = \left\{ f(x, \theta) = \theta^{-1} I_{[0,\theta]}(x), \quad \theta \in \Theta \right\}$$

равномерных распределений на отрезке $[0, \theta]$. Функция правдоподобия в этом случае равна

$$L(\vec{X}, \theta) = \prod_{i=1}^{n} \theta^{-1} I_{[0,\theta]}(X_i) = \theta^{-n} I_{[0,\theta]}(X_{(n)}).$$

Поэтому, из критерия факторизации, получаем, что статистика $T(\vec{X}) = X_{(n)}$ является достаточной статистикой.

Пример 2. Пусть $\Theta = \{(\theta_1, \theta_2),\ \theta_1 < \theta_2;\ \theta_1, \theta_2 \in (-\infty, +\infty)\} \subset R^2$ и статистическая структура является классом $R(\Theta) = \left\{ f(x, \vec{\theta}) = (\theta_2 - \theta_1)^{-1} I_{[\theta_1, \theta_2]}(x), \quad \theta \in \Theta \right\}$ равномерных распределений на отрезке $[\theta_1, \theta_2]$. Функция правдоподобия в этом случае равна

$$L(\vec{X}, \theta) = \prod_{i=1}^{n} (\theta_2 - \theta_1)^{-1} I_{[\theta_1, \theta_2]}(X_i) =$$

$$= (\theta_2 - \theta_1)^{-n} I_{[\theta_1, +\infty)}(X_{(1)}) I_{(-\infty, \theta_2]}(X_{(n)}).$$

Поэтому, на основании критерия факторизации, заключаем, что векторная статистика $\overrightarrow{T(\vec{X})} = (X_{(1)}, X_{(n)})$ будет достаточной двумерной статистикой.

Определение 2. Достаточная статистика $T(\vec{X})$ называется ***минимальной достаточной статистикой****, если эта статистика является функцией любой другой достаточной статистики.*

Фактически это означает *минимально возможную размерность* для минимальной достаточной статистики.

Из определения достаточной статистики, в частности, следует, что сама выборка $\vec{X}$ тоже будет достаточной статистикой (многомерной). В практических задачах удобнее, как правило, иметь дело с минимальными (в смысле размерности) достаточными статистиками.

Пример 3. Пусть $\Theta \subset R^1$, $\alpha = \alpha(\theta)$, $\beta = \beta(\theta) : \alpha(\theta) < < \beta(\theta)$ для всех $\theta \in \Theta$ и статистическая структура является классом

$$R(\Theta, \alpha, \beta) = \Big\{ f(x, \alpha(\theta), \beta(\theta)) =$$

$$= (\beta(\theta) - \alpha(\theta))^{-1} I_{[\alpha(\theta), \beta(\theta)]}(x), \quad \theta \in \Theta \Big\}$$

равномерных распределений на отрезке $[\alpha(\theta), \beta(\theta)]$. Функция правдоподобия равна

$$L(\vec{X}, \theta) = (\beta(\theta) - \alpha(\theta))^{-n} I_{[\alpha(\theta), +\infty)}(X_{(1)}) I_{(-\infty, \beta(\theta)]}(X_{(n)}). \quad (11)$$

Поэтому, как и в предыдущем примере, из критерия факторизации, получим, что векторная статистика $\overrightarrow{T(\vec{X})} = (X_{(1)}, X_{(n)})$ – достаточная двумерная статистика в общем случае.

Рассмотрим теперь частные случаи.

i) Пусть для $\theta \in \Theta$ функции $\alpha(\theta) \uparrow$ и $\beta(\theta) \downarrow$ будут строго монотонны и иметь разные знаки монотонности. Очевидно, что $I_{[\alpha(\theta),+\infty)}(X_{(1)})I_{(-\infty,\beta(\theta)]}(X_{(n)}) = 1$, для подмножества Θ:

$$\left\{\theta \in \Theta\colon \alpha(\theta) \leqslant X_{(1)}\right\} \cap \left\{\theta \in \Theta\colon \beta(\theta) \geqslant X_{(n)}\right\},$$

или, воспользовавшись предположениями о монотонности функций $\alpha(\theta)$ и $\beta(\theta)$ это множество можно записать в виде

$$\left\{\theta \in \Theta\colon \theta \leqslant \alpha^{-1}(X_{(1)})\right\} \cap \left\{\theta \in \Theta\colon \theta \leqslant \beta^{-1}(X_{(n)})\right\} = \\ = \left\{\theta \in \Theta\colon \theta \leqslant \min\left[\alpha^{-1}(X_{(1)}), \beta^{-1}(X_{(n)})\right]\right\}.$$

Воспользовавшись этим, функцию правдоподобия (11) можно записать в виде

$$L(\vec{X}, \theta) = (\beta(\theta) - \alpha(\theta))^{-n} I_{[\theta,+\infty)}\left(\min\left[\alpha^{-1}(X_{(1)}), \beta^{-1}(X_{(n)})\right]\right),$$

откуда заключаем, что статистика

$$U(\vec{X}) = \min\left[\alpha^{-1}(X_{(1)}), \beta^{-1}(X_{(n)})\right]$$

является достаточной, причем минимальной достаточной статистикой, поскольку имеет единичную размерность, в отличии от полученной ранее.

ii) Пусть для $\theta \in \Theta$ функции $\alpha(\theta) \downarrow$ и $\beta(\theta) \uparrow$ будут строго монотонны и также, как и в предыдущем случае, иметь разные знаки монотонности. Аналогичными рассуждениями, как и в *i)*, заключаем, что функция правдоподобия

(11) в этом случае записывается в виде

$$L(\vec{X}, \theta) = (\beta(\theta) - \alpha(\theta))^{-n} I_{(-\infty, \theta]} \left(\max \left[\alpha^{-1}(X_{(1)}), \beta^{-1}(X_{(n)})\right]\right),$$

поэтому одномерная статистика $V(\vec{X}) = \max \left[\alpha^{-1}(X_{(1)}), \beta^{-1}(X_{(n)})\right]$ будет минимальной достаточной статистикой.

Во всех других случаях монотонности и в случае немонотонности хотя бы одной из функций $\alpha(\theta)$ или $\beta(\theta)$ одномерной достаточной статистики не существует.

Пример 4. Пусть имеется статистическая структура Коши $\mathcal{K}(\theta)$ с параметром $\theta \in \Theta = (-\infty, +\infty)$. Функция правдоподобия в этом случае имеет вид

$$L(\vec{X}, \theta) = \pi^{-n} \prod_{i-1}^{n} \frac{1}{1 + (X_i - \theta)^2}.$$

Из критерия факторизации следует, что только выборка $\vec{X}$ будет достаточной статистикой в этом случае и уменьшить ее размерность не предоставляется возможным.

*Определение 3. Статистика $T(\vec{X})$ называется **полной статистикой**, если для любой функции $\mu(x)$ из равенства $E_\theta \mu(T(\vec{X}))$, $\theta \in \Theta$ следует, что $\mu(x) = 0$ для всех $x \in D_T$, где $P_\theta(T(\vec{X}) \notin D_T) = 0$.*

*Определение 4. Статистика $T(\vec{X})$ называется **свободной статистикой**, если распределение этой статистики не зависит от неизвестного параметра $\theta \in \Theta$.*

Теорема. Всякая полная достаточная статистика является минимальной достаточной статистикой.

Это утверждение оставим без доказательства.

Теорема (Басу). Если для модели $\mathfrak{F} = \{F(x, \theta),\ \theta \in \Theta\}$ существует полная достаточная статистика $T(\vec{X})$, а $T_1(\vec{X})$—свободная статистика, то статистики $T(\vec{X})$ и $T_1(\vec{X})$ независимы.

Доказательство. Для любого события A условная $\varphi_A(T) = P_\theta(T_1 \in A|T)$ и безусловная $\gamma_T = P_\theta(T_1 \in A)$ вероятности по условию не зависят от параметра θ, при этом $E_\theta\varphi_A(T) = \gamma_A$, т.е. $E_\theta g(T) = 0$ для любого $\theta \in \Theta$, где $g(T) = \varphi_A(T) - \gamma_a$. Отсюда и из условия полноты статистики T следует, что $\varphi_A(T) \equiv \gamma_A$, т.е. указанные условная и безусловная вероятности совпадают. Но это и означает, что статистики T_1 и T независимы, что и требовалось доказать.

Одной из самых важных теорем в теории точечного оценивания является следующая теорема.

Теорема (Рао—Блэкуэла—Колмогорова). Пусть заданы выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ из статистической структуры $\mathfrak{L}(\Theta)$ и некоторая параметрическая функция $\tau(\theta)$, $\theta \in \Theta$. Пусть $U(\vec{X})$—любая несмещенная оценка для $\tau(\theta)$ и $V(\vec{X})$—любая достаточная статистика на статистической структуре $\mathfrak{L}(\Theta)$. Тогда статистика $T(\vec{X}) = E_\theta(U(\vec{X})|V(\vec{X}))$ (условное математическое ожидание) будет оптимальной оценкой для параметрической функции $\tau(\theta)$.

Доказательство. Как и выше, для простоты, ограничимся доказательством теоремы лишь в дискретном случае.

Определим функцию

$$G(t) = E_\theta(U(\vec{X})|V(\vec{X}) = t) = \\ = \sum_{\vec{x}:V(\vec{x})=t} U(\vec{x})P_\theta(\vec{X} = \vec{x} \mid V(\vec{X}) = t).$$

Ясно, что эта функция не зависит от θ в силу достаточности статистики V. Далее,

$$\begin{aligned}
&E_\theta G(V(\vec{X})) = \int_{-\infty}^{+\infty} E_\theta(U(\vec{X})|V(\vec{X}) = t)P_\theta(V(\vec{X}) = t)dt = \\
&= \int_{-\infty}^{+\infty} \left(\sum_{\vec{x}\in\mathcal{X}} U(\vec{x})P_\theta(\vec{X} = \vec{x}|V(\vec{X}) = t)\right) P_\theta(V(\vec{X}) = t)dt = \\
&= \sum_{\vec{x}\in\mathcal{X}} U(\vec{x}) \int_{-\infty}^{+\infty} \left(P_\theta(\vec{X} = \vec{x}|V(\vec{X}) = t)P_\theta(V(\vec{X}) = t)\right) dt = \\
&= \sum_{\vec{x}\in\mathcal{X}} U(\vec{x})P_\theta(\vec{X} = \vec{x}) = E_\theta U(\vec{X}) = \tau(\theta),
\end{aligned}$$

т.е. получаем, что статистика $G(V(\vec{X}))$ есть несмещенная оценка для параметрической функции $\tau(\theta)$. Дальше, в силу несмещенности статистики $U(\vec{X})$, предыдущего равенства и свойства ковариации, получаем

$$\begin{gathered}
D_\theta U(\vec{X}) = E_\theta \left(U(\vec{X}) - E_\theta U(\vec{X})\right)^2 = E_\theta \left(U(\vec{X}) - \tau(\theta)\right)^2 = \\
= E_\theta \left(U(\vec{X}) - G(V(\vec{X})) + G(V(\vec{X})) - \tau(\theta)\right)^2 = \qquad (12) \\
= E_\theta \left(U(\vec{X}) - G(V(\vec{X}))\right)^2 + E_\theta \left(G(V(\vec{X})) - \tau(\theta)\right)^2 + \\
+2E_\theta \left(U(\vec{X}) - G(V(\vec{X}))\right)\left(G(V(\vec{X})) - \tau(\theta)\right) = \\
= E_\theta \left(U(\vec{X}) - G(V(\vec{X}))\right)^2 + D_\theta G(V(\vec{X})) \geq D_\theta G(V(\vec{X})),
\end{gathered}$$

поскольку

$$\begin{aligned}
&E_\theta \left(U(\vec{X}) - G(V(\vec{X}))\right)\left(G(V(\vec{X})) - \tau(\theta)\right) = \\
&= E_\theta \left(U(\vec{X}) - G(V(\vec{X}))\right) G(V(\vec{X})) =
\end{aligned}$$

$$= \int\limits_{-\infty}^{+\infty} \Big(E_\theta(U(\vec{X}) \mid V(\vec{X}) = t) - G(t) \Big) G(t) P_\theta(V(\vec{X}) = t) \, dt =$$

$$= \int\limits_{-\infty}^{+\infty} (G(t) - G(t)) \, G(t) P_\theta(V(\vec{X}) = t) \, dt = 0.$$

Таким образом, если существует оптимальная оценка, то из (12) следует, что статистика $T(\vec{X}) = E_\theta(U(\vec{X})|V(\vec{X}))$ является оптимальной. Теорема доказана.

Следствие. Пусть $T(\vec{X})$ *– полная достаточная статистика и пусть определена новая статистика* $U(\vec{X}) = g(T(\vec{X}))$*, где* g *– произвольная функция. Тогда, если для произвольного* $\theta \in \Theta$ *справедливо равенство* $\tau(\theta) = E_\theta U(\vec{X})$*, то статистика* $U(\vec{X})$ *является оптимальной оценкой для параметрической функции* $\tau(\theta)$ *для всех* $\theta \in \Theta$*.*

Доказательство. На основании теоремы Рао—Блэкуэлла—Колмогорова получаем, что оптимальная оценка для $\tau(\theta)$ должна быть функцией от достаточной статистики. Единственность следует из условия полноты. Действительно, если для некоторой функции g_1 справедливо равенство $E_\theta \, g_1(T(\vec{X})) = \tau(\theta)$, то в силу полноты достаточной статистики $T(\vec{X})$ (определение 3) получим:

$$E_\theta \left[g_1(T(\vec{X})) - g_1(T(\vec{X})) \right] = 0 \Rightarrow$$

$$\Rightarrow \quad g_1(x) = g(x) \text{ для всех } x \in D_T.$$

Следствие доказано.

Пример 1. Пусть задана выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ из статистической структуры Бернулли с параметром $\theta \in \Theta = (0, 1)$.

Ясно, что статистика $U(\vec{X}) = X_1 \cdots X_i \cdot (1 - X_{i+1}) \cdots (1 - X_{i+j})$, $i + j \leqslant n$ является несмещенной оценкой для параметрической функции

$$\tau(\theta) = \theta^i(1-\theta)^j, \tag{13}$$

поскольку $E_\theta X_i = 1 \cdot \theta + 0 \cdot (1-\theta) = \theta$, $i = 1, \ldots, n$, из определения выборки случайные величины $X_1, X_2, \ldots, X_n$ независимы и из свойств математического ожидания получаем

$$E_\theta U(\vec{X}) = \prod_{k=1}^{i} E_\theta X_k \prod_{l=1}^{j}(1 - E_\theta X_l) = \tau(\theta).$$

Ясно, что $V(\vec{X}) = n\overline{X} = \sum\limits_{l=1}^{n} X_l$ есть достаточная статистика. Действительно, функция правдоподобия для выборки $\vec{X}$ равна

$$L(\vec{X}, \theta) = \prod_{l=1}^{n} \theta^{X_l}(1-\theta)^{1-X_l} = \theta^{n\overline{X}}(1-\theta)^{n-n\overline{X}},$$

поэтому, по критерию факторизации ($h(\vec{X}) = 1$) получаем требуемое.

Будем искать теперь функцию

$$g(t) = E_\theta(U(\vec{X}) \mid V(\vec{X}) = t). \tag{14}$$

Статистика $U(\vec{X})$ может принимать только значения ноль или единица. Ясно, что $U(\vec{X}) = 1$, если все $X_1, \ldots, X_i$ примут значение 1, все $X_{i+1}, \ldots, X_{i+j}$ примут значение 0, а каждая из остальных случайных величин $X_{i+j}, \ldots, X_n$ могут принимать любые значения 0 или 1. Определим событие

$$A = \Big\{ X_1 = 1, \ldots, X_i = 1, X_{i+1} = 0, \ldots, X_{i+j} = 0,$$

$$X_{i+j+1} = x_{i+j+1}, \ldots X_n = x_n \Big\}.$$

Очевидно,

$$g(t) = 1 \cdot P(A \mid \sum_{l=1}^{n} X_l = t) = \frac{P(A, \sum\limits_{l=1}^{n} X_l = t)}{P(\sum\limits_{l=1}^{n} X_l = t)} = \tag{15}$$

$$= \frac{P(X_1=1,\ldots,X_i=1,X_{i+1}=0,\ldots,X_{i+j}=0, \sum\limits_{l=i+j+1}^{n} X_l=t-i)}{P(\sum\limits_{l=1}^{n} X_l=t)}.$$

Из курса теории вероятностей известно, что если $\xi_1, \ldots, \xi_m$ – независимые, одинаково распределенные случайные величины, имеющие распределение Бернулли с параметром p, то есть

$$P(\xi_l = 1) = p \ \text{ и } \ P(\xi_l = 0) = 1 - p, \ \ l = 1, \ldots, m,$$

то их сумма будет иметь биномиальное распределение с параметрами (m, p), т.е.

$$P\left(\sum_{l=1}^{m} \xi_l = k\right) = C_m^k p^k (1-p)^{m-k}, \ \ k = 0, 1, \ldots, m.$$

Поэтому, из (15) получим

$$g(t) = \frac{\theta^i \cdot (1-\theta)^j C_{n-i-j}^{t-i} \theta^{t-i} (1-\theta)^{n-i-j-(t-i)}}{C_n^t \theta^t (1-\theta)^{n-t}} = \frac{C_{n-i-j}^{t-i}}{C_n^t}.$$

Ясно, что функция $g(t)$ не зависит от неизвестного параметра θ, что должно выполняться по определению достаточной статистики. Поэтому по теореме Рао—Блэкуэлла—Колмогорова получаем, что для параметрической функции (13) статистика

$$T(\vec{X}) = C_{n-i-j}^{n\overline{X}-i} / C_n^{n\overline{X}} \tag{16}$$

будет оптимальной (т.е. наилучшей) оценкой в классе несмещенных оценок для параметрической функции (13).

Раньше было показано, что статистика $\overline{X}$ является эффективной для неизвестного параметра θ в статистической структуре Бернулли. Очевидно, что если в (13) положить $i = 1$ и $j = 0$, то $\tau(\theta) = \theta$ и из (16) следует $T(\vec{X}) = \overline{X}$, т.е. получаем тот же результат другим способом. Однако, из теоремы Рао—Блэкуэлла—Колмогорова мы получили оптимальные оценки для широкого класса функций, для которого метод Рао—Крамера не работает.

Попутно отметим, что из утверждения 2 раздела 7, в частности, следует, что если параметрическая функция $\tau(\theta)$ представляет из себя полином порядка не выше n, т.е.

$$\tau(\theta) = \sum_{i=0}^{k} a_i \theta^i, \quad 0 \leqslant k \leqslant n,$$

то оптимальной оценкой этого полинома будет статистика

$$T(\vec{X}) = \sum_{i=0}^{k} a_i C_{n-i}^{n\overline{X}-i} / C_n^{n\overline{X}}.$$

Если

$$(n)_i = n(n-1)\cdots(n-i+1), \quad (n)_0 = 1$$

обозначает число сочетаний из n элементов по i, то последняя статистика переписывется в виде

$$T(\vec{X}) = \sum_{i=0}^{k} a_i \left(n\overline{X}\right)_i / (n)_i.$$

Наконец, покажем, что для параметрической функции

$\tau(\theta) = \theta^k, \quad k > n$ не существует оптимальной оценки. В этом конкретном случае достаточно показать, что не существует даже несмещенной оценки для этой параметрической функции. Предположим от противного, что статистика $T(\vec{X})$ является несмещенной оценкой для $\theta^k,\ k > n$. Тогда

$$\theta^k = E_\theta T(\vec{X}) =$$

$$= \sum_{i_1=0}^{1} \cdots \sum_{i_n=0}^{1} T(i_1, \ldots, i_n)\theta^{i_1}(1-\theta)^{1-i_1} \cdots \theta^{i_n}(1-\theta)^{1-i_n}.$$

Но отсюда получаем, что правая часть последнего равенства представляет из себя полином по θ степени не выше n, а в левой части стоит полином степени $k > n$. Равенство, поэтому, в последнем выражении невозможно. Получили противоречие, связанное с предположением о существовании несмещенной оценки. Окончательно получаем, что для любого полинома по θ степени выше n не существует несмещенной, а как следствие и оптимальной оценки.

