МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

Ю.В. СОЛОДЯННИКОВ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Рекомендовано редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева» в качестве учебного пособия для обучающихся по основной образовательной программе высшего образования по специальности 01.05.01. Фундаментальная математика и механика

САМАРА
Издательство Самарского университета
2023

УДК 519.2(075)
ББК В172я7
С604

Рецензенты: д-р физ.-мат. наук, проф. А.Н. Степанов,
канд. физ.-мат. наук, доц. Н.Д. Голубева

Солодянников, Юрий Васильевич
С604 **Математическая статистика**: учебное пособие / *Ю.В. Солодянников*. – Самара: Издательство Самарского университета, 2023. – 152 с.

ISBN 978-5-7883-1873-8

В пособии излагаются основные результаты теории выборочного метода, методы получения оценок, задачи статистической проверки гипотез, некоторые положения теории статистических решающих правил и оптимальных выводов. Основные положения иллюстрируются рядом примеров и задач.

Пособие предназначено для изучения математической статистики в университетах, пединститутах, а также в технических вузах с повышенной математической подготовкой. Может быть полезно инженерам, аспирантам и научным работникам различных специальностей.

Подготовлено на кафедре функционального анализа и теории функций.

УДК 519.2(075)
ББК В172я7

ISBN 978-5-7883-1873-8

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Всякая наука изучает некоторое конечное число связей между явлениями окружающей нас действительности и устанавливает свойственные этим связям закономерности. Знание закономерностей позволяет предвидеть течение явлений и выбирать рациональные пути поведения в типичных ситуациях.

В основе научного познания лежат наблюдения и факты, которые всегда требуют глубокой обобщающей обработки. Известно, что предметом теории вероятностей является изучение закономерностей массовых случайных явлений, т.е. таких явлений, которые могут многократно наблюдаться при осуществлении одних и тех же условий, и результат наблюдения которых может меняться от опыта к опыту.

К теории вероятностей непосредственно примыкает математическая статистика, задача которой состоит в создании методов сбора и обработки статистических данных (результатов наблюдений) для получения научных и практических выводов. Можно сказать, что математическая статистика является наукой о методах количественного анализа случайных явлений, учитывающей качественное своеобразие этих явлений.

Большой вклад в математическую статистику внесли такие выдающиеся ученые-математики, как П.Л. Чебышев, А.А. Марков, А.М. Ляпунов, А.Н. Колмогоров, Ю.В. Линник, Н.В. Смирнов, К. Гаусс, К. Пирсон, Г. Крамер. Особенно бурное развитие получила математическая статистика в последние годы в связи с запросами практики и совершенствованием основного инструмента использования ее результатов – электронных вычислительных машин.

В настоящее время математическую статистику называют «методом века» – настолько широкое приложение она находит в самых различных областях человеческой деятельности.

Цель данного пособия состоит в систематическом и, по возможности, кратком и доступном изложении ряда разделов основ математической и прикладной статистики студентам, специализирующимся по чистой и прикладной математике. Представленный здесь материал читался автором в качестве спецкурса в 1974–2000 г.г. студентам старших курсов механико-математического факультета Куйбышевского (Самарского) государственного университета.

Первая часть пособия содержит основные положения теории выборочного метода, статистической проверки гипотез, знакомит читателей с методами оценивания параметров, непараметрической статистикой и теорией статистических решений. В приложении рассмотрены применения непараметрического оценивания плотностей и функций распределения, свертки и компонент свертки распределений, статистики (непараметрической) систем массового обслуживания.

Для усвоения изложенного в пособии материала достаточно подготовки, например, по учебнику Коваленко И.Н., Филипповой А.А. «Теория вероятностей и математическая статистика» [7].

Данное учебное пособие «Математическая статистика» предназначено для студентов 4-5 курсов механико-математического факультета.

Глава 1. НЕКОТОРЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1.1. События. Вероятностное пространство. Случайная величина. Закон распределения. Моменты случайной величины

Исчисление вероятностей по аксиоматике Колмогорова строится на исходном пространстве Ω элементов ω (элементарных событий), σ – алгебре *B* подмножеств (событий) пространства Ω и вероятностной мере P на *B*. При этом тройка <Ω, *B* , P > называется вероятностным пространством.

Таким образом, в рамках σ – алгебры к событиям применимы в конечном либо счетном числе теоретико-множественные операции, каждому событию сопоставлено определенное число между 0 и 1 – вероятность осуществления данного события. При этом для событий в <Ω, *B* , P > справедливы следующие формулы (свойства):

а) формула сложения

$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B),$$

где A и B – события;

б) формула (определение) условной вероятности события A при условии наступления события B ($A \cap B$ записывают иначе через AB)

$$P(A / B) = \frac{P(AB)}{P(B)}, \quad P(B) \neq 0;$$

в) формула умножения $P(A) = P(A/B)P(B) = P(B/A)P(A)$

г) формула полной вероятности $P(A) = \sum_i P(B_i)P(A/B_i)$,

где $\bigcup_i B_i \supset A$, причем события B_1 , B_2, ... несовместны, т.е. $\bigcap_i B_i = \varnothing$;

д) формула Байеса для несовместных событий $B_1, B_2, \ldots$:

$$\mathrm{P}\,(B_i/A) = \frac{P(B_i)P(A/B_i)}{\sum\limits_j P(B_j)P(A/B_j)};$$

е) если $\{A_i\}$ есть монотонно убывающая последовательность множеств $A_i \supset A_{i+1}$ и $A = \bigcap\limits_i A_i$, то $P(A) = \lim\limits_{i\to\infty} P(A_i)$;

ж) пусть для $\forall\, i \;\; i = \overline{1,n}$, $P(A_i) = 1$, тогда $P(\bigcap\limits_{i=1}^{n} A_i) = 1$ [5, с. 349];

з) пусть для $\forall\, i$, $i = \overline{1,\infty}$, $P(A_i) = 1$, тогда $P(\bigcap\limits_{i=1}^{n} A_i) = 1$ [5, с. 350].

Вещественнозначная функция X, определенная на Ω, называется *B*-измеримой или случайной величиной (с.в.) в пространстве <Ω, *B* , P>, если для любого борелевского множества Δ числовой прямой множество $\{\omega : X(\omega) \in \Delta\}$ – полный прообраз Δ при отображении X является *B*-измеримым, т.е. событием пространства <Ω, *B* , P>. Если в качестве Δ взять интервал $(-\infty, x]$, то при любом x событие $A_x = \{\omega : X(\omega) < x\}$ имеет определенную вероятность $P\{X < x\} = P(A_x) = F(x)$, при этом $F(x)$ называется функцией распределения (ф.р.) случайной величины X. Говорят, что случайная величина X на <Ω, *B* , P> индуцирует вероятностную меру на вещественной прямой R с борелевской σ – алгеброй *B* . Эта мера полностью определяется функцией распределения $F(x)$.

Производная $F'(x)$ ф.р. $F(x)$, если она существует, называется плотностью $p(x)$ вероятностей (п.в.) с.в. X. При этом

$$F(x) = \int\limits_{-\infty}^{x} p(y)\,dy \;;\; \int\limits_{-\infty}^{\infty} p(x)dx = 1.$$

Моментом 1-го порядка или математическим ожиданием с.в. X называется число $MX = \int\limits_{-\infty}^{\infty} xp(x)dx$, если с.в. X непрерывна. Если же с.в. дискретна, то $MX = \sum\limits_i x_i p_i$, где $p_i = P(X = x_i)$.

Начальным моментом *k*-го порядка с.в. X называется число

$$\mu_k = M[X^k].., \text{ k=1,2,}\ldots .$$

Центральным моментом k-го порядка с.в. X называется число

$$\nu_k = M[X - Mx]^k.$$

Второй центральный момент с.в. X называется дисперсией:

$$DX = M[X - MX]^2 = MX^2 - [MX]^2,$$

а $\sigma = \sqrt{DX}$ – среднеквадратическим отклонением.

Наряду с отмеченными числовыми характеристиками с.в., часто рассматривают следующие характеристики формы кривой плотности распределения:

а) асимметрию $A = \frac{\nu_3}{\sigma^3}$; в) моду $l = \arg\max\ p(x)$;

б) эксцесс $E = \frac{\nu_4}{\sigma^4}$; г) медиану $m \quad (F(m) = 0.5)$.

У симметричного распределения нечетные конечные центральные моменты равны нулю. Если асимметрия положительна (отрицательна), то большая вероятностная масса лежит справа (слева) от медианы распределения; иначе: длинный хвост кривой плотности лежит справа (слева) от медианы. Эксцесс характеризует островершинность кривой плотности. У нормального распределения эта характеристика равна нулю. Положительный (отрицательный) эксцесс указывает на то, что рассматриваемая кривая плотности имеет более (менее) высокую и острую вершину, чем нормальная кривая.

Числовой характеристикой системы двух случайных величин (X,Y) является ковариационной момент (ковариация):

$$co\upsilon\,(X,Y) = \nu_{xy} = M[(X - MX)(Y - MY)].$$

Справедливы следующие свойства математических ожиданий и дисперсий:

а) пусть C = const, тогда $MC = C$, $M[C - MC]^2$=0;

б) $M\,CX = C\,MX,\quad D(CX) = C^2 DX$;

в) $M(X+Y) = MX + MY$, $\mathrm{D(X + Y) = DX} + DY + 2\nu_{x\,y}$.

1.2. Функция распределения и плотность распределения случайного вектора. Условные законы распределения. Регрессия. Независимость случайных величин. Распределение функций от случайных величин. Композиция (свертка) распределений

Под случайным вектором понимается система случайных величин $(X_1, X_2,...,X_n)$. Ф.р. случайного вектора

$$F(x_1, x_2, ..., x_n) = P(X_1<x_1, X_2<x_2, ..., X_n<x_n)$$

является возрастающей по каждому аргументу, причем

$$F(-\infty;...;-\infty)=0\,,\quad F(+\infty\,;...;+\infty)=1.$$

Если $X_1, X_2, ..., X_n$ непрерывны, то плотность распределения вектора

$$p(x_1,x_2,...,x_n)=\frac{\partial^n F(x_1,x_2,...x_n)}{\partial x_1,\partial x_2,...\partial x_n},$$

(если $F(\cdot)$ не только непрерывна, но и дифференцируема по каждому аргументу) называют плотностью совместного распределения с.в. X_1, ..., X_n. Вероятность попадания случайной точки (вектора) в пределы n-мерной области D выражается интегралом

$$P(\{X_1, ..., X_n\} \in D) = \int\limits_D ...\int P(x_1,...,x_n)dx_1...dx_n.$$

Зная распределение n-мерного вектора, можно найти и распределение (маргинальное) $(n\text{-}1)$-мерного вектора. Действительно, пусть $n=2$, тогда ф. р. одной из координат: $F_X(x)=F_{X,\,Y}(x,+\infty)$, а плотность соответственно

$$p(x)=\int\limits_{+\infty}^{-\infty} p(x,y)dy.$$

Ф. р. и п.в. с.в. могут зависеть от некоторых параметров. Ими, в частности, могут являться отмеченные выше числовые характеристики (математическое ожидание, дисперсия и т.п.).

Пусть X_1 и X_2 – независимые с.в., имеющие соответственно ф.р. $F(x_1;\theta_1)$ и $F(x_2;\theta_2)$, где θ_1 и θ_2 – параметры.

Тогда, если ф.р. их суммы $z=X_1+X_2$ равна $F(z;\theta_1+\theta_2\square)$, то ф.р. $F(x;\square\ \theta)$ называется воспроизводимой.

Условным законом распределения случайной величины, входящей в случайный вектор, называется ее закон распределения, вычисленный при условии, что другая с.в. приняла фиксированное значение:

$$F(x/y)=P(X<x/Y=y),$$
$$F(y/x)=P(Y<y/X=x).$$

Соответствующие выражения для п.в. имеют вид:

$$p(x/y)=\frac{p(x,y)}{p(x)},\ \text{где}\ \ p(x)\neq 0,$$

$$p(x/y)=\frac{p(x,y)}{p(y)},\ \text{где}\ \ p(x)\neq 0.$$

Случайные величины независимы, если $F(x,y)=F(x)\,F(y)$.

Часто рассматривают некоторую числовую характеристику (математическое ожидание – среднее значение с.в., моду, медиану и т.п.) условного распределения с.в. (Y) в зависимости от фиксированных значений *x,* принимаемых другой с.в. (X). Кривая такого типа называется *кривой регрессии*, и говорят, что она *изображает регрессию* (упомянутой характеристики) с.в. *Y* на *X*. Если кривая регрессии есть прямая, то это случай *линейной регрессии*. Кривая регрессии для среднего:

$$\bar{y}=\mu_y(x)=M(Y/X=x)=\frac{\int_{-\infty}^{\infty} yp(x,y)dy}{\int_{-\infty}^{\infty} p(x,y)dy}=\frac{\int_{-\infty}^{\infty} yp(x,y)dy}{p(x)}.$$

В этом случае, если X и Y независимы, то

$$\mu_y(x)=\mu_y, \qquad \mu_x(y)=\mu_x,$$

т.е. кривые линейной регрессии – прямые, параллельные координатным осям и проходящие через точку (μ_y,μ_x).

Иногда требуется решать более общую задачу приближенного представления случайной величины $Y \cong g(X)$ [12, 8]. Общепринятым является приближение (при известных условиях) по методу наименьших квадратов. Величина $g(X)$ называется наилучшим приближением величины Y в смысле метода наименьших квадратов, если $M[Y - g(X)]^2$ принимает наименьшее возможное значение. При этом $g(X)$ называется среднеквадратической регрессией величины Y на X.

Пусть $\sigma_x > 0$, $\sigma_y > 0$. Линейная среднеквадратическая регрессия $Y \cong g(X)$ на X имеет вид:

$$g(X) = \mu_y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X),$$

или иначе

$$\frac{g(X) - \mu_Y}{\sigma_Y} = \rho \frac{X - \mu_X}{\sigma_X},$$

где $\rho = \dfrac{\nu_{XY}}{\sigma_X \sigma_Y}$ – коэффициент корреляции величин X и Y, ν_{xy} – ковариация, вычисляемая по формуле:

$$\nu_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - M(X))\,(y - M(Y))\; p(x, y)\, dx\, dy$$

для непрерывных величин и

$$\nu_{XY} = \sum_i \sum_j (x_i - M(X))(y_j - M(Y)) P(x_i, y_j)$$

для дискретных с.в.

Коэффициент $\beta = \rho \dfrac{\sigma_Y}{\sigma_X}$ называется *коэффициентом регрессии* величины Y на X, а прямая, определяемая уравнением

$$\frac{y - \mu_Y}{\sigma_Y} = \frac{\rho(x - \mu_X)}{\sigma_X}, \qquad (1.1)$$

называется *прямой среднеквадратической регрессии* величины Y на X. *Остаточной дисперсией с. в. Y относительно X* называется величина

$$\min_{\alpha, \beta} \; M(Y - \alpha - \beta X)^2 = \sigma_Y^2 (1 - \rho)^2 .$$

Она также характеризует величину ошибки приближения $Y \cong \alpha + \beta \cdot X$. Аналогичное (1.1) соотношение имеет место для регрессии среднего [2]:

$$\bar{y} = \alpha + \beta x = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X).$$

Таким образом, регрессия среднего Y на X совпадает со средне-квадратической регрессией Y на X.

Корреляционный момент определен только для пары с.в. Если определять его для каждой пары с.в. вектора, то получим корреляционную матрицу:

$$\|\sigma_{ij}\| = \begin{Vmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{Vmatrix} = \begin{Vmatrix} \nu_{1} & \nu_{12} & \dots & \nu_{1n} \\ \nu_{21} & \nu_{22} & \dots & \nu_{2n} \\ \dots & \dots & \dots & \dots \\ \nu_{n1} & \nu_{n2} & \dots & \nu_{nn} \end{Vmatrix}.$$

Эта матрица симметрична, диагональные элементы являются дисперсиями с.в., входящих в вектор $\{X_1,...,X_n\}$.

Коэффициент корреляции ρ_{xy} характеризует степень линейности зависимости между случайными величинами X и Y $|\rho_{xy}| \le 1$.

Если X и Y связаны точной функциональной зависимостью вида $Y = \alpha + \beta X$, то ρ_{xy} равен ± 1 (в зависимости от знака β).

Если $\rho_{xy} = 0$, то X и Y называют некоррелированными.

Если X и Y независимы, то $\rho_{xy} = 0$, а значит, X и Y также некоррелированы. Однако из некоррелированности с.в. еще не следует их независимость. Для распределенных по нормальному закону с.в. некоррелированность равносильна их независимости.

Пусть $Y = \varphi(X)$, где $\varphi(X)$ – непрерывная, монотонно возрастающая или монотонно убывающая функция от непрерывной с.в. X.

Иначе говоря, $Y = \varphi[\phi(Y)]$, а, значит, и $y = \varphi[\psi(y)]$, где обратная функция $\psi(y) = \varphi^{-1}(y) = x$. Тогда п.в. с.в. Y находится на формуле:

$$p(y) = p[\psi(y)]\,|\,\psi'(y)\,|. \tag{1.2}$$

Пусть далее $\{X_1, X_2\}$ – непрерывный случайный вектор с п.в. $p(x_1, x_2)$, $Y_i = \varphi_i(X_1, X_2)$, $i = 1,2,...$ – однозначные $(\forall\omega\in\Omega)$ функции, имеющие непрерывные частные производные и допускающие однозначное решение $x_i = \psi_i(y_1, y_2)$, $i = 1,2,...$. Тогда

$$p(y_1, y_2) = p[\psi_1(y_1, y_2), \quad \psi_2(y_1, y_2)] \quad |J|, \tag{1.3}$$

где якобиан преобразования:

$$J = \left(\frac{\partial x_i}{\partial y_j}\right), \quad i = j = 1,2.$$

Частный случай: функция двух независимых с.в – сумма с.в.

$$Y = \phi(X_1, X_2) = X_1 + X_2. \tag{1.4}$$

Уравнение (1.4) имеет единственное решение относительно X_1, а именно $X_1 = \psi\ (Y_1, Y_2) = Y - X_2$, при этом

$$\frac{\partial\,\psi\,(y, x_2)}{\partial y} = 1,$$

а

$$p(y) = \int_{-\infty}^{\infty} p(y - x_2, x_2)dx_2. \tag{1.5}$$

Аналогично $p(y) = \int_{-\infty}^{\infty} p(y - x_1, x_1)dx_1$. Если с.в. X_1, X_2 независимы, то $p(x_1, x_2) = p_X(x_1)p_{X_2}(x_2)$, а из (1.5) имеем

$$p(y) = \int_{-\infty}^{\infty} p_{X_2}(y - x_1)p_{X_1}(x_1)dx_1 = \int_{-\infty}^{\infty} p_{X_1}(y - x_2)p_{X_2}(x_2)dx_2 = \int_{-\infty}^{\infty} p_{X_1}(y - x)p_{X_2}(x)dx.$$

Эта формула называется *формулой композиции* (свертки) плотностей вероятностей и коротко записывается:

$$p_Y(y) = (p_{X_1} * p_{X_2})(y) = p_{X_1} * p_{X_2}(y).$$

Соответственно для ф.р. имеет место формула

$$F_{X_1+X_2}(y) = \int_{-\infty}^{y} p_{X_1+X_2}(t)dt = \int_{-\infty}^{y}\int_{-\infty}^{\infty} p_{X_1}(t - x)p_{X_2}(x)dxdt,$$

т.е. $F_{X_1+X_2}(y) = (F_{X_1} * p_{X_2})(y) = (p_{X_1} * F_{X_2})(y)$.

Аналогично
$$\begin{gathered} p_{X_1} * p_{X_2} * ... * p_{X_n}(y) = (p_{X_1} * p_{X_2} * ... * p_{X_{n-1}}) * p_{X_n}(y), \\ F_{X_1+...+X_n}(y) = (F_{X_1} * p_{X_2} * p_{X_3} * ... * p_{X_n})(y). \end{gathered}$$

Для обобщений свертка предоставляется в виде интеграла Стилтьеса

$$F_{X_1+X_2}(y)=\int_{-\infty}^{\infty}F_{X_1}(y-x)dF_{X_2}(x)=\int_{-\infty}^{\infty}F_{X_2}(y-x)dF_{X_1}(x).$$

Рассмотрим дискретный аналог свертки [20]. Пусть X и Y неотрицательные независимые целочисленные с.в. с распределениями вероятностей $p\{X=j\}=a_j$, $p\{Y=j\}=b_j$, $j=1,2,...$ соответственно. Событие $(X=j,\ Y=k)$ имеет вероятность $a_j\ b_k$. Сумма $Z=X+Y$ есть новая с.в., а событие $Z=r$ есть объединение событий $(X=0,Y=r)$, $(X=1,Y=r-1)$, $(X=2, Y=r-2)$, ..., $(X=r,Y=0)$. Эти события несовместны, и поэтому распределение $C_r=P(Z=r)$, $r=0,1,...$, задается формулой:

$$c_r=a_0b_r+a_1b_{r-1}+a_2b_{r-2}+...+a_rb_0=\sum_{i=0}^{r}a_ib_{r-i}\ . \tag{1.6}$$

1.3. Характеристические и производящие функции, их свойства и формулы обращения. Семиинварианты

Характеристическая функция (х.ф.) с.в. X с ф.р. $F(x)$ однозначно определяется следующим образом:

$$\varphi_X(t)=Me^{itX}=\int_{-\infty}^{\infty}e^{itx}dF(x), \tag{1.7}$$

т.е. для непрерывных величин $\varphi_X(t)=\int_{-\infty}^{\infty}e^{itx}p(x)dx$, а для дискретных

$$\varphi_X(t)=\sum_{k=1}^{n}e^{itx_k}p_k,\quad p_k=P(X=x_k),\quad k=\overline{1,n}..$$

В многомерном варианте (для многомерного непрерывного случайного вектора):

$$\varphi(t)=\varphi(t_1,...,t_n)=\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}e^{i(t_1x_1+...+t_nx_n)}p(x_1,...,x_n)dx_1...dx_n.$$

Необходимо отметить следующие основные свойства характеристических функций:

1. $\varphi_X(0)=1,\quad |\varphi_X(t)|\le 1,\quad \forall t.$

2. $\varphi_{aX+b}(t)=e^{itb}\varphi_X(at).$

3. $\varphi_{X_1+...+X_n}(t)=\varphi_{X_1}(t)\varphi_{X_2}(t)...\varphi_{X_n}(t),$ если с.в. X_1, X_2, ..., X_n независимы.

4. $\varphi(t)$ непрерывна на вещественной оси R.

5. $\overline{\varphi_X(t)}=\varphi_X(-t)=\varphi_{-X}(t).$

6. Пусть $M\,|\,X\,|^k<\infty$, $k=\overline{1,n},$ тогда $\mu_k=(-i)^k\varphi^{(k)}(0).$

В многомерном случае начальный момент порядка $k = k_1 + k_2 + + ... + k_n$ определяется дифференцированием х.ф.

$$m_{k_1k_2...k_n}=\left.\frac{(-i)^k\partial^k\varphi(t)}{\partial t_1^{k_1}\partial t_2^{k_2}...\partial t_n^{k_n}}\right|_{t=0}.$$

7. Если $t\to\infty$, то $\varphi_X(t)\to 0.$

Если начальный момент μ_k существует, то в окрестности нуля справедливо следующее разложение характеристической функции:

$$\varphi(t)=1+\sum_{r=1}^{k}\frac{(it)^r}{r!}\mu_r+o(t^k),\ \text{где}\ \lim_{t\to\infty}\frac{o(t^k)}{t^k}=0. \qquad (1.8)$$

Часто, однако, удобнее рассматривать разложение:

$$\ln\varphi(t)=\sum_{r=1}^{k}\frac{\kappa_r}{r!}(it)^r+o(t^k),$$

где κ_r – семиинварианты (кумулянты) распределения. Такое название числа κ_r получили потому, что все они, кроме κ_1, инвариантны относительно переноса (сдвига) с.в., т.е. прибавления к с.в. постоянной; они складываются, когда складываются независимые с.в.. Любой семиинвариант является полиномом от начальных моментов.

Если в (1.7) положить $e^{it}=z$, тогда имеем производящую функцию $\pi(z)$, которую удобно использовать, если X – дискретная с.в.:

$$\pi(z)=M(z^X), \qquad (1.9)$$

при этом $|z|\le 1$; $\pi(z)$ разлагается в ряд по степеням z и коэффициент при z^x равен вероятности значения x, т.е. $p\,(X=x)$.