Пример 2. Продемонстрируем, как применяется на практике следствие из теоремы Рао—Блекуэлла—Колмогорова.

Пусть имеется выборка $\vec{X} = (X_1, \ldots, X_n) \sim R(0, \theta)$ из равномерного распределения на отрезке $[0, \theta],\ \theta > 0$. Требуется оценить параметр θ.

Выше было показано, что максимальная порядковая статистика $T(\vec{X}) = X_{(n)}$ является достаточной статистикой. Легко проверить, что плотность этой статистики равна

$$f_{(n)}(x, \theta) = \frac{n}{\theta^n} x^{n-1} I_{[0,\theta]}(x).$$

Пусть теперь для функции $\varphi(t),\ t > 0$ выполняется условие

$$E_\theta \varphi(T) = \frac{n}{\theta^n}\int\limits_0^\theta \varphi(t)t^{n-1}dt = 0, \ \ \forall\ \theta > 0.$$

Дифференцируя по θ тождество $\int\limits_0^\theta \varphi(t)t^{n-1}dt \equiv 0$, получаем, что $\varphi(\theta)\theta^{n-1} \equiv 0$, откуда следует, что $\varphi(\theta) = 0\ \ \forall\ \theta > 0$. Таким образом, получаем, что достаточная статистика $X_{(n)}$ является полной. Далее, очевидно, получаем, что

$$E_\theta(T) = \frac{n}{\theta^n}\int\limits_0^\theta t^n dt = \frac{n+1}{n}\theta,$$

откуда следует, что статистика $T^*(\vec{X}) = \frac{n}{n+1}X_{(n)}$ является оптимальной оценкой для параметра θ.

11. Метод максимального правдоподобия получения оценок

Как мы видели в предыдущем разделе, нахождение оптимальных оценок с помощью достаточных статистик является довольно трудоемким. В этом разделе мы опишем другой способ получения оценок, приводящий к приемлемым результатам.

Пусть, как и выше, имеется выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ на заданной статистической структуре $\mathfrak{L}(\Theta)$ и пусть $L(\vec{X}, \theta)$ будет функцией правдоподобия этой выборки.

Определение. Оценкой максимального правдоподобия (сокращенно **о.м.п.***) будем называть такую статистику* $\hat{\theta}(\vec{X}) \in \Theta$*, что*

$$L(\vec{X}, \theta) \leq L(\vec{X}, \hat{\theta}(\vec{X})) \quad \text{для всех } \theta \in \Theta.$$

Если $\Theta \subset R^k$, т.е. $\theta \in \Theta - k$-мерный векторный параметр, функция правдоподобия $L(\vec{X}, \vec{\theta})$ дифференцируема по параметру и для любого $\vec{x}$ из выборочного пространства $\mathcal{X}$ максимальное значение функции $L(\vec{x}, \theta)$ достигается в некоторой внутренней точке множества Θ, то о.м.п. будет удовлетворять уравнениям правдоподобия:

$$\frac{\partial L(\vec{X}, \vec{\theta})}{\partial \theta_i} = 0, \quad i = 1, 2, \ldots, k.$$

Замечание. Из (8) следует, что если статистика $T(\vec{X})$ является эффективной оценкой для параметрической функции $\tau(\theta) = \theta$, то

$$Q_n(\vec{X}, \theta) = \frac{\partial lnL(\vec{X}, \theta)}{\partial \theta} = \alpha^{-1}\left(T(\vec{X}) - \theta\right)$$

и отсюда следует, что $\hat{\theta}(\vec{X}) = T(\vec{X})$, то есть, в этом случае, о.м.п. совпадает с эффективной оценкой.

Пример 1. Пусть имеется выборка $\vec{X}$ из статистической модели нормального распределения $N(\Theta)$, где $\Theta = (\theta, 1)$. В этом случае функция правдоподобия

$$L(\vec{X}, \theta) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} X_i^2 + \theta \sum_{i=1}^{n} X_i - \frac{n}{2}\theta^2\right)$$

и

$$\frac{\partial L(\vec{X}, \theta)}{\partial \theta} =$$

$$= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} X_i^2 + \theta \sum_{i=1}^{n} X_i - \frac{n}{2}\theta^2\right)\left(\sum_{i=1}^{n} X_i - n\theta\right).$$

Приравнивая последнее выражение нулю, получаем, что о.м.п. $\hat{\theta}(\vec{X}) = \overline{X}$.

Пример 2. Пусть теперь выборка имеет нормальное распределение с известным математическим ожиданием a и неизвестной дисперсией $\theta^2 > 0$, т.е. выборка $\vec{X}$ из статистической модели нормального распределения $N(a, \theta^2)$ (здесь неизвестным параметром будет θ^2). Поэтому, функция правдоподобия

$$L(\vec{X}, \theta^2) = (2\pi\theta^2)^{-n/2} \exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n} (X_i - a)^2\right)$$

и

$$\frac{\partial L(\vec{X}, \theta^2)}{\partial (\theta^2)} = (2\pi)^{-n/2} \exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n} (X_i - a)^2\right) \frac{1}{2}\left(\theta^2\right)^{-\frac{n}{2}-2} \times$$

$$\times \left(-n\theta^2 + \sum_{i=1}^{n} (X_i - a)^2\right).$$

Эта производная будет равна нулю, если последнее выражение в скобках будет равно нулю, откуда получаем, что о.м.п.

$$\hat{\theta}^2(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2 .$$

Пример 3. Предположим, наконец, что выборка имеет нормальное распределение с неизвестным математическим ожиданием θ_1 и неизвестной дисперсией $\theta_2^2 > 0$, т.е. выборка $\vec{X}$ из статистической модели нормального распределения $N(\theta_1, \theta_2^2)$. Поэтому, функция правдоподобия имеет вид

$$L(\vec{X}, \theta_1, \theta_2^2) = (2\pi\theta_2^2)^{-n/2} \exp\left(-\frac{1}{2\theta_2^2}\sum_{i=1}^{n}(X_i - \theta_1)^2\right)$$

и

$$\ln L(\vec{X}, \theta_1, \theta_2^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\theta_2^2) - \frac{1}{2\theta_2^2}\sum_{i=1}^{n}(X_i - \theta_1)^2 .$$

Имеем

$$\frac{\partial \ln L(\vec{X}, \theta_1, \theta_2^2)}{\partial \theta_1} = \frac{1}{\theta_2^2}\sum_{i=1}^{n}(X_i - \theta_1) = 0$$

и

$$\frac{\partial \ln L(\vec{X}, \theta_1, \theta_2^2)}{\partial (\theta_2^2)} = -\frac{n}{2\theta_2^2} + \frac{1}{2\theta_2^4}\sum_{i=1}^{n}(X_i - \theta_1)^2 = 0,$$

откуда

$$\hat{\theta}_1(\vec{X}) = \bar{X} = \frac{1}{n}\sum_{i=1}^{n}X_i \quad \text{и} \quad \hat{\theta}_2^2(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 .$$

Пример 4. Пусть имеется выборка $\vec{X}$ из статистической модели равномерного распределения $R(0, \theta)$, где $\theta > 0$ – неизвестный параметр. Как было показано раньше, функция правдоподобия равна (формула (11)) $L(\vec{X}, \theta) = \theta^{-n} I_{[0,+\infty)}(X_{(1)}) I_{(-\infty,\theta]}(X_{(n)})$. Отсюда следует, что функция правдоподобия $L(\vec{X}, \theta)$ монотонно убывает по $\theta \geq X_{(n)}$, причем максимум достигается в точке $\theta = X_{(n)}$. Следовательно, о.м.п. $\hat{\theta}(\vec{X}) = X_{(n)}$. Заметим, что оценку максимального правдоподобия в этом случае дифференцированием найти нельзя, поскольку функция правдоподобия разрывна.

Пример 5. Пусть имеется выборка $\vec{X}$ из статистической модели равномерного распределения $R(\theta - 1, \theta)$, где $\theta > 1$ – неизвестный параметр. Функция правдоподобия в этом случае равна $L(\vec{X}, \theta) = I_{[\theta-1,+\infty)}(X_{(1)}) I_{(-\infty,\theta]}(X_{(n)})$. Ясно, что эта функция принимает только значения 0 и 1. Максимальное значение будет в случае $X_{(n)} \leq \theta \leq X_{(1)} + 1$. Поэтому, в этом случае будет бесконечное число оценок максимального правдоподобия: $\hat{\theta}(\vec{X}, \lambda) = X_{(n)} + \lambda(1 + X_{(1)} - X_{(n)})$, где параметр $\lambda \in [0, 1]$.

Отметим несколько свойств о.м.п.

i) Если существуют эффективная *оценка и* о.м.п. *для скалярного параметра* θ*, то эти оценки совпадают.*

Это свойство есть прямое следствие равенства (8).

ii) Пусть для скалярного параметра θ *существует и единственна о.м.п. Тогда эта оценка есть функция от достаточной статистики* $\hat{\theta}(\vec{X}) = g(T(\vec{X}))$, *где* g *какая то функция, а* $T(\vec{X})$ – достаточная *статистика.*

Действительно, из критерия факторизации (10) следует, что для максимизации функции правдоподобия нужно искать максимум по θ функции, зависящей от достаточной статистики, откуда и следует утверждение.

iii) ***Принцип инвариантности для о.м.п.*** *Пусть* $\psi = \psi(\theta)$ *– такая параметрическая функция, что отоб-*

ражение $\psi : \Theta \to \Psi$ *взаимно однозначно, где* Θ*—параметрическое множество, а* Ψ *— некоторое множество. Тогда* $\hat{\psi}(\vec{X}) = \psi(\hat{\theta}(\vec{X}))$.

В силу того, что функция ψ взаимно однозначна, $\theta = \theta(\psi)$ будет обратной функцией и $\sup\limits_{\theta\in\Theta} L(\vec{X}, \theta) = \sup\limits_{\psi\in\Psi} L(\vec{X}, \theta(\psi))$. Поэтому, если по θ максимум функции $L(\vec{X}, \theta)$ достигается в точке $\hat{\theta}$, то максимум $L(\vec{X}, \theta(\psi))$ будет в точке, удовлетворяющей уравнению $\theta(\hat{\psi}) = \hat{\theta}$, т.е. в точке $\hat{\psi} = \psi(\hat{\theta})$. В этом заключается смысл принципа инвариантности.

Замечание. Напомним, что понятие супремума (**sup**) вводится аналогично инфинуму как точная верхняя граница, но только для ограниченных сверху числовых множеств. Супремум совпадает с наибольшим элементом, если такой в этом множестве есть. Если такого элемента в множестве нет, то супремум все равно существует, но этому множеству не принадлежит. Например, в множестве $[0, 1)$ супремум равен 1, хотя наибольшего числа в этом множестве нет.

Кратко остановимся на *асимптотических свойствах о.м.п.* Итак, пусть имеется выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ из класса распределений на заданной статистической структуре $\mathfrak{L}(\Theta)$. Будем предполагать, что объем этой выборки может неограниченно увеличиваться, т.е. $n \to \infty$ и все распределения из заданной статистической структуры регулярны (см. выше ...).

Для простоты, рассмотрим только скалярный случай неизвестных параметров, т.е. $\Theta \subset R^1$. Пусть статистика $\hat{\theta}_n(\vec{X})$—о.м.п. для параметра θ (индекс у статистики означает лишь то, что она построена по выборке переменного объема). Пусть, как обычно, функция $f(x, \theta)$ обозначает распределение ($f(x, \theta) = P_\theta(X_1 = x)$), если выборка дискретна или плотность распределения ($f(x,\theta) = (P_\theta(X_1 < x))'_x$), если выборка абсолютно непрерывна.

Теорема (асимптотическая нормальность о.м.п.).

j) Статистика $\hat{\theta}_n(\vec{X})$ *является асимптотически несмещенной и состоятельной оценкой параметра* θ.

jj) Если функция $f(x,\theta)$ *трижды дифференцируема по* θ *и при этом*

$$E_\theta \left| \frac{\partial^3 \ln f(X_1,\theta)}{\partial \theta^3} \right| = E_\theta V(X_1) < \infty,$$

где $V(x)$*—некоторая функция, то*

$$\sqrt{n}\left(\hat{\theta}_n(\vec{X}) - \theta\right) \xrightarrow[n\to\infty]{d} \eta \sim N(0, 1/i(\theta)),$$

где $\eta_n \xrightarrow[n\to\infty]{d} \eta$ *означает* $\lim\limits_{n\to\infty} P_\theta(\eta_n < x) = P_\theta(\eta < x)$*—сходимость по распределению, а* $i(\theta)$ *обозначает, как обычно,* информацию Фишера.

Более того, если $\hat{\tau}_n(\vec{X})$ *— о.м.п. для дифференциируемой параметрической функции* $\tau(\theta)$*, то*

$$\sqrt{n}\left(\hat{\tau}_n(\vec{X}) - \tau(\theta)\right) \xrightarrow[n\to\infty]{d} \zeta \sim N(0, (\tau'(\theta))^2/i(\theta)).$$

jjj) О.м.п. $\hat{\theta}_n(\vec{X})$ *является* асимптотически эффективной *оценкой для параметра* θ*, т.е. при* $n \to \infty$ *дисперсия* $\hat{\theta}_n(\vec{X})$ *стремится к правой части неравенства Рао—Крамера (см. [2]).*

Эту теорему оставим без доказательства.

Отметим, в заключение, что о.м.п. из примеров 4 и 5 не будут асимптотически нормальными, поскольку распределения выборок не *регулярны*.

12. Метод моментов получения точечных оценок. Их свойства

Предположим что статистическая структура зависит от r неизвестных параметров $\theta_1, \theta_2, ..., \theta_r$ и существуют, по крайней мере, первые r моментов распределения $\alpha_k = \alpha_k(\theta_1, ..., \theta_r)$, $k = 1, 2, ..., r$. Пусть в нашем распоряжении имеется выборка объема n из этой статистической структуры. Приравняв выборочные моменты k-го порядка

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

к теоретическим моментам того же порядка, получим систему из r уравнений с r неизвестными θ_k

$$A_k = \alpha_k, \qquad k = 1, 2, ..., r.$$

Решая эту систему уравнений, найдем оценки $\hat{\theta}_1, \ldots, \hat{\theta}_r$ неизвестных параметров $\theta_1, \ldots, \theta_r$. Они, естественно, будут функциями выборочных моментов.

Поскольку статистики A_k сходятся по вероятности к α_k, т.е. $A_k \xrightarrow{P} \alpha_k$, то $\hat{\theta}_j \xrightarrow{P} \theta_j$, $j = 1, 2, ..., r$ при $n \to \infty$, т.е. оценки, полученные методом моментов, являются ***состоятельными***.

Пример. Пусть задана выборка $\vec{X} = (X_1, X_2, ..., X_n)$ объема n из абсолютно непрерывного распределения с плотностью распределения

$$f(x, \theta_1, \theta_2) = \begin{cases} \frac{1}{\theta_1}\exp\left(-\frac{x-\theta_2}{\theta_1}\right), & \text{если } x \geq \theta_2 \\ 0, & \text{если } x < \theta_2. \end{cases}$$

Найдем оценки для неизвестных параметров θ_1 и θ_2. Для этого вычислим первые два момента распределения:

$$\alpha_1 = \frac{1}{\theta_1}\int\limits_{\theta_2}^{+\infty} x \exp\left(-(x-\theta_2)/\theta_1\right) dx = \theta_1 + \theta_2,$$

$$\alpha_2 = \frac{1}{\theta_1}\int\limits_{\theta_2}^{+\infty} x^2 \exp\left(-(x-\theta_2)/\theta_1\right) dx = 2\theta_1^2 + 2\theta_1\theta_2 + \theta_2^2.$$

Приравнивая теоретические моменты выборочным и решая систему уравнений

$$\begin{cases} A_1 = \hat\theta_1 + \hat\theta_2 \\ A_2 = 2\hat\theta_1^2 + 2\hat\theta_1\hat\theta_2 + \hat\theta_2^2, \end{cases}$$

найдем нужные оценки:

$$\begin{gathered} \hat\theta_1 = \sqrt{A_2 - A_1^2}, \\ \hat\theta_2 = A_1 - \sqrt{A_2 - A_1^2} \end{gathered}$$

или в развернутом виде:

$$\hat\theta_1 = \sqrt{\tfrac{1}{n}\sum_{i=1}^{n} X_i^2 - \tfrac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2},$$

$$\hat\theta_2 = \tfrac{1}{n}\sum_{i=1}^{n} X_i - \sqrt{\tfrac{1}{n}\sum_{i=1}^{n} X_i^2 - \tfrac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2}.$$

Исходя из изложенного выше, заключаем, что полученные оценки являются *состоятельными*.

13. Доверительные интервалы для неизвестных параметров и параметрических функций

Для упрощения изложения ограничимся случаем скалярных параметров.

Предположим, что в нашем распоряжении имеется статистическая структура

$$\{(\Omega, \mathfrak{A}, P_\theta),\ \theta \in \Theta \subset R^1\}$$

со скалярным неизвестным параметром θ. Пусть $\vec{X} = (X_1, X_2, ..., X_n)$—выборка объема n, распределенная на заданной статистической структуре с функцией распределения $F(x, \theta)$.

Определение 1. Пусть для некоторой константы $\alpha \in (0, 1)$ *существуют две статистики* $T_i = T_i(\vec{X}, \alpha),\ i = 1, 2$*, такие, что* $P_\theta(T_1 < \theta < T_2) = 1 - \alpha$*. Тогда интервал* (T_1, T_2) *называется* ***доверительным интервалом*** *для параметра* θ *с уровнем доверия* $1 - \alpha$*.*

В этом разделе будут рассмотрены два основных метода получения доверительных интервалов.

13.1. Метод центральных статистик для получения доверительных интервалов

Определение 2. Функция $G = G(\vec{X}, \theta)$, от выборки и неизвестного параметра называется ***центральной*** *статистикой, если: 1) функция распределения этой функции G не зависит от неизвестного параметра θ и 2) функция G является непрерывной и строго монотонной по θ.*

Определение 3. Если для функции $G = G(\vec{X}, \theta)$ выполняется только условие (1), то такая функция называется ***свободной*** *статистикой.*

Замечание. Функция G не является статистикой в определенном выше смысле, поскольку зависит не только от выборки, но еще и от неизвестного параметра θ. Название «центральная статистика» или «свободная статистика» используется лишь для функции с указанными выше свойствами.

Выберем и зафиксируем уровень значимости α (число в интервале $0 < \alpha < 1$). Для любой центральной статистики, очевидно, найдется пара констант β_1 и β_2, такая, что

$$P_\theta\left(\beta_1 \le G(\vec{X}, \theta) \le \beta_2\right) \ge 1 - \alpha.$$

Пусть, для определенности, функция $G(\vec{X}, \theta)$ будет строго возрастающей по θ. Тогда обратная функция $G^{-1}(\vec{X}, y)$ для функции $G(\vec{X}, \theta)$ также будет строго возрастающей. Поэтому, неравенства $\beta_1 \le G(\vec{X}, \theta) \le \beta_2$ и $G^{-1}(\vec{X}, \beta_1) \le \theta \le \le G^{-1}(\vec{X}, \beta_2)$ являются эквивалентными. Таким образом, получаем доверительный интервал для θ с уровнем доверия не меньше, чем $1 - \alpha$:

$$P_\theta\left(T_1(\vec{X}) \le \theta \le T_2(\vec{X})\right) \ge 1 - \alpha,$$

где $T_1(\vec{X}) = G^{-1}(\vec{X}, \beta_1)$ и $T_2(\vec{X}) = G^{-1}(\vec{X}, \beta_2)$.

В случае строгого убывания по θ функции $G(\vec{X}, \theta)$, обратная функция $G^{-1}(\vec{X}, y)$ будет также строго убывающей и, поэтому, неравенства

$$\beta_1 \le G(\vec{X}, \theta) \le \beta_2 \quad \text{и} \quad G^{-1}(\vec{X}, \beta_2) \le \theta \le G^{-1}(\vec{X}, \beta_1)$$

будут эквивалентными и снова, как и выше, получим, что статистика $T_1(\vec{X}) = G^{-1}(\vec{X}, \beta_2)$ будет левой границей, а статистика $T_2(\vec{X}) = G^{-1}(\vec{X}, \beta_1)$ – правой границей доверительного интервала с уровнем доверия не меньше, чем $1 - \alpha$.

Очевидно, что в случае, когда распределение центральной статистики абсолютно непрерывно (например, когда исходная выборка $\vec{X}$ из абсолютно непрерывного распределения), можно выбрать константы β_1 и β_2 так, чтобы $P_\theta\Big(\beta_1 \le G(\vec{X}, \theta) \le \beta_2\Big) = 1 - \alpha$. Поэтому, в этом случае, получим $P_\theta\Big(T_1(\vec{X}) \le \theta \le T_2(\vec{X})\Big) = 1 - \alpha$, где статистики T_1 и T_2 определяются так же, как описано выше.

Нахождение центральной статистики является часто не простой задачей и, зачастую, основывается на свойствах распределения выборки. В частности, если функция распределения $F(x, \theta)$ каждого элемента X_j выборки $\vec{X}$ является непрерывной и строго монотонной функцией по θ, то функция $G(\vec{X}, \theta) = -\sum\limits_{i=1}^{n} \ln F(X_i, \theta)$ будет центральной статистикой. Это следует из того, что случайные величины $F(X_i, \theta)$ имеют равномерное распределение на отрезке $[0, 1]$; $-\ln F(X_i, \theta)$–независимые и одинаково показательно распределенные случайные величины с параметром 1, а сумма таких случайных величин, как известно из курса теории вероятностей, имеет гамма распределение с параметрами $(1, n)$. К сожалению, приведенная центральная статистика не имеет широкого применения в случае, когда вид функции распределения $F(x, \theta)$ достаточно сложный, например, не выражается в элементарных функциях.

Для выбора констант β_1, β_2 иногда используют следующий прием: в качестве β_1 выбирается наибольшее число, такое что $P_\theta\left(G(\vec{X}, \theta) < \beta_1\right) < \alpha/2$, а в качестве β_2—наименьшее число, такое что $P_\theta\left(G(\vec{X}, \theta) > \beta_2\right) < \alpha/2$. Очевидно, что при таком выборе пара чисел β_1 и β_2 будет единственной и

$$P_\theta\left(\beta_1 \leq G(\vec{X}, \theta) \leq \beta_2\right) \geq 1 - \alpha/2.$$

Такой прием оправдан в случае, когда распределение центральной статистики симметрично (или близко к симметричности), т.е. $P_\theta\left(G(\vec{X}, \theta) < x\right) = 1 - P_\theta\left(G(\vec{X}, \theta) > x\right)$ и, кроме этого, величина $\left|G^{-1}(\vec{X}, \beta_2) - G^{-1}(\vec{X}, \beta_1)\right|$ прямо пропорциональна величине $\beta_2 - \beta_1$. В этом случае получается наикратчайший (или близкий к такому) доверительный интервал. Доверительный интервал, полученный при таком выборе констант β_1 и β_2 часто называют ***центральным***.

Метод центральных статистик получения интервальных доверительных оценок, описанный выше в общем случае, зависит от выбора пары констант β_1, β_2, что, вообще говоря, может привести к большому числу доверительных интервалов. А это обстоятельство, в свою очередь, приводит к необходимости получения оптимального доверительного интервала. Понятно, что оптимизировать доверительный интервал следует по его длине. *Наилучший* интервал должен быть *наикратчайшим* из всех возможных доверительных интервалов. Поэтому указанные константы разумно выбирать из условия

$$\beta_1, \beta_2 : \left|G^{-1}(\vec{X}, \beta_2) - G^{-1}(\vec{X}, \beta_1)\right| =$$

$$= \inf_{\gamma_1, \gamma_2} \left|G^{-1}(\vec{X}, \gamma_2) - G^{-1}(\vec{X}, \gamma_1)\right|.$$

Замечание. Мы оптимизируем длину интервала только выбирая соответствующие константы для фиксированной, выбранной заранее, центральной статистики и, в силу большой сложности, не рассматриваем оптимизацию по классу всех центральных статистик для имеющейся в нашем распоряжении выборки.

Пример 1. Пусть задана выборка $\vec{X} = (X_1, \ldots, X_n)$ из статистической модели нормального распределения $N(\theta, \sigma^2)$. Здесь будем считать, что математическое ожидание θ нам не известно, а дисперсия σ^2 имеет известное значение. Выберем произвольное число α из интервала $(0, 1)$ и построим доверительный интервал по заданной выборке $\vec{X}$ уровня значимости $\gamma = 1 - \alpha$.

Как известно из курса теории вероятностей, сумма независимых случайных величин, имеющих нормальные распределения будет иметь нормальное распределение с математическим ожиданием, равным сумме математических ожиданий исходных случайных величин и с дисперсией, равной сумме дисперсий тех же исходных случайных величин. Поэтому, в силу независимости, одинаковой распределенности и нормального распределения компонент нашей выборки, очевидно, получаем:

$$n\bar{X} = \sum_{i=1}^{n} X_i \sim N(n\theta, n\sigma^2),$$

где значок $' \sim '$ (тильда), как обычно, означает принадлежность случайной величины классу распределения, в нашем случае — нормальному. Воспользовавшись свойствами математического ожидания и дисперсии, получим:

$$n\bar{X} - n\theta \sim N(0, n\sigma^2)$$

и

$$\left(n\bar{X} - n\theta\right)/\sqrt{n\sigma^2} = \sqrt{n}\frac{\bar{X} - \theta}{\sigma} \sim N(0, 1).$$

Теперь заметим, что функция

$$G(\vec{X}, \theta) = \sqrt{n}\frac{\bar{X} - \theta}{\sigma}$$

является центральной статистикой, поскольку она монотонно убывающая по θ и имеет стандартное нормальное распределение, т.е. ее распределение не зависит от неизвестного параметра. Пусть $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2} du$ функция распределения стандартного нормального распределения. Известно, что для этого распределения выполняется равенство симметрии функции распределения: для любого $x > 0$ справедливо равенство $\Phi(-x) = 1 - \Phi(x)$. Поэтому, если выбрать константу $\beta > 0$ так, чтобы $1 - \Phi(\beta) = \alpha/2$, т.е. $\beta = \Phi^{-1}(1 - \alpha/2)$ — *квантиль* стандартного нормального распределения уровня $1 - \alpha/2$, то в силу свойства симметричности, получаем справедливость равенства

$$P_\theta(G(\vec{X}, \theta) < -\beta) = P_\theta(G(\vec{X}, \theta) > \beta) = \alpha/2$$

и, следовательно, $P_\theta(-\beta \leq G(\vec{X}, \theta) \leq \beta) = 1 - \alpha$ или, эквивалентно

$$P_\theta(-\beta \leq \sqrt{n}\frac{\bar{X} - \theta}{\sigma} \leq \beta) = \gamma.$$

Из последнего выражения, очевидно, получаем

$$P_\theta\left(\bar{X}-\frac{\beta\sigma}{\sqrt{n}}\leq\theta\leq\bar{X}+\frac{\beta\sigma}{\sqrt{n}}\right)=\gamma,$$

т.е. $\left(\bar{X}-\frac{\beta\sigma}{\sqrt{n}},\bar{X}+\frac{\beta\sigma}{\sqrt{n}}\right)$ является γ—доверительным *центральным* интервалом (по способу выбора постоянной β) для неизвестного математического ожидания θ.

В общем случае, если выбрать пару констант $\beta_1,\ \beta_2$ из условия

$$\frac{1}{\sqrt{2\pi}}\int\limits_{\beta_1}^{\beta_2}e^{-x^2/2}dx=\Phi(\beta_2)-\Phi(\beta_1)=\gamma,$$

то проводя выкладки, аналогичные выше приведенным, получим, что γ—доверительным интервалом для θ будет интервал $\left(\bar{X}-\frac{\beta_2\sigma}{\sqrt{n}},\bar{X}-\frac{\beta_1\sigma}{\sqrt{n}}\right)$.

Применим метод Лагранжа для нахождения *наикратчайшего* интервала. Функция Лагранжа для нашей задачи равна

$$L(\beta_1,\beta_2,\lambda)=\frac{\sigma(\beta_2-\beta_1)}{\sqrt{n}}+\lambda\left(\gamma-\Phi(\beta_2)+\Phi(\beta_1)\right).$$

Приравнивая нулю частные производные функции Лагранжа по $\beta_1,\ \beta_2,$ и λ, получаем систему уравнений

$$\lambda\varphi(\beta_1)=\sigma/\sqrt{n},\quad \lambda\varphi(\beta_2)=\sigma/\sqrt{n},\quad \Phi(\beta_2)-\Phi(\beta_1)=\gamma,$$

где $\varphi(x)=\Phi'(x)=\frac{1}{\sqrt{2\pi}}exp(-x^2/2)$—плотность распределения стандартного нормального распределения. В силу того, что плотность - четная функция, $\beta_1\neq\beta_2$ и свойства симметрии функции $\Phi(x)$, из системы получим $\beta_2=-\beta_1=$ $=\Phi^{-1}\left(\frac{1+\gamma}{2}\right)=\beta$.

Таким образом, ясно, что для этой задачи *наикратчайший* доверительный интервал совпадает с *центральным* доверительным интервалом. Это, как отмечалось выше, есть следствие симметрии центральной статистики.

В случае, если требуется получить доверительный интервал для параметрической функции $\tau(\theta)$, то можно воспользоваться тем же методом центральных статистик. Нужно только, чтобы центральная статистика зависела от нужной параметрической функции: $G = G(\vec{X}, \tau(\theta))$ и была строго монотонной по τ.

Пример 2. Пусть имеется выборка $\vec{X} = (X_1, \ldots, X_n)$ из статистической модели нормального распределения $N(a, \theta^2)$. На этот раз математическое ожидание известно, а дисперсия $\tau(\theta) = \theta^2$ нет. Задача состоит в получении доверительного интервала для неизвестной дисперсии.

Из курса теории вероятностей известно, что сумма из n квадратов независимых, одинаково распределенных стандартных нормальных случайных величин дает случайную величину χ_n^2 (хи-квадрат с n степенями свободы), имеющую плотность распределения

$$p_{\chi_n^2}(x) = \frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)} I_{[0,+\infty)}(x).$$

Поскольку случайные величины $\frac{X_i - a}{\sigma} \sim N(0,1)$, $i = 1, \ldots, n$ и независимы, то функция $G(\vec{X}, \theta^2) = \frac{1}{\theta^2}\sum\limits_{i=1}^{n}(X_i - a)^2 \sim \chi_n^2$. Выберем теперь пару констант β_1, β_2 таких, чтобы выполнялось равенство $\int\limits_{\beta_1}^{\beta_2} p_{\chi_n^2}(x)dx = \gamma$. Очевидно, имеем

$$\gamma = P_\theta\left(\beta_1 \le G(\vec{X}, \theta^2) \le \beta_2\right) =$$

$$= P_\theta\left(\frac{1}{\beta_2}\sum_{i=1}^{n}(X_i - a)^2 \le \theta^2 \le \frac{1}{\beta_1}\sum_{i=1}^{n}(X_i - a)^2\right).$$

Итак, получаем, что интервал $(T_1(\vec{X}), T_2(\vec{X}))$, где $T_1(\vec{X}) = \frac{1}{\beta_2}\sum\limits_{i=1}^{n}(X_i - a)^2$ и $T_2(\vec{X}) = \frac{1}{\beta_1}\sum\limits_{i=1}^{n}(X_i - a)^2$ будет $\gamma-$доверительным интервалом для неизвестной дисперсии θ^2. Опустим нахождение *оптимального* доверительного интервала, поскольку это сопряжено с достаточно большими аналитическими трудностями.

Для нахождения *центрального* доверительного интервала выберем константы β_1 и β_2, как указывалось выше, из равенств

$$P_\theta\left(G(\vec{X}, \theta^2) < \beta_1\right) = \int\limits_0^{\beta_1} p_{\chi^2_n}(x)dx = \frac{1-\gamma}{2} \text{ и}$$

$$P_\theta\left(G(\vec{X}, \theta^2) > \beta_2\right) = \int\limits_{\beta_2}^{+\infty} p_{\chi^2_n}(x)dx = \frac{1-\gamma}{2}.$$

Если $\chi^2_{\varepsilon,n}$ обозначает ε-квантиль хи-квадрат распределения с n степенями свободы, т.е. $\int\limits_0^{\chi^2_{\varepsilon,n}} p_{\chi^2_n}(x)dx = \varepsilon$, то, очевидно, $\beta_1 = \chi^2_{(1-\gamma)/2,n}$ и $\beta_2 = \chi^2_{(1+\gamma)/2,n}$. Несмотря на то, что центральный доверительный интервал в этом случае не оптимален, он достаточно просто находится в практических задачах (имеются подробные и точные таблицы квантилей хи-квадрат распределений [1]) и при достаточно больших значениях объема n выборки (больше 10) мало отличается от оптимального доверительного интервала.

13.2. Получение доверительных интервалов методом распределения точечных оценок

Если в нашем распоряжении имеется абсолютно непрерывная структура и точечная оценка $T(\vec{X})$ параметра θ, то будем строить *центральный доверительный интервал* для этого неизвестного параметра, основываясь на функции распределения $F_T(x, \theta)$ нашей статистики. Предположим, что эта функция распределения непрерывна и строго монотонна по параметру θ. Тогда для каждого $\theta \in \Theta$ выберем два числа $t_1(\theta)$ и $t_2(\theta)$ такие, что

$$\begin{aligned} P_\theta(T(\vec{X}) < t_1(\theta)) = F_T(t_1(\theta), \theta) = \tfrac{1-\gamma}{2}, \\ P_\theta(T(\vec{X}) > t_2(\theta)) = 1 - F_T(t_2(\theta), \theta) = \tfrac{1-\gamma}{2}, \end{aligned} \tag{17}$$

где $0 < \gamma < 1$ заранее выбранная и зафиксированная доверительная вероятность. Отсюда, очевидно, получаем

$$P_\theta\left(t_1(\theta) \leq T(\vec{X}) \leq t_2(\theta)\right) = \gamma. \tag{18}$$

Далее, в силу предположения о строгой монотонности функции $F_T(x, \theta)$ по θ, замечаем, что обе функции $t_i(\theta), \;\; i = 1, 2$ будут тоже строго монотонны одного типа (убывающие или возрастающие). Поэтому, например, в случае монотонного возрастания $t_i(\theta)$, получаем

$$t_1(\theta) \leq T(\vec{X}) \;\Rightarrow\; \theta \leq t_1^{-1}(T(\vec{X})) \quad \text{и}$$

$$T(\vec{X}) \leq t_2(\theta) \;\Rightarrow\; \theta \geq t_2^{-1}(T(\vec{X}))$$

и из (18), очевидно, получаем

$$P_\theta\left(t_2^{-1}(T(\vec{X})) \leq \theta \leq t_1^{-1}(T(\vec{X}))\right) = \gamma.$$

В случае монотонно убывающих $t_i(\theta)$, аналогично, получим

$$P_\theta\left(t_1^{-1}(T(\vec{X})) \le \theta \le t_2^{-1}(T(\vec{X}))\right) = \gamma.$$

Резюмируя сказанное, кратко алгоритм нахождения доверительного интервала для неизвестного параметра θ можно описать просто следующим образом. Если выборка $\vec{X}$ приняла конкретное значение $\vec{x}$, то статистика примет значение $t = T(\vec{x})$. Из двух уравнений

$$F_T(t,\theta) = \frac{1-\gamma}{2} \quad \text{и} \quad F_T(t,\theta) = \frac{1+\gamma}{2} \qquad (19)$$

находим два числа $\theta_1 < \theta_2$ и получаем $\theta_1 < \theta < \theta_2$.

В случае, если выборка из дискретной статистической модели, то функция распределения $F_T(x,\theta)$ будет ступенчатой. Поэтому в текст выше нужно внести очевидные следующие изменения: (17) нужно заменить на

$$\begin{array}{l} P_\theta(T(\vec{X}) < t_1(\theta)) = F_T(t_1(\theta),\theta) < \frac{1-\gamma}{2}, \\ P_\theta(T(\vec{X}) > t_2(\theta)) = 1 - F_T(t_2(\theta),\theta) < \frac{1-\gamma}{2} \end{array}$$

и вместо (18)

$$P_\theta\left(t_1(\theta) \le T(\vec{X}) \le t_2(\theta)\right) \ge \gamma,$$

а уравнения (19) остаются без изменения.

Пример 3. Пусть имеется выборка $\vec{X} = (X_1, X_2)$ объема 2, имеющая распределение Пуассона $\Pi(\theta)$ с параметром $\theta > 0$. Пусть, во время некоторого эксперимента, выборка приняла значение $\vec{x} = (3,1)$. Требуется найти доверительный интервал для параметра θ с уровнем значимости $\gamma = 0,9$.

Поскольку статистика $T(\vec{X}) = \bar{X} = (X_1 + + X_2)/2$—эффективная оценка для неизвестного параметра θ, то возьмем ее за основу. Известно, что $2\bar{X} \sim \Pi(2\theta)$. Очевидно, что

$$F_{\bar{X}}(t) = P_\theta\left(\frac{X_1+X_2}{2} \le t\right) = P_\theta\left(X_1+X_2 \le 2t\right) = F_{2\bar{X}}(2t) =$$

$$= \sum_{i=0}^{2t} e^{-2\theta}\frac{(2\theta)^i}{i!}.$$

Несложно проверить, что функция распределения $F_{\bar{X}}(t)$ строго монотонно убывает по $\theta > 0$. В нашем случае, очевидно, $t = T(\vec{x}) = \frac{3+1}{2} = 2$. Поэтому, из (19) получаем:

$$e^{-2\theta}\left(1 + 2\theta + 2\theta^2 + \tfrac{4}{3}\theta^3 + \tfrac{2}{3}\theta^4\right) = 0,1;$$
$$e^{-2\theta}\left(1 + 2\theta + 2\theta^2 + \tfrac{4}{3}\theta^3 + \tfrac{2}{3}\theta^4\right) = 0,9.$$

Численно решая эти уравнения, из первого получим $\theta_2 = 3,997$, а из второго – $\theta_1 = 1,216$. Следовательно, построен доверительный интервал $1,216 \le \theta \le 3,997$ с доверительной вероятностью $\gamma = 0.8$.