Если с.в. X непрерывна, то справедлива формула обращения характеристических функций (обратное преобразование Фурье):

$$p_X(X)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\varphi_X(t)dt.$$

Для получения и обращения характеристических функций положительных случайных величин удобно пользоваться преобразованием Лапласа; существуют таблицы преобразований Лапласа и Фурье.

1.4. Основные виды сходимости. Некоторые неравенства

Говорят, что последовательность с.в. $X_1, ..., X_n$ сходится по распределению к с.в. X $(X_n \Rightarrow X)$, если последовательность их ф.р. сходится к $F(x)$ во всех точках x, за исключением, быть может, точек разрыва последней.

С понятием сходимости по распределению связана следующая теорема непрерывности П. Леви.

Теорема Леви. Если $\{X_n\}$ сходится при $n \to \infty$ к X по распределению, то соответствующая последовательность характеристических функций $(\varphi_n(t))_{n=1}^{\infty}$ равномерно сходится $(\varphi_n(t) \to \varphi_X(t))$ в любом конечном интервале значений t.

Говорят, что $\{X_n\}$ сходится по вероятности к с.в. X $\left(X_n \xrightarrow[P]{} X\right)$, если $P(|X_n - X| > \varepsilon) \to 0$ при $n \to \infty$, $\forall \varepsilon > 0$. Последовательность $\{X_n\}$ сходится к X почти наверное (п.н.) (или с вероятностью 1 $(P\{X_n \to X\} = 1)$, если $X_n(\omega) \to X(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$, за исключением, быть может, множества нулевой вероятности.

Из сходимости п.н. следует сходимость по вероятности, из сходимости по вероятности следует сходимость по распределению.

Если начальный момент $\mu_n = MX^k$ существует и конечен, то справедливы:

а) неравенство Коши-Буняковского

$$M|X|^2|Y|^2 \leq M|X|^2 \times M|Y|^2; \tag{1.10}$$

б) неравенство Шварца

$$M^2|XY| \leq M|X|^2 \times M|Y|^2; \tag{1.10'}$$

в) неравенство Чебышева

$$P[|X| \geq \varepsilon] \leq \frac{M|X|^2}{\varepsilon^2}, \text{ при } \varepsilon > 0; \tag{1.11}$$

г) неравенство Иенсена

$$g(MX) \le Mg(X), \tag{1.12}$$

если g – выпуклая функция.

1.5. Некоторые важные распределения

1.5.1. Дискретные распределения

а) распределение Бернулли Br(p), биномиальное Bi(n,p), мультиномиальное (полиномиальное) распределение M(n, p_1 ..., p_k).

Рассмотрим серию независимых экспериментов (испытаний) Бернулли, каждый из которых может приводить к появлению события A с вероятностью $P(A)=p$ или события $\overline{A}$ с вероятностью $P(\overline{A}) = 1 - p = q$.

Введем случайные величины

$$X_i(\omega) = \begin{cases} 1, & если \quad \omega = A \\ 0, & если \quad \omega = \overline{A} \end{cases},$$

где индекс $i=1, 2, ..., n$ есть номер эксперимента[*] в серии из n испытаний.

Каждая случайная величина X_t, таким образом, имеет следующее распределение (распределение Бернулли $Br(p)$):

$$\begin{cases} P(X_i = 1) = p, \\ P(X_i = 0) = q. \end{cases} \tag{1.13}$$

Отметим, что $MX_i = p, \quad DX_i = pq, \quad х.ф. \quad \varphi(t) = pe^{it} + q.$

Так как с.в. X_i $(i=1, 2, ..., n)$ независимы, то распределение вероятностей случайных величины $S_n = \sum_{i=1}^{n} X_i$, дающее число m появлений исхода A в n экспериментах, имеет вид

$$P_{m,n} = C_n^m p^m q^{n-m}, \quad m = 0, 1, ..., n. \tag{1.14}$$

---

[*] Здесь знак $\in$ означает «имеет распределение» (такое соглашение удобно и не приводит к недоразумениям), в подобных случаях этот знак будет употребляться и далее.

При этом коэффициент $C_n^m = \frac{n!}{m!(n-m)!}$, так как событие, состоящее в появлении исхода A m раз в n экспериментах может осуществляться C_n^m различными способами в зависимости от порядка следования A и $\overline{A}$ в серии.

Распределение (1.13) называется биномиальным *Bi(n,p):*

$$M(S_n) = np; \quad D(S_n) = npq; \quad \varphi(t) = \sum_{m=1}^{n} C_n^m p^m q^{n-m} e^{itm}.$$

Если S_{n_1} и S_{n_2} – независимые с. в., то $S_{n_1} + S_{n_2} \in B_i(n_1 + n_2, p)^*$.

Мультиномиальное распределение $M(n, p_1, p_2, ..., p_{k+1})$ является обобщением биномиального распределения в том смысле, что, в отличие от схемы испытаний Бернулли, исходом испытания может явиться одно из нескольких (>2) взаимоисключающих событий ($A_1, ..., A_{k+1}$). Вероятности, соответствующие этим событиям, равны $p_1, ..., p_{k+1}$, где $p_1 + ... + p_{k+1} = 1$.

Пусть осуществляется n испытаний. Введем $(k+1)$-мерную случайную величину $(X_1, ..., X_{k+1})$, компоненты которой $X_1, ..., X_{k+1}$ означают числа испытаний, приводящих соответственно к $A_1, ..., A_{k+1}$. Компоненты $X_1, ..., X_{k+1}$ линейно зависимы, т. к. $X_1 + ... + X_{k+1} = n$.

Распределение с. в. $(X_1, ..., X_{k+1})$ дается формулой

$$P(x_1,...,x_k) = \frac{n\,!}{x_1!...x_{k+1}!} p_1^{x_1} ... p_{k+1}^{x_{k+1}}, \tag{1.15}$$

где необходимо положить $x_{k+1} = n - x_1 - ... - x_k$ и $p_{k+1} = 1 - p_1 - ... - p_k$.

Это выражение представляет собой общий член в разложении мультинома $(p_1 + ... + p_{k+1})^n$ по степеням $p_1, p_2, ..., p_{k+1}$.

Х.ф. мультиномиального распределения:

$$\varphi(t_1,...,t_k) = (p_1 e^{it_1} + ... + p_k e^{it_k} + p_{k+1} e^{it_{k+1}})^n;$$

$$MX_i = np_i; \quad DX_i = np_i(1 - p_i);$$
$$cov\,(X_i, X_j) = -np_i p_j;$$

$M(n; p_1, ..., p_k)$ является воспроизводящим по n;

б) распределение Пуассона $P_0(\lambda)$:

$$P\{X=m\}=\frac{\lambda^m}{m!}e^{-\lambda}, \qquad (1.16)$$

$m\geq 0, \quad \lambda>0,$ m – целые числа.

Если независимые с.в. $X_k \in P_0(\lambda_k), \quad k=1, 2. \quad то \quad x_1+x_2 \in P_0(\lambda_1+\lambda_2).$

1.5.2. Распределения абсолютно непрерывного типа:

а) равномерное распределение R(1/2, 1) на интервале *(0, 1),* его п.в.

$$p(x)=\begin{cases}1, & x\in(0,1),\\ 0, & x\notin(0,1),\end{cases}$$

$$MX=\frac{1}{2}; \quad DX=\frac{1}{12}; \quad \text{х. ф. } \varphi(t)=\frac{e^{it}-1}{it}.$$

В условном обозначении *R(a, h)* параметр *a* означает математическое ожидание, *h* – размах, т.е. длина (мера) области значений с.в., где *p(x) >0*;

б) нормальное распределение N(a,σ^2 *)* задается ф.р.

$$F(x)=\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{x}e^{-\frac{(u-a)^2}{2\sigma^2}}du, \quad -\infty<x<\infty, \text{ где } a=MX, \quad \sigma^2=DX.$$

Характеристическая функция $\varphi(t)=e^{ita-\frac{1}{2}\sigma^2t^2}$.

Пусть X_1 и X_2 – независимые с.в., имеющие нормальное распределение:

$$X_i\in N(a_i,\sigma_i^2), \quad i=1, 2; \quad \text{тогда}$$

$$X_1+X_2\in N(a_1+a_2, \sigma_1^2+\sigma_2^2);$$

в) гамма-распределение G(λ *)* задается ф.р.

$$F(x)=\frac{1}{\Gamma(\lambda)}\int_0^x u^{\lambda-1}e^{-u}du, \quad x\geq 0, \qquad (1.17)$$

где гамма-функция $\Gamma(\lambda)=\int_0^\infty u^{\lambda-1}e^{-u}du$ при $\lambda>0$, в частности, $\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi},$ а $\Gamma\,(n+1)=n!,$ х.ф. $\phi(t)=(1-it)^{\lambda}, \quad MX=DX=\lambda$.

Пусть $X_1 \in G(\lambda_1)$ и $X_2 \in G(\lambda_2)$ – независимые с.в., тогда

$$X_1 + X_2 \in G(\lambda_1 + \lambda_2);$$

г) бета-рапределение $Be\,(\lambda_1, \lambda_2), \quad \lambda_1 > 0, \lambda_2 > 0,$ его ф.р.

$$F(x) = \frac{1}{B(\lambda_1, \lambda_2)} \int_0^x u^{\lambda_1 - 1}(1-u)^{\lambda_2 - 1} du, \quad 0 \le x \le 1, \tag{1.18}$$

где бета-функция $B(\lambda_1, \lambda_2) = \int_0^1 x^{\lambda_1 - 1}(1-x)^{\lambda_2 - 1} dx$ имеет связь с гамма-функцией $B(\lambda_1, \lambda_2) = \dfrac{\Gamma(\lambda_1)\Gamma(\lambda_2)}{\Gamma(\lambda_{1+\lambda_2})}$.

Х.ф. у бета-распределения обычно не пользуются.

$$MX = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \quad DX = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2 (\lambda_1 + \lambda_2 + 1)}.$$

Пусть $X_1 \in G(\lambda_1), \quad X_2 \in G(\lambda_2),$ тогда $\dfrac{X_1}{X_1 + X_2} \in Be(\lambda_1, \lambda_2);$

д) хи – квадрат распределение обозначается *P(n)* или χ^2 *(n).*

Это есть распределение с. в. $\chi_n^2 = X_1^2 + X_2^2 + ... + X_n^2$, где X_1 , X_2 , ..., X_n – независимые стандартные нормальные с. в..

Параметр *n* называют числом степеней свободы.

$$\chi_n^2 = 2Y, \text{ где } Y \in G\left(\frac{n}{2}\right), \ M\chi_n^2 = n, \quad D\chi_n^2 = 2n.$$

П.в. с.в. χ_n^2 имеет вид

$$P_n(x) = \begin{cases} [1/(2^{n/2}\Gamma(n/2))] x^{(n/2)-1} e^{-x/2} & , \textit{при } x > 0, \\ 0, & \textit{при} \quad x \le 0; \end{cases} \tag{1.19}$$

е) распределение Стьюдента S (n) (*t* – распределение) есть распределение с.в.

$$t = \frac{X}{\sqrt{\chi^{2/n}}}, \quad \textit{где} \quad X \in N(0,1)$$

и $\chi^2 = \chi_n^2 \in \chi^2(n)$, *X* не зависит от χ^2.

П.в. с.в. t имеет вид

$$P_n(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}}\left(1+\frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)}. \qquad (1.20)$$

Параметр n так же, как и в $\chi^2(n)$, называют числом степеней свободы. $Mt=0, \quad Dt=\frac{n}{n-2}, \quad n>2;$

ж) распределение Фишера-Снедекора $S(n_1, n_2)$ есть распределение с.в.

$$F=\frac{X_1}{n_1}:\frac{X_2}{n_2},$$

где $X_1\in\chi^2(n_1)$ и $X_2\in\chi^2(n_2)$ – независимы.

з) распределение Дирихле – упорядоченное $D^*(\lambda_1,\lambda_2,...,\lambda_k,\lambda_{k+1})$ и неупорядоченное $D(\lambda_1,\lambda_2,...,\lambda_k,\lambda_{k+1})$ [20, 16].

K-мерное распределение (неупорядоченное) Дирихле имеет п.в.

$$p(x_1,...,x_k)=\frac{\Gamma\left(\lambda_1+...+\lambda_{k+1}\right)}{\Gamma\left(\lambda_1\right)...\Gamma\left(\lambda_{k+1}\right)}x_1^{\lambda_1-1}...x_k^{\lambda_k-1}(1-x_1-...-x_k)^{\lambda_{k+1}-1}, \quad \lambda_k\geq 0 \ (1.21)$$

в любой точке симплекса $S_k=\{(x_1,...,x_k):x_i\geq 0, \quad i=1,...,k, \quad \sum_{i=1}^{k}x_i\leq 1\}$ и равную нулю в других точках R_k.

$$M(X)=\frac{\lambda_i}{\lambda_1+...+\lambda_{k+1}} \quad i=1,...,k;$$

$$D(X_i)=\frac{\lambda_i(\lambda_1+...+\lambda_{k+1}-\lambda_i)}{(\lambda_1+...+\lambda_{k+1})^2(\lambda_1+...+\lambda_{k+1}+1)}, \quad i=1,...,k;$$

$$co\upsilon\left(X_i,X_j\right)=-\frac{\lambda_i\lambda_j}{\left(\lambda_1+...+\lambda_{k+1}\right)^2\left(\lambda_1+...+\lambda_{k+1}+1\right)}, \quad i=j=1,...,k.$$

Х.ф. для $D(\lambda_1,\lambda_2,...,\lambda_k;\lambda_{k+1})$ так же, как и для $Be(\lambda_1,\lambda_2)$, не пользуются.

Пусть $(X_1, ..., X_k)$ – векторная с.в., имеющая k-мерное распределение Дирихле $D(\lambda_1,\lambda_2,...,\lambda_k.\lambda_{k+1})$.

Введем новые величины:

$$Y_1 = X_1,$$
$$Y_2 = X_1 + X_2,$$
$$\dots\dots\dots\dots\dots\dots\dots$$
$$Y_k = X_1 + \dots + X_k.$$

Так как якобиан этого преобразования равен единице, п.в. вектора $(Y_1, \dots, Y_k)$ имеет вид

$$p(y_1,\dots,y_k) = \frac{\Gamma(\lambda_1 + \dots + \lambda_{k+1})}{\Gamma(\lambda_1)\dots\Gamma(\lambda_{k+1})} y_1^{\lambda_1 - 1}(y_2 - y_1)^{\lambda_2 - 1}\dots(y_k - y_{k-1})^{\lambda_k - 1}(1 - y_k)^{\lambda_{k+1} - 1},$$

где $(0 < y_1 < \dots < y_k < 1)$.

Это распределение называется упорядоченным распределением Дирихле $D^*(\lambda_1, \lambda_2, \dots, \lambda_k . \lambda_{k+1})$. При $k=1$ оно совпадает с $Be(\lambda_1, \lambda_2)$.

Если $(X_1, X_2, \dots, X_n)$ – векторная с.в., имеющая k-мерное распределение Дирихле $D(\lambda_1, \lambda_2, \dots, \lambda_k . \lambda_{k+1})$, то сумма $X_1 + \dots + X_n$ имеет бета-распределение $Be(\lambda_1 + \dots + \lambda_k; \lambda_{k+1})$.

Если

$$(X_1,\dots,X_k) \in D(\lambda_1,\dots,\lambda_k;\lambda_{k+1}), \quad \text{то} \quad (X_1,\dots,X_{k_1}) \in D(\lambda_1,\dots,\lambda_{k_1};\lambda_{k_1} + \dots + \lambda_{k+1}),$$
$$k_1 < k_2$$

Если

$$(Y_1,\dots,Y_k) \in D(\lambda,\dots,\lambda_k \lambda_{k+1}), \quad \text{то} \quad (Y_k, Y_{k_1+k_2},\dots,Y_{k_1+\dots+k_s}) \in D^*(\lambda_{(1)},\dots,\lambda_{(s)};\lambda_{(s+1)}),$$

где $\lambda_{(1)} = \lambda_1 + \dots + \lambda_{k_1}, \dots, \lambda_{(s)} = \lambda_{k_1+\dots+k_{s-1}+1} + \dots + \lambda_{k_1+\dots+k_s}$, $\lambda_{(s+1)} = \lambda_{k_1+\dots+k_s+1} + \dots + \lambda_{k+1}$;
$k_1 + \dots + k_s \le k$;

и) многомерное нормальное распределение $N(\{\mu_i\}, \|\sigma_{ij}\|)$, $i, j = 1,\dots,k$ задается п.в.

$$p(x_1,\dots,x_k) = \frac{\sqrt{\|\sigma^{ij}\|}}{(2\pi)^{(k/2)}} e^{-\frac{1}{2}Q(x_1,\dots,x_k)}, \tag{1.22}$$

где $(x_1, \dots, x_k)$ — любая точка в R_k и $Q(x_1,\dots,x_k) = \sum_{i,j=1}^{k} \sigma^{ij}(x_i - \mu_i)(x_j - \mu_j)$,

причем $\|\sigma^{ij}\|$ — матрица, обратная ковариационной матрице $\|\sigma_{ij}\|$, и μ_i – среднее соответствующих с.в., входящих в k-мерный вектор $\{x_1, \dots, x_k\}$.

Х.ф. распределения $N\left(\{\mu_i\},\ \left\|\sigma^{ij}\right\|\right),\quad i, j = 1,\ldots,k$ имеет вид

$$\varphi(t_1,\ldots,t_k) = \exp\left\{i\sum_{j=1}^{k}\mu_j t_j - \frac{1}{2}\sum_{i,j=1}^{k}\sigma_{ij}t_i t_j\right\}$$

$$MX_i = \mu_i,\quad DX_i = \sigma_{ii},\quad cov\left\{X_i, X_j\right\} = \sigma_{ij}.$$

1.6. Предельные теоремы

Законы больших чисел

Пусть $\{X_k\}_{k=1}^{\infty}$ – последовательность независимых одинаково распределенных случайных величин, имеющих математическое ожидание $M\{X_k\} = a$. Обозначим $S_n = \sum\limits_{k=1}^{n} X_k$. Тогда справедливы следующие теоремы.

Теорема 1 (закон больших чисел в форме Хинчина).

Для любого $\eta > 0$ и любого $\varepsilon > 0$ существует столь большое *n*, что $P\left\{\left|\frac{S_n}{n} - a\right| > \varepsilon\right\} < \eta$. Иначе говоря, $\frac{S_n}{n} \xrightarrow[P]{} a$.

Теорема 2 (усиленный закон больших чисел Колмогорова).

$P\left\{\lim\limits_{n\to\infty}\frac{S_n}{n} = a\right\} = 1$. Иначе говоря, среднее арифметическое $\frac{S_n}{n}$ не только становится (для достаточно большого *n*), но и остается достаточно близким к математическому ожиданию *a*.

Теорема 3 (закон больших чисел для схемы испытаний Бернулли) – частный случай теоремы 1. Если $P\{X_k = 1\} = p$ и $P\{X_k = 0\} = 1 - p$, то $\frac{m}{n} \xrightarrow[p]{} p$, где $m = S_n$. Иначе говоря, $P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} \to 1$ при $n \to \infty$ для любого $\varepsilon > 0$.

Теорема 3′ (усиленный закон больших чисел Бореля для схемы испытаний Бернулли). Пусть *m* – число наступлений события *A* в *n* испытаниях Бернулли, *p* – вероятность успеха, тогда $P\left\{\frac{m}{n} \to p\right\} = 1$.

При дополнительном предположении, что у слагаемых X_k существуют дисперсии, теоремы, аналогичные теоремам I и 2, были установлены еще Чебышевым как следствия неравенства Чебышева.

Локальные и интегральные предельные теоремы

Пусть дана последовательность независимых случайных величин $\{X_k,\ \ k=1,2,...\}$, имеющих конечные математические ожидания $a_k = M\{X_k\}$ и дисперсии $b_k^2 = D\{X_k\}$. Обозначим $B_k^2 = \sum_{k=1}^{n} b_k^2$. Тогда справедливы следующие теоремы (центральные) о сходимости к нормальному распределению нормированной центрированной суммы большого числа с.в. X_k.

Теорема 4 (теорема Линдеберга). Если последовательность $\{X_k,\ k=1,2,...\}$ независимых с.в., имеющих ф.р. $F_k(x)$ при $\forall \tau > 0$, удовлетворяет условию Линдеберга:

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int\limits_{|x-a_k|>\tau B_n} (x-a_k)dF_k(x) = 0, \tag{L}$$

то при $n \to \infty$ равномерно относительно x

$$P\left\{\frac{1}{B}\sum_{k=1}^{n}(x_k - a_k) < x\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \equiv \Phi(x).$$

Условие (L) является достаточным.

Теорема 5 (теорема Ляпунова) – следствие теоремы 4.

Если для последовательности с.в. $\{X_k, k=1,2,...\}$ можно подобрать такое положительное число $\delta > 0$, что при $n \to \infty$

$$\frac{1}{B_n^{\ 2+\delta}} \sum_{k=1}^{n} M\left\{|X_k - a_k|^{2+\delta}\right\} \to 0, \tag{Л}$$

то равномерно по x

$$P\left\{\frac{1}{B_n}\sum_{k=1}^{n}(X_k - a_k) < x\right\} \to \Phi(x). \tag{1.23}$$

Условие (Л) Ляпунова также является достаточным и более удобным для практического использования.

Теорема 6 (теорема Линдеберга-Леви**).** Если независимые с.в. X_k одинаково распределены и имеют конечную, отличную от нуля дисперсию, то при $n \to \infty$ выполняется (1.23).

Теорема 7 (интегральная теорема Муавра-Лапласа для последовательности испытаний Бернулли) – частный случай предыдущей теоремы.

Пусть $P(X_k = 1) = p; \quad P(X_k = 0) = q = 1 - p, \quad S_n = \sum_{k=1}^{n} X_k,$ тогда при

$\max(|a|, |b|) = o(n^{1/6}) \;\; P(a \le \frac{S_n - np}{\sqrt{npq}} < b) \xrightarrow[n\to\infty]{} \Phi(b) - \Phi(a).$

Как известно, при вычислении вероятностей конкретных локальных значений сумм $\sum_{k=1}^{n} X_k$ дискретных с.в. возникают трудности в вычислении факториалов при больших значениях n. Обойти эти трудности позволяют следующие локальные теоремы.

Теорема 8 [5] (локальная теорема Муавра-Лапласа для последовательности испытаний Бернулли). В обозначениях теоремы 7

$$P(S_n = m) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi npq}} e^{-x^2/2},$$

где $x = \frac{m - np}{\sqrt{npq}}, \;\; m - np = \mathrm{O}(n^{2/3}).$

Теорема 9 (локальная теорема Пуассона)**.** В обозначениях теоремы 7 пусть p – мало и изменяется с ростом n так, что $np = \lambda = const,$

тогда $P\{S_n = m\} \xrightarrow[n\to\infty]{} \frac{\lambda^m}{m!} e^{-\lambda}$, m=0,1,… .

Глава 2. ТЕОРИЯ ВЫБОРОЧНОГО МЕТОДА

2.1. Основные понятия математической статистики. Теоремы Гливенко и Колмогорова

Пусть X – одномерная с.в., заданная в R^1 и имеющая ф.р. $F(x)$. *Случайной выборкой* объема n на совокупности S возможных значений x *(генеральной совокупности)* с.в. X с ф.р. $F(x)$ генеральной совокупности называется n – мерная с.в. $\tau_n = (x_1, \ldots, x_n)$ (исход выборки или эксперимента), имеющая ф.р. $\prod_{j=1}^{n} F(x_j)$ и определенная в выборочном пространстве (пространстве исходов) $\mathrm{R}^n = \mathrm{R}^1_{(1)} \times \mathrm{R}^1_{(2)} \times \ldots \times \times \mathrm{R}^1_{(n)}$, где $\mathrm{R}^1_{(j)}$ – одномерное выборочное пространство (вещественная прямая), снабженном своей σ-алгеброй подмножеств $B(\mathrm{R}^n)$.

Необходимо подчеркнуть, что элементы $x_1, \ldots, x_n$ выборки взаимно независимы и имеют одинаковые ф.р. и поэтому такую случайную выборку называют *простой (повторной)*.

Замечание. Далее часто, если это не приводит к недоразумениям, будем одинаково обозначать через τ_n (а иногда просто через x) выборку и ее наблюденное (фиксированное) значение.