Понятно, что чем больше доверительная вероятность, тем шире будет доверительный интервал. Например, для $\gamma = 0,9$ доверительный интервал будет $0,985 \le \theta \le 4,577$.

13.3. Асимптотические доверительные интервалы

Определение 4. Пусть для некоторой константы $\gamma \in (0,1)$ *существуют две статистики* $T_i = T_i(\vec{X}_n, \gamma)$, $i = 1,2$*, такие, что*

$$\lim_{n\to\infty} P_\theta(T_1(\vec{X}_n,\gamma) < \theta < T_2(\vec{X}_n,\gamma)) = \gamma.$$

Тогда интервал (T_1, T_2) *называется* ***асимптотическим доверительным интервалом*** *для параметра* θ *с уровнем доверия* γ.

Из теоремы об асимптотической нормальности о.м.п., приведенной раньше, следует, что, при некоторых не обременительных условиях, оценки максимального правдоподобия $\hat{\theta}_n(\vec{X})$ будут асимптотически эффективными и нормальными. Поэтому, при больших объемах выборки ($n \to \infty$) будем иметь (в случае непрерывности информации Фишера $i(\theta)$):

$$\lim_{n\to\infty} P_\theta\left(\left|\hat{\theta}_n(\vec{X}) - \theta\right| \sqrt{ni(\hat{\theta}_n(\vec{X}))} \le c_\gamma\right) =$$

$$= \Phi(c_\gamma) - \Phi(-c_\gamma) = 2\Phi(c_\gamma) - 1 = \gamma,$$

где $c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$ —квантиль стандартного нормального распределения. Отсюда, если положить

$$T_1(\vec{X}_n,\gamma) = \hat{\theta}_n(\vec{X}) - \frac{c_\gamma}{\sqrt{ni(\hat{\theta}_n(\vec{X}))}}$$

и

$$T_2(\vec{X}_n,\gamma) = \hat{\theta}_n(\vec{X}) + \frac{c_\gamma}{\sqrt{ni(\hat{\theta}_n(\vec{X}))}},$$

получаем требуемый асимптотический доверительный интервал. Заметим, что этот интервал будет еще и наикратчайшим в силу центральности построенного интервала и симметричности стандартного нормального распределения (как это отмечалось выше).

Пример 4. Требуется построить доверительный интервал для неизвестного параметра $\theta \in (0,1)$ по выборке $\vec{X} = (X_1, X_2, \ldots, X_n)$ из распределения Бернулли $Bi(1,\theta)$. Будем считать, что объем выборки неограниченно возрастает.

Очевидно, что функция правдоподобия в этом случае равна

$$L(\vec{X},\theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{n\bar{X}}(1-\theta)^{n-n\bar{X}} =$$

$$= exp\left(n\bar{X}\ln\theta + (n-n\bar{X})\ln(1-\theta)\right).$$

Поэтому, из равенства

$$\frac{\partial \ln L(\vec{X},\theta)}{\partial\theta} = \frac{n\bar{X}(1-\theta)-(n-n\bar{X})\theta}{\theta(1-\theta)} = \frac{n\bar{X}-n\theta}{\theta(1-\theta)} = 0,$$

получаем, что о.м.п. для параметра θ равна $\hat{\theta}_n(\vec{X}) = \bar{X}$. Информация Фишера для распределения Бернулли $i(\theta) = 1/(\theta(1-\theta))$ была найдена раньше. Поэтому, асимптотический γ-доверительный интервал для параметра θ имеет вид

$$\left(\bar{X} - \frac{c_\gamma}{\sqrt{n}}\sqrt{\bar{X}(1-\bar{X})},\ \ \bar{X} + \frac{c_\gamma}{\sqrt{n}}\sqrt{\bar{X}(1-\bar{X})}\right),$$

$$c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right).$$

14. Задачи проверки статистических гипотез. Непараметрические гипотезы. Критерии согласия и однородности

Гипотеза – это допущение, которое выдвигается на основании каких-либо соображений и может быть как истинным, так и ложным. Основной задачей статистики здесь является вопрос о том, следует ли, по имеющейся в нашем распоряжении информации, признать гипотезу верной или ее следует отклонить.

Приведем один практический пример ситуации, когда можно сказать, что мы «проверяем гипотезу».

Предположим, что ставится вопрос о том, какую оценку студент Иванов получит на предстоящем экзамене. Выдвигается такая гипотеза: студент Иванов на экзамене получит оценку «отлично». Как можно рассуждать? Если известно, что этот студент учится хорошо, на экзаменах получает преимущественно высокие баллы, то, скорее всего, эту гипотезу следует принять. Если же студент Иванов слабый студент, неоднократно пересдавал экзамены, то гипотезу следует отклонить. Описанные рассуждения можно назвать проверкой гипотезы. Конечно, чтобы быть абсолютно уверенным, нужно дождаться результата экзамена. Но если так поступить, то это уже больше не «проверка» гипотезы, а констатация факта – ложна гипотеза или нет.

Статистическая гипотеза – это гипотеза, которая выдвигается на основании поставленной задачи и проверяется согласно статистическим данным. Такие задачи могут возникать, где угодно. Например,

i) эта игральная кость правильная;

ii) сведения поставщика деталей о своей продукции ложны;

iii) новый метод производства детали лучше, чем старый;

iv) требования стандарта на чистоту воздуха в Красноярске не выдерживаются.

Перечисленные гипотезы являются *статистическими*. Сформулируем их более точно:

i) вероятности выпадения каждой грани такой игральной кости равны, т.е. это означает, что имеется равновероятное распределение дискретной случайной величины, которое представляет собой число очков на каждой грани данной кости;

ii) средняя длина детали, поступившей от поставщика, больше, чем он заявлял (или меньше, или отличается от заявленной);

iii) средняя оценка качества деталей, изготовленных новым методом выше, чем изготовленных старым методом;

iv) значения параметров, характеризующих чистоту воздуха в Красноярске, больше, чем установлено стандартом.

Эти примеры, в частности, показывают, что *статистическая гипотеза* – это некоторое утверждение относительно законов распределений случайных величин или параметров с ними связанных. Отметим, что бывают *параметрические* и *непараметрические* статистические гипотезы. Об этом мы еще будет говорить ниже.

Гипотеза называется *простой*, если она полностью определяет распределение случайной величины. Например, значение некоторого параметра $\boldsymbol{\theta}$ в точности равно заданной величине a_0. Во всех остальных случаях гипотеза называется *сложной*. Например, параметр $\boldsymbol{\theta} > a_0$ — сложная гипотеза. По-другому можно сказать, что сложной называют гипотезу, котороя состоит из конечного или бесконечного числа простых гипотез.

Вернемся к нашим примерам. Гипотеза (i) является простой: здесь проверяется то, что вероятность $\boldsymbol{\theta}$ выпадения каждой грани равна $1/6$. Следующие три гипотезы являются сложными, поскольку значения параметров не конкретизируются.

Поскольку любой статистической гипотезе присущ элемент случайности, то требуется установить, не противоречит ли высказанная нами гипотеза имеющимся выборочным данным. Эта процедура носит название *статистической проверки гипотез*.

Результат сопоставления гипотезы с выборочным (наблюдаемыми) данными может быть либо *отрицательным* (данные наблюдения противоречат высказанной гипотезе, а поэтому гипотезу надо отклонить), либо *положительными* (данные наблюдения не противоречат гипотезе, а поэтому ее можно принять в качестве одного из возможных решений).

Если для изучения определенного явления сформулирована некоторая гипотеза, которую будем называть *основной* или иногда *нулевой*, то эту гипотезу принято обозначать $\boldsymbol{H}_0$. Задача проверки гипотез состоит в том, чтобы создать такое правило, по которому, используя результаты наблюдений (статистических данных) можно было бы принять или отвергнуть основную гипотезу. Это правило обычно называют *статистическим критерием*, а иногда *статистическим тестом* проверки гипотезы $\boldsymbol{H}_0$.

14.1. Основные типы статистических непараметрических гипотез

Некоторые из появляющихся в ходе решения различных задач гипотез перечислены ниже.

(i) Гипотеза о типе закона распределения исследуемой случайной величины. Пусть производится n наблюдений за некоторой случайной величиной ξ с неизвестной функцией распределения $F_\xi(x)$. Задача состоит в подборе некоторой функции распределения $F_0(x)$, при помощи которой можно описать исследуемую функцию распределения $F_\xi(x)$ случайной величины ξ. Проверяемая гипотеза имеет вид $H_0 : F_\xi(x) = F_0(x)$, где $F_0(x)$—известная функция (*простая гипотеза*) или $H_0 : F_\xi(x) \in \mathfrak{F}(\Theta)$, где $\mathfrak{F}(\Theta) = \{F(x,\theta),\ \theta \in \Theta\}$ —параметрическое семейство функций распределения (*сложная гипотеза*).

(ii) Гипотезы об однородности двух или нескольких выборок. Производится l серий наблюдений $(x_{i1}, x_{i2}, \ldots, x_{in_i})\,, i = 1, 2, \ldots, l$. Обозначим $F_i(x)$ неизвестную функцию распределения, которой подчиняются наблюдения i-й выборки. Тогда гипотезу однородности можно записать в виде: $H_0 : F_1(x) \equiv F_2(x) \equiv \cdots \equiv F_l(x)$. Отметим, что количество наблюдений в каждой серии, вообще говоря, разное.

(iii) Гипотезы о числовых значениях параметров исследуемой совокупности. Задачи такого типа возникают, например, когда нужно проверить точность некоторого прибора, инструмента и т. п. Они сводятся к проверке гипотезы о параметрах исследуемой статистической структуры. Формально гипотезы такого рода принимают вид: $H_0 : \Theta = \Delta_0$, где Θ—некоторое параметрическое множество исследуемой статистической структуры, а Δ_0—область конкретных его значений. В качестве таких па-

раметров часто рассматриваются такие характеристики распределения, как среднее значение, медиана и дисперсия.

(iv) Гипотеза независимости. Производятся наблюдения за двумерной случайной величиной (ξ, η) с неизвестной совместной функцией распределения $F_{(\xi,\eta)}(x, y)$. Если есть предположение, что с.в. ξ и η—независимы, то требуется проверить гипотезу $H_0 : F_{(\xi,\eta)}(x, y) = F_\xi(x)F_\eta(y)$. Понятно, что эта задача легко обобщается на любое число случайных величин.

(v) Гипотеза случайности. Предположим, что результат некоторого теста описывается многомерной случайной величиной $\vec{X} = (X_1, X_2, \ldots, X_n)$ с функцией распределения $F_{\vec{X}}(\vec{x})$, которая неизвестна. Можно ли считать, что все компоненты вектора $\vec{X}$ есть независимые, одинаково распределенные случайные величины? Для ответа на этот вопрос требуется проверить гипотезу $H_0 : F_{\vec{X}}(\vec{x}) = F_\xi(x_1) \cdot F_\xi(x_2) \cdots F_\xi(x_n)$, где $F_\xi(x)$—какая то одномерная функция распределения, а $\vec{x} = (x_1, x_2, \ldots, x_n)$.

Конечно, эти примеры статистических гипотез далеко не полностью охватывают всего многообразия возникающих задач. Позже будут отдельно рассматриваться *параметрические* гипотезы, т.е. гипотезы относительно неизвестных параметров распределений.

14.2. Критерии согласия (проверка соответствия выбранной модели распределения исходных данных)

Общий принцип построения критериев согласия (решающего правила) достаточно прост. Пусть для распределения выборки $\vec{X} = (X_1, \ldots, X_n)$ имеется некоторая гипотеза H_0. Находят такую неотрицательную статистику $T(\vec{X})$, которая учитывает отклонение наблюдаемых в процессе эксперимента данных от данных, соответствующих гипотезе H_0. При этом, если гипотеза H_0—сложная, то распределение статистики $T(\vec{X})$ должно быть одним и тем же для всех простых гипотез, из которых состоит H_0.

Допустим, что такую статистику мы нашли и ее распределение при выполнении нулевой гипотезы известно. Пусть $\mathcal{T} = \{t = T(\vec{x}),\ \vec{x} \in \mathcal{X}\}$ множество всех значений статистики $T(\vec{X})$. Выберем некоторое, достаточно малое число $\alpha > 0$ и определим подмножество $\mathcal{T}_{1\alpha} \subset \mathcal{T}$ так, чтобы выполнялось неравенство $P\left(T(\vec{X}) \in \mathcal{T}_{1\alpha} \mid H_0\right) \leq \alpha$, где $P\left(A \mid H_0\right)$ обозначает вероятность события A при выполнении гипотезы H_0. Тогда статистический критерий проверки гипотезы H_0 можно сформулировать таким образом: *если для реализации $\vec{x}$ выборки $\vec{X}$, полученной в результате эксперимента, будем иметь $T(\vec{x}) \in \mathcal{T}_{1\alpha}$, то гипотезу H_0 мы вынуждены отклонить как маловероятную, а в противном случае $T(\vec{x}) \notin \mathcal{T}_{1\alpha}$ должны будем принять*.

В указанном правиле статистику T обычно называют *статистикой критерия*; число α называется *уровнем значимости* критерия, а множество $\mathcal{T}_{1\alpha}$ называется *критическим множеством*. Величину $W(F) = P\left(T(\vec{X}) \in \mathcal{T}_{1\alpha} \mid F\right)$ называют *функцией мощности кри-*

терия. Для распределений из гипотезы $\boldsymbol{H_0}$ ($\boldsymbol{F} \in \boldsymbol{H_0}$), очевидно, $\boldsymbol{W(F)} \leq \boldsymbol{\alpha}$.

14.3. Критерий согласия Колмогорова о виде распределения

Пусть имеется выборка $\vec{X} = (X_1, X_2, \ldots, X_n)$ с функцией распределения $F_\xi(x)$, т.е. с.в. X_j, $j = 1, \ldots, n$—независимы, одинаково распределены и имеют распределение, такое же, как и ξ. Пусть рассматривается простая гипотеза $H_0 : F_\xi(x) = F(x)$, где $F(x)$—некоторая непрерывная функция распределения. Статистикой критерия является выражение

$$D_n(\vec{X}) = \sup_{-\infty<x<+\infty} |F(x) - F_n(x)|,$$

которое представляет из себя максимальное отклонение гипотетической функции от выборочной функции распределения $F_n(x)$ (см. раздел 6). Приятной особенностью этой статистики, как было показано А.Н.Колмогоровым, является то обстоятельство, что распределение D_n не зависит от конкретной функции распределения $F(x)$ (главное только ее непрерывность) и при больших n ($n \geq 20$) ее распределение задается распределением Колмогорова:

$$P(\sqrt{n}D_n \leq t) \approx K(t) = \sum_{k=-\infty}^{+\infty} (-1)^k \exp(-2k^2t^2). \tag{20}$$

Поэтому, для выбранного уровня значимости $\alpha > 0$ критический уровень можно взять равным $t_\alpha(n) = \mu_\alpha/\sqrt{n}$, где $K(\mu_\alpha) = 1 - \alpha$, поскольку

$$P\left(D_n(\vec{X}) \in \mathcal{T}_{1\alpha} \mid H_0\right) = P\left(\sqrt{n}D_n(\vec{X}) \geq \mu_\alpha \mid H_0\right) \approx$$

$$\approx 1 - K(\mu_\alpha) = \alpha.$$

Приведем несколько значений квантилей распределения

Колмогорова: при $\alpha = 0,01,\ \mu_\alpha = 1,63$; при $\alpha = 0,05,\ \mu_\alpha = 1,36$; при $\alpha = 0,1,\ \mu_\alpha = 1,22$; при $\alpha = 0,2,\ \mu_\alpha = 1,07$.

Таким образом, критерий Колмогорова можно сформулировать следующим образом: *если по результатам наблюдений статистика критерия Колмогорова приняла значение $t = D_n(\vec{x})$ (при $n \geq 20$) и удовлетворяет неравенству $\sqrt{t} \geq \mu_\alpha$, то гипотезу H_0 следует отклонить, в противном случае, когда $\sqrt{t} < \mu_\alpha$ – принять*. Отметим, что практическое применение критерия Колмогорова весьма трудоемкая задача.

Пример 1. Эмпирическая функция распределения была определена по 20 значениям случайной величины. При этом наибольшая разность $D_{20}(\vec{x})$ между эмпирической и теоретической функциями распределения оказалась равной 0,33. Находим $\sqrt{n}D_n = \sqrt{20} \cdot 0,33 \approx 1,476$. При уровне значимости $\alpha = 0,05$ нулевая гипотеза (результаты наблюдения подчиняются выбранному распределению) отклоняется, поскольку вычисленное значение статистики критерия 1,476 больше $\mu_{0,05} = 1,36$.

14.4. Критерий согласия χ^2 К.Пирсона о виде распределения для простой гипотезы

В отличие от критерия Колмогорова, критерий χ^2 применим для любых распределений.

Пусть некоторая случайная величина ξ принимает значения в области $\mathcal{E}$. Разобьем это множество на N непересекающихся частей $\mathcal{E}_1, \dots, \mathcal{E}_N \; : \; \mathcal{E} = \bigcup\limits_{i=1}^{N} \mathcal{E}_i, \; \mathcal{E}_i \bigcap \mathcal{E}_j = \emptyset \quad \forall i \neq j$. Пусть ν_i — число компонентов реализации выборки $\vec{x} = (x_1, \dots, x_n)$, попавших в область $\mathcal{E}_i$; $\nu_1 + \dots + \nu_N = n$.

Как обычно, все компоненты выборки $\vec{X} = (X_1, \dots, X_n)$ есть независимые, одинаково распределенные случайные величины, имеющие такое же распределение как и у случайной величины ξ. Пусть $p_j^0 = P(\xi \in \mathcal{E}_j \mid H_0)$, $j = 1, \dots, N$—вероятность попадания с.в. ξ в область $\mathcal{E}_j$ при выполнении нулевой гипотезы. Ясно, что $p_1^0 + \dots + p_N^0 = 1$. Введем статистику

$$X_n^2 = X_n^2(\vec{\nu}) = \sum_{j=1}^{N} \frac{(\nu_j - np_j^0)^2}{np_j^0} = \sum_{j=1}^{N} \frac{\nu_j^2}{np_j^0} - n. \qquad (21)$$

Справедлива следующая теорема.

Теорема 1. Если все $0 < p_j^0 < 1, \; j = 1, \dots, N$ *то*

$$\lim_{n\to\infty} P(X_n^2 \le u \mid H_0) = P(\chi^2_{N-1} \le u) =$$

$$= \int_0^u \frac{x^{(N-1)/2-1} e^{-x/2}}{2^{(N-1)/2}\Gamma((N-1)/2)} dx. \qquad (22)$$

Эту теорему оставим без доказательства.

Замечание. Равенство (22) есть сходимость по распределению последовательности случайных величин X_n^2 к случайной величине χ^2_{N-1} при выполнении гипотезы H_0. Это выражение, как и выше, можно записать кратко, например, так:

$$X_n^2|_{H_0} \xrightarrow[n\to\infty]{d} \chi^2_{N-1}.$$

Практически, когда объем наблюдений $n \geq 50$ и $\min\limits_{1\leq j\leq N} \nu_j \geq 5$, то распределение статистики X_n^2 можно считать имеющим распределение χ^2_{N-1}. Поэтому, для выбранного уровня значимости α критическая граница

$$t_\alpha = \chi^2_{1-\alpha,N-1}$$

есть квантиль распределения χ^2_{N-1}, поскольку

$$P(X_n^2 \in \mathcal{T}_{1\alpha} \mid H_0) = P(X_n^2 \geq t_\alpha \mid H_0) \approx$$

$$\approx 1 - F_{\chi^2_{N-1}}(\chi^2_{1-\alpha,N-1}) = \alpha.$$

Таким образом, для проверки простой гипотезы

$$H_0 : F_\xi(x) = F(x),$$

где $F(x)$—некоторая заданная функция распределения, после проведения описанной выше процедуры группировки, решающее правило выглядит таким образом: *пусть для заданного уровня значимости α и исходного объема выборки $n \geq 50$ наблюдавшиеся значения $\vec{h} = (h_1, \ldots, h_N)$ вектора частот $\vec{\nu} = (\nu_1, \ldots, \nu_N)$ удовлетворяют условиям $\min\limits_{1\leq j\leq N} h_j \geq 5$; тогда, если наблюдавшееся значение $t = X_n^2(\vec{h})$ статистики (21) удовлетворяет неравенству $t > \chi^2_{1-\alpha,N-1}$, то нулевую гипотезу отвергают, а в противном случае — принимают.*

Пример 2. При подбрасывании 4040 раз монеты Бюффон получил 2048 выпадений «герба» и 1992 выпадений «решки». Проверить гипотезу H_0 о том, что подбрасываемая монета была симметричной.

Решение. Выберем уровень значимости критерия $\alpha = 0,05$. Очевидно, гипотеза имеет вид $H_0 : P(\text{герб}) = P(\text{решка}) = 0,5$, а исходные данные здесь таковы: $n = 4040$, $N = 2$, $h_1 = 2048$, $h_2 = 1992$, $p_1^0 = 0.5$, $p_2^0 = 0,5$. Поэтому, применяя формулу (21), получим $t = X_n^2(\vec{h}) = (2048 - 4040/2)^2/(4040/2) + (1992 - 4040/2)^2/(4040/2) \approx 0,776$. По таблицам распределения χ^2 найдем $\chi^2_{095,1} \approx 3,841$. В силу того, что $0,776 < 3,841$, гипотеза H_0 принимается.

14.5. Критерий согласия χ^2 К.Пирсона о виде распределения для сложной гипотезы

Метод, описанный в предыдущем разделе применим и в случае сложной гипотезы:

$$H_0 : F_\xi(x) \in \mathfrak{F}(\Theta), \text{ где } \quad \mathfrak{F}(\Theta) = \left\{ F(x, \theta), \ \theta \in \Theta \subset R^k \right\} -$$

заданное параметрическое семейство функций распределения. Как обычно, в нашем распоряжении имеется выборка $\vec{X} = (X_1, \ldots, X_n)$ из распределения некоторой неизвестной случайной величины ξ и наша задача состоит в проверке гипотезы H_0 относительно распределения этой случайной величины. Как отмечалось выше, в случае если с.в. ξ имеет непрерывное распределение, задачу сводим к дискретной схеме, предварительно группируя данные по N интервалам $\mathcal{E}_1, \ldots, \mathcal{E}_N$ и в качестве событий выбираем $A_j = \{\xi \in \mathcal{E}_j\}$ и $p_j(\vec{\theta}) = P_\Theta(A_j), \ j = 1, \ldots, N$.

Зададим, как и в предыдущем разделе, функцию

$$X_n^2 = X_n^2(\vec{\nu}, \vec{\theta}) = \sum_{j=1}^{N} \frac{(\nu_j - np_j(\vec{\theta}))^2}{np_j(\vec{\theta})},$$

где вероятности $p_j(\vec{\theta}) = P(\xi \in \mathcal{E}_j \mid H_0), \ j = 1, \ldots, N$, а, следовательно и X^2, будут функциями от неизвестного векторного (в общем случае) параметрп $\vec{\theta}$. Чтобы исключить зависимость X_n^2 от неизвестных параметров, возьмем оценку $\vec{\tilde{\theta}}_n = \vec{\tilde{\theta}}_n(\vec{X})$ и тогда вместо функции X_n^2 получаем статистику

$$\tilde{X}_n^2 = X_n^2(\vec{\nu}, \vec{\tilde{\theta}}_n) = \sum_{j=1}^{N} \frac{(\nu_j - np_j(\vec{\tilde{\theta}}_n))^2}{np_j(\vec{\tilde{\theta}}_n)},$$

представляющую из себя функцию только от выборки $\vec{X}$. Понятно, что при таком подходе, распределение статистики $\tilde{X}_n^2$ может зависеть от выбора конкретной оценки неизвестных параметров и, поэтому, теорема 1 предыдущего раздела не будет справедлива. Однако, если выбрать специальные оценки неизвестных параметров $\vec{\theta}$, то можно утверждать, что асимптотическое распределение (т.е. при $n \to \infty$) имеет распределение χ^2, а именно, справедлива

Теорема 2 (Фишера). Пусть функции $p_j(\vec{\theta}) = P(\xi \in \in \mathcal{E}_j \mid H_0)$, $j = 1, \ldots, N$, $\vec{\theta} = (\theta_1, \ldots, \theta_k)$, $k < N - 1$, *удовлетворяют следующим условиям:*

j) $\sum\limits_{j=1}^{N} p_j(\vec{\theta}) = 1$ *для всех* $\vec{\theta} \in \Theta$;

jj) $p_j(\vec{\theta}) \geq c > 0, \quad j = 1, \ldots, N$, *и существуют непрерывные производные* $\frac{\partial p_j(\vec{\theta})}{\partial \theta_l}$ *и* $\frac{\partial^2 p_j(\vec{\theta})}{\partial \theta_l \partial \theta_m}$, $l, m = 1, \ldots, k$;

jjj) матрица $\left\| \frac{\partial p_j(\vec{\theta})}{\partial \theta_l} \right\|$ *размера* $N \times k$ *имеет ранг* k *для всех* $\vec{\theta} \in \Theta$.

Тогда, если $\tilde{\vec{\theta}}_n = \hat{\vec{\theta}}_n$*—мультиномиальная оценка максимального правдоподобия (т.е., о.м.п., полученная по сгруппированным, как описывалось выше, данным) для параметра* $\vec{\theta}$ *и* $\hat{X}_n^2 = X_n^2(\vec{\nu}, \hat{\vec{\theta}}_n)$, *то* $\lim\limits_{n\to\infty} P(\hat{X}_n^2 \leq u \mid H_0) = = P(\chi^2_{N-k-1} \leq u)$.

Эту теорему мы также оставляем без доказательства. Используя утверждение теоремы 2, приведем схему использования решающего правила χ^2 для проверки сложной нулевой гипотезы.

Пусть во время некоторого эксперимента наблюдается одно из N *несовместных событий* $A_1, \ldots, A_N$ *и о вероятностях появления этих событий для некоторого фиксированного* k*-мерного параметрического множе-*

ства $\Theta \subset R_k$ *ставится гипотеза*

$$H_0 : p_j = p_j(\vec{\theta}),\ j = 1, \ldots, N,\ \text{где } \vec{\theta} = (\theta_1, \ldots, \theta_k) \in \Theta.$$

Пусть число наблюдений $n \geq 50$ *и наблюдавшиеся частоты* $h_1, \ldots, h_N$ *событий удовлетворяют условию* $\min\limits_{1 \leq j \leq N} h_j \geq 5$. *Определим значение оценки* $\hat{\vec{\theta}}_n$ *неизвестного параметра* $\vec{\theta}$ *решая систему*

$$\sum_{j=1}^{N} \frac{h_j}{p_j(\vec{\theta})} \frac{\partial p_j(\vec{\theta})}{\partial \theta_l} = 0, \quad l = 1, \ldots, k \tag{23}$$

относительно $\theta_1, \ldots, \theta_k$. *Далее, вычисляем* $\hat{p}_j = p_j(\hat{\vec{\theta}}_n)$, $j = 1, \ldots, N$ *и находим значение статистики* $\hat{X}_n^2 = \sum\limits_{j=1}^{N} (h_j - n\hat{p}_j)^2 / (n\hat{p}_j)$. *Для заданного уровня значимости* α *определяем по таблице распределения с.в.* χ^2_{N-k-1} *значение квантили* $\chi^2_{1-\alpha, N-k-1}$ *и в случае* $\hat{X}_n^2 > \chi^2_{1-\alpha, N-k-1}$ *гипотеза* H_0 *отвергается, а в противном случае – принимается.*

Пример 3. Пусть производится n независимых наблюдений над некоторой неотрицательной целочисленной случайной величиной. Требуется проверить гипотезу о том, что эта случайная величина имеет распределение Пуассона.

Решение. Пусть $\mathcal{E}_j = \{j - 1\}$, $j = 1, \ldots, N - 1$ – одноточечные множества и $\mathcal{E}_N = \{N - 1, N, N + 1, \ldots$ – многоточечное множество, $\varphi(l, \theta) = P_\theta(\xi = l) = e^{-\theta}\theta^l / l!$. Тогда $p_j(\theta) = P_\theta(\xi \in \mathcal{E}_j) = \varphi(j - 1, \theta)$, $j = 1, \ldots, N - 1$ и $p_N(\theta) = P_\theta(\xi \in \mathcal{E}_{N-1}) = \sum\limits_{l=N-1}^{\infty} \varphi(l, \theta)$. Ищем оценку $\hat{\theta}_n$.

Поскольку $k = 1$ из (23) получаем:

$$\sum_{j=0}^{N-2}\left(\frac{j}{\theta}-1\right)h_{j+1} + h_N \sum_{l=N-1}^{\infty}\left(\frac{l}{\theta}-1\right)\varphi(l,\theta)\times$$

$$\times\left(\sum_{l=N-1}^{\infty}\varphi(l,\theta)\right)^{-1} = 0,$$

откуда следует

$$\theta = \frac{1}{n}\left[\sum_{j=0}^{N-2} jh_{j+1} + h_N \sum_{l=N-1}^{\infty} l\varphi(l,\theta)\left(\sum_{l=N-1}^{\infty}\varphi(l,\theta)\right)^{-1}\right].$$

В последнем выражении первое слагаемое в квадратных скобках равно сумме всех наблюдений за случайной величиной ξ, в которых каждое наблюдение не превосходит значения $N-2$, а второе слагаемое приблизительно равно сумме всех наблюдений, каждое из которых не меньше $N-1$. Поэтому, все последнее выражение приблизительно равно среднему арифметическому из всех n значений, наблюдаемых в эксперименте. Поэтому получаем, что $\hat{\theta}_n \approx \bar{x} = \sum\limits_{i=1}^{n} x_i$. Дальше, находим

$$\hat{p}_j = p_j(\bar{x}) = \begin{cases} e^{-\bar{x}}\frac{\bar{x}^{j-1}}{(j-1)!}, & \text{при} \;\; j = 1,\ldots,N-1; \\ e^{-\bar{x}}\sum\limits_{l=N-1}^{\infty}\frac{\bar{x}^l}{l!}, & \text{при} \;\; j = N. \end{cases}$$

Окончательно получаем, что гипотезу H_0 о распределении наблюдений по закону Пуассона отвергаем, если

$$\sum_{j=1}^{N}\frac{(h_j - n\hat{p}_j)^2}{n\hat{p}_j} \geq \chi^2_{1-\alpha,N-2}.$$

Пример 4. Проверялся размер $n = 200$ деталей, изготовленных на токарном станке–автомате. В таблице

$\mathcal{E}_j$	h_j	p_j	np_j
(–20, –15)	7	0,0239	4,78
(–15, –10)	11	0,0469	9,38
(–10, –5)	15	0,0977	19,54
(–5, 0)	24	0,1615	32,30
(0, 5)	40	0,1979	39,58
(5, 10)	41	0,1945	38,90
(10, 15)	26	0,1419	28,38
(15, 20)	17	0,0831	16,62
(20, 30)	10	0,0526	10,52

приведены границы интервалов $\mathcal{E}_j$ отклонения от номинального размера, h_j—число деталей со значением отклонения, попадающим в данный интервал, p_j—теоретическая вероятность попадания отклонений в интервалы, np_j—теоретическая частота попадания отклонений в интервалы. Требуется оценить с помощью критерия χ^2 согласие выборочного распределения с нормальным распределением при уровне значимости $\alpha = 0,05$.

Решение. Вычисляем теоретические вероятности p_j попадания отклонений в интервалы $\mathcal{E}_j = (y_j, y_{j+1})$ по формуле $p_j = \Phi(z_{j+1}) - \Phi(z_j)$, где $\Phi(\cdot)$—функция распределения стандартного нормального закона, $z_j = (y_j - \bar{y})/\bar{\sigma}$, где $\bar{y}$ и $\bar{\sigma}$– выборочные оценки среднего и среднего квадратического отклонения, $\bar{y} = 4,30, \quad \bar{\sigma} = 9.71$. Находим значение $X^2 \approx 7,09$. Число степеней свободы расределения статистики X^2 равно 9 – 2 – 1 = 6. Критическое значение статистики X^2, т.е. квантиль случайной величины χ^2_6 уровня $1 - \alpha = 0,95$ $\chi^2_{0,95,6} \approx 14,45$. Следовательно, поскольку $7,09 < 1445$, гипотеза о нормальности отклонений размеров деталей от номинальных размеров не противоречит наблюдениям.

14.6. Критерий однородности Смирнова

На практике, например, в системах контроля качества производства, или в различных задачах страхования, а также во многих других прикладных областях, часто приходится рассматривать задачи, связанные со сравнением распределений двух и более различных выборок. Например, для различных выборок гипотезу о равенстве соответствующих функций распределений принято называть *гипотезой однородности*.

Одним из критериев проверки гипотезы однородности для двух выборок из *непрерывных* распределений является критерий Смирнова. Статистикой этого критерия является величина

$$\mathcal{D}_{nm} = \sup_{-\infty<x<+\infty} |F_{1n}(x) - F_{2m}(x)|\,,$$

где $F_{1n}(x)$ и $F_{2m}(x)$ – две выборочные функции, построенные по выборкам $\vec{X} = (X_1, \ldots, X_n)$ с некоторой неизвестной функцией распределения $F_1(x)$ и $\vec{Y} = (Y_1, \ldots, Y_m)$ с некоторой неизвестной функцией распределения $F_2(x)$ соответственно. В силу того, что эмпирическая функция распределения является оптимальной оценкой теоретической функции распределения, гипотеза однородности $H_0 : \; F_1(x) \equiv F_2(x)$ означает, что эмпирические функции распределения $F_{1n}(x)$ и $F_{2m}(x)$ оценивают одну и ту же неизвестную теоретическую функцию распределения. Таким образом, по крайней мере для больших n и m, статистика $\mathcal{D}_{nm}$ не должна сильно отклоняться от нуля при справедливости нулевой гипотезы. Значительное отклонение от нуля этой статистики означает что нулевая гипотеза не верна. Поэтому, критическую область (т.е. область отклонения нулевой гипотезы) следует выбрать в виде $\mathcal{T}_{1\alpha} = \{t \geq t_\alpha(n,m)\}$. Смирновым было получено предельное (при n и $m \to \infty$) распределение статистики $\mathcal{D}_{nm}$. На основании теоремы Смирнова, положим $t_\alpha(n,m) = \sqrt{\frac{n+m}{nm}}\lambda_\alpha$ при больших n и m, где $K(\lambda_\alpha) = 1-\alpha$ и функция распределения Колмогорова $K(t)$ определена равенством (20). При таком выборе критической области, получаем

$$P(\mathcal{D}_{nm} \in \mathcal{T}_{1\alpha} \mid H_0) = P(\mathcal{D}_{nm} \geq t_\alpha(n,m) \mid H_0) =$$

$$= P\left(\sqrt{\frac{nm}{n+m}}\mathcal{D}_{nm} \geq \lambda_\alpha \mid H_0\right) \approx 1 - K(\lambda_\alpha) = \alpha.$$

Используя описанные выше идеи, сформулируем критерий однородности Смирнова: *если объемы выборок удовлетворяют условиям $n \geq 50$ и $m \geq 50$, то вычислив значение t статистики $\mathcal{D}_{nm}$, нулевую гипотезу отвергают в случае $t \geq \sqrt{\frac{n+m}{nm}}\lambda_\alpha$.*

Замечание. Кроме технических ограничений критерия Смирнова, вычисление статистики $\mathcal{D}_{nm}$ представляет собой достаточно трудоемкую процедуру, поэтому этот критерий сравнительно редко применяется на практике.

14.7. Критерий однородности χ^2

Этот критерий применяется как для дискретных распределений, так и для непрерывных после предварительной группировки данных. Кроме этого этот метод позволяет одновременно анализировать любое конечное число выборок.

Предположим, что проведено m серий наблюдений, независимых в каждой серии, состоящих из $n_1, \ldots, n_m$ наблюдений соответственно. Считаем, что при каждом опыте наблюдается одно из k различных значений исходов. Пусть ν_{ij} – число i-го исхода в j-ой серии, причем $\sum\limits_{i=1}^{k} \nu_{ij} = n_j, \ \ j = 1, \ldots, m.$

Проверяется нулевая гипотеза H_0, о том, что все серии наблюдений проводятся над одной и той же случайной величиной. Эту гипотезу в математическом виде можно сформулировать, например, следующим образом. Пусть p_{ij}–вероятности того, что появится i-й исход в j–й серии ($i = 1, \ldots, k$ и $j = 1, \ldots, m$). Тогда $(p_{1j}, \ldots, p_{kj})$–вероятностное распределение j-й серии и нулевая гипотеза состоит из m равенств

$$H_0 : (p_{1j}, \ldots, p_{kj}) = (p_1, \ldots, p_k), \ \ j = 1, \ldots, m,$$

где $(p_1, \ldots, p_k)$–некоторое фиксированное, вообще говоря, неизвестное вероятностное распределение, $p_1 + \ldots + p_k = 1$. Положим $\nu_i = \sum\limits_{j=1}^{m} \nu_{ij}, \ i = 1, \ldots, k$. Общее число наблюдений равно $n = \sum\limits_{i=1}^{k} \sum\limits_{j=1}^{m} \nu_{ij}$. Определим статистику

$$\hat{X}_n^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(\nu_{ij} - n_j \nu_i / n)^2}{n_j \nu_i / n}. \qquad (24)$$

Доказано, что в случае, когда нулевая гипотеза H_0 справедлива, то статистика $\hat{X}_n^2$, определенная равенством (24), сходится по распределению к случайной величине, имеющей распределение χ^2 с $(k-1)(m-1)$ степенями свободы. Поэтому, критерий однородности χ^2 может быть сформулирован следующим образом: *если вычисленная по практическим данным статистика (24) приняла значение большее или равное, чем квантиль $\chi^2_{1-\alpha,(k-1)(m-1)}$ для некоторого выбранного уровня значимости α, то нулевая гипотеза отвергается*.