Случайная величина X может быть k-мерной с.в. $(X_1, \ldots, X_k)$ с ф.р. $F(x_1, \ldots, x_k)$, определенной в R^k . В этом случае выборка τ_n есть kn-мерная с.в. $(x_{1j}, \ldots, x_{kj}; j = 1, \ldots, n)$ с ф.р. $\prod_{j=1}^{n} F(x_{1j}, \ldots, x_{kj})$, определенной в $\mathrm{R}^{nk} = \mathrm{R}^k_{(1)} \times \mathrm{R}^k_{(2)} \times \ldots \times \mathrm{R}^k_{(n)}$, где $\mathrm{R}^k_{(j)}$ – *выборочное пространство* вектора $(x_{1j}, \ldots, x_{kj})$, $j = 1, \ldots, n$. Если τ_{n_1} и τ_{n_2} – выборки объемов n_1 и n_2 с ф.р. $F_1(x)$ и $F_2(x)$, то мы рассматриваем $(x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2})$ как $(n_1 + n_2)$ – мерную с.в. с ф.р. $\prod_{j_1=1}^{n_1} F_1(x_{1j_1}) \prod_{j_2=1}^{n_2} F_2(x_{2j_2})$ в $\mathrm{R}^{n_1+n_2}$.

Основная задача теории выборочного метода состоит в нахождении ф.р. от функции n с.в., входящих в выборку $\tau_n = (x_1, ..., x_n)$. Любая функция $g(\tau_n) = g(x_1, ..., x_n)$, которая сама является с.в., называется *статистикой.*

Примерами статистик являются выборочная сумма $z = \sum_{j=1}^{n} x_j$, выборочное среднее значение $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, выборочная дисперсия $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$, наименьший элемент выборки $\min(x_1, ..., x_n)$, наибольший элемент выборки $\max(x_1, ..., x_n)$ и т.п.

Выборку можно представить геометрически множеством точек $x_1, ..., x_n$ на оси x.

Последовательность элементов выборки, расположенных в возрастающем порядке, носит название *вариационного ряда*: $x_{(1)}, ..., x_{(n)}$, где $x_{(1)} \le ... \le x_{(n)}$.

Вариационный ряд также является статистикой (так называемой *порядковой статистикой*).

Эмпирической функцией распределения $F_n(x)$ мы назовем функцию, определенную следующими равенствами:

$$F_n(x) = \begin{cases} 0 & \textit{при} \qquad\qquad x \le x_{(1)} \\ \dfrac{k}{n} & \textit{при} \quad x_{(k)} < x \le x_{(k+1)}, \quad k = 2, n-1, \\ 1 & \textit{при} \qquad\qquad x > x_{(n)}. \end{cases}$$

Ясно, что эмпирическая функция распределения монотонна, непрерывна слева и имеет точки разрыва только при значениях аргумента, равных членам вариационного ряда. Величины скачков в точках разрыва являются целыми кратными от $\frac{1}{n}$. Если обозначить через K_x число выборочных значений, не превосходящих x, то

$$F_n(x) = \frac{K_x}{n},$$

так что $F_n(x)$ представляет собой частоту события $X < x$ в последовательности из n наблюдений с.в. X.

Относительно $F_n(x)$ справедливы следующие теоремы.

Теорема 10 (теорема Гливенко). Если *F(x)* – ф.р. с.в. X, а $F_n(x)$ – эмпирическая ф.р. n значений величины, то при $n \to \infty$

$$P\left\{ \sup_{-\infty \le x \le \infty} \left| F_n(x) - F(x) \right| \to 0 \right\} = 1.$$

Доказательство [3,12,15]. Предположим, что $F(x)$ является функцией распределения абсолютно непрерывного типа. Обозначим через $x_{r,k}$ наименьшее x, удовлетворяющее неравенствам

$$F(x-0) = F(x) \le \frac{k}{r} \le F(x+0) \quad (k = 1,2,...,r).$$

Пусть A означает событие, состоящее в том, что $X < x_{r,k}$. Ясно, что

$$P(A) = F(x_{r,k}).$$

Так как частота появления события A равна $F_n(x_{r,k})$, то по теореме Бореля (т. 3')

$$P\left\{ F_n(x_{r,k}) \underset{n\to\infty}{\to} F(x_{r,k}) \right\} = 1. \tag{2.1}$$

Пусть теперь E_k^r есть событие, состоящее в том, что при $n \to \infty$

$$F_n(x_{r,k}) \to F(x_{r,k}) \quad (k = 1,2,...,r),$$

$$E^r = E_1^r E_2^r ... E_r^r.$$

Ясно, что событие E^r равносильно тому, что при $n \to \infty$

$$\max_{1 \le k \le r} \left| F_n(x_{r,k}) - F(x_{r,k}) \right| \to 0.$$

Так как согласно (2.1) $P(E_1^r) = P(E_2^r) = ... = P(E_r^r) = 1$, то в силу свойства ж) (см. п. 1.1) $P\left\{ E^r \right\} = 1$.

Пусть далее $E = E^1 E^2 E^3 ...$.

В силу свойства з) того же раздела $P\left\{ E \right\} = 1$.

Обозначим, наконец, через В событие, состоящее в том, что при $n \to \infty$ $\sup\limits_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0$.

Для любого x, заключенного между $x_{r,k}$ и $x_{r,k+1}$, выполняются неравенства:

$$F_n(x_{r,k}+0) \le F_n(x) \le F_n(x_{r,k+1}),$$

$$F(x_{r,k}+0) \le F(x) \le F(x_{r,k+1}),$$

причем $0 \le F(x_{r,k+1}) - F(x_{r,k}+0) \le \frac{1}{r}$.

Отсюда мы заключаем, что

$$F_n\left(x_{r,k}+0\right) - F\left(x_{r,k+1}\right) \le F_n(x) - F(x) \le F_n\left(x_{r,k+1}\right) - F\left(x_{r,k}+0\right),$$

т.е. что $|F_n(x) - F(x)| \le \max\limits_{1\le k\le r} \left|F_n\left(x_{r,k}\right) - F\left(x_{r,k}\right)\right| + \frac{1}{r}$.

Поскольку *r* произвольно, то из последнего неравенства вытекает, что $E \subset B$, откуда и следует утверждение теоремы. Для произвольной функции распределения доказательство аналогично.

Теорема Гливенко носит название фундаментальной теоремы математической статистики.

Теорема 11 (теорема Колмогорова). Если *F(x)* непрерывна, то при $n \to \infty$

$$P(\sqrt{n}\sup_x |F_n(x) - F(x)| < z) \to K(z),$$

$$K(z) = \begin{cases} 0 & при\ z \le 0, \\ \sum\limits_{\kappa=-\infty}^{\infty} (-1)^k e^{-2k^2z^2} & при\ z > 0. \end{cases}$$

Таким образом, согласно приведенным теоремам 10 и 11, эмпирическая ф.р. $F_n(x) \underset{p}{\to} F(x)$, а ф.р. максимума отклонения *F*$_n$*(x)* от *F(x)* на всем протяжении оси *x* при $n \to \infty$ имеет пределом некоторую функцию, не зависящую от вида функции *F(x)*.

Мы опустим здесь довольно сложное доказательство теоремы Колмогорова, но рассмотрим следующие основные положения теории выборочного метода для порядковых статистик, которые лежат в основе этого доказательства и представляют самостоятельный интерес.

2.2. Выборочные распределения порядковых статистик

Пусть $\tau_n = (x_1,...,x_n)$ – выборка из совокупности, имеющей непрерывную ф.р. $F(x)$, $x_{(k)}$ – *к-я* порядковая статистика вариационного ряда $\tau_{(n)} = (x_{(1)}, x_{(2)},..., x_{(n)})$.

Интервалы $(\infty, x_{(1)}]$, $(x_{(1)}, x_{(2)}],...,(x_{(n)}, +\infty)$ называются выборочными блоками, а функции $F(x_{(1)})$, $F(x_{(2)}) - F(x_{(1)}),...,1- F(x_{(n)})$ этих блоков – выборочными долями $u_1,...u_{n+1}$ соответственно. Таким образом, доля для данного выборочного блока есть сумма вероятностей в распределении элементов выборки, попадающих в этот блок, и является случайной величиной.

Для дальнейшего изложения нам понадобится следующая важная теорема.

Теорема 12. Если X – с.в., имеющая непрерывную ф.р. $F(x)$, то с.в. $Y = F(X)$ имеет равномерное (см. п. 1.5.2., г)) распределение $R(1/2,1)$ независимо от $F(x)$.

Доказательство этой теоремы тривиально, так как ф.р. с.в. Y равна:

$$F_Y(y) = P\{F(X) < y\} = P(X < F^{-1}(y)) = \begin{cases} 1, & y > 1, \\ y, & 0 < y \le 1, \\ 0, & y \le 0. \end{cases}$$

Пусть теперь $y_i = F(x_i)$, $i = 1,......, n$. Тогда $(y_1,...,y_n)$ есть выборка объема n из совокупности, имеющей равномерное распределение $R(1/2,1)$. Ей соответствует элемент вероятности $1\times dy_1 ... dy_n$ в выборочном пространстве R^n внутри единичного «куба»

$\{ (y_1,...,y_n) : 0 < y_i < 1, \quad i = \overline{1,n} \}$ и *0* вне этого «куба».□

Элемент вероятности для порядковых статистик выборки

$y_{(1)} \le y_{(2)} \le ... \le y_{(n)}$ может быть получен в том же единичном «кубе» путем перестановки координат y_i $n!$ способами.

Поэтому элемент вероятности $\left(y_{(1)},...,y_{(n)}\right)$ есть $n! \times d_{(1)} \ldots d_{(n)}$ внутри $\left\{(y_1,...,y_n): 0<y_i<1,\ i=\overline{1,n}\right\}$ и *0* вне этой области. Это есть *n*-мерное упорядоченное распределение Дирихле D^* *(1,...;1)* (п.1.5.2., з)).

Итак, мы получили следующий результат:

Теорема 13. Если $(x_{(1)},...,x_{(n)})$ – порядковые статистики выборки с непрерывной ф.р. *F(x)*, то с.в. $F(x_{(1)}),...,F(x_{(n)})$ имеют упорядоченное *n*-мерное распределение Дирихле *D*(1,1,...,1;1)*.

Теорема 14. С.в. $F(x_{(k)})$, $1\le k\le n$ имеет бета-распределение *Be(k,n-k+1)*.

Доказательство этого утверждения следует непосредственно из (п.1.5.2., з)), если заметить, что каждое $y_{(k)}(k=1,2,...n)$ из $\left(y_{(1)},y_{(2)},...,y_{(n)}\right)\in D^*(\lambda_1,...\lambda_n;\lambda_{n+1})$ может быть представлено в виде $y_{(k)}=x_1+...+x_k$, где $(x_1+...+x_k)\in D(\lambda_1,\lambda_2,...,\lambda_n;\lambda_{n+1})$. Но согласно (п.1.5.2.,з)) $x_1+...+x_k$ имеет бета – распределение $Be(\lambda_1,...,\lambda_k;\lambda_{n-k+1})$, причем $\lambda_1=\lambda_2=...=\lambda_k=\lambda_{k+1}=1$. Отсюда следует утверждение теоремы.

В более часто встречающейся форме результат теоремы 14 можно сформулировать так: вероятность $P_k^n(x)$ равенства $F_n(x)=\dfrac{k}{n}$, где *k* – число элементов выборки $x_k<x$ представляется биномиальным членом

$$P_k^n(x)=\frac{n!}{k!(n-k)!}\{\,F(x)\,\}^k\{1-F(x)\,\}^{(n-k)}. \tag{2.2}$$

Действительно, обозначим через $A_x^{(k)}$ событие, равносильное осуществлению неравенства $x_k<x$. Тогда события $A_x^{(k)}(k=1,2,...,n)$ при данном *x* независимы и $P\left\{A_x^{(k)}\right\}=F(x)$. Обозначим через $F_n(x)$ частоту положительных исходов *n* независимых испытаний относительно событий $A_x^{(1)},A_x^{(2)},\ldots,A_x^{(n)}$ с вероятностью успеха в каждом испытании, равной *F(x)*. Так как мы находимся, очевидно, в условиях классической схемы Бернулли (см. п.1.5.,а), то имеем (2.2).

Замечание. В связи с (1) уже самые элементарные соображения, основанные на теореме Бернулли (теоремы 3,3’), подсказывают предположение, что чем больше объем выборки, тем более точно аппроксимирует $F_n(x)$ ф.р. $F(x)$ (см. теоремы 10 и 11). Это предположение мы полностью оправдаем в дальнейшем.

Теорема 15 (о распределении *к*-й порядковой статистики). Элемент вероятности для $x_{(k)}, 1 \le k \le n$ при условии, что $F(x)$ непрерывна и имеет п.в. $p(x)$, есть

$$\frac{n!}{(k-1)!(n-k)!}\left[F(x_{(k)})\right]^{k-1} \quad \left[1-F(x_{(k)})\right]^{n-k} P(x_{(k)})\, dx_{(k)} \tag{2.3}$$

и выборочное пространство для $x_{(k)}$ есть R_1.

Действительно, пусть

$$\Phi_{nk}(x) = P\{\, x_{(k)} < x\}; \tag{2.4}$$

с другой стороны,

$$P\{\, x_{(n)} < x\} = P\left\{F_n(x) \ge \frac{k}{n}\right\}, \tag{2.5}$$

так как для осуществления неравенства $x_{(k)} < x$ необходимо и достаточно, чтобы не менее k членов в исходной последовательности с.в. $x_i (i=1,2,...,n)$ были меньше x.

Из (2.2), (2.4) и (2.5) следует, что

$$P\left\{F_n(x) \ge \frac{k}{n}\right\} = \sum_{m=k}^{n} P_m^{(n)}(x) = \sum_{m=k}^{n} \frac{n!}{m!(n-m)!}\{F(x)\}^m \{\, 1-F(x)\}^{n-m} \tag{2.6}$$

и отсюда также

$$\Phi_{nk}(x) = \frac{n!}{(k-1)!(n-k)!}\int_0^{F(x)} x^{k-1}(1-x)^{n-k}\, dx\,. \tag{2.7}$$

Справедливость последней формулы легко проверяется последовательным интегрированием по частям в (6), в силу чего получаются суммы (5). Так как X имеет плотность $p(x) = F'(x)$, то из (6) получаем плотность $P_{n,k}(x) = \Phi'_{n,k}(x)$ *к-го* члена вариационного ряда, которая будет иметь вид (2). Теорема доказана.

При *k=1* из (2) можно получить элемент вероятности для наименьшего элемента выборки, при $k = n$ – для наибольшего элемента выборки.

Рассмотрим вопрос о выборочном *распределении одномерных долей*.

Сделаем следующее преобразование:

$$
\begin{aligned}
&y_{(1)} = u_1,\\
&y_{(2)} = u_1 + u_2,\\
&\dots\dots\dots\dots\dots\dots\\
&y_{(n)} = u_1 + \dots + u_n.
\end{aligned}
$$

Элемент вероятности с.в. $u_1,\dots,u_n$ будет равен $n!\, du_1 \cdot \dots \cdot du_n$ в симплексе $S_n : \left\{ (u_1,\dots u_n) : u_i > 0,\ i = 1,\dots,n,\ \ \sum_{i=1}^{n} u_i < 1 \right\}$ и 0 вне S_n. А это есть n-мерное распределение Дирихле $D(1,1\dots,1;1)$. Но $u_1,\dots,u_n$ – доли и их распределение не зависит от $F(x)$, которая по предположению непрерывна. Отсюда является справедливой следующая теорема.

Теорема 16. Пусть $(X_{(1)},\dots,X_{(n)})$ – порядковые статистики выборки с непрерывной ф.р. $F(x)$. Тогда доли

$$u_1 = F(x_{(1)}),\quad u_2 = F(x_{(2)}) - F(x_{(1)}),\dots,\quad u_n = F(x_{(n)}) - F(x_{(n-1)})$$

есть с.в., имеющие n-мерное распределение Дирихле $D(1,\dots,1;1)$.

Теорема 17. Сумма любых k долей имеет бета-распределение $Be(k,n-k+1)$. Действительно, нетрудно заметить, что теорема 14 есть частный случай теоремы 17, так как $F(x_{(k)})$ есть сумма первых k долей $u_1,\dots,u_n$. К тому же распределение долей полностью симметрично относительно своих переменных. Но, в соответствии с п.1.5., з, сумма любых $k \le n$ с.в., имеющих совместное распределение Дирихле, имеет бета-распределение с соответствующими параметрами.

Учитывая п.1.5.з, заметим также, что справедлив следующий более общий результат.

Теорема 17'. Случайные величины $F(x_{(k_1)}), F(x_{(k_1+k_2)}),\dots, F(x_{(k_1+\dots+k_s)})$ имеют упорядоченное s-мерное распределение Дирихле $D^*(k_1,\dots,k_s; n-k_1-\dots-k_s+1)$.

Замечание. В настоящем разделе основное внимание было уделено эмпирической ф.р. и связанным с ней статистикам. Обычно используемые в математической статистике сводные характеристики эмпирического распределения, например, среднее, выборочная дисперсия, асимметрия и т.п. (см. п.2.3.2) представляют собой некоторые функционалы, зависящие от лестничных кривых $F_n(x)$. Так, для среднего $\bar{x}$ и эмпирической дисперсии s^2 имеем

$$\bar{x} = \int_{-\infty}^{\infty} x dF_n(x), \qquad s^2 = \int_{-\infty}^{\infty} (x-\bar{x})^2 dF_n(x).$$

Действительно, пусть (без ограничения общности) все элементы выборки $\tau_n=\{x_1,...,x_n\}$ различны. Тогда, например,

$$\int_{-\infty}^{\infty} x dF_n(x) = \sum_{k=1}^{n} x_k \frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k = \bar{x}.$$

Поэтому изучение поведения эмпирической кривой распределения представляет весьма большой интерес для статистики.

2.3. Выборочные характеристики и основные понятия теории статистической оценки параметров

Нахождение неизвестной ф.р. $F(x)$ по $F_n(x)$ часто не является практически необходимым. Обычно бывает заранее известно, что ф.р. $F(x)$ принадлежит к определенному классу функций (семейству P={ $P_\theta, \Theta=\{\theta_1,\theta_2,...,\theta_k\}$} распределений, определенных на соответствующем измеримом пространстве <R^n, B (R^n)>), зависящему от одного или нескольких параметров $F_X(x)=F(x,\theta_1,...,\theta_k)$. В этом случае определение неизвестной ф.р. сводится к оценке неизвестных параметров по исходной выборке $\tau_n=\{x_1,...,x_n\}$, т.е. к определению характеристик выборки. Тройку <R^n,B , P > называют параметрической статистической структурой (подробнее см. в гл.5).

Определение. Оценкой параметра θ называется приближенное значение этого параметра, полученное по выборке.

Всякая оценка неизвестного параметра является статистикой $g_\theta(\tau_n)$, не содержащей этот параметр.

Возникает вопрос о выборе вида функции $g\theta(\tau n)$, для оценки неизвестного параметра θ ф.р. F(x, θ), т.е. вопрос о качестве той или иной оценки.

2.3.1. Классификация оценок

Определение. Оценка $g_\theta(\tau_n)$, обладающая следующим свойством:

$M\ g_\theta(\tau_n)=\theta$ называется несмещенной оценкой для θ. В противном случае оценка называется смещенной.

Определение. Оценка $g_\theta(\tau_n)$ называется состоятельной, если она при $n\to\infty$ сходится по вероятности к θ, т.е. если $g_\theta(\tau_n)\underset{p}{\to}\theta$ при $n\to\infty$.

Определение. Пусть $\sigma^2(g_\theta(\tau_n))$ – дисперсия оценки $g_\theta(\tau_n)$. Оценка $g_\theta(\tau_n)$ называется эффективной в данном классе оценок, если она обладает минимальной дисперсией по сравнению со всеми оценками данного класса.

Определения достаточной статистики (оценки). Теорема факторизации

Пусть дана статистическая структура $<R^n, B, P>$.

Определение 1. Статистика (оценка) $y = g_\theta(\tau_n)$ называется достаточной для параметра θ ф.р. $F(x;\ \theta)=P_\theta\ (X<x)\ \equiv P(X<x;\ \theta)$, если $P_\theta(\tau_n\in A/t)=P(\tau_n\in A/t)$, то есть условная вероятность любого события $A\in R^n$, вычисленная при данном значении t достаточной статистики, не зависит от параметра θ.

Замечание. Статистику $g_\theta(\tau_n)$; называют достаточной для семейства P функций распределения, если условная вероятность $P(\tau_n\in A/t)$ является одной и той же для всех ф.р. из P.

Определение 2. $y = g(\tau_n)$, $\tau_n=\{x_1,\ldots,x_n\}$ называется достаточной (для параметра), если

$$p(\tau_n;\theta)=\varphi(g(\tau_n);\theta)h(\tau_n), \tag{2.8}$$

где $h(\tau_n)$ не зависит от параметра θ, то есть плотность выборочного распределения $p(\tau_n;\theta)$ можно разбить на два множителя (при этом множитель, зависящий от параметра, зависит от исхода только через достаточную статистику, а множитель, зависящий от самого исхода, не содержит параметра θ).

Замечание. Доказательство эквивалентности определений 1 и 2 составляет содержание так называемой «теоремы факторизации». Таким образом, теорема факторизации позволяет конструктивно определить достаточную статистику выполнением условия (1) (так называемого *«критерия факторизации»*).

Доказательство теоремы факторизации [9]. Пусть, например, с.в. X имеет дискретное распределение. Если выполнено (1) и, поскольку $p_\theta(\tau_n; g(\tau_n)=t) = p_\theta(\tau_n)$, то

$$p_\theta\{\tau_n / g(\tau_n) = t\} = \frac{P(\tau_n;\theta)}{P_\theta(g(\tau_n)=t)} = \frac{P(\tau_n;\theta)}{\sum\limits_{x:\,g(x)=t} p(x;\theta)} = \frac{\varphi(t;\theta)h(\tau_n)}{\varphi(t;\theta)\sum\limits_{x:\,g(x)=t} h(x)} = \frac{h(\tau_n)}{\sum\limits_{x:\,g(x)=t} h(x)}$$

не зависят от θ. С другой стороны, если $p_\theta(\tau_n / g(\tau_n)=t) = h(\tau_n, t)$ не зависит от θ, то $p_\theta(\tau_n) = p_\theta(g(\tau_n)=t)h(\tau_n,t)$. Отсюда следует определение 2. В непрерывном случае доказательство аналогично.

Пример 1. Пусть $p_X(x;\theta) = \theta\, e^{-\theta x}, \quad x \ge 0, \quad \theta > 0,$ отсюда $p(\tau_n,\theta) = \prod\limits_{i=1}^{n} \theta\, e^{-\theta x_i} = \theta^n e^{-\theta \sum\limits_{i=1}^{n} x_i} = \theta^n e^{-\theta y}$, тогда $y = \sum\limits_{i=1}^{n} x_i$ – достаточная статистика $(h(\tau_n)=1)$.

Пример 2. Пусть $X \in N(\theta,\sigma^2), \quad -\infty < \theta < \infty, \quad \sigma^2$ – известна.

Поскольку можно представить

$$p(\tau;\theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\sum_{i=1}^{n}(x_i-\theta)^2 / 2\sigma^2\right\} =$$

$$= \left\{\exp\left[-n(\bar{x}-\theta)^2/2\sigma^2\right]\right\} \times \left\{\frac{\exp\left[-\sum\limits_{i=1}^{n}(x_i - x)^2/2\sigma^2\right]}{(\sigma\sqrt{2\pi})^n}\right\},$$

то $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ является достаточной статистикой для θ.

Иначе: «Статистики, содержащие всю информацию, доставленную экспериментом, называются достаточными статистиками. Когда нам известны достаточные статистики, то сам исход не нужен (не необходим, не несет большой информации)» Р. Фишер.

$\tau_n=\{x_1,...,x_n\}$ также является достаточной статистикой для любого распределения (и параметра) и называется тривиальной.

Таким образом, не всякая достаточная статистика является практически необходимой.

Достаточная статистика минимальной размерности называется *минимальной достаточной статистикой* (более строгое определение см. п.5.2).

Пример 3. В примерах 1,2 статистики являются минимальными достаточными (размерность каждой из них равна 1).

Замечание. Из доказательства теоремы факторизации можно заметить, что если статистика $g(\tau_n)$ является достаточной, то для любой функции $f(x)$, задающей взаимно-однозначное отображение, статистика $f(g(\tau_n))$ также является достаточной. Заметим также, что каждая эффективная оценка является одновременно и достаточной (см. также гл. 5).

2.3.2. Некоторые выборочные характеристики, их свойства, выборочные распределения и числовые характеристики

Каждая характеристика $g(x_1,...,x_n)$ выборки может рассматриваться как наблюденное значение с.в. $g(x_1,...,x_n)$, где $x_1,...,x_n$ – независимые с.в., имеющие то же распределение, что и с.в. X. Распределение с.в. $g(x_1,...,x_n)$ называется *выборочным распределением характеристики* $g(x_1,...,x_n)$.

Выборочное среднее

Пусть $\tau_n = (x_1,...,x_n)$ – выборка с.в. X, имеющей конечные математическое ожидание и дисперсию $MX=a$. $DX=\sigma$.

Оценка вида $\bar{x}=\frac{1}{n}(x_1+...+x_n)$ называется выборочным средним для τ_n.

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

Но x_i – независимые с.в., имеющие одну и ту же ф.р.

Тогда $Mx_1 = \ldots = Mx_n = a$. Отсюда $\bar{x}$ имеет среднее значение $M(\bar{x}) = a$, т.е. оценка $\overline{X}$ — несмещенная для a.

$$D(\bar{x}) = \frac{1}{n^2} \sum_{i=1}^{n} Dx_i = \frac{\sigma^2}{n} .$$

Выборочная дисперсия.

Выборочная дисперсия определяется следующей формулой:

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

Иначе

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} [(x_i - a) - (\bar{x} - a)]^2 = \frac{1}{n-1} (\sum_{i=1}^{n} (x_i - a)^2 - 2(\bar{x} - a) \sum_{i=1}^{n} (x_i - a) + (\bar{x} - a)^2 \times n) .$$

Учитывая, что

$$(\bar{x} - a) \sum_{i=1}^{n} (x_i - a) = (\bar{x} - a) [\sum_{i=1}^{n} x_i - na] = (\bar{x} - a) \times (n\bar{x} - na) = n(\bar{x} - a)^2,$$

так как $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, получаем следующую формулу для s^2:

$$s^2 = \frac{1}{n-1} [\sum_{i=1}^{n} (x_i - a)^2 - n(\bar{x} - a)^2].$$

Тогда

$$Ms^2 = \frac{1}{n-1} M \left[\sum \left(x_i - a \right)^2 - n\left(\bar{x} - a \right)^2 \right] =$$

$$= \frac{\sum \left(x_i - a \right)^2 - nM\left(\bar{x} - a \right)^2}{n-1} = \frac{n\sigma^2}{n-1} - \frac{\sigma^2}{n-1} = \sigma^2 .$$

Тем самым мы показали, что s^2 – оценка, несмещенная для σ^2.

Поскольку $D(s^2) = M[(s^2)^2] - [M(s^2)]^2$, после простых преобразований можно получить, что

$$D(s^2) = \frac{1}{n} (\nu_4 - \frac{n-3}{n-1} \sigma^4),$$

где ν_4 – четвертый центральный момент с.в. X, если он существует.

***K*-статистики Фишера, старшие центральные моменты и семиинварианты.**

Все моменты и семиинварианты выборки конечны, и соотношения между моментами и семиинвариантами остаются справедливыми, если характеристики генеральной совокупности (вероятностные характеристики) заменить выборочными характеристиками.

Начальные и центральные выборочные моменты определяются соответственно формулами:

$$\alpha_s = \frac{1}{n}\sum_{i=1}^{n} x_i^s \ ; \ \ m_s = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^s.$$

Если обозначить $\varphi_n(t)$ характеристическую функцию выборки

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} e^{itx_j} \ ,$$

то семиинвариантами выборки будут величины, определяемые разложением

$$\log \varphi_n(t) = \sum_{s=1}^{\infty} \frac{k_s}{s!}(it)^s \ .$$

Те же правила будут употребляться и для выборок из многомерных совокупностей.

В силу теоремы Хинчина (закон больших чисел) (см. теорему 1) и соотношения $M(a_s) = \frac{1}{n}\sum_{i=1}^{n} M(x_i^s) = \mu_s$, выборочный момент сходится по вероятности к соответствующему начальному моменту при $n \to \infty$, если μ_s существует.

Подобным же образом все центральные m_s и семиинварианты k_s выборки являются состоятельными оценками соответствующих величин v_s и κ_s, но при $s>1$ не являются несмещенными. Добиться несмещенности этих оценок можно введением некоторых поправок, таких, например, как коэффициент $\frac{n}{n-1}$ в формуле выборочной дисперсии. Так, например, для $s=2,3,4$ получим следующие исправленные оценки M_s для v_s

$$M_2 = \frac{n}{n-1} m_2,$$

$$M_3 = \frac{n^2}{(n-1)(n-2)} m_3,$$

$$M_4 = \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} m_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} m_2^2.$$

При больших n часто бывает безразлично, используются M_s или m_s, но для малых n смещение весьма существенное.

Фишер ввел функции $K_s(x_1,...,x_n),\ s=1,2,...,$ являющиеся наиболее общими однородными полиномами степени s от переменных $x_1,...,x_n$ и подчиняющиеся при этом следующим условиям:

1. Функция $K_s(x_1,...,x_n)$ симметрична относительно $x_1,...,x_n$.

2. $M(K_s)=\kappa_s$.

Такие функции называются κ-статистиками. Первые три κ-статистики, очевидно, имеют вид:

$$K_1 = a_1 \sum_i x_i,$$
$$K_2 = a_{21} \sum_i x_i^2 + a_{22} \sum_{i \neq j} x_i x_j, \tag{2.9}$$
$$K_3 = a_{31} \sum_i x_i^3 + a_{32} \sum_{i \neq j} x_i^2 x_j + a_{33} \sum_{i \neq j,\, j \neq r} x_i x_j x_r,$$

Беря средние значения и приравнивая их первым трем семиинвариантам распределения совокупности

$$\kappa_1 = \mu_1 = M, \quad \kappa_2 = \mu_2 - (\mu)^2 = \sigma^2,$$
$$\kappa_3 = \mu_3 - 3\mu_1\mu_2 - 2(\mu_1)^3.$$

определим искомые коэффициенты.

Подставив эти коэффициенты в (2.9), получим для первых трех κ-статистик следующие формулы:

$$k_1 = \frac{1}{n} z_1 = \bar{x}, \quad k_2 = \frac{1}{n(n-1)}(nz_2 - z_1^2) = s^2,$$

$$k_3 = \frac{1}{n(n-1)(n-2)}(n^2 z_3 - 3n z_1 z_2 + 2 z_1^3).$$

где $z_j = \sum_i x_i^j, \quad j=1,2,3, \quad i=\overline{1,n}$.

Благодаря комбинаторным особенностям вычисление *к*-статистик проще вычисления оценок моментов.

Замечание. Резкое увеличение производительности компьютеров в последние годы стимулирует использование более сложных в вычислительном отношении методов получения несмещенных оценок (оценок «складного ножа», «бутстреп»-оценок, робастных оценок на основе функций влияния и т.п. [34])

2.4. Распределения выборочных сумм и средних значений

Случайные величины X_i $(i=1,2,...,n)$ независимы и одинаково распределены, следовательно, ф.р. их суммы $z_n = x_1+...+x_n$, $F_{x_1+x_2+...+x_n}(x) = (F_{x_1} * p_{x_2} * ... * p_{x_n})(x) = F^{*n}(x)$, плотность распределения $p_{x_1+x_2+...+x_n}(x) = p_{x_1} * ... * p_{x_n})(x) = p^{*n}(x)$, характеристическая функция z_n равна $\varphi_{z_n} = (\varphi(t))^n$, а характеристическая функция выборочного среднего равна $[\varphi(t/n)]^n$.

Пусть X_1 и X_2 – независимые с.в., имеющие соответственно ф.р. $F(x_1;\theta_1)$ и $F(x_2;\theta_2)$ и характеристические функции $\varphi(t;\theta_1)$ и $\varphi(t;\theta_2)$ где θ_1 и θ_2 – параметры. Из теории вероятностей известна следующая теорема.

Теорема 18. Для того чтобы $F(x,\theta)$ была воспроизводимой по θ, необходимо и достаточно выполнения условия:

$$\varphi(t;\theta_1)\varphi(t;\theta_2) = \varphi(t;\theta_1 + \theta_2).$$

Доказательство теоремы 18 представляется читателю в качестве упражнения.

Если ф.р. $F_n(x;\theta)$ выборки является воспроизводимой по отношению к θ, то $[\varphi(t, \theta)]^n = \varphi(t;n\theta)$, то есть z_n имеет ф.р. $F(z,n\theta)$.

Таким образом, в качестве следствия теоремы 18 мы получим следующее утверждение:

Следствие. Пусть $\tau_n = (x_1,...,x_n)$ – выборка из генеральной совокупности с ф.р. $F(x;\theta)$ и х.ф. $\varphi(t;\theta)$. Тогда, если $[\varphi(t, \theta)]^n = \varphi(t;n\theta)$, то ф.р. выборочной суммы z_n и выборочного среднего $\bar{x}$ есть $F(z,n\theta)$ и $F(n\bar{x},n\theta)$ соответственно.

Пример 1. Пусть $\tau_n = (x_1,...,x_n)$ – выборки из биномиального распределения $Bi(nm,p)$. Действительно,

$$\varphi_{x_k}(t) = (g + e^{it}p)^m, \quad k = \overline{1,n}, \text{ а } \quad \varphi_{z_n}(t) = (g + e^{it}p)^{mn}.$$

Пример 2. Пусть $\tau_n = (x_1,...,x_n)$ – выборки из $N(\mu,\sigma^2)$, тогда распределение выборочной суммы z есть $N(n\mu, n\sigma^2)$, а выборочного среднего $\bar{x}$ есть $N(\mu,\sigma^2/n)$.

Замечание. Если ф.р. выборки не воспроизводимая, то х.ф. для нахождения распределения выборочных сумм лучше не пользоваться.

2.5. Распределение квадратичных и полиномиальных форм нормальной выборки

2.5.1. Теория выборочного метода для некоторых квадратичных форм в выборках из нормального распределения. Распределение χ^2 и Стьюдента

Теорема 19. Если $(x_1,...,x_n)$ — выборка из нормального распределения $N(\mu,\sigma^2)$, то выборочное распределение суммы квадратов $\frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k - \mu)^2$ есть $\chi^2_{(n)}$распределение (см. п.1.5, д).

Доказательство. Пусть $y_k = \dfrac{x_k - \mu}{\sigma} \in N(0,1)$, тогда

$$\chi^2_n = \sum_{k=1}^{n} y_k, \; y_k \text{ – независимы.}$$

Найдем (см. п. 1.2) распределение с.в. $z = y^2$ как монотонно возрастающей функции $y_{1,2} = \phi(z) = \pm\sqrt{z}$, т.к. $z>0$.

Тогда $p(y) = \dfrac{1}{\sqrt{2\pi}} e^{-y^2/2}, \quad |\phi'(z)| = \dfrac{1}{2\sqrt{z}}$

Отсюда получаем

$$p(z) = \frac{1}{\sqrt{2\pi}}\left[e^{-\frac{(\sqrt{z})^2}{2}} + e^{-\frac{(-\sqrt{z})^2}{2}}\right] \times \frac{1}{2\sqrt{z}} = \begin{cases} \dfrac{1}{\sqrt{2\pi z}} \times e^{-(z/2)}, & z > 0, \\ 0 & , z \le 0. \end{cases} \quad (2.10)$$

Это распределение является частным случаем гамма-распределения, плотность которого имеет вид (см. п.1.5, в)

$$p(x,\alpha,\lambda)=\begin{cases}\dfrac{\alpha^{\lambda}}{\Gamma(\lambda)}x^{\lambda-1}e^{-\alpha x} & \text{при} \quad x>0,\\ 0 & \text{при} \quad x\le 0.\end{cases} \qquad (2.11)$$

Действительно, если $\alpha=\lambda=\dfrac{1}{2}$, то получим (2.11). Найдем х.ф.

$$\varphi(t,\alpha,\lambda)=\int_{-\infty}^{\infty}e^{itx}p(x;\alpha,\lambda)dx=\frac{\alpha^{\lambda}}{\Gamma(\lambda)}\int_{0}^{\infty}x^{\lambda-1}e^{-(\alpha-it)x}dx,$$

но $\Gamma(p)=\int_{0}^{\infty}x^{p-1}e^{-x}dx, \quad \forall \quad p>0$.

Если к тому же положить $(\alpha\text{-}it)x = y$, то

$$x^{\lambda-1}e^{-(\alpha-it)x}dx=\frac{y^{\lambda-1}}{(\alpha-it)^{\lambda-1}}e^{-y}\frac{1}{\alpha-it}dy.$$

Отсюда получаем

$$\varphi(t;\alpha\lambda)=\frac{\alpha^{\lambda}}{\Gamma(\lambda)}\cdot\frac{\Gamma(\lambda)}{(\alpha-it)^{\lambda}}=\frac{1}{(1-\frac{it}{\alpha})^{\lambda}}. \qquad (2.12)$$

Таким образом, х.ф. распределения с.в. $z_k=y_k^2$ представится в виде

$$\varphi_{z_k}(t)=\varphi\left(t;\frac{1}{2},\frac{1}{2}\right)=\frac{1}{\sqrt{1-2it}}=(1-2it)^{-1/2}.$$

Соответственно, ввиду независимости z_k, х.ф. с.в. $\sum_{k=1}^{n}z_k$ равна

$$\varphi_{\chi_n^2}=(1-2it)^{-n/2}.$$

Но $\varphi_{\chi_n^2}$ может быть получена из (2.12) при $\alpha=\dfrac{1}{2}$ и $\lambda=\dfrac{1}{2}n$, т.е.

$$P_n(x)=P(x,\frac{1}{2},\frac{1}{2}n).$$

Итак, мы показали, что при любом n (см. п.1.5,д)

$$P_n(x)=\begin{cases}\dfrac{1}{2(n/2)\Gamma(n/2)}(x)^{(n/2)}e^{-(\lambda/2)} & при \quad x>0,\\ 0 & при \quad x\le 0.\end{cases}$$

Теорема доказана.

Число n степеней свободы в распределении $\chi^2_{(n)}$ есть ранг квадратичной формы $Q=xAx=\sum_{i,k=1}^{n}a_{i,k}x_ix_k$, т.е. наименьшее число независимых величин, к которым может быть приведена форма Q с помощью неособенного линейного преобразования (ранг матрицы A).

Замечание. Квадратичная форма Q не обязательно имеет распределение χ^2_n.

Теорема 20. Для того чтобы квадратичная форма $x'Ax$ имела распределение χ^2, необходимо и достаточно выполнение условия $A^2=A$, то есть A – идемпотентная матрица, в этом случае число степеней свободы χ^2_n равно $R(A)=TrA$. (Доказательство этого утверждения можно найти в книге Рао С.Р. [16, с. 164]).

Теорема 21. Если $(x_1,...,x_n)$ – выборка из нормального распределения $N(\mu,\sigma^2)$, то выборочное среднее $\bar{x}=\frac{1}{n}\sum_{k=1}^{n}x_k$ и выборочная дисперсия $s^2=\frac{1}{n-1}\sum_{k=1}^{n}(x_k-\bar{x})^2$ статистически независимы.

Величина $\sqrt{n}(\bar{x}-\mu)/\sigma \in N(0{,}1)$, а величина $(n-1)s^2/\sigma^2 \in \chi^2_{n-1}$.

Доказательство. Мы уже знаем (пример 2 п.2.4), что $\bar{x}$ имеет распределение $N(\mu,\sigma^2/n)$, что эквивалентно утверждению о том, что $\sqrt{n}(\bar{x}-\mu)/\sigma$ имеет распределение $N(0{,}1)$. Докажем оставшиеся утверждения.

s^2 не изменится, если все x_k заменить на $x_k^I=x_k-\mu$. Иначе говоря, можно без ограничения общности считать, что $\mu=0$, то есть, что величины $x_k\in N(0,\sigma)$. Перейдем от x_k к величинам y_k, используя формулы,

$$y_k=\sum_{j=1}^{n}a_{kj}x_j \quad (k=1,2,...,n)\,, \tag{2.13}$$

где числа a_{kj} образуют квадратную матрицу A порядка n.

Пусть все элементы последней строки матрицы A равны $\frac{1}{\sqrt{n}}$, т.е. $a_{nj} = \frac{1}{\sqrt{n}}$, $j = 1,2,...,n$, тогда $y_n = \frac{1}{n}\sum_{j=1}^{n} x_j = \sqrt{n}\ \bar{x}$ и $\sum_{j=1}^{n} a_{nj}^2 = 1$.

Пусть при этом будут выполнены условия ортогональности (что всегда можно сделать):

$$\sum_{j=1}^{n} a_{kj} a_{ij} = \begin{cases} 1 & при \quad k = i, \\ 0 & при \quad k \neq i, \end{cases} \qquad k, i = 1,2,...,n.$$

Ортогональное преобразование означает вращение n-мерного пространства около начала координат, причем длины векторов и углы между ними не меняются, т.е. $\sum_{k=1}^{n} x_n^k = \sum_{k=1}^{n} y_k^2$.

Но распределение y_k, $k=1,2,...,n$ как суммы независимых нормальных с. в. нормально, причем $My_k = Mx_k = 0$;

$$Dy_k = My_k^2 = \sum_{j=1}^{n} a_{kj}^2 Mx_j^2 = \sigma^2 \sum_{j=1}^{n} a_{kj}^2 = \sigma^2.$$

В силу условия ортогональности при $k \neq i$

$$My_k y_i = \sum_{j=1}^{n} a_{kj} a_{ij} Mx_j^2 = \sigma^2 \sum_{j=1}^{n} a_{kj} a_{ij} = \sigma^2 = 0.$$

Это означает, что нормально распределенные с.в. $y_1, ..., y_n$ некоррелированы, а следовательно, и независимы (см. п.1.2).

Для величины s^2 имеем

$$(n-1)s^2 = \sum_{k=1}^{n} (x_k - \bar{x})^2 = \sum_{k=1}^{n} x_k^2 - n\bar{x}^2 = \sum_{k=1}^{n} y_k^2 - y_n^2 = \sum_{k=1}^{n-1} y_k^2.$$

Соответственно, $\frac{n-1}{\sigma^2} s^2 = \sum_{k=1}^{n-1} \left(\frac{y_k}{\sigma}\right)^2$, где $\frac{y_k}{\sigma}$, $(k = 1,2,...,n-1)$ − независимые нормально распределенные с. в. с параметрами 0,1. Но сумма квадратов таких величин имеет распределение χ^2 с $(n\text{-}1)$ степенями свободы (см. п.1.5, д).

Покажем [19, с.222], что характеристическая функция $\varphi(t_1,t_2)$ двух с. в. $\bar{x}$ и s^2 равна произведению их характеристических функций:

$$\varphi(t_1, t_2) = \varphi_1(t_1)\,\varphi_2(t_2). \tag{2.14}$$

При этом очевидно

$$\varphi_1(t_1) = e^{-\frac{1}{2}t_1^2},$$

$$\varphi_2(t_2) = (1-2it_2)^{-\frac{1}{2}(n-1)},$$

$$\varphi(t_1,t_2) = M\exp\left(it_1\right)(\bar{x}-\mu)\frac{\sqrt{n}}{\sigma} + it_2\frac{(n-1)s^2}{\sigma^2} =$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \int\limits_{\mathbf{R}_n} \exp\left\{-\frac{1}{2}[Q(x_1,...x_n) - 2it_1(\bar{x}-\mu)\frac{\sqrt{n}}{\sigma}]\right\} dx_1...dx_n, \tag{2.15}$$

где

$$Q(x_1,...,x_n) = \frac{1}{\sigma^2}\sum_{k=1}^{n}(x_k-\mu)^2 - \frac{2it_2}{\sigma^2}\sum_{k=1}^{n}(x_k-\bar{x})^2 .$$

Обозначив

$$\tau^{kk} = \frac{1}{\sigma^2}\left[1-2it_2\left(1-\frac{1}{n}\right)\right], \quad k=\overline{1,n},$$

$$\tau^{kj} = \tau^{jk} = \frac{2it_2}{n\sigma^2}, \quad k \neq j = \overline{1,n}, \tag{2.16}$$

получим $Q(x_1,...,x_n) = \sum_{k,j}^{n}\tau^{kj}(x_k-\mu)(x_j-\mu).$

Нетрудно заметить, что

$$Q(x_1,...,x_n) - \frac{2it_1}{\sigma}(\bar{x}-\mu)\sqrt{n} = Q'(x_1,...,x_n) + \left(\sum_{k,j}^{n}\tau_{kj}\right)\frac{t_1^2}{n\sigma^2}, \tag{2.17}$$

где $Q'(x_1,...,x_n) = Q(x_1 - ig_1,...,x_n - ig_n)$,

причем $\|\tau_{kj}\| = \|\tau^{kj}\|^{-1}$; $g_k = \frac{t_1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\tau_{kj}$, $k=\overline{1,n}.$

Учитывая простой вид матрицы $\|\tau^{kj}\|$, непосредственно находим элементы матрицы $\|\tau_{kj}\|$:

$$\tau_{kk} = \frac{\sigma^2(1-2i\frac{t_2}{n})}{(1-2it_2)}, \quad k=\overline{1,n},$$

$$\tau_{kj} = -\frac{\sigma^2(2i\frac{t_2}{n})}{(1-2it_2)}, \quad k \neq j = \overline{1,n}, \tag{2.18}$$

а также $\sum_{k,j=1}^{n}\tau_{kj} = n\sigma^2.$

Подставляя (2.18) в (2.17), а затем полученное выражение в (2.15), имеем

$$\varphi(t_1,t_2)=e^{\frac{1}{2}t_1^2}\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n\int\limits_{\mathbf{R}_n} e^{-\frac{1}{2}Q'(x_1,\dots x_n)}dx_1\dots dx_n.$$

Но $\int\limits_{\mathbf{R}_n} e^{-\frac{1}{2}Q(x_1,\dots x_n)}dx_1\dots dx_n=\sqrt{(2\pi)^n}\left|\tau^{kj}\right|^{-\frac{1}{2}}.$

Поэтому, замечая (2.16), что $\left|\tau^{kj}\right|=\frac{1}{\sigma^{2n}}(1-2it_2)^{n-1}$, убеждаемся в справедливости (2.14), т.е. независимости $\bar{x}$ и s^2.

Теорема доказана.

Теорема 22. Если *$(x_1,\dots,x_n)$* – выборка из *$N(\mu,\sigma^2)$,* то $t=\sqrt{n}(\bar{x}-\mu)/s$ имеет распределение Стьюдента с *n-1* степенями свободы.

Доказательство. Пусть $u=\sqrt{n}(\bar{x}-\mu)/\sigma,\quad v=(n-1)s^2/\sigma^2$, тогда

$$t=\frac{u}{\sqrt{v/(n-1)}}=\sqrt{n}(\bar{x}-\mu)/s.$$

Но $u\in N(0,1)$, $v\in\chi^2_{n-1}$, они независимы, их совместная плотность

$$p\,(u,v)=\frac{1}{2\sqrt{2\pi}\Gamma(n/2)}\left(\frac{v}{2}\right)^{\frac{1}{2}(n-2)}e^{-\frac{1}{2}(u^2+v)}.$$

Применим к с.в. *(u,v)* преобразование

$$s=v,\ \ t=\frac{u}{\sqrt{v/(n-1)}}.$$

Якобиан преобразования равен $\sqrt{s/(n-1)}$ и, следовательно,

$$p(s,t)=\frac{1}{2\sqrt{2\pi(n-1)}\Gamma(n/2)}\left(\frac{s}{2}\right)^{\frac{1}{2}(n-1)}e^{-\frac{1}{2}\left[1+\frac{t^2}{n-1}\right]s}.$$

Но маргинальная плотность

$$\int\limits_0^\infty p(s,t)ds=\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1+\frac{t^2}{k}\right)^{-\frac{1}{2}(k+1)},$$

где $k=n-1$.

А это есть п.в. распределения Стьюдента $S(n\text{-}1)$ (см. п.1.5,е.).

Замечание к использованию распределения Стьюдента.

Распределение Стьюдента не содержит μ и σ. Как только известно μ, можно вычислить t по выборочным значениям и сравнить вычисленное значение t с теоретическим, т.е. определить, насколько выборочное среднее $\bar{x}$ отклоняется от μ.

Распределение Стьюдента применяется и для вычисления отклонения между двумя средними значениями двух независимых выборок (критерий значимости разности двух средних значений). Делается это следующим образом.

Пусть $x_1,...,x_n$ и $y_1,...,y_n$ – две независимые выборки, извлеченные из некоторой нормальной совокупности. Без ограничения общности пусть $\mu = 0$.