14.8. Критерий независимости χ^2

Как отмечалось выше, если имеется двумерная выборка $((X_1, Y_1), \ldots, (X_n, Y_n))$ из двумерной случайной величины (ξ, η) с неизвестной совместной функцией распределения $F_{(\xi,\eta)}(x, y)$, то требуется проверить гипотезу H_0 : $F_{(\xi,\eta)}(x, y) = F_\xi(x) F_\eta(y)$.

Будем считать, что случайная величина ξ принимает k различных значений $a_1, \ldots, a_k$, а случайная величина η — m различных значений $b_1, \ldots, b_m$, в случае, когда эти случайные величины имеют дискретные распределения. Если же с.в. ξ и η не дискретны, то множество значений, принимающих с.в. ξ разбиваем на k интервалов $\mathcal{E}_1^{(1)}, \ldots, \mathcal{E}_k^{(1)}$ и для η — на m интервалов $\mathcal{E}_1^{(2)}, \ldots, \mathcal{E}_m^{(2)}$ и, соответственно, двумерная область значений случайного вектора (ξ, η) разбивается на km прямоугольных областей $\mathcal{E}_i^{(1)} \times \mathcal{E}_j^{(2)}$.

Пусть ν_{ij} обозначает число наблюдений пары (a_i, b_j) в реализации выборки. Положим $\nu_{i\cdot} = \sum\limits_{j=1}^{m} \nu_{ij},\ i = 1, \ldots, k$ и

$\nu_{\cdot j} = \sum\limits_{i=1}^{k} \nu_{ij},\ j = 1. \ldots, m$. Выберем статистику

$$\hat{X}_n^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(\nu_{ij} - \nu_{i\cdot}\nu_{\cdot j}/n)^2}{\nu_{i\cdot}\nu_{\cdot j}/n}. \tag{25}$$

Если нулевая гипотеза верна, то значение этой статистики будет близко к нулю. Кроме того, доказано, что в случае справедливости гипотезы H_0 при $n \to \infty$ статистика $\hat{X}_n^2$ по распределению стремится к случайной величине, имеющей распределение χ^2 с $(k-1)(m-1)$ степенями свободы.

Поэтому сформулируем статистическое правило: *гипотезу H_0 о независимости отвергают тогда, когда по* *фактическим данным вычисленное значение по формуле (25) будет больше или равно, чем квантиль $\chi^2_{1-\alpha,(k-1)(m-1)}$ для выбранного уровня значимости α.*

Пример 5. По переписи населения Швеции 1936 г. из совокупности всех супружеских пар была получена выборка (данные в таблице), содержащая распределение годовых доходов (в тыс. крон) и количества детей в семье.

	доходы (тыс. крон)			
К-во детей	0-1	1-2	2-3	≥ 3
0	2161	3577	2184	1636
1	2755	5081	2222	1052
2	936	1753	640	306
3	225	419	96	38
≥ 4	39	98	31	14

Требуется установить при уровне значимости $\alpha = 0,05$ являются ли зависимыми количество детей в семье (величина X) и уровень годового дохода этой семьи (величина Y).

Решение. Для проверки гипотезы H_0 о независимости количества детей в семье и семейного дохода вычислим значение статистики $\hat{X}^2_n$ по формуле (25), где $k = 5,\ m = 4$, используя данные из таблицы. Получаем $\hat{X}^2_{20} \approx 568,6$. По таблице распределения χ^2 находим квантиль $\chi^2_{095,12} \approx \approx 21,0$. Из этих результатов (568,6>21,0) следует, что гипотеза H_0 о независимости доходов от количества детей в семье отвергается.

15. Параметрические гипотезы

Выше было введено понятие статистической гипотезы. Напомним, что статистическая гипотеза – это предположение о свойствах случайных величин и их распределений, которые можно проверить статистическими методами по результатам наблюдений.

Бывают случаи, когда закон распределения известен, но неизвестны его параметры. Если есть основания предположить, что неизвестный параметр θ равен определенному значению θ_0, то выдвигают гипотезу $\theta = \theta_0$. Эта гипотеза называется гипотезой о *значении параметра распределения*.

Наряду с выдвигаемой основной гипотезой H_0 будем рассматривать противоположную ей гипотезу H_1, которую будем называть *альтернативной гипотезой*.

Пример: если нулевая гипотеза состоит в предположении, что математическое ожидание нормального распределения a равно 0, то альтернативная гипотеза может быть, например, $a \neq 0$. Коротко это можно записать таким образом: $H_0 : \theta = 0$; $H_1 : \theta \neq 0$. Напомним, что *простой* называют гипотезу, которая содержит только одно предположение (например, определенная выше H_0 – простая гипотеза); *сложной* называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез (H_1 – сложная гипотеза).

Как и выше, для проверки статистической гипотезы H_0 против альтернативной гипотезы H_1 будем использовать специальную случайную величину, распределение $f(x, \theta)$ которой известно. Такую величину, как и раньше, будем называть *статистикой критерия*.

Наблюдаемым значением критерия называют значение статистики критерия, вычисленное с использованием выборочных (наблюдаемых) данных. Множество всех возможных значений статистики критерия (или короче критерия) разбивают на два непересекающихся подмножества: одно из них соответствует значениям критерия, при которых нулевая гипотеза принимается и отвергается первая гипотеза, а другое – при которых принимается первая гипотеза и соответственно нулевая гипотеза отвергается.

Критической областью будем называть совокупность значений выборки (т.е. такое $n-$мерное подмножество выборочного множества), при которых значения статистики критерия принадлежат области принятия альтернативы $\boldsymbol{H}_1$.

Если рассматривается задача проверки нулевой гипотезы против альтернативной первой гипотезы, то для этого достаточно определить только критическую область. Критерий проверки нулевой гипотезы $\boldsymbol{H}_0$ против альтернативной гипотезы $\boldsymbol{H}_1$ можно сформулировать так: *если реализация выборки $\vec{x} = (x_1, \ldots, x_n)$ принадлежит* критической *области, то принимается альтернативная гипотеза $\boldsymbol{H}_1$ и отвергается нулевая $\boldsymbol{H}_0$, а в противном случае наоборот принимается нулевая гипотеза $\boldsymbol{H}_0$ и отвергается альтернативная $\boldsymbol{H}_1$.*

При реализации описанной выше процедуры проверки статистической гипотезы $\boldsymbol{H}_0$ против альтернативы $\boldsymbol{H}_1$ возможно принятие неверных решений, а именно могут быть допущены ошибки двух видов. *Ошибка первого рода* состоит в том, что отклоняется правильная нулевая гипотеза, а принимается неверная первая гипотеза; *ошибка второго рода* — принимается неверная нулевая гипотеза и отклоняется верная первая гипотеза. Конечно, хотелось бы использовать такую статистику критерия, при которой будут минимальными вероятности ошибки и первого, и второго рода. Однако этого сделать не удается, поскольку, как правило, уменьшение одной ошибки приводит к увеличению другой. Поэтому обычно поступают так: выбирают значение вероятности ошибки первого рода или фиксируют максимальный уровень границы вероятности ошибки первого рода и статистику критерия выбирают так, чтобы минимизировать вероятность ошибки второго рода.

Итак, пусть $\vec{X} = (X_1, \ldots, X_n)$—выборка объема n с функцией распределения из параметрического семейства $\mathfrak{L}(\Theta) = \{F(x, \vec{\theta}),\ \vec{\theta} \in \Theta \subset R^k\}$ и проверяется гипотеза $H_0 : \vec{\theta} \in \Theta_0$ против альтернативы $H_1 : \vec{\theta} \in \Theta_1$, где $\Theta_i \subset \Theta,\ i = 0, 1$—два подмножества, таких, что $\Theta_0 \bigcap \bigcap \Theta_1 = \emptyset$.

Пусть $\mathcal{X} \subset R^n$—выборочное множество, т.е. множество всех значений, которые принимает выборка $\vec{X}$. Предположим, что $\mathcal{X}_1 \subset \mathcal{X}$—критическое множество. Как понятно из приведенной выше процедуры, выбор *критерия* для задачи проверки нулевой гипотезы против альтернативной гипотезы будет определение критического множества $\mathcal{X}_1$. Назовем $W(\theta) = P_\theta(\vec{X} \in \mathcal{X}_1),\ \theta \in \Theta$ ***функцией мощности критерия***. Понятно, что для $\theta \in \Theta_0$ значение $W(\theta)$ будет вероятностью ошибки *первого* рода и, соответственно, значение $1 - W(\theta)$ для $\theta \in \Theta_1$ будет вероятностью ошибки *второго* рода.

Пусть $\alpha = \sup\limits_{\theta \in \Theta_0} W(\theta)$ —максимальная вероятность ошибки первого рода. Это число α будем называть ***уровнем значимости критерия***.

Предположим, что имеются два критерия одинакового уровня значимости. Скажем, что первый критерий $\mathcal{X}_{1\alpha}^{(1)}$ ***мощнее*** второго критерия $\mathcal{X}_{1\alpha}^{(2)}$, *если максимальная вероятность ошибки второго рода первого критерия строго меньше максимальной вероятности ошибки второго ро-*

да второго критерия. Иначе это можно определить так: $\inf\limits_{\theta\in\Theta_1} W^{(1)}(\theta) > \inf\limits_{\theta\in\Theta_1} W^{(2)}(\theta)$, где $W^{(i)}(\theta) = P_\theta(\vec{X} \in \mathcal{X}_{1\alpha}^{(i)}),\ i = 1, 2$.

Критерий $\mathcal{X}_{1\alpha}^{*}$ *уровня значимости* α *называют* ***наиболее мощным*** *критерием, если для любого другого критерия* $\mathcal{X}_{1\alpha}$ *уровня значимости* α *справедливо неравенство* $\inf\limits_{\theta\in\Theta_1} W^{*}(\theta) > \inf\limits_{\theta\in\Theta_1} W(\theta)$.

16. Критерий Неймана—Пирсона для простых гипотез

Как и выше, считаем, что в нашем распоряжении имеется выборка $\vec{X} = (X_1, \dots, X_n)$ объемом n с функцией распределения $F_\xi(x, \theta)$ и неизвестным параметром θ. Относительно случайной величины ξ будем предполагать, что она имеет либо дискретное распределение с $f(u_i, \theta) = P_\theta(\xi = u_i),\ i = 1, 2, \dots$ и, поэтому, ее функция распределения имеет вид $F_\xi(x, \theta) = \sum\limits_{i:\ u_i \le x} f(u_i, \theta)$; либо абсолютно-непрерывное распределение с плотностью $f(u, \theta)$ и тогда функция распределения будет $F_\xi(x, \theta) = \int\limits_{-\infty}^{x} f(u, \theta) du$.

Предположим, что каждое из множеств Θ_0 и Θ_1, рассматриваемых в предыдущем разделе, являются одноточечными, т.е. будем рассматривать простую гипотезу $H_0 : \theta = \theta_0$ против простой альтернативы $H_1 : \theta = \theta_1$. То есть, задачу, которую будем рассматривать в этом разделе, можно сформулировать так: *исходя из n наблюдений за одной и той же случайной величиной ξ, нужно выбрать, какая из двух функций распределения $F_\xi(x, \theta_0)$ или $F_\xi(x, \theta_1)$ является правильной; при этом требуется, чтобы вероятность ошибки первого рода не превосходила (или в точности была бы равна) заранее выбранного, достаточно малого значения $\alpha > 0$, а ошибка второго рода была бы наименьшей*. Ответ на поставленный вопрос дает следующее утверждение.

Теорема (Неймана—Пирсона). Для сформулированной задачи существует наиболее мощный критерий уровня значимости α, который задается критической областью

$$\mathcal{X}_{1\alpha} = \{\vec{x} \in \mathcal{X} :\ l(\vec{x}, \theta_0, \theta_1) \ge c_\alpha\},$$

где константа c_α определяется из равенства

$$P_{\theta_0}(l(\vec{X}, \theta_0, \theta_1) \geq c_\alpha) = \alpha,$$

где

$$l(\vec{X}, \theta_0, \theta_1) = \left(\prod_{i=1}^{n} f(X_i, \theta_1)\right) / \left(\prod_{i=1}^{n} f(X_i, \theta_0)\right)$$

является статистикой критерия, которую называют статистикой отношения правдоподобий.

Эту теорему оставляем без доказательства.

Замечание. Если требуется проверить простую гипотезу $H_0 : F(x) = F_0(x)$ против простой альтернативы $H_1 : F(x) = F_1(x)$, то можно воспользоваться теоремой Неймана—Пирсона. Для этого достаточно взять параметрическую функцию

$$G(x, \theta) = (1 - \theta)F_0(x) + \theta F_1(x)$$

и выбрать простую гипотезу $H_0 : \theta = 0$ против простой альтернативы $H_1 : \theta = 1$.

Пример 1. Пусть имеется выборка $\vec{X} = (X_1, \ldots, X_n) \sim$ $\sim N(\theta, \sigma^2)$ из нормального распределения с математическим ожиданием θ и дисперсией $\sigma^2 > 0$. Требуется проверить гипотезу $H_0 : \theta = \theta_0$ против альтернативы $H_1 : \theta = \theta_1$, где $\theta_0 > \theta_1$, и найти ошибку второго рода.

Ясно, что функция правдоподобия равна

$$L(\vec{X}, \theta) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \theta)^2}{2\sigma^2}\right) =$$

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

откуда получаем статистику отношения правдоподобий

$$l(\vec{X},\theta_0,\theta_1)=\exp\left(\frac{n(\theta_1-\theta_0)}{\sigma^2}\bar{X}-\frac{n(\theta_1^2-\theta_0^2)}{2\sigma^2}\right)=$$

$$=\exp\left(\sqrt{n}\frac{(\theta_1-\theta_0)}{\sigma}\left(\sqrt{n}\frac{\bar{X}-\theta_0}{\sigma}-\sqrt{n}\frac{\theta_1-\theta_0}{2\sigma}\right)\right).$$

Если справедлива гипотеза H_0, то $\sqrt{n}\frac{\bar{X}-\theta_0}{\sigma}\sim N(0,1)$, поэтому

$$P_{\theta_0}(l(\vec{X},\theta_0,\theta_1)\geq c_\alpha)=$$

$$=P_{\theta_0}\left(\sqrt{n}\frac{(\theta_1-\theta_0)}{\sigma}\left(\sqrt{n}\frac{\bar{X}-\theta_0}{\sigma}-\sqrt{n}\frac{\theta_1-\theta_0}{2\sigma}\right)\geq\right.$$

$$\left.\geq\ln(c_\alpha)\right)=P_{\theta_0}\left(\sqrt{n}\frac{\bar{X}-\theta_0}{\sigma}\leq-\frac{\sigma\ \ln(c_\alpha)}{\sqrt{n}(\theta_0-\theta_1)}-\frac{\sqrt{n}(\theta_0-\theta_1)}{2\sigma}\right)=$$

$$=P_{\theta_0}\left(\sqrt{n}\frac{\bar{X}-\theta_0}{\sigma}\leq-\tilde{c}_\alpha\right)=\Phi(-\tilde{c}_\alpha)=\alpha,$$

где

$$\tilde{c}_\alpha=\frac{\sigma\ \ln(c_\alpha)}{\sqrt{n}(\theta_0-\theta_1)}+\frac{\sqrt{n}(\theta_0-\theta_1)}{2\sigma}=-\Phi^{-1}(\alpha).$$

Из теоремы Неймана—Пирсона получаем критическую область наиболее мощного критерия:

$$\mathcal{X}_{1\alpha}=\left\{\vec{x}\in\mathcal{X}:\ \sum_{i=1}^{n}x_i\leq n\theta_0+\sqrt{n}\sigma\Phi^{-1}(\alpha)\right\},$$

которая, как видно, не зависит от конкретного значения параметра θ_1. Таким образом, если реализация выборки принадлежит критической области $\mathcal{X}_{1\alpha}$, то принимается гипотеза H_1, в противном случае – H_0.

Функция мощности при этом равна

$$W(\theta) = P_\theta\left(\sum_{i=1}^{n} X_i \le n\theta_0 - \sqrt{n}\sigma\Phi^{-1}(\alpha)\right) =$$

$$= P_\theta\left(\left(\sum_{i=1}^{n} X_i - n\theta\right)/\sqrt{n\sigma} \le \sqrt{n}(\theta_0 - \theta)/\sigma - \Phi^{-1}(\alpha)\right) =$$

$$= \Phi\left(\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - \Phi^{-1}(\alpha)\right).$$

Исходя из равенства $\Phi(-\gamma) = 1 - \Phi(\gamma)$, справедливого для произвольного $\gamma > 0$, найдем вероятность ошибки второго рода:

$$\beta = 1 - W(\theta_1) = \Phi\left(\Phi^{-1}(\alpha) - \frac{\sqrt{n}(\theta_0 - \theta_1)}{\sigma}\right).$$

Пример 2. Пусть выборка $\vec{X} = (X_1, \ldots, X_n) \sim \Pi(\theta)$ из распределения Пуассона с параметром $\theta > 0$. Требуется проверить простую гипотезу $H_0 : \theta = \theta_0$ против простой альтернативы $H_1 : \theta = \theta_1$, где $\theta_0 > \theta_1$.

Критерий будем строить предполагая, что уровень значимости (ошибка первого рода) не будет превышать заданного, достаточно малого уровня $\alpha > 0$.

Функция правдоподобия равна

$$L(\vec{X}, \theta) = \prod_{j=1}^{n} \frac{e^{-\theta}\theta^{X_j}}{X_j!} = e^{-n\theta}\theta^{n\bar{X}} / \prod_{j=1}^{n} X_j!\,,$$

поэтому статистика отношения правдоподобий равна

$$l(\vec{X}, \theta_0, \theta_1) = e^{n(\theta_0 - \theta_1)} \left(\frac{\theta_1}{\theta_0}\right)^{n\bar{X}}.$$

Если справедлива гипотеза H_0, то, очевидно, $n\bar{X} \sim \Pi(n\theta_0)$ и из условия $\theta_0 > \theta_1$ получим

$$P_{\theta_0}(l(\vec{X}, \theta_0, \theta_1) \geq c_\alpha) = P_{\theta_0}\left(\left(\frac{\theta_1}{\theta_0}\right)^{n\bar{X}} \geq c_\alpha e^{n(\theta_1 - \theta_0)}\right) =$$

$$= P_{\theta_0}\left(n\bar{X} \leq \log_{\frac{\theta_1}{\theta_0}}\left(c_\alpha e^{n(\theta_1-\theta_0)}\right)\right) =$$

$$= P_{\theta_0}\left(n\bar{X} \leq \tilde{c}_\alpha\right) = e^{-n\theta_0}\sum_{k=0}^{[\tilde{c}_\alpha]}\frac{n^k\theta_0^k}{k!} \leq \alpha,$$

где $\tilde{c}_\alpha = \log_{\frac{\theta_1}{\theta_0}}\left(c_\alpha e^{n(\theta_1-\theta_0)}\right)$ и $[\tilde{c}_\alpha]$ означает целую часть числа $\tilde{c}_\alpha$.

Фактически, в качестве константы $\tilde{c}_\alpha$ нужно выбрать такое наибольшее целое число, чтобы функция распределения случайной величины $n\bar{X} = \sum\limits_{j=0}^{n} X_j$ в этой точке была бы не больше заданного уровня значимости α. Таким образом, получаем, что критической областью для наиболее мощного критерия с уровнем значимости не превышающим заданное значения α в нашей задаче будет множество

$$\mathcal{X}_{1\alpha} = \left\{\vec{x} \in \mathcal{X} : \sum_{i=1}^{n} x_i \leq \tilde{c}_\alpha\right\}.$$

Ясно, что ошибка второго рода $\beta = P_{\theta_1}\left(n\bar{X} > \tilde{c}_\alpha\right) = e^{-n\theta_1}\sum\limits_{k=[\tilde{c}_\alpha]+1}^{+\infty}\frac{n^k\theta_1^k}{k!}$.