Обозначим $\bar{x} = \frac{1}{n}\sum_1^{n_1} x_i$ и $s_1^2 = \frac{1}{n}\sum_1^{n_1}(x_i - \bar{x})^2$ – среднее и дисперсия первой выборки, а $\bar{y}$ и s_2^2 – второй. Заменим все $n_1 + n_2$ величины новыми величинами $z_1,...,z_{n_1+n_2}$. $z_1 = \sqrt{n_1}\,\bar{x}$ и $z_2 = \sqrt{n_2}\,\bar{y}$. Тогда квадратичная форма

$$Q = n_1 s_1^2 + n_2 s_2^2 = \sum_1^n x_i^2 + \sum_1^n y_i^2 - n_1\bar{x}^2 - n_2\bar{y}^2$$

преобразуется в $Q = \sum_3^{n_1+n_2} z_i^2$, откуда видно, что ранг (или число степеней свободы) формы Q равен $n_1 + n_2 - 2$.

Определим с.в. u соотношением

$$u = \sqrt{\frac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}}\,\frac{\bar{x}-\bar{y}}{\sqrt{Q}} = \sqrt{\frac{n_1 n_2(n_1+n_2-2)}{(n_1+n_2)(n_1 s_1^2 + n_2 s_2^2)}}\left(\bar{x}-\bar{y}\right) =$$

$$= \frac{\sqrt{\dfrac{n_2}{n_1+n_2}}z_1 - \sqrt{\dfrac{n_2}{n_1+n_2}}z_2}{\sqrt{\dfrac{1}{n_1+n_2-2}\sum_3^{n_1+n_2} z_i^2}} = \frac{\omega}{\sqrt{\dfrac{1}{n_1+n_2-2}\sum_3^{n_1+n_2} z_i^2}},$$

где числитель $\omega \in N(0,1)$, а знаменатель $\in \chi^2_{(n_1+n_2-2)}$. Так как u не содержит параметров μ и σ, то полученное по выборке значение u можно сравнить с теоретическим табулированным распределением Стьюдента с $n_1 + n_2 - 2$ степенями свободы и тем самым определить, насколько одно выборочное значение отличается от другого.

2.5.2. Распределение полиномиальных форм нормальной выборки

Рассмотрим полиномиальные статистики $P(x_1,...,x_n)$ повторной нормальной выборки $\tau_n = (x_1,...,x_n)$.

Ясно, что такие статистики имеют все моменты:

$$MP^r=\mu_r,\ r=1,2,\ldots$$

Для полиномиальных статистик нормальной выборки $P(x_1,...,x_n)$ общего вида пока неизвестно, определяются ли их распределения моментами, то есть решается ли в этом случае проблема моментов [22,с.134]. Для однородных полиномиальных статистик (форм) любой степени указанный вопрос решается следующей теоремой [6, теорема 11.1.5.], которую мы приведем без доказательства.

Теорема 22'. Пусть $P(x_1,...,x_n)$ – однородный полином степени $m \geq 1$. Распределение полинома $P(x_1,...,x_n)$ при нормальной выборке определяется заданием его первых $R_0=R_0\,(m)$ моментов

$$\mu_r = MP^r,\ r=1,2,\ldots,R_0.$$

Замечание. Х.ф. распределений линейных, квадратичных и некоторых полиномиальных форм нормальной выборки являются решениями дифференциальных уравнений весьма специального вида (см. [6, § 11.2]).

2.6. Асимптотическая теория выборочного метода для больших выборок

2.6.1. Сходимость и предельные распределения выборочных сумм и средних значений

Элементы выборки рассматривались нами как случайные величины, поэтому основные предельные теоремы для одномерного случая, упоминавшиеся в обзоре по теории вероятностей, могут быть легко сформулированы в терминах теории выборочного метода. Например, теорема Хинчина будет звучать так:

Теорема 23. Если $(x_1,...,x_n)$ – выборка из совокупности с ф.р. $F(x)$, для которой MX имеет конечное значение μ, , то выборочное среднее $\bar{x}$ сходится по вероятности к μ при $n \to \infty$.

Типичная центральная предельная теорема имеет следующий вид.

Теорема 24. Пусть $(x_1,...,x_n)$ – выборка из распределения с конечным средним μ и дисперсией σ^2. Тогда выборочная сумма z и выборочное среднее $\bar{x}$ распределены асимптотически нормально $N(n\mu,\sigma^2)$ и $N(\mu\sigma^2/n)$ соответственно.

Доказательство полностью аналогично известному из курса теории вероятностей доказательству центральной предельной теоремы для независимых одинаково распределенных случайных величин. Приведем его.

Пусть $\varphi_n(t)$ – характеристическая функция $\frac{z-n\mu}{\sqrt{n}\sigma}$, тогда

$$\varphi_n(t) = M\left[e^{it(z-n\mu)\sqrt{n}\sigma}\right] = M\left[\exp\left(it\sum_{k=1}^{n}(x_k-\mu)/\sqrt{n}\sigma\right)\right] = [\varphi(t)]^n,$$

где $\varphi(t)$ характеристическая функция $(x-\mu)/\sqrt{n}\sigma$. Воспользовавшись свойством 7° (см. п.1.3) характеристических функций, получаем

$$\varphi(t) = 1 + iM\left(\frac{x-\mu}{\sqrt{n}\sigma}\right)t - \frac{1}{2}M\left(\frac{x-\mu}{\sqrt{n}\sigma}\right)^2 t^2 + O(t^2/n),$$

где $n \times O(t^2/n) \to 0$ при $n \to \infty$ $\forall$ $t \neq 0$. Если $t=0$, то $O(t^2/n) = 0$.

Следовательно,

$$\lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty} [\,1 - \frac{t^2}{2n} + 0\left(\frac{t^2}{n}\right)]^n = e^{-t^2/2},$$

что является характеристической функцией нормального распределения $N(0,1)$.

Таким образом, (см. теорему непрерывности п.1.4) функция распределения $\frac{(z - n\mu)}{\sqrt{n\sigma}}$ сходится при $n\to\infty$ к функции распределения $N(0,1)$, что эквивалентно, учитывая свойство 2^0 (см. п. 1.3), утверждению теоремы 24.

Теорема доказана.

Соответствующая теорема Муавра-Лапласа для схемы испытаний Бернулли формулируется как следующее следствие теоремы 24.

Следствие. Если z – выборочная сумма, имеющая биномиальное распределение $Bi(n,p)$, то z асимптотически распределена согласно $N(np,npq)$.

Рассмотрим κ-мерный случай. Тогда имеет место следующий аналог центральной предельной теоремы.

Теорема 25. Допустим, что $(x_{1e},...,x_{ke};\ e=1,2,...,n)$ — выборка объема n из κ-мерного распределения, имеющего конечные средние значения μ_i $(i=1,....k)$ и положительно определенную ковариационную матрицу $\|\sigma_{ij}\|$, $i,j=1,2,...,k$.

Пусть $(z_1 = \sum_{e=1}^{n} x_{1e},..., z_k = \sum_{e=1}^{n} x_{ke})$ – выборочные суммы и $(\overline{x_1},...,\overline{x_k})$ – выборочные средние значения. Тогда $(z_1,...,z_k)$ и $(\overline{x_1},...,\overline{x_k})$ в качестве своих асимптотических распределений имеют κ–мерные распределения $N(\{n\mu_i\},\ \ \|n\sigma_{ij}\|)$ и $N\left(\{\mu_i\},\ \ \left\|\frac{\sigma_{ij}}{n}\right\|\right)$, $(i,j=1,...,k)$ соответственно.

Следствием этой теоремы является следующий многомерный аналог теоремы Муавра – Лапласа.

Теорема 26. Если $(x_{1e},...,x_{ke};\ e=1,...,n)$ – выборка объема n из мультиномиального распределения $M(1;p_1,...,p_k)$, то выборочные суммы $(z_1,...,z_k)$ имеют при больших n в качестве асимптотических распределений $N\left(\{np_i\},\ \|n(p_i\sigma_{ir}-p_ip_r)\|\right)$, где σ_{ir} – символ Кронекера.

Доказательство тривиально, так как в мультиномиальном распределении $M(1;p_1,...,p_k)$ (см. п.1.5,а):

$$M(x_i)=p_i, \qquad \sigma^2(x_i,x_j)=\sigma_{ij}=p_i\sigma_{ij}-p_ip_j.$$

2.6.2. Асимптотическое распределение функций от выборочных средних значений

Существует следующий общий результат.

Теорема 27. Пусть $(x_1,...,x_n)$ – выборка из распределения с конечными средними μ и дисперсией σ^2. Пусть $g(x)$ – функция, которая имеет в некоторой окрестности точки $x=\mu$ первую производную такую, что $g'(\mu)\neq 0$. Тогда $g(\bar{x})$ асимптотически распределена при больших n в соответствии с $N\left\{g(\mu),\ [\sigma g'(\mu)]^2/n\right\}$.

Доказательство теоремы 27 [19, с.270]. Пусть $V(\mu)$ – окрестность точки $x=\mu$, в которой $\exists\, g'(x)$ для $\forall x\in V(\mu)$. Согласно теореме 2 (усиленный закон больших чисел) для любого $\varepsilon>0$ существует такое n_ε, что при $n>n_\varepsilon$

$$p(\bar{x}\in V(\mu) \text{ для всех } n>n_\varepsilon)>1-\varepsilon.$$

Для любой точки $x=\bar{x}$ из $V(\mu)$ можно записать

$$g(\bar{x})=g(\mu)+g'(x^*)*(\bar{x}-\mu),$$

где x^* – случайная величина такая, что $|x^*-\mu|\le|\bar{x}-\mu|$.

Иначе

$$(g(\bar{x})-g(\mu))\sqrt{n}=g'(x^*)(\bar{x}-\mu)\sqrt{n}\,. \qquad (2.19)$$

Обозначим левую часть (2.19) через u_n, а правую – через v_n $p(u_n>v_n)>1-\varepsilon$ для всех $n>n_\varepsilon$ при $\forall\varepsilon$. Последовательности $(u_1,u_2,...)$ и $(v_1,v_2,...)$ сходятся друг к другу по распределению, если одна из них сходится по распределению к некоторой случайной величине; так как σ^2 конечна, то, в соответствии с теоремой 24, последовательность –

$\sqrt{n}(\bar{x}-\mu)$, $n=1,2,...$ сходится к случайной величине с нормальным законом $N(0,\sigma^2)$. Поскольку $g'(x)$ существует во всех точках $V(\mu)$, то она непрерывна в $V(\mu)$ и, в частности, при $x=\mu$. Но тогда и соответствующая последовательность $g'(x_n)$ сходится по вероятности к $g'(\mu)$.

Поэтому

$$\lim_{n\to\infty} P(v_n < w) = P(g'(\mu)s < w,$$

где $s \in N(0,\sigma^2)$.

Отсюда $\lim_{n\to\infty} P(u_n < w) = P(t < w),$ где $t \in N(0,(\sigma g(\mu))^2)$, что как раз и эквивалентно утверждению теоремы.

Теорема доказана.

Рассмотрим асимптотические распределения некоторых квадратичных функций выборочных средних значений и сумм в выборках из нормального распределения (теоремы 28' и 29). Ранее мы показали (п.2.4), что $\frac{1}{\sigma^2}\sum_{k}^{n}(x_k-\mu)^2$ имеет распределение $\chi^2_{(n)}$ с n степенями свободы.

Справедлив и следующий (более общий) результат.

Теорема 28. Квадратичная форма

$$Q_k = \sum_{i,j}^{k}\frac{1}{\sigma_{ij}}(x_i-\mu_i)(x_j-\mu_j) = \sum_{i,j=1}^{k}\sigma^{ij}(x_i-\mu_i)(x_j-\mu_j),$$

где $\left(\left\|\sigma^{ij}\right\|\right)$ – матрица, обратная ковариационной $\left(\left\|\sigma_{ij}\right\|\right)$, имеет $\chi^2_{(k)}$ – распределение с k степенями свободы, если все величины x_i $(i=1,k)$ нормальны $N\left(\{\mu_i\},\left\|\sigma_{ij}\right\|\right)$.

Доказательство теоремы 28.

Пусть $Q(x_1,...x_k) = \sum_{i,j=1}^{k}\sigma_i^{ij}(x_i-\mu_i)(x_j-\mu_j)$.

Рассмотрим характеристическую функцию формы $\frac{1}{2}Q(x_1,...x_k)$

$$\phi(t) = \frac{\sqrt{\left|\sigma^{ij}\right|}}{(2\pi)^{(1/2)k}}\int_{R_k} e^{-\frac{1}{2}(1-it)Q(x_1,...x_k)}dx_1...dx_k.$$

Сделав преобразование $y_i = \sqrt{(1-it)}(x_i - \mu_i), \qquad i = \overline{1,k}$, получим

$$\varphi(t) = \frac{\sqrt{|\sigma^{ij}|}}{(2\pi)^{(1/2)k}}(1-it)^{-\frac{1}{2}k} \int\limits_{R_k} e^{-\frac{1}{2}k} \sum_{i,j=1}^{k} \sigma^{ij} y_i y_j dy_1 ... dy_k .$$

Поскольку $\int\limits_{R_k} e^{-\frac{1}{2}k} \sum_{i,j=1}^{k} \sigma^{ij} y_i y_j dy_1 ... dy_k = \frac{(\pi)^{-\frac{1}{2}k}}{\sqrt{|\sigma^{ij}|}}$, имеем

$$\varphi(t) = (1-it)^{-\frac{1}{2}k},$$

что является характеристической функцией гамма–распределения *G(k/2),* откуда, учитывая п.1.5.2, д, следует утверждение теоремы.

Теорема 28'. Квадратичная форма

$$Q_n = \sum_{i,j}^{k} \sigma^{ij} \left(\frac{z_i - n\mu_i}{\sqrt{n}} \right) \left(\frac{z_j - n\mu_j}{\sqrt{n}} \right) = n \sum_{i,j=1}^{k} \sigma^{ij} (\bar{x}_i - \mu_i)(\bar{x}_j - \mu_j),$$

где $z_i = \sum_{e=1}^{n} x_{ie}$, также сходится по распределению к $\chi^2_{(k)}$ при $n \to \infty$. Утверждение теоремы 28' следует из теорем 25 и 28 непосредственно.

Иначе этот результат можно сформулировать так: если $(x_{1e},...,x_{ke})$, $e = \overline{1,n}$ – выборка из κ – мерного распределения с конечными средними значениями μ_i, $i=1,...,k$ и конечной положительно определенной ковариационной матрицей $\|\sigma^{ij}\|$, то Q_n вида (2) имеют при больших n в качестве асимптотического распределения $\chi^2_{(k)}$.

Распределение статистики χ^2 - Пирсона

Пусть теперь выборка берется из мультиномиального распределения $M(1;p_1,p_2,....p_k)$. Средние значения и ковариационная матрица даются в этом случае выражениями: (см. теорему 26)

$$\mu(x_i) = p_i, \qquad \sigma^2(x_i, x_j) = \sigma_{ij} = p_i \sigma_{ij} - p_i p_j .$$

Тогда прямым вычислением можно показать, что

$$\|\sigma^{ij}\| = \|p_i \sigma_{ij} - p_i p_j\|^{-1} = \left\| \frac{\sigma_{ij}}{p_i} + \frac{1}{p_{k+1}} \right\|.$$

Подставляя $\left\| \frac{\sigma_{ij}}{p_i} + \frac{1}{p_{k+1}} \right\|$ вместо σ_{ij} и p_i вместо μ_i в (2), находим, что

$$Q_n = \sum_{i=1}^{k+1} \frac{(z_i - np_i)^2}{np_i},$$

где
$$z_{k+1} = \sum_{i=1}^{n} x_{(k+1),i} = n - \sum_{i=1}^{k} z_i .$$

Действительно, так как по выборке $(x_{1l}, \dots, x_{kl};\ l = \overline{1,n})$

$$\bar{x}_i = \frac{\sum_{i,j=1}^{n} x_{ij}}{n}; \qquad z_i = \sum_{j=1}^{n} x_{ij}\text{ , то}$$

$$n\sum_{i,j=1}^{k}\left(\frac{\sigma_{ij}}{p_i} + \frac{1}{p_{k+1}}\right)\left(\bar{x}_i - p_i\right)\left(\bar{x}_j - p_j\right) = n\sum_{i=1}^{k}\frac{\left(\bar{x}_i - p_i\right)^2}{p_i} + \frac{n}{p_{k+1}}\sum_{i,j=1}^{k}\left(\bar{x}_i - p_i\right)\left(\bar{x}_j - p_j\right) =$$
$$= \sum_{i=1}^{k}\frac{\left(z_i - np_i\right)^2}{np_i} + \frac{n}{p_{k+1}}[\sum_{i=1}^{k}\left(\bar{x}_i - p_i\right)]^2 = \sum_{i=1}^{k}\frac{\left(z_i - np_i\right)^2}{np_i} + \frac{n}{p_{k+1}}[\sum_{i=1}^{k}\bar{x}_i - 1]^2 =$$
$$= \sum_{i=1}^{k}\frac{\left(z_i - np_i\right)^2}{np_i} + \frac{1}{np_{k+1}}[n\sum_{i=1}^{k}\bar{x}_i - 1]^2 = \sum_{i=1}^{k+1}\frac{\left(z_i - np_i\right)^2}{np_i},$$

где $z_{k+1} = \sum_{l=1}^{n} x_{k+1,l} = n - \sum_{i=1}^{k}\ \sum_{j=1}^{n} x_{ij} = n - \sum_{i=1}^{k}\bar{x}_i$.

Но мультиномиальное распределение $M(1;\ p_1, p_2, \dots, p_k)$ является воспроизводимым по n, т.е. выборочные суммы из $M(1; p_1, p_2, \dots, p_k)$ имеют распределение $M(n; p_1, p_2, \dots, p_k)$. Таким образом, оказывается справедливой следующая теорема.

Теорема 29. Если $(z_1, \dots, z_k)$ – $к$-мерная с.в., имеющая мультиномиальное распределение $M(n; p_1, p_2, \dots, p_k)$, то

$Q_n = \sum_{i=1}^{k+1} \frac{(z_i - np_i)^2}{np_i}$ асимптотически при $n \to \infty$ имеет хи-квадрат распределение с k степенями свободы. Утверждение теоремы 29 является частным случаем утверждения теоремы 28 при $\mu_I = p_i$.

2.6.3. Асимптотические распределения, связанные с порядковыми статистиками

Непосредственное отношение к свойствам порядковых статистик (см. п.2.2) имеет вопрос об асимптотическом распределении сумм выборочных долей.

Имеет место следующая теорема.

Теорема 29'. Если ($x_{(1)},...,x_{(n)}$) – вариационный ряд выборки из совокупности с непрерывной ф.р. $F(x)$, то при (фиксированном k $nF(x_{(k)})$, $n=k, k+1,...$ есть последовательность с.в., сходящаяся по распределению к с.в., имеющей гамма-распределение $G(k)$.

Доказательство. Пусть, как и ранее, сумма первых k долей $y_{(k)}=F(x_{(k)})$. Обозначим $w_k=ny_{(k)}$, имеющую распределение $H_n(w_k)$, тогда

$$H_n(w_k) = D_n\left(\frac{w_k}{n}\right) = \int_0^{w_k} h_n(y)dy,$$

где $D_n(\bullet)$ и $h(\bullet)$ обозначены ф.р. Дирихле и его плотность соответственно, так как (см. п. 1.5)

$$h(y) = \frac{\Gamma(n+1)}{n^k\Gamma(k)\Gamma(n-k+1)} y^{k-1}(1-\frac{y}{n})^{n-k}$$

и при $n\to\infty$ последовательность $h_n(y)$, $n=k, k+1,...$, стремится к гамма-распределению $G(k)$ (см. п.1.5.) равномерно в интервале $(0,w_k)$

$$\lim H_n(w_k) = \frac{1}{\Gamma(k)}\int_0^{w_k} y^{k-1}e^{-y}dy.$$

Отметим, что результат теоремы сохранится и на тот случай, если $F(x_{(k)})$ заменить на сумму некоторых (любых) k долей

$$F(x_{(1)}), F(x_{(2)})-F(x_{(1)}),...,1-F(x_{(n)}), \quad k\le n.$$

Предельные распределения самих порядковых статистик детально изучены Смирновым, их рассмотрение выходит за рамки настоящего пособия (см. [19, 15]).

Глава 3. МЕТОДЫ ПОЛУЧЕНИЯ ОЦЕНОК

Целью настоящей главы является ознакомление с существом и достоинствами того или иного метода оценивания. При этом из методических соображений утверждения, связанные с качеством той или иной оценки, даются преимущественно без доказательств. Некоторые из этих доказательств приводятся в гл. 6, посвященной теории оптимальных статистических решений.

3.1. Точечное оценивание

Пусть $F_n(x_1,...,x_n;\theta)$ – ф.р. выборки $\tau_n = (x_1,...,x_n)$, где θ – одномерный действительный параметр из параметрического пространства Θ. Пусть $\hat{\theta}$ $(x_1,...,x_n)$, или кратко $\hat{\theta}$ – функция от τ_n. Если полученное по τ_n (наблюденное) значение $\hat{\theta}$ используется вместо θ_0, истинного значения θ, то случайная переменная $\hat{\theta}$ называется точечной оценкой или просто оценкой параметра θ.

При точечном оценивании широко используются понятия (см. п.2.3.1) несмещенной, состоятельной, эффективной, достаточной, а также *асимптотически эффективной,* оценки, т.е. оценки, имеющей наибольший (равный 1) показатель эффективности $e_0(\hat{\theta}) = \lim\limits_{n\to\infty} e(\hat{\theta}),$ где эффективность $e(\hat{\theta}) = \dfrac{D_{\min}(\hat{\theta})}{D(\hat{\theta})}$ – отношение минимально возможной дисперсии оценки данного параметра к фактическому значению дисперсии оценки. Очевидно, что

$$0 \le e_0(\hat{\theta}) \le 1.$$

3.1.1. Метод моментов

Пусть $x_1,\ldots,x_n$ – независимые наблюдения над с.в., распределение которой зависит от неизвестных параметров $\theta_1, \theta_2,\ldots, \theta_k$.

Предположим, что первые r моментов распределения существуют и явно выражаются функциями $\mu_r(\theta_1, \theta_2,\ldots, \theta_k)$; $r=1,\ldots,k$ неизвестных параметров.

Если обозначить $a_r = \frac{1}{n}\sum_i^n x_i^r$ выборочные моменты, то метод моментов состоит в приравнивании наблюденных значений a_{r0} выборочных моментов теоретическим $\mu_r(\theta_1, \ldots, \theta_k)= a_{r0}$; $r=1,2,\ldots,k$. Решая эти уравнения относительно $\theta_1,\ldots, \theta_k$, получим оценки параметров по методу моментов.

Пример. Пусть п.в. с.в. X есть $p(x)=\lambda e^{-\lambda x}$, $\lambda > 0$, $\tau_n = (x_1,\ldots,x_n)$. Приравняем математическое ожидание μ_1 с.в. выборочному среднему:

$$\mu_1 = \int_0^\infty t\lambda e^{-\lambda t}dt = \frac{\sum_{i=1}^n x_i}{n}.$$

Проведя в этом соотношении интегрирование по частям, получим

$$\lambda = \frac{n}{\sum_{i=1}^n x_i}.$$

Это и есть искомая оценка $\hat{\lambda}$ параметра λ.

В пункте 2.3.2 уже отмечалось, что такие оценки являются состоятельными (если $\mu|x_i|^2 < \infty$, то, согласно усиленному закону больших чисел, $a_r \to \mu_r(\theta_1,\ldots, \theta_k)$ с вероятностью 1, когда $n \to \infty$), смещение этих оценок удается исключить с помощью простых поправок. Кроме того, при довольно общих условиях распределение оценки при больших n асимптотически нормально, и стандартное отклонение имеет вид $\frac{c}{\sqrt{n}}$.

Однако асимптотическая *эффективность* часто значительно меньше единицы, то есть в смысле эффективности оценки не являются «наилучшими» из возможных. Часто, ввиду простоты метода, оценки, получаемые по методу моментов, можно принять в качестве первого приближения, по которому определяют другими методами оценки более эффективные. Исключение составляют лишь оценки $\bar{x}$ и s^2 в случае нормального распределения, которые являются асимптотически эффективными.

Замечание к использованию метода моментов к группированной выборке. На практике выборки часто подвергают группировке. При этом индивидуальные выборочные значения не даются, а называется лишь число выборочных значений, попавших в интервалы некоторого определенного разбиения. Затем каждый интервал принимается за основание элементарной фигуры (например, прямоугольника или трапеции), площадь которой равна частоте появления соответствующего интервала. В этом случае общая фигура, составленная из прямоугольников, называется гистограммой, а из трапеций – полигоном.

Пусть $\tau_n(x_1,...,x_n)$ из $p(x)$ и пусть выборочные значения сгруппированы по интервалам длины h со средними точками $\xi_I = \xi_n + ih$, где $i=0;\pm1;\pm2;...$. Тогда при вычислении моментов и других выборочных характеристик обычно предполагается, что все выборочные значения, принадлежащие некоторому интервалу, совпадают со средней точкой интервала. Тогда в действительности мы производим выбор из распределения дискретного типа, в котором с.в. принимает значения $\xi_i = \xi_0 + ih$ с вероятностью

$$p_i = \int_{\xi_i-(1/2)h}^{\xi_i+(1/2)h} p(x)dx .$$

Таким образом, имеем, например, исправленные моменты «группированного распределения» $\overline{a_r} = \sum_{-\infty}^{\infty} p_i \xi_i^r$. Но нас интересует $\mu_r = \int_{-\infty}^{\infty} x^r p(x)dx$. Можно показать, что при некоторых условиях состоятельные и несмещенные оценки для $\theta_1, ..., \theta_\gamma$ можно получить, прибавляя некоторые поправки (так называемые поправки Шеппарда) к группированным моментам $\overline{a_r}$.

3.1.2. Метод максимума правдоподобия

Для данной выборки $(x_1,...,x_n)$ из совокупности с непрерывным распределением определим функцию правдоподобия L:

$$L(x_1,...,x_n;\ \theta)=p(x_1;\ \theta)...p(x_n;\ \theta)$$

и, соответственно, для дискретного случая:

$$L(x_1,...,x_n;\ \theta)=p_{i_1}(\theta)...p_{i_n}(\theta),$$

где $i_1, i_2, ..., i_n$ – наблюденные значения с.в.

Оценкой $\hat{\theta}$ максимального правдоподобия для параметра θ называется то значение θ, которое доставляет максимум для функции L (или, что однозначно, $logL$), то есть является решением уравнения правдоподобия

$$\frac{\partial \log L}{\partial \tau}=0.$$

Для двух параметров аналогичная функция правдоподобия

$$L(x_1,...,x_n;\theta_1,\theta_2)=\prod_{i=1}^{n} p(x_i;\theta_1,\theta_2),$$

а оценки максимального правдоподобия получаются при совместном решении уравнений

$$\frac{\partial \log L}{\partial \theta_1}=0;\ \frac{\partial \log L}{\partial \theta_2}=0.$$

Ценность оценок максимального правдоподобия обусловлена следующими их свойствами, которые выполняются при достаточно общих условиях:

1. Уравнение правдоподобия имеет решение, которое является состоятельной и асимптотически аффективной оценкой для θ_0.

2. Решение (оценка в) имеет асимптотическое нормальное распределение при $n\to\infty$.

3. Если для θ существует эффективная оценка, то уравнение правдоподобия имеет единственное решение, совпадающее с этой оценкой.

Оценка максимального правдоподобия может, однако, оказаться смещенной. На практике метод максимального правдоподобия иногда приводит к сложным системам уравнений.

Пример 1. Для *n* испытаний Бернулли

$$L(x_1,...,x_n;p) = p^m (1-p)^{n-m},$$

где $m = \sum_{k=1}^{n} x_k$.

Тогда $\frac{\partial \log L}{\partial p} = \frac{m}{p} - \frac{n-m}{1-p} = 0$.

Единственное решение: $\hat{p}(x_1,...x_n) = \frac{m}{n}$. Отметим, что в соответствии с законом больших чисел (теорема 3) $\frac{m}{n} \xrightarrow[p]{} p$ при $n \to \infty$, что характеризует состоятельность данной оценки, а ее асимптотическая нормальность является следствием теоремы 8. Несмещенность имеет место, поскольку $M \sum_{k=1}^{n} x_k = np$.

Пример 2. (непрерывный случай). Пусть $p(x,\lambda) = \lambda e^{-\lambda x}, x \ge 0, \lambda > 0$.

В этом случае $L(x_1,...,x_n;\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$. Тогда $\frac{\partial \log L}{\partial L} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$.

Отсюда $\hat{\lambda}(x_1,...x_n) = \frac{n}{\sum_{i=1}^{n} x_i}$.

Целую систему различных методов нахождения точечных оценок можно получить из следующих соображений. Зададим каким-либо образом меру *D* отклонений эмпирической ф.р. $F_n(x)$ от теоретической $F(x, \theta)$. Эта мера зависит от θ и выборочных значений $x_1,...,x_n$

$$D=D(x_1,...,x_n;\ \theta).$$

Естественно принять за точечную оценку параметра θ значение θ, которое минимизирует меру *D*. Например, меру χ^2 Пирсона (см.

замечание к теореме 32) или так называемую величину ω^2 (критерий Мизеса):

$$D(x_1,\dots x_n;\theta) = \omega^2 = \int\limits_{-\infty}^{\infty} [F_n(x) - F(x,\theta)]^2 \, dF(x,\theta).$$

Точечная оценка по методу минимума χ^2 при весьма общих условиях является состоятельной, асимптотически эффективной и асимптотически нормальной.

Оценка по методу минимума ω^2 также асимптотически нормальна, но не будет, вообще говоря, асимптотически эффективной.

3.2. Оценивание с помощью интервалов

Определение доверительного интервала. Пусть $(x_1,\dots,x_n)$ – случайная выборка с ф.р. $F(x_1,\dots,x_n;\ \theta)$. Пусть $\underline{\theta}(x_1,\dots,x_n)$, $\overline{\theta}(x_1,\dots,x_n)$ – две функции выборочных значений случайных величин, такие, что $\underline{\theta}<\overline{\theta}$. Если $\underline{\theta}$ и $\overline{\theta}$ могут быть выбраны так, что при заданном γ $P(\underline{\theta}<\theta\leq\overline{\theta})=\gamma$, где γ означает вероятность события, написанного в скобках, то $(\underline{\theta};\overline{\theta})$ называется 100γ – процентным доверительным интервалом для θ; $\underline{\theta}$ и $\overline{\theta}$ – нижней и верхней границами для θ, γ – доверительным коэффициентом.

Введение понятия доверительного интервала дает возможность указать точность и надежность полученного по выборке значения оценки $\hat{\theta}$ для параметра θ.

Различают два метода оценивания с привлечением понятия доверительного интервала: так называемый классический (байесовский) метод и метод доверительных интервалов, предложенный Нейманом.

3.2.1. Классический метод

Этот метод применим в том случае, когда заранее известно априорное распределение θ, например, плотность $\varphi(x)$; п.в. $g(x/\theta)$ является условной п.в. оценки $\hat{\theta}$ при заданном значении θ. Сущность метода заключается в следующем. По формуле Байеса условная п.в. с.в. θ при заданном значении $\hat{\theta}$:

$$h(x/\hat{\theta}) = \frac{\phi(x)g(\hat{\theta}/x)}{\int_{-\infty}^{\infty}\phi(x)g(\hat{\theta}/x)dx}.$$

Отсюда условная вероятность события $\underline{\theta} < \theta \leq \bar{\theta}$ при заданном значении $\hat{\theta}$,

$$P(\underline{\theta} < \theta \leq \bar{\theta}/\hat{\theta}) = \int_{\underline{\theta}}^{\bar{\theta}} h(x/\hat{\theta})dx.$$

Эта вероятность называется апостериорной вероятностью события ($\underline{\theta} < \theta \leq \bar{\theta}$), в отличие от априорной, равной $p(\underline{\theta} < \theta \leq \bar{\theta}/\theta) = \int_{\underline{\theta}}^{\bar{\theta}} \varphi(x)dx$.

Итак,

$$\int_{\underline{\theta}}^{\bar{\theta}} h(x/\theta)dx = P(\underline{\theta} < \theta \leq \bar{\theta}/\hat{\theta}) = \gamma.$$

На практике классический метод часто неприменим, так как θ является просто неизвестной постоянной, а не случайной величиной.

Если же это с.в., то обычно ее распределение неизвестно.

3.2.2. Метод доверительных интервалов

Этот метод применим и когда θ – с. в., и когда θ – фиксированное число. При этом методе γ задается постоянным, достаточно близким к единице. Доверительные интервалы, отвечающие коэффициенту доверия γ, можно строить различными способами, так что будут справедливы два эквивалентных утверждения [8, 12]:

$$\theta_H(\theta,\gamma) < \hat{\theta} < \theta_B(\theta,\gamma) \quad \text{и} \quad \underline{\theta}(\hat{\theta},\gamma) < \theta < \bar{\theta}(\hat{\theta},\gamma).$$

Первое из этих соотношений означает, что с.в. $\hat{\theta}$ заключена между пределами θ_H и θ_B, а второе, что θ (которая может быть случайной или неслучайной) заключена между случайными пределами $\underline{\theta}$ и $\bar{\theta}$.

Из самой общности метода доверительных интервалов вытекает, что этот метод не приводит к вероятностным утверждениям типа:

«Вероятность того, что θ заключена между такими-то фиксированными пределами, равна γ».

Действительно, это утверждение не имеет смысла, если θ не является с.в. Утверждения, следующие из метода доверительных интервалов, являются утверждением типа:

«Вероятность того, что такие-то пределы (которые могут меняться от выборки к выборке) заключают между собой значение параметра θ, соответствующее фактической выборке, равняется γ».

При методе доверительных интервалов по каждому множеству выборочных значений вычисляем величины $\underline{\theta} = \underline{\theta}(\hat{\theta}, \gamma)$ и $\overline{\theta} = \overline{\theta}(\hat{\theta}, \gamma)n$, используя заранее заданное значение доверительной вероятности γ.

3.2.3. Доверительные интервалы для μ в случае нормального распределения генеральной совокупности

Пусть $\tau_n = (x_1,\dots,x_n)$ – выборка из нормального распределения.

1. Пусть $\sigma_x > 0$ известно, тогда $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \in \quad N(\mu, \frac{\sigma_x}{\sqrt{n}})$,

$$P\left\{\left|\overline{x} - \mu\right| < \varepsilon\right\} = P\left\{\frac{\left|\overline{x} - \mu\right|}{\sigma_{\overline{x}}} < \frac{\varepsilon}{\sigma_{\overline{x}}}\right\} = F\left(\frac{\varepsilon}{\sigma_{\overline{x}}}\right) - F\left(-\frac{\varepsilon}{\sigma_{\overline{x}}}\right) =$$

$$= F\left(\frac{\varepsilon\sqrt{n}}{\sigma_x}\right) - F\left(-\frac{\varepsilon\sqrt{n}}{\sigma_x}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma_x}\right),$$

где $F(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt; \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{0}^{x} e^{-t^2/2}dt$.

Зададимся коэффициентом доверия γ. Тогда, определив корень уравнения $\gamma = 2\Phi(t_\gamma)$ (по таблицам функций Лапласа или нормальной ф.р.), можем считать, что с вероятностью γ

$$\overline{x} - t_\gamma \frac{\sigma_x}{\sqrt{n}} < \mu < \overline{x} + t_\gamma \frac{\sigma_x}{\sqrt{n}},$$

то есть $(\underline{\theta}, \overline{\theta}) = (\overline{x} - t_\gamma \frac{\sigma_x}{\sqrt{n}}, \overline{x} + t_\gamma \frac{\sigma_x}{\sqrt{n}})$ – доверительный интервал для μ_x.

2. σ_x – неизвестно.

Мы знаем, что $\frac{\sqrt{n}}{s}(\overline{x} - \mu) \in S_{(n-1)}$ (имеет распределение Стьюдента с *(n-1)* степенями свободы).

Зададимся γ и пусть t_γ – корень уравнения $\gamma = \int_{t_1}^{t_2} S_{(n-1)}(x)\, dx.$

Тогда $P\left\{t_1 < \frac{\sqrt{n}(\overline{x} - \mu)}{s} < t_2\right\} = \gamma,$ что эквивалентно

$$P\left\{\overline{x} - t_\gamma \frac{\sigma_x}{\sqrt{n}} < \mu < \overline{x} + t_\gamma \frac{\sigma_x}{\sqrt{n}}\right\} = \gamma.$$

Но t_1, t_2 можно выбрать различными способами. Существуют таблицы распределения Стьюдента, где принято $|t_1| = |t_2| = t_\gamma$. В этом случае $(\overline{x} - t_\gamma \frac{\sigma_x}{\sqrt{n}} < \mu < \overline{x} + t_\gamma \frac{\sigma_x}{\sqrt{n}})$.

3.2.4. Доверительные интервалы для σ в случае нормального распределения генеральной совокупности

Мы знаем, что $\frac{(n-1)s^2}{\sigma_x^2} \in \chi^2_{(n-1)}$ (см. теорему 21). Зададимся γ, тогда

$$P\left\{\chi_H^2 < \frac{(n-1)s^2}{\sigma_x^2} < \chi_B^2\right\} = \int_{\chi_N^2}^{\chi_B^2} p_{n-1}(x)dx = \gamma\,. \tag{3.1}$$

Но χ_H^2 и χ_B^2 можно выбрать различными способами. Пусть

$$\int_0^{\chi_H^2} p_{n-1}(x)dx = \frac{1-\gamma}{2}; \int_{\chi_B^2}^{\infty} p_{n-1}(x)dx = \frac{1-\gamma}{2}.$$

Тогда справедливо следующее эквивалентное (3.1) утверждение:

$$P\left\{\frac{\sqrt{n-1}s}{\chi_B^2} < \sigma_x < \frac{\sqrt{n-1}s}{\chi_H^2}\right\} = \gamma,$$

то есть доверительный интервал для σ_x:

$$(\underline{\sigma}_x, \overline{\sigma}_x) = \left\{ \frac{\sqrt{n-1}s}{\chi_B^2} < \sigma_x < \frac{\sqrt{n-1}s}{\chi_H^2} \right\}$$ при коэффициенте доверия γ,

χ_B^2 и χ_H^2 находятся по таблицам распределения $\chi^2_{(n-1)}$ [2].

Замечание об интервальном оценивании при помощи больших выборок

Способ построения доверительных интервалов следует выбирать так, чтобы при данном коэффициенте доверия γ они оказывались возможно более короткими. Асимптотическая теория интервального оценивания по большим выборкам непосредственно связана с теорией точечного оценивания, так как асимптотически эффективные точечные оценки порождают для больших выборок асимптотически наикратчайшие доверительные интервалы.

3.3. Непараметрическое статистическое оценивание

Непараметрические оценки – это оценки, не зависящие от функциональной формы (вида) ф.р. генеральной совокупности, из которой извлечена выборка. Основные случайные переменные в непараметрическом оценивании – это определяемые выборкой порядковые статистики (вариационный ряд) и определяемые ими доли.

3.3.1. Доверительные интервалы для квантилей

Случай малых выборок

Пусть мы имеем бесконечную совокупность с непрерывной ф.р. *F(x)*. Для любого числа *p*, лежащего в интервале (0,1), *p* – я квантиль $\underline{x}_p$ непрерывной случайной величины X, имеющей ф.р. *F(x)*, определяется как наименьшее число $\underline{x}_p$, для которого $F(\underline{x}_p) = p$. Например, $\underline{x}_{0,5}$ есть медиана; $\underline{x}_{0,25}$ и $\underline{x}_{0,75}$ – соответственно нижняя квартиль и верхняя квартиль величины.

Пусть $(x_1,...,x_n)$ – выборка из *F(x)* и $(x_{(1)},...,x_{(n)})$ – вариационный ряд этой выборки.

Пусть $x_{(k_1)}$ и $x_{(k_1+k_2)}$ две любые порядковые статистики, мы знаем (т.17′), что случайные переменные $u = F(x_{(k_1)}), v = F(x_{(k_1+k_2)})$, равные суммам долей, подчиняются двумерному распределению Дирихле *D*(k₁, k₂, n-k₁-k₂+1)* с плотностью вероятности

$$f(u,v) = \left\{ \begin{array}{ll} \dfrac{\Gamma(n+1)}{\Gamma(k_1)\Gamma(k_2)\Gamma(k_1+k_2)} u^{k_1-1}(v-u)^{k_1-1}(1-v)^{n-k_1-k_2+1} & \text{при} \quad 0 < u < v-1, \\ 0 & \text{при других значениях} \quad u \quad \text{и} \quad v \end{array} \right\}.$$

Рассмотрим неравенство

$$F(x_{(k_1)}) < p < F(x_{(k_1+k_2)}). \tag{3.2}$$

Так как *F(x)* непрерывна, то оно выполняется тогда и только тогда, когда

$$x_{(k_1)} < \underline{x}_p < x_{(k_1+k_2)}. \tag{3.3}$$

Следовательно, вероятность выполнения (3.3) равна вероятности выполнения (3.2). Поэтому

$$P(x_{(k_1)} < \underline{x}_p < x_{(k_1+k_2)}) = \int_p^1 \int_0^p f(u,v)dudv.$$

Но

$$\int_p^1 \int_0^p f(u,v)dudv = \int_0^p \int_u^1 f(u,v)dudv - \int_0^p \int_0^v f(u,v)dudv \tag{3.4}$$

Делая замену переменных *u=r, v=1-s(1-r)* в первом интеграле правой части (3.4) и *u=rs, v=s* во втором, находим

$$P(x_{(k_1)} < \underline{x}_p < x_{(k_1+k_2)}) = \Im_p\left(k_1; n-k_1+1\right) - \Im_p\left(k_1+k_2; n-k_1-k_2+1\right),..., \tag{3.5}$$

где $\Im_p(v_1,v_2) = \dfrac{\Gamma(v_1,v_2)}{\Gamma(v_1)\Gamma(v_2)} \int_0^p x^{v_1-1}(1-x)^{v_2-1}dx.$

Вероятность того, что интервал $(x_{(k_1)}, x_{(k_1+k_2)})$ содержит квантиль $\underline{x}_p$, не зависит от *F(x)*, поэтому этот интервал является доверительным интервалом для $\underline{x}_p$ при доверительной вероятности, равной правой части (3.5). Обычно для данного γ выбираются две порядковые статистики с наиболее близкими номерами, то есть так, чтобы k_2 было как можно меньше.

Случай больших выборок

Пусть $(x_1,...,x_n)$ – выборка из совокупности с непрерывной ф.р. $F(x)$ и $\underline{x}_p$ – квантиль, отвечающий уровню p. Если n_1 – число компонент выборки, меньших $\underline{x}_p$, то $n_1 \in Bi(n,p)$. Мы знаем (следствие теоремы 24), что для больших n статистика n_1 распределена асимптотически по закону $N(np, np(1-p))$.

Следовательно, для данной доверительной вероятности γ

$$\lim_{n\to\infty} P\left(-y_\gamma < \frac{n_1 - np}{\sqrt{np(1-p)}} < +y_\gamma\right) = \gamma, \tag{3.6}$$

где $\frac{1}{\sqrt{2\pi}}\int_{-y_\gamma}^{+y_\gamma} e^{-\frac{1}{2}y^2}dy = \gamma.$

Таким образом, для больших n приближенный 100γ – процентный доверительный интервал $\left(\underline{p}_\gamma, \overline{p}_\gamma\right)$ для p образован множеством всех значений p, удовлетворяющих (3.6) для фиксированных n_1, n, y_γ, то есть $\underline{p}_\gamma$ и $\overline{p}_\gamma$ являются решениями уравнения $\frac{(n_1 - np)^2}{np(1-p)} = y_\gamma^2$ относительно p.

Поэтому для больших n $P(\underline{p}_\gamma < p < \overline{p}_\gamma) \cong \gamma$, что эквивалентно $P(x_{\left[n\underline{p}_\gamma\right]} < \underline{x}_p < x_{\left[n\overline{p}_\gamma\right]}) \cong \gamma$, то есть две порядковые статистики $(x_{\left[n\underline{p}_\gamma\right]}, x_{\left[n\overline{p}_\gamma\right]})$ образуют приближенный 100γ – процентный доверительный интервал для $\underline{x}_p$, где $\left[n\underline{p}_\gamma\right]$ и $\left[n\overline{p}_\gamma\right]$ – ближайшие целые соответственно к $n\underline{p}_\gamma$ и $n\overline{p}_\gamma$.

3.3.2. Односторонние границы для непрерывной функции распределения. Теорема Смирнова

Пусть $x_1,...,x_n$ – выборка из непрерывной ф.р. $F(x)$. Тогда функции выборки $F_n^+(x)$ и $F_n^-(x)$ такие, что для данного значения γ

$$P[F_n^+(x) \ge F(x) \text{ для всех } x] = \gamma,$$

$$P[F_n^-(x) \le F(x) \text{ для всех } x] = \gamma$$

называют соответственно верхней и нижней 100 γ – процентной доверительной границей для $F(x)$.

Пусть $x_{(1)},\ldots,x_{(n)}$ – вариационный ряд. Тогда эмпирическая ф.р. $F_n(x)$

$$F_n(x)=\begin{cases}0, & x\le x_{(1)},\\ \dfrac{k}{n}, & x_{(k)}<x\le x_{(k+1)}, \quad k=\overline{1,n}.\\ 1, & x>x_{(n)}.\end{cases}$$

Для $0<d<1$ и любого x пусть

$$F_n^+(x,d)=\min[F_n(x)+d;1],$$

$$F_n^-(x,d)=\max[F_n(x)-d;0],$$

$$D_n^+(d)=\sup_x[F_n^+(x,d)-F(x)],$$

$$D_n^-(d)=\inf_x[F_n^-(x,d)-F(x)].$$

Заметим, что $P(D_n^+(d)>0)$ – вероятность того, что график $F_n(x)+d$ нигде не имеет общих точек с графиком $F(x)$, аналогично $P(D_n^-(d)<0)$ – вероятность того, что график $F_n(x)$-d не имеет общих точек с графиком $F(x)$.

Бирнбаум и Тинджи получили следующий результат.

Теорема 30 [19, с. 346]**.**

$$P(D_n^+(d)>0)=P(D_n^-(d)<0)=P_n(d),$$

где $P_n(d)=1-d\sum\limits_{i=1}^{[n(1-d)]}\binom{n}{i}(1-d-\frac{i}{n})^{n-i}(d+\frac{i}{n})^{i-1}$ и $[n(1-d)]$ – целая часть числа $n(1-d)$.

Доказательство. Положим $y_{(i)}=F(x_{(i)})$, $i=\overline{1,n}$. Так как величины $y_i=F(x_i)$ равномерно распределены на отрезке $(0,1)$, то вероятность попадания величин $\{y_i\}$, $i=\overline{1,n}$ в элемент пространства $dy_1\ldots dy_n$ равняется $1\,dy_1\ldots dy_n$ при $0<y_i<1$ и 0 при остальных значениях y_i. В таком случае вероятность попадания в элемент пространства $dy_1\ldots dy_n$ для случайных величин $y_{(i)}$, $i=\overline{1,n}$ равна (см. п.2.1., а) $n!\ dy_1\ldots dy_n$ при

0 < y(i) <1, $i=\overline{1,n}$ и 0! $dy_1...dy_n$ в противном случае. Очевидно, что неравенство $D_n^+(d)>0$ имеет место тогда и только тогда, когда $y_{(i)}$, $i=\overline{1,n}$ удовлетворяют неравенствам

$$
\begin{aligned}
&y_{(i-1)} < y_{(i)} < \frac{i-1}{n}+d, \qquad i=\overline{1,k+1},\\
&y_{(i-1)} < y_{(i)} < 1, \qquad i=\overline{1,k+2,n}.
\end{aligned}
\tag{3.7}
$$

где k – целая часть числа $n(1-d)$; $y_0=0$. Аналогично $D_n^-(d)<0$ выполняется в том случае, если

$$
\begin{aligned}
&\frac{i}{n}-d < y_{(i)} < y_{(i+1)}; \qquad i=\overline{n-k,n},\\
&0 < y_{(i)} < y_{(i+1)}; \qquad i=\overline{1,n-k+1},
\end{aligned}
\tag{3.8}
$$

где $y_{(n+1)}=1$.

Если сделать замену переменных $y_{(i)}=1-y_{(n-i+1)}$, $i=\overline{1,n}$, то неравенство (3.8) сведется к неравенству (1) и, следовательно,

$$P(D_n^+(d)>0)=P(D_n^-(d)<0)=P_n(d).$$

Общее значение $P_n(d)$ этих вероятностей получается интегрированием вероятностного элемента по области, определяемой неравенством (3.7):

$$P_n(d)=n!\int_0^d \int_{y_{(1)}}^{1/n+d} \ldots \int_{y_{(k)}}^{k/n+d} \int_{y_{(k+1)}}^{1} \ldots \int_{y_{(n-1)}}^{1} dy_{(n)}\ldots dy_{(1)}.$$

Интегрируя по $y_{(n)},\ldots,y_{(k+2)}$, получим

$$G_n(k,d)=\int_0^d \int_{y_{(1)}}^{1/n+d} \ldots \int_{y_{(k)}}^{k/n+d} \frac{\left(1-y_{(k+1)}\right)^{n-k-1}}{(n-k-1)!}\,dy_{(k+1)}\ldots dy_{(1)}.$$

Найдем значение $G_n(k,d)$, используя метод математической индукции.