17. Сложные параметрические гипотезы

Предположим, что, как и в предыдущем разделе, требуется построить критерий проверки гипотезы против некоторой альтернативы. Отличие от рассмотренной выше задачи заключается в том, что альтернативная гипотеза является сложной гипотезой. Итак, пусть имеется выборка $\vec{X} = (X_1, \ldots, X_n)$ с функцией распределения $F_\xi(x, \theta)$, где, как и в предыдущем разделе предполагаем, что распределение либо абсолютно непрерывное либо дискретное. Пусть требуется проверить простую гипотезу $H_0 : \theta = \theta_0$ против сложной альтернативы $H_1 : \theta \in \Theta_1$, причем $\theta_0 \notin \Theta_1$.

Для того, чтобы построить критерий проверки гипотез для этой задачи нужно применить описанный выше критерий Неймана-Пирсона для всех $\theta_1 \in \Theta_1$. Если полученное критическое множество явно зависит от конкретного значения альтернативной гипотезы, то это означает, что для сложной альтернативы не существует равномерно (т.е. одновременно для всех точек альтернативной гипотезы) наиболее мощного критерия. Но есть распределения, для которых критическая область явно не использует конкретные значения параметра из альтернативной гипотезы. Так, например, в первом примере в предыдущем разделе задачу можно сформулировать следующим образом: для выборки $\vec{X} = (X_1, \ldots, X_n) \ \sim N(\theta, \sigma^2)$ проверяется гипотеза

$$H_0 : \theta = \theta_0 \text{ против альтернативы } H_1 : \theta < \theta_0.$$

Повторяя дословно вышеприведенные выкладки, получаем, что множество

$$\mathcal{X}_{1\alpha} = \left\{ \vec{x} \in \mathcal{X} : \ \sum_{i=1}^{n} x_i \le n\theta_0 + \sqrt{n}\sigma\Phi^{-1}(\alpha) \right\}$$

является критической областью равномерно наиболее мощного критерия уровня значимости α для этой задачи. Если для этой же выборки заданы гипотезы

$$H_0 : \theta = \theta_0 \text{ и } H_1 : \theta > \theta_0,$$

то несложно получить, что

$$\mathcal{X}_{1\alpha} = \left\{ \vec{x} \in \mathcal{X} : \sum_{i=1}^{n} x_i \geq n\theta_0 - \sqrt{n}\sigma\Phi^{-1}(\alpha) \right\}.$$

Это были примеры ***односторонних*** альтернатив.

В общем случае, для скалярного параметра можно сформулировать достаточное условие существования *равномерно наиболее мощного* критерия для односторонних альтернативных гипотез. Пусть альтернативная гипотеза имеет вид $H_1 : \theta < \theta_0$ либо $H_1 : \theta > \theta_0$. Предположим, что статистика $T(\vec{X})$ является *достаточной* для выборки с функцией распределения $F(x, \theta)$, $\theta \in \Theta$. Тогда, по критерию факторизации (см. выше) получаем, что функция правдоподобия равна $L(\vec{X}, \theta) = g(T(\vec{X}), \theta)h(\vec{X})$. Поэтому, статистика отношения правдоподобия равна $l(\vec{X}, \theta_0, \theta_1) = g(T(\vec{X}), \theta_1)/g(T(\vec{X}), \theta_0)$ и, следовательно, критическая область по критерию Неймана—Пирсона будет выражаться с использованием функции $T(\vec{X})$. Если статистика $l(\vec{X}, \theta_0, \theta_1)$ монотонна по $T(\vec{X})$, то в задаче проверки простой гипотезы $H_0 : \theta = \theta_0$ против простой альтернативы $H_1 : \theta = \theta_1$ критическая область, полученная критерием Неймана—Пирсона, не зависит явно от θ_1. Тогда для этого же критерия возможно простую альтернативу заменить на сложную одностороннюю альтернативу, а это означает, что этот же критерий будет *равномерно наиболее мощным* для простой нулевой гипотезы против сложной альтернативы. Итак, это возможно, если выполняется монотонность отношения правдоподобия.

Пример. Пусть имеется выборка $\vec{X} = (X_1, \ldots, X_n) \sim$ $\sim N(a, \theta^2)$ из нормального распределения с математическим ожиданием a и дисперсией $\theta^2 > 0$. Требуется проверить гипотезу $H_0 : \theta^2 = \theta_0^2$ против альтернативы $H_1 : \theta^2 > \theta_0^2$.

Ясно, что для произвольного $\theta_1 > \theta_0$ статистика отношения правдоподобий равна

$$l(\vec{X}, \theta_0, \theta_1) = \exp\left\{\left(\frac{1}{2\theta_0^2} - \frac{1}{2\theta_1^2}\right)\sum_{i=1}^{n}(X_i - a)^2 - \frac{1}{2} ln\left(\frac{\theta_1}{\theta_0}\right)\right\}.$$

Ясно, что эта статистика монотонно возрастающая по $T(\vec{X}) = \sum\limits_{i=1}^{n}(X_i - a)^2$. Наилучший критерий задается критической областью $\mathcal{X}_{1\alpha} = \left\{\vec{x} \in \mathcal{X} : T(\vec{X}) \geq c_\alpha\right\}$. Ясно, что при выполнении гипотезы H_0 будет справедливо распределение $T(\vec{X})/\theta_0^2 \sim \chi_n^2$ (хи-квадрат с n степенями свободы) и поэтому $c_\alpha = \theta_0^2\chi_{1-\alpha,n}^2$ где, как и выше, $\chi_{1-\alpha,n}^2$ обозначает квантиль с.в. хи-квадрат с n степенями свободы уровня $1 - \alpha$.

18. Элементы корреляционного анализа

При исследовании зависимости случайных величин желательно исследовать эти случайные величины на сам факт наличие такой связи, а затем измерить, насколько она сильна. Обычно это делается при помощи различных коэффициентов корреляции.

18.1. Простой коэффициент корреляции

Этот коэффициент применяется для измерения величины линейной связи между двумя случайными величинами.

Определение 1. *Простым коэффициентом корреляции двух случайных величин* ξ_1 *и* ξ_2 *называют число* $\rho(\xi_1, \xi_2) = cov(\xi_1, \xi_2)/\sqrt{D\xi_1 D\xi_2}$.

Из курса теории вероятностей известно, что для простого коэффициента корреляции выполняются следующие свойства.

1) $|\rho(\xi_1, \xi_2)| \leq 1$.

2) Если ξ_1 *и* ξ_2 *некоррелированы, то* $\rho(\xi_1, \xi_2) = 0$.

3) $\rho(a_1\xi_1 + b_1,\ a_2\xi_2 + b_2) = sign(a_1 a_2)\ \rho(\xi_1, \xi_2)$.

4) Если $\rho(\xi_1, \xi_2) = +1$*, то* $\xi_2 = a\xi_1 + b$*, где* $a > 0,\ -\infty < b < +\infty$ *и,*

если $\rho(\xi_1, \xi_2) = -1$*, то* $\xi_2 = a\xi_1 + b$*, где* $a < 0,\ -\infty < b < +\infty$*.*

Определение 2. *Если* $\rho(\xi_1, \xi_2) > 0$*, то случайные величины* ξ_1 *и* ξ_2 *называются положительно коррелированы, а в случае* $\rho(\xi_1, \xi_2) < 0$ *– отрицательно коррелированы.*

Пусть имеются две случайные величины X и Y и необходимо узнать поведение случайной величины Y с помощью случайной величины X. Для этого определим наилучшее для Y линейное приближение $\hat{Y} = \gamma + \beta X$ и

$e = Y - \hat{Y}$. Для поиска коэффициентов γ и β воспользуемся равенствами $E(e \cdot 1) = EY - \gamma - \beta EX = 0$ и $E(e \cdot X) = EYX - \gamma EX - \beta EX^2 = 0$. Решая эти уравнения, получим

$$\beta = \frac{cov(X, Y)}{DX} = \rho \frac{\sigma_Y}{\sigma_X},$$

$$\gamma = EY - \rho EX \frac{\sigma_Y}{\sigma_X},$$

где $\rho = \rho(X, Y)$. Таким образом, $\hat{Y} - EY = (X - EX)\frac{\sigma_Y}{\sigma_X}\rho$, откуда, воспользовавшись свойствами дисперсий, получаем $D\hat{Y} = \rho^2 DY$. Далее, в силу некоррелированности случайных величин $\hat{Y}$ и e, получаем, что $DY = D\hat{Y} + De$. Поэтому $De = (1 - \rho^2)DY$. Таким образом, получаем, что *величина ρ^2 показывает, какая часть дисперсии задается линейным влиянием случайной величины X*. Аналогично, *величина $1 - \rho^2$ показывает какая часть дисперсии случайной величины Y не определяется линейным влиянием случайной величины X и требуется применение других факторов*.

Для того, чтобы искать линейную аппроксимацию Y через X нужно предварительно убедиться, что $\rho(X, Y) \neq 0$, т.е. X и Y коррелированы. Чтобы было возможно проверить это предположение, дополнительно потребуем, чтобы вектор (X, Y) имел нормальное распределение (как известно из курса теории вероятностей, двумерное нормальное распределение явно зависит от коэффициента корреляции). Пусть имеется двумерная нормальная выборка $(X_1, Y_1), \ldots, (X_n, Y_n)$ объема n (здесь (X_i, Y_i)—независимые копии вектора (X, Y)). Проверяем гипотезу $H_0 : \rho(X, Y) = 0$ против альтернативной гипотезы $H_1 : \rho(X, Y) \neq 0$. Для проверки гипотезы H_0 естественно использовать выборочный коэффициент корреляции

$$R = R(\vec{X}, \vec{Y}) = \frac{S_{XY}}{S_X S_Y},$$

где $S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$—выборочная ковариация, $S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ и $S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ - выборочные дисперсии для X и Y соответственно.

Определение 3. *Случайеая величина* $\xi/\sqrt{\eta_1^2 + \cdots + \eta_m^2}$ *называется случайной величиной Стьюдента с* m *степенями свободы, если* $\xi, \eta_1, \ldots, \eta_m$ *независимые одинаково распределенные нормальные стандартные случайные величины.*

Определим величину $t_{n-1} = R\sqrt{n-1}/\sqrt{1-R^2}$. Справедливо следующее утверждение.

Теорема. *В случае справедливости гипотезы* H_0 *случайная величина* t_{n-1} *имеет распределение Стьюдента с* $n-1$ *степенью свободы* (см. дополнение).

Эту теорему принемаем без доказательства.

Далее, как обычно, по заданному значению α выбираем из таблицы распределения Стьюдента константу $t_{n-1}(\alpha)$ так, чтобы $P(|t_{n-1}| > t_{n-1}(\alpha)) = \alpha$. Тогда, если для экспериментально полученных данных значение $|t_{n-1}|$ будет больше числа $t_{n-1}(\alpha)$, то гипотеза H_0 неверна и случайные величины рассматриваются как зависимые. В противном случае говорят, что *экспериментальные данные не показывают коррелированности* X *и* Y.

Если нормальность распределения случайного вектора (X, Y) не может быть гарантирована, то можно воспользоваться тем, что случайная величина t_{n-1} асимптотически нормальна в случае независимости X и Y, т.е. при больших значениях объема наблюдений n можно считать t_{n-1} нормальной.

Если было выяснено, что корреляция есть, то желательно найти величину коэффициента корреляции, например построить доверительный интервал. Р.Фишер показал, что случайная величина $Z = \frac{1}{2}\ln\left(\frac{1+R}{1-R}\right)$ даже при небольших значениях n имеет приближенно нормальное распределение с математическим ожиданием $\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right) + \frac{\rho}{2(n-1)}$ и дисперсией $\frac{1}{n-3}$. Это позволяет проверять гипотезы и строить доверительные интервалы для ρ.

18.2. Множественный коэффициент корреляции

Множественный коэффициент корреляции применяется для измерения величины линейной связи одной случайной величины Y от нескольких случайных величин $X_1, \ldots, X_m$. Как и выше, ищем наилучшее линейное приближение $\hat{Y} = \gamma + \beta_1 X_1 + \cdots + \beta_m X_m$ для случайной величины Y при помощи случайных величин $X_1, \ldots, X_m$.

Определение. *Множественным коэффициентом корреляции случайной величины Y с набором случайных величин $X_1, \ldots, X_m$ называется величина*

$$\rho_{Y \cdot X_1, \ldots, X_m} = \rho(Y, \hat{Y}),$$

где $\hat{Y}$ есть наилучшее линейное приближение Y при помощи случайных величин $X_1, \ldots, X_m$.

Положим $e = Y - \hat{Y}$. В силу некоррелированности e и $\hat{Y}$ получаем, что $DY = D\hat{Y} + De$. Поскольку $cov(e, \hat{Y}) = 0$, то пользуясь свойствами ковариации, получаем, что $cov(Y, \hat{Y}) = D\hat{Y}$ и $\rho(Y, \hat{Y}) = \frac{D\hat{Y}}{\sqrt{DY D\hat{Y}}}$ или $\rho^2(Y, \hat{Y}) = \frac{D\hat{Y}}{DY}$. Таким образом, квадрат множественного коэффициента корреляции показывает, какую часть дисперсии Y можно объяснить совокупным линейным влиянием случайных величин $X_1, \ldots, X_m$. Аналогично, $1 - \rho^2(Y, \hat{Y})$ показывает, какую часть дисперсии Y нельзя объяснить влиянием случайных величин $X_1, \ldots, X_m$.

Некоторые свойства множественного коэффициента корреляции приводятся в следующем утверждении.

Теорема.

j) $|\rho_{Y \cdot X_1, \ldots, X_m}| \leq 1$*;*

jj) $|\rho_{Y \cdot X_1, \ldots, X_m}|$ *равен максимальному по модулю простому коэффициенту корреляции между Y и $\tilde{Y} = c_0 + c_1 X_1 +$*

$+\dots+c_m X_m$ *и достигается на случайной величине* $\hat{Y}$*, которая является наилучшим линейным приближением;*

jjj) если $\rho_{Y\cdot X_1,\dots,X_m}=0$*, то* $\hat{Y}=EY$*; если* $\rho_{Y\cdot X_1,\dots,X_m}=1$*, то* $\hat{Y}=Y$*.*

Свойство *j)*, очевидно, следует из определения. Свойства *jj)* и *jjj)* оставим без доказательства.

18.3. Частный коэффициент корреляции

Часто в практических задачах возникает необходимость установить величину влияния на Y одного из факторов X_j, когда устранено влияние всех остальных факторов. Для этого используется частный коэффициент корреляции, который измеряет силу линейной связи между двумя переменными, когда устранено влияние других.

Пусть имеются случайные величины $Y, X_1, \ldots, X_m$. Выберем некоторый фактор X_j и пусть $\mathbb{C}$ обозначает все остальные факторы.

Пусть $Y^{\mathbb{C}}$ и $X_j^{\mathbb{C}}$ обозначают наилучшие линейные приближения с помощью случайных величин из набора $\mathbb{C}$ и положим $Z_Y = Y - Y^{\mathbb{C}}$ и $Z_{X_j} = X_j - X_j^{\mathbb{C}}$.

Определение. *Частным коэффициентом корреляции случайных величин Y и X_j называется парный коэффициент корреляции случайных величин Z_Y и Z_{X_j}:*

$$\rho_{YX_j\cdot\mathbb{C}} = \rho(Z_Y, Z_{X_j}).$$

Некоторые свойства частного коэффициента корреляции приведены в следующем утверждении.

Теорема

j) $|\rho_{YX_j\cdot\mathbb{C}}| \le 1$*;*

jj) $1 - \rho^2_{Y\cdot X_1,\ldots X_j} = (1 - \rho^2_{Y\cdot X_1,\ldots X_{j-1}})(1 - \rho^2_{YX_j\cdot X_1,\ldots X_{j-1}})$*;*

jjj) пусть $\hat{Y} = \gamma + \beta_1 X_1 + \cdots + \beta_m X_m$ есть наилучшее линейное приближение Y с помощью $X_1, \ldots, X_m$. Тогда $\beta_j = \rho_{YX_j\cdot\mathbb{C}}\, \frac{\sigma_{Y\cdot\mathbb{C}}}{\sigma_{X_j\cdot\mathbb{C}}}$, где $\sigma^2_{Y\cdot\mathbb{C}}$ и $\sigma^2_{X_j\cdot\mathbb{C}}$—условные дисперсии случайных величин X_j и Y, когда устранено влияние случайных величин из набора $\mathbb{C}$;

jv) если X_i и X_j некоррелированы с набором $\mathbb{C}$, то $\rho_{X_iX_j\cdot\mathbb{C}} = \rho_{X_iX_j}$.

Свойство *j)* очевидно. Остальные три свойства оставляем без доказательства.

Кратко обсудим вопрос о более эффективном вычислении множественного и частного коэффициентов корреляции.

Назовем парные коэффициенты корреляции ρ_{ij} *коэффициентами корреляции нулевого порядка*. Выделим номера i, j, k и пусть $\mathbb{C}$ набор из случайных величин всех оставшихся номеров. Можно показать, что коэффициенты корреляции первого порядка

$$\rho_{ij\cdot k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)(1-\rho_{jk}^2)}}$$

и

$$\rho_{ij\cdot k,\mathbb{C}} = \frac{\rho_{ij\cdot\mathbb{C}} - \rho_{ik\cdot\mathbb{C}}\rho_{jk\cdot\mathbb{C}}}{\sqrt{(1-\rho_{ik\cdot\mathbb{C}}^2)(1-\rho_{jk\cdot\mathbb{C}}^2)}}\,.$$

Далее будем использовать парные коэффициенты корреляции между различными случайными величинами $Y, X_1, \ldots, X_m$ и обозначим $CORR = \|\rho_{ij}\|$ матрицу парных коэффициентов корреляций. Все диагональные элементы этой матрицы $\rho_{ii} = 1,\ i = 0, 1, \ldots, m$ (здесь предполагается $X_0 = Y$).

Еще один метод вычисления множественного и частного коэффициентов корреляции, использует определители матриц. Пусть $CORR_{ij}$ обозначает алгебраическое дополнение к ρ_{ij} в матрице $CORR$. Тогда

$$\rho_{Y\cdot X_1,\ldots,X_m} = 1 - \frac{|CORR|}{CORR_{00}}$$

и

$$\rho_{ij\cdot\mathbb{C}} = -\frac{CORR_{ij}}{\sqrt{CORR_{ii}\,CORR_{jj}}}\,.$$

В практических задачах теоретические коэффициенты корреляции неизвестны и заменяются на выборочные аналоги.

18.4. Пример использования коэффициентов корреляции

По итогам года на нескольких профильных предприятиях регистрировались показатели их работы:

$\boldsymbol{Y} = \boldsymbol{X_0}$ – среднемесячная характеристика качества продукции;

$\boldsymbol{X_1}$ – среднемесячное число профилактик оборудования;

$\boldsymbol{X_2}$ – среднемесячное число краткосрочных остановок производственного оборудования.

По накопленным данным за определенный период были найдены выборочные коэффициенты корреляции: $\boldsymbol{R_{01} = 0,105}$, $\boldsymbol{R_{02} = -0,024}$, $\boldsymbol{R_{12} = -0,996}$. Эти данные показывают слабую связь между качеством продукции и числом профилактических наладок и числом краткосрочных остановок, что противоречит практическому опыту работы предприятий. Но если вычислить частные коэффициенты корреляции $\boldsymbol{R_{01\cdot 2} = 0,907}$, $\boldsymbol{R_{02\cdot 1} = -0,906}$, то мы видим сильную корреляционную связь. Полученные результаты обусловлены тем, что факторы $\boldsymbol{X_1}$ и $\boldsymbol{X_2}$ сильно отрицательно коррелированны и их взаимное влияние маскирует вклад каждого из факторов. Общий вклад факторов является значительным, что показывает величина множественного коэффициента корреляции $\boldsymbol{R^2_{Y\cdot X_1X_2} = 0,823}$.

19. Элементы регрессионного анализа

Выше, когда рассматривались задачи в терминах *выборки* всегда предполагалось, что статистические данные являются реализацией случайного вектора $\vec{X} = (X_1, \ldots, X_n)$, где компонентами являются независимые и одинаково распределенные случайные величины. Однако, в некоторых задачах математической статистики это предположение не выполняется, в частности, в задачах *линейной регрессии*, рассматриваемых в этом разделе.

19.1. Метод наименьших квадратов

Пусть на плоскости в декартовой прямоугольной системе координат Oxy заданы n точек $(x_1, y_1), \ldots, (x_n, y_n)$. Требуется провести прямую линию так, чтобы эти точки, по отношению к ней, располагались «наилучшим» образом. Понятие «наилучший» здесь, конечно, требует разъяснений. В задачах *линейной регрессии* под этим понимают минимум некоторого выражения, характеризующего близость расположения точек к искомой прямой.

Введем понятие невязок $e_j = y_j - Ax_j - B, \quad j = 1, \ldots, n$; $y = Ax + B$ – уравнение искомой прямой. Выражение, минимум которого ищется, может быть задано по-разному. Рассмотрим три основных варианта, наиболее часто используемых на практике:

i) сумма квадратов невязок $U(e_1, \ldots, e_n) = \sum\limits_{i=1}^{n} e_i^2 = \sum\limits_{i=1}^{n} (y_i - Ax_i - B)^2$;

ii) сумма модулей невязок $U(e_1, \ldots, e_n) = \sum\limits_{i=1}^{n} |e_i| = \sum\limits_{i=1}^{n} |y_i - Ax_i - B|$;

iii) Функция Хьюбера $U(e_1, \ldots, e_n) = \sum\limits_{i=1}^{n} \{e_i^2 I_{[0,c]}(|e_i|) + (2c|e_i| - c^2) I_{(c,+\infty]}(|e_i|)\}$.

Из этих трех вариантов наиболее часто используется вариант i), известный как *метод наименьших квадратов (МНК)*. Остановимся подробнее на этом методе. Сумма квадратов невязок в нашем случае зависит от двух параметров A и B, что обозначим явно $U = U(A, B)$. Запишем необходимые условия экстремума функции двух переменных:

$$\frac{\partial U(A,B)}{\partial A} = -2\sum_{i=1}^{n}(y_i - Ax_i - B)x_i = 0,$$

$$\frac{\partial U(A,B)}{\partial B} = -2\sum_{i=1}^{n}(y_i - Ax_i - B) = 0,$$

откуда получаем систему уравнений

$$\begin{cases} A\sum\limits_{i=1}^{n} x_i + nB = \sum\limits_{i=1}^{n} y_i \\ A\sum\limits_{i=1}^{n} x_i^2 + B\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i \ . \end{cases} \tag{26}$$

Решая эту систему, получаем $A = S_{xy}/S_x^2$, $B = \overline{y} - \overline{x} S_{xy}/S_x^2$, где

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

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y}; \ S_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2.$$

Докажем, что найденные значения A и B доставляет минимум функции $U(A, B)$. Положим

$$a = \frac{\partial^2 U(A,B)}{\partial A^2} = 2\sum_{i=1}^{n} x_i^2 > 0, \quad b = \frac{\partial^2 U(A,B)}{\partial A \partial B} = 2\sum_{i=1}^{n} x_i,$$

$$c = \frac{\partial^2 U(A,B)}{\partial B^2} = 2n.$$

Далее, $d = ac - b^2 = 4\left(n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2\right)$. Покажем, что $d > 0$. Тем самым будет установлен факт минимума. Воспользуемся индукцией по числу n. По смыслу $n \geq 2$. При $n = 2$, имеем $2(x_1^2 + x_2^2) - (x_1 + x_2)^2 = (x_1 - x_2)^2 > 0$, так как $x_1 \neq x_2$.

Допустим, что при $n \geq 2$ выполняется неравенство $n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2 > 0$. Заменим n на $n+1$. В результате получим:
$(n+1)\sum\limits_{i=1}^{n+1} x_i^2 - \left(\sum\limits_{i=1}^{n+1} x_i\right)^2 = (n+1)\left(\sum\limits_{i=1}^{n} x_i^2 + x_{n+1}^2\right) - \left(\sum\limits_{i=1}^{n} x_i + x_{n+1}\right)^2 = n\sum\limits_{i=1}^{n} x_i^2 + \sum\limits_{i=1}^{n} x_i^2 + (n+1)x_{n+1}^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2 - 2x_{n+1}\sum\limits_{i=1}^{n} x_i - x_{n+1}^2$.

Выражение $n\sum\limits_{i=1}^{n} x_i^2 - \left(\sum\limits_{i=1}^{n} x_i\right)^2 > 0$ согласно индуктивному предположению. Для оставшихся слагаемых, имеем

$$\sum_{i=1}^{n} x_i^2 + nx_{n+1}^2 - 2x_{n+1}\sum_{i=1}^{n} x_i = \left(x_1^2 + x_{n+1}^2 - 2x_1 x_{n+1}\right) + \left(x_2^2 + x_{n+1}^2 - 2x_2 x_{n+1}\right) + \ldots \left(x_n^2 + x_{n+1}^2 - 2x_n x_{n+1}\right) > 0,$$

так как в каждой скобке выражение положительное. Таким образом, установлен достаточный признак минимума.

19.2. Модель линейной регрессии

Предположим, что $x_1, \dots, x_n$ некоторые заданные неслучайные числа. Пусть $Y_i = \alpha + \beta x_i + \varepsilon_i$, где α и β - некоторые постоянные, а $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$ - независимые нормальные одинаково распределенные случайные величины. Отсюда следует, что $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$, где α, β и σ^2 три параметра, которые необходимо будет оценивать. Для того, чтобы оценить параметры α и β, воспользуемся МНК для минимизации выражения

$$U = \sum_{i=1}^{n} (Y_i - \alpha - \beta x_i)^2.$$

Решение этой задачи, как показано выше, дает

$$\hat{\alpha} = \overline{Y} - \overline{x} S_{xY} / S_x^2, \quad \hat{\beta} = S_{xY} / S_x^2,$$

где

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i, \; S_{xY} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y}),$$

$$S_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Утверждение 1. *Оценки МНК $\hat{\alpha}$ и $\hat{\beta}$ являются несмещенными оценками параметров α и β соответственно.*

Доказательство. Положим $z_j = (x_j - \bar{x}) / \sum\limits_{i=1}^{n} (x_i - \bar{x})^2$. Ясно, что

$$\sum_{j=1}^{n} z_j = \frac{1}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \sum_{i=1}^{n} (x_i - \bar{x}) =$$

$$= \frac{1}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\left(\sum_{i=1}^{n} x_i - n\bar{x}\right) = 0 \qquad (27)$$

и

$$\sum_{j=1}^{n} z_j^2 = \left(\frac{1}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right)^2 \sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\,. \qquad (28)$$

Поэтому,

$$\hat{\beta} = \sum_{i=1}^{n} z_i(Y_i - \overline{Y}) = \sum_{i=1}^{n} z_i Y_i = \sum_{i=1}^{n} z_i(\alpha + \beta x_i + \varepsilon_i) =$$

$$= \beta \sum_{i=1}^{n} z_i x_i + \sum_{i=1}^{n} z_i \varepsilon_i = \beta \sum_{i=1}^{n} z_i(x_i - \bar{x}) + \sum_{i=1}^{n} z_i \varepsilon_i = \beta + \sum_{i=1}^{n} z_i \varepsilon_i$$

и, аналогично,

$$\hat{\alpha} = \overline{Y} - \overline{x} S_{xY}/S_x^2 = \sum_{i=1}^{n}\left(\frac{1}{n} - z_i\overline{x}\right) Y_i =$$

$$= \sum_{i=1}^{n}\left(\frac{1}{n} - z_i\overline{x}\right)(\alpha + \beta x_i + \varepsilon_i) = \alpha + \sum_{i=1}^{n}\left(\frac{1}{n} - z_i\overline{x}\right)\varepsilon_i.$$

Отсюда, используя свойства математических ожиданий, получаем:

$$E\hat{\alpha} = \alpha + \sum_{i=1}^{n}\left(\frac{1}{n} - z_i\overline{x}\right) E\varepsilon_i = \alpha \quad \text{и} \quad E\hat{\beta} = \beta + \sum_{i=1}^{n} z_i E\varepsilon_i = \beta,$$

поскольку α, β, $\bar{x}$ и z_i—постоянные и $E\varepsilon_i = 0$.

Утверждение 2. *Случайный вектор* $(\hat{\alpha}, \hat{\beta})$ *имеет кова-*

риационную матрицу, равную

$$\frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}\begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix},$$

где $\overline{x^2}=\frac{1}{n}\sum\limits_{i=1}^{n}x_i^2$.

Доказательство. Используя свойства дисперсий, аналогично тому, как это сделано в утверждении 1, вычислим элементы ковариационной матрицы:

$$\begin{aligned} &cov(\hat{\alpha},\hat{\alpha}) = D\hat{\alpha} = \sum_{i=1}^{n}\left(\tfrac{1}{n}-z_i\overline{x}\right)^2\sigma^2 == \\ &= \sigma^2\left(\tfrac{1}{n}-2\tfrac{1}{n}\bar{x}\sum_{i=1}^{n}z_i+\bar{x}^2\sum_{i=1}^{n}z_i^2\right) = \\ &= \sigma^2\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2+n\bar{x}^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sigma^2\frac{\frac{1}{n}\sum_{i=1}^{n}x_i^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sigma^2\frac{\overline{x^2}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}, \end{aligned}$$

$$cov(\hat{\beta},\hat{\beta}) = D\hat{\beta} = \sum_{i=1}^{n}z_i^2\sigma^2 = \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}$$

и

$$\begin{aligned} &cov(\hat{\alpha},\hat{\beta}) = cov(\hat{\beta},\hat{\alpha}) = \sum_{i=1}^{n}z_i\left(\tfrac{1}{n}-z_i\overline{x}\right)\sigma^2 = \\ &= \sigma^2\bar{x}\sum_{i=1}^{n}z_i^2 = -\sigma^2\frac{\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\ . \end{aligned}$$

Утверждение 3. *Пусть* $\frac{\overline{x^2}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\longrightarrow 0$ *и* $\sum\limits_{i=1}^{n}(x_i-\bar{x})^2\longrightarrow$ $\longrightarrow\infty$ *при* $n\longrightarrow\infty$. *Тогда оценки МНК* $\hat{\alpha}$ *и* $\hat{\beta}$ *являются состоятельными оценками параметров* α *и* β *соответственно.*

Доказательство. Из *утверждения 1* следует, что $E\hat{\alpha}=\alpha$ и $E\hat{\beta}=\beta$. Из предположений доказываемого утверждения и *утверждения 2* следует, что $\boldsymbol{D}\hat{\boldsymbol{\alpha}} \longrightarrow 0$ и $\boldsymbol{E}\hat{\boldsymbol{\beta}} \longrightarrow 0$ при $n \longrightarrow \infty$, откуда и следует справедливость утверждения.

Определение. *Оценка $\hat{\boldsymbol{\gamma}}$ для параметра $\boldsymbol{\gamma}$ называется оптимальной в среднем квадратическом в некотором классе оценок $\boldsymbol{\Gamma}$, если для любой другой оценки $\tilde{\boldsymbol{\gamma}} \in \boldsymbol{\Gamma}$ из этого же класса имеет место неравенство*

$$\boldsymbol{E}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^2 \leq \boldsymbol{E}(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma})^2 \, .$$

Теорама *(Гаусса—Маркова). Оценки МНК $\hat{\boldsymbol{\alpha}}$ и $\hat{\boldsymbol{\beta}}$ являются оптимальными в среднем квадратическом в классе всех линейных несмещенных оценок параметров $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$.*

Эту теоремы оставим без доказательства.

19.3. Доверительные интервалы для модели линейной регрессии

Предположим, как и выше, $Y_i = \alpha + \beta x_i + \varepsilon_i$, где $x_1, \dots, x_n$ некоторые заданные неслучайные числа, α и β—некоторые постоянные, а $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$—независимые нормальные одинаково распределенные случайные величины.

Для построения доверительных интервалов для параметров α и β недостаточно утверждений, полученных ранее. В этом разделе будут рассмотрены дополнительные утверждения, использующие заданные распределения остатков ε_i.

Утверждение 1. *Вектор* $(\hat{\alpha}, \hat{\beta})$ *оценок МНК параметров* (α, β) *имеет двумерное нормальное распределение со средним* (α, β) *и ковариационной матрицей*

$$\frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \begin{pmatrix} \overline{x^2} & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}.$$

Доказательство. Вектор $\vec{Y} = (Y_1, \dots, Y_n)$ имеет нормальное распределение с независимыми компонентами, а компоненты вектора $(\hat{\alpha}, \hat{\beta})$, являющиеся линейными функциями от вектора Y, как известно из курса теории вероятностей, тоже имеют нормальное распределение. Поэтому, дальнейшее доказательство непосредственно следует из утверждений 1 и 2 предыдущего раздела *ii)*.

Утверждение 2. *Векторы* $(\hat{\alpha}, \hat{\beta})$ *и* $\vec{e} = (e_1, \dots, e_n)$*, где* $e_i = \hat{Y}_i - Y_i$ *независимы.*

Доказательство. Поскольку векторы $(\hat{\alpha}, \hat{\beta})$ и $\vec{e}$ имеют нормальные распределения, то, как известно из курса теории вероятностей, достаточно показать покомпонентную некоррелированность этих векторов. Покажем это на примере случайных величин $\hat{\beta}$ и Y_k.

В доказательстве утверждения 1 предыдущего раздела *ii* показано, что $\hat{\beta} = \beta + \sum\limits_{i=1}^{n} z_i\varepsilon_i$. В силу того, что оценки МНК $\hat{\alpha}$ и $\hat{\beta}$ удовлетворяют системе уравнений (26), имеем $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$. Поэтому

$$e_k = Y_k - \hat{Y}_k = \alpha + \beta x_k + \varepsilon_k - \hat{\alpha} - \hat{\beta}x_k = \alpha + \beta x_k + \varepsilon_k - \bar{Y} + \hat{\beta}\bar{x} - \hat{\beta}x_k =$$

$$= \alpha + \beta x_k + \varepsilon_k - \alpha - \beta\bar{x} - \bar{\varepsilon} + \hat{\beta}\bar{x} - \hat{\beta}x_k = (\beta - \hat{\beta})(x_k - \bar{x}) + (\varepsilon_k - \bar{\varepsilon}) =$$

$$= -\sum_{i=1}^{n} z_i\varepsilon_i(x_k - \bar{x}) + (\varepsilon_k - \bar{\varepsilon}).$$

Отсюда, очевидно, следует, что $Ee_k = 0$ и, используя (27) и (28), получаем

$$cov(\hat{\beta}, e_k) = -\sum_{i=1}^{n}\sum_{j=1}^{n} z_i z_j \; cov(\varepsilon_i, \varepsilon_j)(x_k - \bar{x}) +$$

$$+ \sum_{i=1}^{n} z_i \; cov(\varepsilon_i, \varepsilon_k - \bar{\varepsilon}) =$$

$$- \sum_{i=1}^{n} z_i^2 \sigma^2 (x_k - \bar{x}) - \frac{1}{n}\sum_{i=1}^{n} z_i\sigma^2 = -z_k\sigma^2 + z_k\sigma^2 = 0.$$

Утверждение 3. *Случайная величина* $(n-2)S^2/\sigma^2 = \sum\limits_{i=1}^{n} e_i^2/\sigma^2$ *имеет распределение* χ^2 *с (n-2) степенями свободы.*

Это утверждение оставим без доказательства.

Теперь можем заняться построением доверительных интервалов для параметров α и β. Из утверждения 1 имеем, что оценка $\hat{\beta}$ имеет нормальное распределение с математическим ожиданием β и дисперсией $\sigma^2/\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$. Из утверждений 1 и 2 следует, что случайная величина $\frac{\hat{\beta}-\beta}{\sigma}\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}/(S/\sigma)$ имеет распределение Стьюдента с $n-2$ степенями свободы. (Напомним, что если случайные величины $\xi\sim N(0,1)$ и $\eta\sim\chi^2_m$ независимы, то случайная величина $\xi/\sqrt{\eta}$ имеет распределение Стьюдента с m степенями свободы.) Аналогично, можно показать, что случайная величина $(\hat{\alpha}-\alpha)\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}/\left(S\sqrt{\overline{x^2}}\right)$ имеет распределение Стьюдента с $n-2$ степенями свободы.

Используя эти факты, можно доказать справедливость следующего утверждения.

Теорема. *Доверительные интервалы уровня значимости γ для параметров α и β имеют вид*

$$\hat{\alpha}-\sqrt{\overline{x^2}}t_{n-2}(\gamma)\ \tilde{S}<\alpha<\hat{\alpha}+\sqrt{\overline{x^2}}t_{n-2}(\gamma)\ \tilde{S},$$

$$\hat{\beta}-t_{n-2}(\gamma)\ \tilde{S}<\beta<\hat{\beta}+t_{n-2}(\gamma)\ \tilde{S},$$

где константа $t_{n-2}(\gamma)$ выбирается из таблицы квантилей распределения Стьюдента с $n-2$ степенями свободы уровня γ и $\tilde{S}=S/\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2}$.

Контрольные вопросы

1. Что представляет из себя конкретное статистическое наблюдение?
2. Что такое статистика?
3. Что называется вариационным рядом.
4. Что такое порядковая статистика?
5. Какая статистика называется несмещенной?
6. Какая статистика называется состоятельной?
7. Является ли выборочная функция распределения несмещенной и состоятельной оценкой?
8. Что называется функцией правдоподобия?
9. Что такое доверительный интервал?
10. Что называется статистической гипотезой?
11. Что такое ошибки первого и второго рода?
12. Что такое мощность критерия?
13. Что такое критическая область?
14. Что такое непараметрические и параметрические гипотезы?

Список рекомендуемой литературы

1. **Большев, Л.Н.** Таблицы математической статистики./Л.Н. Большев, Н.В. Смирнов – М., 1983.

2. **Ивченко** Г.И. Введение в математическую статистику./Г.И.Ивченко, Ю.И. Медведев – М., 2010.

3.**Ивченко** Г.И. Теория вероятностей и математическая статистика в задачах./Г.И.Ивченко, Ю.И. Медведев – М., 2005.

4. **Прохоров** Ю.В., Лекции по теории вероятностей и математической статистике./Ю.В. Прохоров, Л.С. Пономаренко – М., 2004.