При $(k+1)$ имеем

$$G(k+1,d)=\int_0^d \int_{y_{(1)}}^{1/n+d} \ldots \int_{y_{(k+1)}}^{k+1/n+d} \frac{\left(1-y_{(k+2)}\right)^{n-k}}{(n-k)!}\,dy_{(k+2)}\ldots dy_{(1)} =$$

$$=\int_0^d \int_{y_{(1)}}^{1/n+d} \dots \int_{y_{(k)}}^{k/n+d} \frac{\left(1-y_{(k+1)}\right)^{n-k-1}}{(n-k-1)} dy_{(k+1)} \dots dy_{(1)} -$$

$$-\ \frac{\left(1-d-\frac{(k+1)}{n}\right)^{n-k-1}}{(n-k-1)!} \times \int_0^d \int_{y_{(1)}}^{1/n+d} \dots \int_{y_{(k)}}^{k/n+d} dy_{(k+1)} \dots dy_{(1)},$$

$$\int_0^d \int_{y_{(1)}}^{1/n+d} \dots \int_{y_{(k)}}^{k/n+d} dy_{(k+1)} \dots dy_{(1)} = -\frac{d}{(k+1)!} \sum_{i=1}^{k+1} (-1)^i \left(\frac{(k+1)!}{(k+1-i)!\,i!} \left(d + \frac{k+1-i}{n} \right)^k \right) =$$
$$= -\frac{d}{(k+1)!} \frac{\partial^k}{\partial t^k} \left\{ e^{(d+\frac{(k+1)}{n})t} \left[(1-e^{-\frac{t}{n}})^{k+1} - 1 \right] \right\}_{t=0}.$$

Так как

$$\left\{ \frac{\partial^k}{\partial t^k} \left[e^{(\frac{d}{k+1}+\frac{1}{n})t} - e^{\frac{d}{k+1}t} \right]^{k+1} \right\}_{t=0} = 0,$$

то

$$\int_0^d \int_{y_{(1)}}^{(1/n)+d} \dots \int_{y_{(k)}}^{(k/n)+d} dy_{(k+1)} \dots dy_{(1)} = \frac{d}{(k+1)!} \left\{ \frac{\partial^k}{\partial t^k} e^{(d+\frac{(k+1)}{n})t} \right\}_{t=0} = \frac{d}{(k+1)!} (d + \frac{k+1}{n})^k.$$

Таким образом,

$$G_n(k+1,d) = G_n(k,d) - \frac{\left(1-d-\frac{k+1}{n}\right)^{n-k-1}}{(n-k-1)!} \frac{d}{(k+1)!} \left(d + \frac{k+1}{n} \right)^k.$$

Учитывая, что $G_n(0,d) = \frac{1-(1-d)^n}{n}$ и что k – целая часть числа *n(1-d)*, получим

$$P_n(d) = n! G_n(k,d) = 1 - d \sum_{i=1}^{[n(1-d)]} \frac{n!}{(n-i)!\,i!} (1-d-\frac{i}{n})^{n-1} (d+\frac{i}{n})^{i-1} =$$
$$= 1 - d \sum_{i=1}^{[n(1-d)]} C_n^i (1-d-\frac{i}{n})^{n-i} (d+\frac{i}{n})^{i-1}, \tag{3.9}$$

Теперь рассмотрим предел $P_n(d)$ при $n \to \infty$. Положив $d = \frac{\lambda}{\sqrt{n}}$, перепишем (3.9) в следующем виде:

$$P_n\left(\frac{\lambda}{\sqrt{n}}\right)=1-\lambda\left\{\sum_{i=1}^{\left[n-\sqrt{n}\lambda\right]}\sqrt{n}C_n^i(1-\frac{\lambda}{\sqrt{n}}-\frac{i}{n})^{n-i}(\frac{\lambda}{\sqrt{n}}+\frac{i}{n})^{i-1}\Delta y\right\}, \qquad (3.10)$$

где

$$\Delta y=\frac{1}{n}.$$

Используя формулу Стирлинга для факториалов, убеждаемся в том, что сумма в фигурных скобках в (3.10) сходится к интегралу

$$\frac{1}{\sqrt{2\pi}}\int_0^1\frac{1}{\sqrt{y(1-y)}}e^{-\frac{\lambda^2}{2y(1-y)}}dy$$

при $n\to\infty$, который после интегрирования дает $\frac{1}{\lambda}e^{-2\lambda^2}$. Таким образом, в условиях теоремы 30 имеем

$$\lim_{n\to\infty}P_n\left(\frac{\lambda}{\sqrt{n}}\right)=1-e^{-2\lambda^2}.$$

Этот результат известен под названием теоремы Смирнова.

3.3.3. Доверительные полосы для непрерывной функции распределения

Задача нахождения двусторонних границ, или доверительной полосы, для $F(x)$ представляет собой задачу определения вероятности $P(D_n < d)$, где

$$D_n=\sup_x\left|F_n(x)-F(x)\right|.$$

Если n конечно и d кратно $1/n$, то известен следующий результат Масси.

Теорема 31 [19, с. 349]**.** Пусть $x_{(1)},\dots,x_{(n)}$ – вариационный ряд выборки объема n из совокупности с непрерывной ф.р. $F(x)$ и $D_n=\sup_x\left|F_n(x)-F(x)\right|$.

Тогда

$$P\left(D_n<\frac{k}{n}\right)=\frac{n!}{n^n}u(k,n), \qquad k=1,2,\dots,n-1,$$

где $u(j,m+1)$ удовлетворяют системе уравнений

$u(j,m+1)=\sum_{i=1}^{j+1}\frac{u(j,m)}{(j+1-i)!}$ с граничными условиями

$$\begin{cases} u(i,m)=0, & i\ge m+k, \\ u(i,0)=0, & i=1,...,k-1, \\ u(k,0)=0, & \end{cases}$$

Доказательство. Пусть $(x_{(1)},...,x_{(n)})$ – вариационный ряд выборки $(x_{(1)},...,x_{(n)})$ из распределения с непрерывной функцией распределения $F(x)$. Произведем замену переменной $y = F(x)$.

Случайная величина Y имеет [см. т.12] равномерное распределение на $(0,1)$. Пусть $G(y)$ – функция распределения случайной величины Y и $G_n(y)$ — эмпирическая функция распределения, построенная по вариационному ряду $(y_{(1)},...,y_{(n)})$, $y_{(i)}=F(x_{(i)})$. Очевидно, что

$$P(D_n<\frac{k}{n})=P(\sup_y\left|G_n(y)-G(y)\right|<\frac{k}{n}).$$

Разделим интервал $(0,1)$ на n равных интервалов $I_\xi=\left(\frac{\xi-1}{n},\frac{\xi}{n}\right)$, $\xi=\overline{1,n}$. Пусть $(r_1,...,r_n)$ – частоты попадания элементов выборки $y_1,...,y_n$ в интервалы разбиения.

Так как

$$P\left(\frac{\xi-1}{n}<y_i<\frac{\xi}{n}\right)=\frac{1}{n},\qquad i=\overline{1,n}\qquad \xi=\overline{1,n},$$

то по известной формуле полиномиального распределения (см.п. 1.5.1.,а) получим распределение частот $(r_1,...,r_n)$

$$P(r_1,...,r_n)=\frac{n!}{r_1!...r_n!n^n},$$

поэтому $P(D_n(y)<\frac{k}{n})$ находится суммированием $P(r_1,...,r_n)$ по всем точкам выборочного пространства $(r_1,...,r_n)$, для которых $|G_n(y)-G(y)|<\frac{k}{n}$ для всех y.

Если двигаться слева направо по графику $G_n(y)$, лежащему полностью внутри полосы E, для которой $|G_n(y)-G(y)|<\frac{k}{n}$ при

$y = \frac{1}{n}, \frac{2}{n}, ..., \frac{m+1}{n}$, то график должен проходить через одну из точек $\left(\frac{m}{n}, \frac{m-k+i}{n}\right)$, $i = 1,...,2k-1$. Обозначим эти точки *A (i, m),* $i = \overline{1,2k-1}$.

Пусть $U(i,m) = \sum_{(i)} \frac{1}{r_1!...r_n!}$, где суммирование ведется по всему множеству значений $(r_1,...,r_n)$, для которых график $G_n(y)$ приходит в $A\ (i,\ m)$, оставаясь внутри полосы E. Так как $G_n(y)$ не убывает, график может достичь $A\ (j,\ m+1)$, только проходя через одну из точек $A(1,m)$, $A(2,m),...,A(j+1,m)$, причем r_{m+1} примет соответственно значения $j,j-1,...,1,0$. Следовательно, должно быть

$$U(j,m+1) = \frac{U(1,m)}{j!} + \frac{U(2,m)}{(j-1)!} + ... + \frac{U(j+1,m)}{0!} = \sum_{i=1}^{j+1} \frac{U(i,m)}{(j+1-i)!},$$
$$(j = \overline{1,2k-1}; \qquad m = \overline{0,n-1}),$$

причем

$$U(i,m) = 0, \qquad i \ge m+k,$$
$$U(i,0) = 0, \qquad i = \overline{1,k-1}.$$
$$U(k,0) = 0,$$

что и требовалось доказать.

Мы уже знаем, что асимптотический результат Колмогорова (теорема 11) в условиях теоремы 31 имеет вид

$$\lim_{n\to\infty} P(D_n \frac{\lambda}{\sqrt{n}}) = \begin{cases} \sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2\lambda^2} & \text{при} \quad \lambda > 0, \\ 0 & \text{при} \quad \lambda \le 0. \end{cases}$$

Проведение этого предельного перехода выходит за рамки настоящего пособия.

Глава 4. СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

4.1. Проверка параметрических статистических гипотез

Пусть $\tau_n(x_1,...,x_n)$ – независимая повторная выборка из совокупности с ф.р. $F(x;\theta)$, θ- r-мерный параметр, принадлежащий открытому множеству Θ евклидова пространства R_n

$$F_n(x_1,...,x_n;\theta) = \prod_{i=1}^{n} F(x_i,\theta).$$

Пусть A – подмножество Θ. Если θ_o– истинное значение параметра, то *статистической гипотезой* $\mathrm{H} = \mathrm{H}(\mathrm{A},\Theta)$ называется утверждение: $\theta_o \in \mathrm{A}$ против альтернативы $\theta_o \in \Theta \setminus \mathrm{A}$. Говорят, что гипотеза H истинна, если $\theta_o \in \mathrm{A}$, и ложна, если $\theta_o \in \Theta \setminus \mathrm{A}$. Утверждение $\theta_o \in \mathrm{A}$ называют *нулевой* гипотезой.

Далее будем для простоты обозначать θ вместо θ_o. Если A содержит одну точку θ, гипотеза A называется *простой*, в противном случае – *сложной*.

Пример.

$\{\mu = \mu_o, \sigma = \sigma_o\}$ – простая гипотеза,

$\{\mu = \mu_o, \sigma \neq 0\}$ – сложная гипотеза.

Пусть W_n – множество в выборочном пространстве R^n, не зависящее от θ, такое, что H отвергается, когда $\tau_n \in W_n$, и принимается в противном случае. Здесь возможны ошибки двух типов: *ошибка I-го рода* происходит при отклонении гипотезы H, когда она истинна, *ошибка 2-го рода* происходит при принятии H, когда она ложна.

Выбор множества W_n и отклонение гипотезы H при попадании τ_n в W_n называют статистическим критерием (тестом) для проверки гипотезы H или иначе критерием гипотезы H, W_n называется *крити-*

ческой областью теста. Коротко говоря, W_n является критерием гипотезы H.

Вероятности (риски) ошибок I-го и 2-го рода равны соответственно:

$$\mathrm{P}(W_n/\theta), \theta \in \mathrm{A},$$

$$1-\mathrm{P}(W_n/\theta), \theta \in \Theta \setminus \mathrm{A}.$$

Ошибки I-го и 2-го родов, по возможности, стараются свести к минимуму. При этом желательно обеспечить несмещенность W_n, которая для заданного *уровня значимости* α, 0<α<1 определяется условиями:

$$\mathrm{P}(W_n/\theta) \le \alpha, \theta \in \mathrm{A}, \tag{4.1}$$

$$\mathrm{P}(W_n/\theta) > \alpha, \theta \in \Theta \setminus \mathrm{A}. \tag{4.2}$$

Тест, не обладающий свойством несмещенности, называется смещенным.

Пусть W_n таков, что $\mathrm{P}(W_n/\theta) \le \alpha, \theta \in \mathrm{A}$, то говорят, что он имеет размерность (объем) α и записывают $W_{n\alpha}$ или просто W_α.

Величина $P(W_n/\theta) = \int\limits_{W_n} dF_n$, когда $\theta \in \Theta \setminus A$ при фиксированном θ, называется мощностью критерия W_n, она представляет собой функцию $\beta(\theta)$ (*функцию мощности)* со значениями в интервале [0,1]. Очевидно, мощность есть дополнение ошибки 2-го рода до единицы.

4.1.1. Примеры использования введенных выше понятий

В качестве 1-го примера рассмотрим *проверку гипотезы относительно средних арифметических.*

Пример 1. Пусть $(x_1,...,x_n) = \tau_n$ – выборка из $N(\mu,\sigma^2)$. Пусть $\mathrm{H}(\mathrm{A},\Theta)$ – сложная гипотеза, где Θ – множество точек (μ,σ^2) с $\sigma^2 > 0, \mu \in R, \mathrm{A} \subset \Theta$ с $\mu = \mu_0$, то есть A – полупрямая $\mu = \mu_o, \sigma^2 > 0$.

Если H истинна, то, согласно теореме 22, $\dfrac{\sqrt{n}(\bar{x}-\mu)}{s}$ имеет распределение Стьюдента S_{n-1}. Пусть W_α – множество тех точек *Rn*, в которых (см. рис. 1) $|t|=\left|\frac{(x-\mu_0)\sqrt{n}}{s}\right|>t_\alpha$, где $1-\int\limits_{-t_\alpha}^{+t_\alpha} f_{n-1}(t)dt=\alpha$ и $f_{n-1}(t)$ – плотность S_{n-1}.

Тогда $\mathrm{P}\left(W_n \mid \left(\mu,\sigma^2\right)\in \mathrm{A}\right)=\mathrm{P}\{|t|>t_\alpha\}=\alpha$ и, следовательно, тест W_α гипотезы $\mathrm{H}(\mathrm{A},\Theta)$ имеет объем α (т.е. выполняется (4.1)). Кроме того, для всякой точки $\left(\mu_1,\sigma_1{}^2\right)\in\Theta\setminus\mathrm{A}$:

$$(P(W_\alpha \mid (\mu_1,\sigma_1^2))=P\left(\left|\frac{(\bar{x}-\mu_0)\sqrt{n}}{s}\right|>t_\alpha / (\mu_1,\sigma_1^2))=P\left(|t+\frac{\delta}{\sqrt{u}}|>t_\alpha\right),$$

где $\delta=(\mu_1-\mu_o)\sqrt{n(n-1)}/\sigma_1$, а t и u – случайные величины, для которых совместная плотность

$$f(t,u)=\frac{(u/2)^{\frac{1}{2}(n-3)}}{2\sqrt{\pi(n-1)}\Gamma((n-1)/2)}e^{\frac{1}{2}(1+t^2(n-1))u} \text{ при } u\geq 0.$$

Но при $\forall\, u\in(0,\infty)$ (см. вид $f(t,u)$)

$$\int\limits_{-t_\alpha-\frac{\delta}{\sqrt{n}}}^{t_\alpha-\frac{\delta}{\sqrt{n}}} f(t,u)dt<\int\limits_{-t_\alpha}^{+t_\alpha} f(t,u)\,dt.$$

Отсюда $P(|t+\frac{\delta}{\sqrt{u}}|>P(|t|>t_\alpha)=\alpha,\ \ \delta\neq 0$

(т.е. выполняется (4.2)), что доказывает несмещенность W_α.

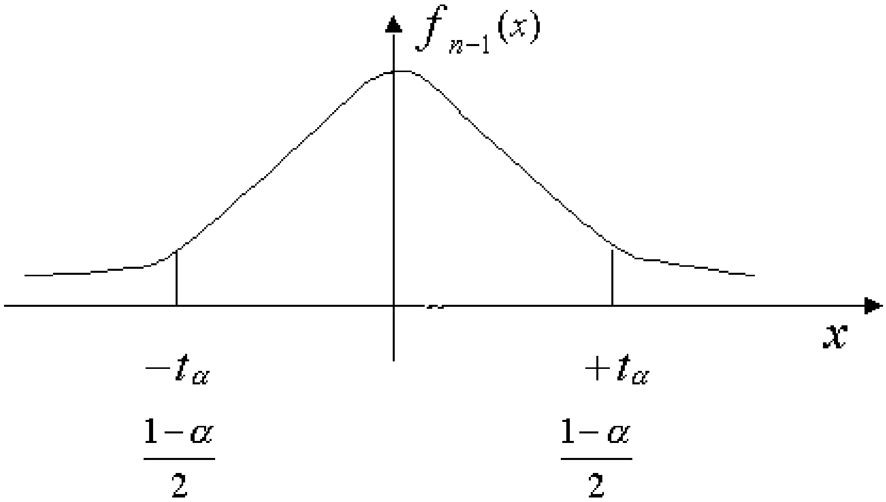

Рисунок 1

Пример 2. Построение функции мощности для проверки гипотезы о математическом ожидании μ относительно среднего арифметического при известной дисперсии σ^2.

Пусть $(x_1,...,x_n)$ – выборка из $N(\mu,\sigma^2)$. Пусть $\mathrm{H}(\mathrm{A},\Theta)$ – простая гипотеза $\mu=\mu_o$, где $\Theta=\{\mu\}$, $\mathrm{A}\in\Theta$.

Если Н истинна, то $\dfrac{(\overline{x}_1-\mu_0)\sqrt{n}}{\sigma}\in N(0,1)$.

Пусть W_n – множество тех точек R^n, в котором

$$|\,t\,|=\left|\;\frac{(\overline{x}_1-\mu_0)\sqrt{n}}{\sigma}\;\right|<t_\alpha.$$

Тогда $\mathrm{P}\{|\,t\,|>t_\alpha\}=\alpha$.

Для всякой точки (альтернативы) $\mu_1\in\Theta/A$

$$\beta(\mu_1)=P\{W_\alpha/\mu_1\}=P_\alpha\left(\left|\frac{(\bar{x}-\mu_1)\sqrt{n}}{\sigma}+d\right|>t_\alpha\right)=P\left(\frac{(\bar{x}-\mu_1)\sqrt{n}}{\sigma}>t_\alpha-d\right)+$$

$$+P\left(\frac{(\bar{x}-\mu_1)\sqrt{n}}{\sigma}<-t_\alpha-d\right)=1-P\left(\frac{(\bar{x}-\mu_1)\sqrt{n}}{\sigma}<t_\alpha-d\right)+P\left(\frac{(\bar{x}-\mu_1)\sqrt{n}}{\sigma}<-t_\alpha-d\right)=$$

$$=1-\Phi(t_\alpha-d)+\Phi(-t_\alpha-d)=1-\Phi(t_\alpha-d)-\Phi(t_\alpha+d)=P_{t_\alpha}(d),$$

где $d=\dfrac{(\mu_0-\mu_1)\sqrt{n}}{\sigma}$

t_α находится по таблицам [2] по доверительной вероятности α.

Если $d=0$, то $\mathrm{P}(W_\alpha/\mu_1)=\mathrm{P}(W_\alpha/\mu_o)=\beta(\mu_1=\mu_o)=(1-\alpha)/2=\beta$.

Если $d\neq 0$, то $\beta\le\mathrm{P}(W_\alpha/\mu_1)\le 1$.

Построим график (рис.2) ($\beta(\mu_1)=P(d)$).

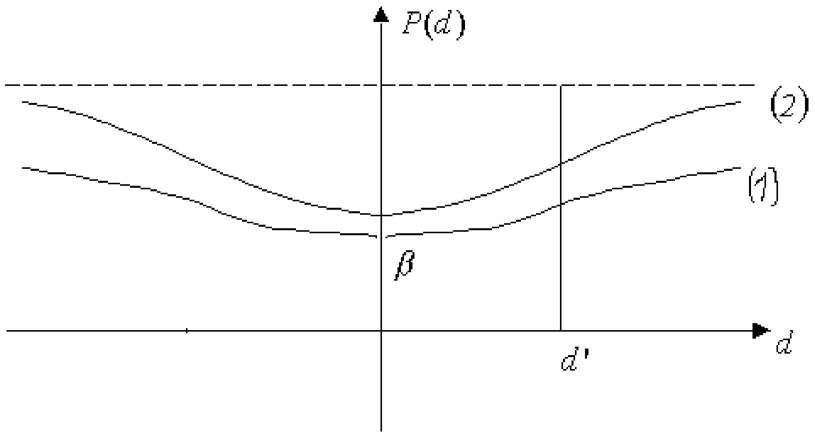

Рисунок 2

Если $d = d'$, то вероятность того, что мы отвергнем гипотезу, будет $\mathrm{P}(d')$, а примем – $1 - \mathrm{P}(d')$.

Пусть мы имеем другой критерий и соответственно другую функцию мощности. Тогда, если при одном и том же d' $\mathrm{P}_2(d') > \mathrm{P}_1(d')$, то есть вероятность отвергнуть неверную гипотезу по 2-му критерию больше, то и точность его выше; то есть тот критерий лучше, у которого функция мощности расположена выше.

Пусть W_α и $W_\alpha{}^*$ – два критерия объема α гипотезы $\mathrm{H}(\mathrm{A}, \Theta)$ таких, что

$$\begin{aligned} &\mathrm{P}\left(W_\alpha{}^* / \theta\right) \le \mathrm{P}(W_\alpha / \theta), \theta \in \mathrm{A}, \\ &\mathrm{P}\left(W_\alpha{}^* / \theta\right) > \mathrm{P}(W_\alpha / \theta), \theta \in \Theta \setminus \mathrm{A}. \end{aligned} \tag{4.3}$$

Тогда критерий $W_\alpha{}^*$ равномерно мощнее критерия W_α, т.е. следует предпочесть $W_\alpha{}^*$ перед W_α. Если (4.3) выполняется для всякого (борелевского) подмножества $W_\alpha \subset R_n$, то $W_\alpha{}^*$ называется равномерно *наиболее мощным критерием* (РНМК) для проверки гипотезы Н. К сожалению, несмещенные и равномерно наиболее мощные критерии существуют только в некоторых специальных случаях.

Нейманом и Пирсоном была поставлена задача определения критической области таким образом, чтобы при заданном уровне значимости ошибка 2-го рода была как можно меньше, т.е. функция мощности была как можно выше.

4.1.2. Лемма Неймана-Пирсона (случай простых гипотез)

1. Пусть с.в. X имеет плотность вероятностей $\rho(x, \theta)$. Проверяется простая гипотеза $\theta = \theta_o$ против простой альтернативы $\theta = \theta_1$. Обозначим как ранее $\rho_o(x_1) = \rho(x, \theta), \ldots; \rho_o(\tau_n) = \rho(\tau_n, \theta_o)$; $\rho_1(x) = \rho(x_1, \theta_1), \ldots;$ $\rho_1(\tau_n) = \rho(\tau_n, \theta_1)$, $\tau_n = (x_1, \ldots, x_n)$.

Лемма утверждает: для любого α, $0<\alpha<1$, можно выбрать константу k такую, что

$$\rho(W_\alpha / \theta_o) = \alpha, \tag{4.4}$$

где W_α – множество точек выборочного пространства, для которых

$$\rho_1(\tau_n) > k\rho_o(\tau_n). \qquad (4.5)$$

При этом W_α будет несмещенным и равномерно наиболее мощным критерием гипотезы $\theta = \theta_o$ против $\theta = \theta_1$.

Доказательство. Обозначим E (c) – множество точек выборочного пространства *R*, в которых $\rho(x_1,...,x_n;\theta_1) \geq cp(x_1,...,x_n;\theta_o)$. Рассмотрим функцию

$$g(c) = \mathrm{P}(E(c)/\theta_o) = \int\limits_{E(c)}....\int (x_1,...,x_n;\theta)dx_1...dx_n.$$

Функция *g(c)* является непрерывной и монотонно убывающей, *g(0) = 1,* $\lim\limits_{c\to\infty} g(c) = 0$, следовательно, для любого α (0<α<1) существует *c = k*, что, если положить $W_n = E(k)$, то $\mathrm{P}(W_n/\theta_o) = \alpha$. Пусть W_n' – любая другая критическая область объема α, то есть $\mathrm{P}\left(W_n'/\theta_o\right) = \alpha$. Тогда

$$\mathrm{P}\left(W_n \setminus \left(W_n \cap W_n'\right)/\theta_o\right) = \mathrm{P}\left(W_n'/\left(W_n \cap W_n'\right)/\theta_o\right).$$

Во всякой точке множества *Wn* по построению выполняется условие

$$p(x_1,...,x_n;\theta_1) > kp(x_1,...,x_n;\theta_o),$$

поэтому $\mathrm{P}\left(W_n \setminus \left(W_n \cap W_n'\right)/\theta_1\right) > k\,\mathrm{P}\left(W_n/\left(W_n \cap W_n'\right)/\theta_1\right)$.

Для всякой точки, не принадлежащей *Wn* ,

$$p(x_1,...,x_n;\theta_1) > kp(x_1,...,x_n;\theta_o)$$

и, следовательно,

$$k\,\mathrm{P}\left(W_n'/\left(W_n \cap W_n'\right)/\theta_o\right) > \mathrm{P}\left(W_n' \setminus \left(W_n \cap W_n'\right)\theta_1\right).$$

Это означает, что

$$\mathrm{P}(W_n \setminus (W_n \cap W_n)/\theta_1) > \mathrm{P}\left(W_n'/\left(W_n \cap W_n'\right)/\theta_1\right).$$

Множества $\left(W_n \setminus \left(W_n \cap W_n'\right)\right)$ и $\left(\left(W_n \cap W_n'\right)\right)$ не пересекаются так же, как и множества $\left(W_n' \setminus \left(W_n \cap W_n'\right)\right)$ и $\left(\left(W_n \cap W_n'\right)\right)$. Поэтому неравенство

$$\mathrm{P}\left(W_n \setminus \left(W_n \cap W_n'\right)/\theta_1\right) + \mathrm{P}\left(W_n \cap W_n' /\theta_1\right) > \mathrm{P}\left(W_n' /\left(W_n \cap W_n'\right)/\theta_1\right) + \mathrm{P}\left(W_n \quad \cap W_n' /\theta_1\right)$$

влечет $\mathrm{P}(W_n/\theta_1) > \mathrm{P}\left(W_n'/\theta_1\right)$.

Таким образом, W_n – более мощный критерий, чем любой W_n'. Несмещенность его очевидна (см. [19, с. 407]). Лемма доказана. Детали доказательства см. в [19, с. 406].

На основе леммы Неймана-Пирсона для независимых испытаний, учитывая несмещенность W_n (см. п.4.1.), (4.5) записывается в виде

$$\frac{p_1(\tau_n)}{p_0(\tau_n)} = \frac{p_1(x_1)...p_1(x_n)}{p_0(x_1)...p_0(x_n)} > k,\ \tau_n \in W_n,\ \ \frac{p_1(\tau_n)}{p_0(\tau_n)} = \frac{p_1(x_1)...p_1(x_n)}{p_0(x_1)...p_0(x_n)} \le k,\ \tau_n \in R_n \setminus W_n.$$

Таким образом, задавшись уровнем значимости α, зная распределение статистики отношения правдоподобия $\dfrac{p_1(\tau_n)}{p_0(\tau_n)}$, мы можем найти критическую зону W_n, то есть построить критерий $W_{n\alpha}$, так называемый *критерий отношения правдоподобия*.

Пример.

I. $X \in N(\theta, \sigma = 1)$; θ – неизвестный параметр.

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

H_0: $\theta = \theta_o$; $\mathrm{H}_1 : \theta = \theta_1 : \theta_1 > \theta_o$. Зададим α=0,05.

$$p_1(\tau_n) = \frac{1}{(2\pi)^{n/2}} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta_1)^2\right\} = \frac{1}{(2\pi)^{n/2}} exp\left\{-\frac{1}{2}\sum_i x_i^2 + \theta_1 \sum_i x_i - n\frac{\theta_1^2}{2}\right\},$$

$$p_0(\tau_n) = \frac{1}{(2\pi)^{n/2}} exp\left\{-\frac{1}{2}\sum_i x_i^2 + \theta_0 \sum_i x_i - n\frac{\theta_0^2}{2}\right\},$$

$$W_n = \left\{\tau_n : \frac{p_1(\tau_n)}{p_0(\tau_n)} > k\right\},\quad ln\frac{p_1(\tau_n)}{p_0(\tau_n)} = (\theta_1 - \theta_0)\sum_{i=1}^{n} x_i + \frac{n}{2}(\theta_0^2 - \theta_1^2) > ln\,k,$$

$$\sum_{i=1}^{n} x_i > \frac{ln\,k}{(\theta_1 - \theta_0)} - \frac{n}{2}\frac{(\theta_0^2 - \theta_1^2)}{(\theta_1 - \theta_0)} = k'.$$

Критическая область W_n состоит из $(x_1,...,x_n)$, сумма которых превосходит порог k', иначе говоря,

$$\frac{\sum(x_i-\theta_0)}{n} > \frac{k'-n\theta_0}{n} = k''.$$

$$P_0\{(\tau_n)\in W_n\} = P_o\left(\frac{\sum(x_i-\theta_0)}{n} > k''\right) \le \alpha.$$

Поскольку $\frac{1}{n}\sum_{i=1}^{n}(x_i-\theta_0)\in N(0,1/n),$ то по таблицам нормального распределения [2] имеем для α *=0,05,* k''=1,64$\sqrt{n}$ (сравни п. 3.2.3.1.).

4.1.3. Проверка сложной гипотезы

При заданном уровне α для разных θ критерии максимальной мощности различные. Желательно знать, когда существуют равномерно наиболее мощные критерии (РНМК) и для сложной гипотезы.

Справедливо следующее утверждение.

Теорема 32. Пусть X имеет $p(x,\theta)=G(\theta)h(x)\exp\{Q(\theta)T(x)\}$ – однопараметрическую плотность; θ – действительный параметр; $Q(\theta)$ – строго монотонная функция θ (либо возрастает, либо убывает); $\tau_n=\{x_1,...,x_n\}$ – выборка.

Для проверки гипотезы $\mathrm{H}_o:\theta\le\theta_o$ против альтернативы $\mathrm{H}_1:\theta\le\theta_o$ существует РНМК:

1. Если $Q(\theta)$ – возрастающая функция, то H_o отклоняется, когда

$$\sum_{j=1}^{n}T(x_j) > \gamma. \tag{4.6}$$

H_0 принимается в случае противоположного неравенства.

Критическая область W_n подмножества R_n

$$P_{\theta_1}(\tau_n(x)\in W_n) = P_{\theta_1}\left\{\sum_{j=1}^{n}T(x_j) > \gamma\right\} = \int_{\gamma}^{\infty} g(r,\theta)dr = \alpha,$$

где $g(z,\theta)$ – п.в. статистики $z=\sum T(x_j)$.

2. Если $Q(\theta)$ убывающая, то неравенство (4.6) меняет знак

$$W_n\{(x_1,...,x_n):\sum_{j=1}^{n}T(x_j) < \gamma\}.$$

Доказательство. Согласно лемме Неймана-Пирсона (см.п.4.I.2.) существует *k* такое, что неравенство

$$p\left(\tau_n, \theta_1\right) > k\ p\left(\tau_n, \theta_o\right) \qquad (4.7)$$

определяет равномерно наиболее мощный критерий для проверки гипотезы $\theta = \theta_1$ против альтернативы $\theta = \theta_o$.

В силу статической независимости элементов выборки имеем

$$p(\tau_n, \theta_1) = \prod_{i=1}^{n} [G(\theta_1) h(x_i)] exp\{Q(\theta_1) \sum_{i=1}^{n} T(x_i)\},$$

$$p(\tau_n, \theta_0) = \prod_{i=1}^{n} [G(\theta_0) h(x_i)] exp\{Q(\theta_0) \sum_{i=1}^{n} T(x_i)\}.$$

Тогда неравенство (2) примет вид

$$exp\,\{[Q(\theta_1) - Q(\theta_0)] \sum_{i=1}^{n} T(x_i)\} > k \left[\frac{G(\theta_0)}{G(\theta_1)}\right]^n.$$

Или

$$[Q(\theta_1) - Q(\theta_0)] \sum_{i=1}^{n} T(x_i) > ln \left[\frac{Q(\theta_0)}{Q(\theta_1)}\right]^n.$$

Если $Q(\theta)$ – строго возрастающая функция, то из $\theta > \theta_o$ следует, что

$$\sum_{i=1}^{n} T(x_i) > \frac{ln\, k \left[\frac{G(\theta_o)}{G(\theta)}\right]^n}{Q(\theta) - Q(\theta_o)}.$$

Таким образом, можно подобрать γ, что из $\theta > \theta_0$ следует

$$\sum_{j=1}^{n} T(x_j) > \gamma.$$

Аналогичный результат получается при строгой убываемости функции $Q(\theta)$. Теорема доказана.

Пример 1. Пусть $X \in N\left(0, \sigma^2\right), \sigma^2$ – неизвестный параметр, то есть

$$p(x, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2\sigma^2} x^2\right\}, \sigma > 0, x \in (-\infty, \infty).$$

Здесь $G(\sigma) = \dfrac{1}{\sigma\sqrt{2\pi}};\ h(x) = 1;\ T(x) = x^2;\ Q(\sigma) = -\dfrac{1}{2\sigma^2}$, то есть $Q(\sigma)$ – строго возрастающая функция. Рассмотрим соответствующие гипотезы $\mathrm{H}_o : \sigma \le \sigma_o, \mathrm{H}_1 : \sigma > \sigma_o$. В качестве статистики критерия возьмем

$$z = \sum_{i=1}^{n} x_j^2,$$

тогда

$$g(z,\sigma^2) = \frac{z^{n/2-1} e^{-z/2\sigma^2}}{(2\sigma^2)^{n/2-1}\Gamma\left(\frac{n}{2}\right)},$$

то есть z имеет $\chi^2{}_{(n)}$ – распределение. Тогда

$$P_{\sigma_0}\{\tau_n(x) \subset W_n\} = P_{\sigma_0}\{z > \gamma\} = \int_{\gamma}^{\infty} g(z,\sigma^2)dz = \alpha$$

По таблицам $\chi_{(n)}{}^2$ – распределения с n степенями свободы по α определяем γ [2].

Пример 2. Пусть случайная величина X имеет распределение Бернулли : $\rho(X = I) = 0, p(X = 0) = I - \theta$, где θ – неизвестный параметр. Рассмотрим гипотезы $\mathrm{H}_o : \theta \le \theta_o, \mathrm{H}_1 : \theta > \theta_o$. Тогда

$$p(\tau_n;\theta) = \theta^{\sum_{j=1}^{n} x_j} (1-\theta)^{n-\sum_{j=1}^{n} x_j}.$$

В этом случае $m = \sum_{j=1}^{n} x_j$, отсюда $G(\theta) = (1-\theta)^n$, $T(x) = m$, иначе говоря,

$$p(\tau_n;\theta) = \theta^m (1-\theta)^{n-m} = (1-\theta)^n \left(\frac{\theta^m}{1-\theta^m}\right)^m = (1-\theta)^n exp\left\{m \, ln \frac{\theta}{1-\theta}\right\},$$

где *Q(θ)* – возрастающая функция. Гипотеза H_0 отклоняется, если m>γ, где при фиксированном α $\sum_{m=\gamma}^{n} C_n^m \, \theta^m \, (1-\theta)^{n-m} = \alpha.$

Имеются соответствующие таблицы, позволяющие по α определить γ биномиального распределения [2].

Замечание. При проверке сложной гипотезы $\mathrm{H}_o : \theta \in \mathrm{A}$ против $\mathrm{H}_1 : \theta \in \Omega \setminus \mathrm{A}$ отношение правдоподобия будет иметь вид:

$$\lambda_n = \frac{\sup\limits_{\theta \in A} p(\tau_n, \theta_0)}{\sup\limits_{\theta \in \Omega \setminus A} p(\tau_n, \theta_1)}.$$

Пример.

Пусть $\tau_n = (x_1,...,x_n)$ – выборка из совокупности $N(\mu,\sigma^2)$. Определим критерий отношения правдоподобия для гипотезы $H_o : \{\mu = \mu_o, \sigma^2 > 0\}$ против альтернативы

$$H_1 : \{-\infty < \mu < \infty, \sigma^2 > 0\}.$$

Имеем

$$p(\tau_n,\theta) = L(x_1,...,x_n;\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k-\mu)^2\right\},$$

где $\theta = (\mu,\sigma^2)$.

Максимизируя $L(x_1,...,x_n,\theta)$ по $(\mu,\sigma^2)\in A$ и $(\mu,\sigma^2)\in\Omega\setminus A$ и беря отношение этих максимумов, получим:

$$\lambda_n = (s_o / s_1)^{n/2},$$

где $s_0 = \sum_1^n (x_k-\mu)^2$, $s_1 = \sum_1^n (x_k - \bar{x})^2$.

При этом λ_n можно записать в виде:

$$\lambda_n = (1+\frac{t^2}{n-1})^{-\frac{n}{2}},$$

где $t = \sqrt{n(x-\mu)/s}$, $s^2 = \frac{1}{n-1}\sum_1^n (x_k-\bar{x})^2$, с.в. t имеет распределение Стьюдента $S_{(n-1)}$, поэтому область объема α имеет вид

$$W_{o,\alpha} = \{(x_1,...,x_n) : \lambda_n > z_n\},$$

где $z_\alpha = (1+\frac{t_\alpha^2}{n-1})^{-\frac{n}{2}}$ и t_α определяется из

$$\int_{-t_\alpha}^{t_\alpha} f_{n-1}(x)dx = 1-\alpha,$$

где $f_{(n-1)}(x)$ – *п.в.* распределения $S_{(n-1)}$.

Метод χ^2. Этот метод используется для построения критериев проверки гипотез с помощью статистики типа χ^2– Пирсона (см. теорему 29).

Пусть, например, проверяется простая гипотеза H_o, согласно которой задается фиксированный вектор вероятностей $p = p(p_1, p_2, ..., p_n)$ исходов $E_i (i = 1, N)$ мультиномиального распределения $M(1; p_1, ..., p_n)$ (см. п.I.5.1, а) внутри сложной гипотезы H_1(против альтернативы H_1), согласно которой $p \in P$, где P-симплекс $P = \{p : p_i \geq 0, i = 1, N, p_1 + ... + p_n = 1\}$. Тогда статистика

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - np_i)^2}{np_i},$$

где n – число появления события E_i в n испытаниях, имеет предельное распределение $\chi^2{}_{(N-1)}$, если верна гипотеза H_o. Интересно сравнить в данном случае метод отношения правдоподобия и метод χ^2. Следуя первому, рассмотрим статистику

$$\lambda_n = \frac{L(p)}{\max\limits_{p \in P} L(p)}.$$

Ясно, что максимум $L(p)$ достигается во внутренней точке P, поскольку на границе $L(p)=0$. Решая задачу на условный экстремум функции $lnL(p)$ при условии $\sum_{i=1}^{N} p_i = 1$, легко найти, что экстремум достигается в единственной точке

$$\hat{p} = (\hat{p}_1, ..., \hat{p}_N) = (\frac{n_1}{n} ..... \frac{n_N}{n}).$$

Эта точка дает максимум $lnL(p)$, следовательно, р является оценкой максимального правдоподобия для p. Иначе говоря,

$$\lambda_n = \frac{L(p)}{L(\hat{p})} = \prod_{i=1}^{N} \left(\frac{np_i}{n_i} \right)^{n_i}.$$

Отсюда

$$-2 \ln \lambda_n = 2 \sum_{i=1}^{N} n_i ln \frac{n_i}{np_i}. \tag{4.8}$$

Преобразуем правую часть (4.8). Положим

$$n_i = np_i + z_i \sqrt{np_i},$$

где $z=\dfrac{n_i-np_i}{\sqrt{np_i}}$. Поскольку при $n\to\infty, n_i\in N(0,q_in), q_i=1-p_i$, то при достаточно больших n с вероятностью, близкой к 1, будут выполняться неравенства $|n_i|<T$, где T – достаточно большое постоянное число. Но тогда

$$ln\frac{n_i}{np_i}=ln(1+\frac{z_i}{\sqrt{np_i}})=\frac{z_i}{\sqrt{np_i}}-\frac{z_i^2}{2np_i}+\mathrm{O}(\frac{T^3}{n^{3/2}}).$$

Поэтому

$$2\sum_{i=1}^{N}n_i\,ln\frac{n_i}{np_i}=2\sum_{i=1}^{N}(np_i+z_i\sqrt{np_i})\left(\frac{z_i}{\sqrt{np_i}}-\frac{z_i^2}{2np_i}+\mathrm{O}(\frac{T^3}{n^{3/2}})\right)=$$
$$=2\sum_{i=1}^{N}z_i\sqrt{np_i}+\sum_{i=1}^{N}z_i^2+\mathrm{O}(\frac{T^3}{\sqrt{n}}).$$

Поскольку $\sum_{i=1}^{N}z_i\sqrt{np_i}=\sum_{i=1}^{N}(n_i-np_i)=0$, то окончательно получим, что при $n\to\infty$, с.в. $2\ln\lambda_n$ и $\chi^2=\sum_{i=1}^{N}\dfrac{(n_i-np_i)^2}{np_i}$ с вероятностью, стремящейся к единице, совпадают. Отсюда следует, что совпадают и их предельные распределения, а также, что критические области

$$\{\ \sum_{i=1}^{N}\frac{(n_i-np_i)^2}{np_i}>t_\alpha\ \}\ ,\ \{\ 2\sum_{i=1}^{N}n_i\,ln\frac{n_i}{np_i}<t_\alpha\ \},$$

где $P\{\chi^2_{(N-1)}>t_\alpha\}=\alpha$, эквивалентны.

Оба метода обладают одинаковыми асимптотически оптимальными свойствами (в частности, состоятельностью).

Имеются и другие методы проверки гипотез.

4.1.4. Последовательный анализ и последовательный критерий Вальда

Специфической особенностью рассмотренных критериев является то, что объем наблюдений определяется до начала эксперимента. В этом случае размерность выборочного пространства известна заранее. Выборочное пространство разбивается на два непересекающихся множества W_0 и W_1.

Вальд в 1940 г. предложил последовательный анализ. Объем выборки заранее не фиксируется, и, следовательно, размерность выборочного пространства заранее не определена. Выборочное пространство разбивается на три непересекающихся подмножества: множество принятия гипотезы; множество отклонения гипотезы; множество принятия решения провести еще одно наблюдение (зона неуверенности)

$$R_n = W_0^n \bigcup W_1^n \bigcup W_2^n ,$$

где – n случайное число.

Если $\tau_n \in W_0^n \leftrightarrow d_0 : H_0$,

$\tau_n \in W_1^n \leftrightarrow d_1 : H_1$,

$\tau_n \in W_2^n \leftrightarrow d_2$ – промежуточное решение,

где $\leftrightarrow$ – знак принятия решения (например, отвергнуть гипотезу H_0).

Рассмотрим простую гипотезу, простую альтернативу:

$$H_0 : \theta = \theta_0 , H_1 : \theta = \theta_1 .$$

Определяются два числа $L_0 < L_1$,

$L_0 < \dfrac{p(\tau_k , \theta_1)}{p(\tau_k , \theta_0)} < L_1$, принимается d_2.

Если $\dfrac{p(\tau_k , \theta_1)}{p(\tau_k , \theta_0)} > L_1$, то $\leftrightarrow d_1$.

Если $\dfrac{p(\tau_k , \theta_1)}{p(\tau_k , \theta_0)} \le L_0$, то $\leftrightarrow d_0$.

Конечно, и при последовательном анализе можно предложить множество критериев (по-разному определив L_0 и L_1). Наилучший критерий определяется следующим образом: задаются вероятности α_0 отклонения гипотезы, когда $\theta = \theta_0$ и α_1 принятия гипотезы, когда $\theta = \theta_1$. Из всех правил с одинаковыми α_0 и α_1 выбирается то, которое минимизирует средний объем наблюдений при $\theta = \theta_0$ и $\theta = \theta_1$,

$$\begin{cases} \min \ M(k | \theta = \theta_0), \\ \min \ M\left(k | \theta = \theta_1\right). \end{cases}$$

Такой критерий называется последовательным критерием отношения правдоподобия. В нем представляет некоторую трудность определение L_0 и L_1. Можно показать, что

$$L_0 \le \frac{\alpha_1}{1-\alpha_0};$$

$$L_1 \ge \frac{\alpha_1}{1-\alpha_0}.$$

Поэтому L_0 и L_1 можно заменить этими отношениями. При этом интервал (L_0, L_1) расширяется, ошибка аппроксимации L_0 и L_1 невелика и завышение числа испытаний будет малым.

Обозначим $p_{0k} = p(\tau_k, \theta_0)$; $p_{1k} = p(\tau_k, \theta_1)$,

$\ln \frac{p_{1k}}{p_{0k}} = \sum_{i=1}^{k} \ln \frac{p_{1i}}{p_{0i}}$, тогда

$$\begin{cases} d_2 : \ln \dfrac{\alpha_1}{1-\alpha_0} < \sum\limits_{i=1}^{k} \ln \dfrac{p_{1i}}{p_0} < \ln \dfrac{1-\alpha_1}{\alpha_0}, \\ d_0 : \sum\limits_{i=1}^{k} \ln \dfrac{p_{ij}}{p_{0i}} \le \ln \dfrac{\alpha_1}{1-\alpha_0}, \\ d_1 : \sum\limits_{i=1}^{k} \ln \dfrac{p_{1i}}{p_{0i}} > \ln \dfrac{1-\alpha_1}{\alpha_0}. \end{cases}$$

Поскольку k – с.в., то $\sum_{i=1}^{k} \ln \frac{p_{1i}}{p_{0i}}$ также является с.в.

Математическая задача последовательной проверки гипотезы сводится к задаче блужданий (т.е. ищется момент первого выхода статистики из зоны неуверенности).

Пример.

$$\begin{cases} P(x=1) = \theta, \\ P(x \ne 1) = 1-\theta. \end{cases}$$

$H_0 : \theta = \theta_0$,

$H_1 : \theta = \theta_1$.

$$\begin{cases} \theta_0 = \dfrac{1}{2}, \\ \theta_1 = \dfrac{3}{4}, \end{cases}$$

$$\frac{p_{1k}}{p_{0k}}=\left(\frac{\theta_1}{\theta_0}\right)^{\sum x_i}\left(\frac{1-\theta_1}{1-\theta_0}\right)^{k-\sum x_i}=\left(\frac{1-\theta_1}{1-\theta_0}\right)^{k}\left(\frac{\theta_1}{\theta_0}\right)^{\sum x_i}\left(\frac{1-\theta_0}{1-\theta_1}\right)^{\sum x_i},$$

$$\ln\frac{p_{1k}}{p_{0k}}=\sum_{i=1}^{k}x_i\cdot\ln\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}+k\ln\frac{1-\theta_1}{1-\theta_0}.$$

Зададим $\alpha_0=\alpha_1=0{,}01$, тогда для удобства берем логарифм по основанию 2:

$$\log_2\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}=\log_2\frac{\frac{3}{4}\frac{1}{2}}{\frac{1}{2}\frac{1}{4}}=\log_2 3\cong 1{,}58,$$

$$\log_2\frac{(1-\theta_1)}{(1-\theta_0)}=\log_2\frac{\frac{1}{4}}{\frac{1}{2}}=-1,$$

$$\log_2\frac{1-\alpha_1}{\alpha_0}=6{,}63,$$

$$\log_2\frac{\alpha_1}{1-\alpha_0}=-6{,}63.$$

Итак, окончательно получаем

$$\begin{cases}d_2:-6{,}63<1{,}58\sum x_i-k<6{,}63,\\ d_0:1{,}58\sum x_i-k\le -6{,}63,\\ d_1:1{,}58\sum x_i-k>6{,}63.\end{cases}$$

4.2. Непараметрические критерии проверки гипотез

Характерным для непараметрической статистики является то, что в ее задачах распределения вероятностей считаются полностью известными, а сами задачи формируются в терминах различий между классами распределений или распределениями внутри заданного класса. Непараметрическая статистика устанавливает некоторые непараметрические факты – свойства выборки (или ее преобразований), которые не зависят от вида распределений генеральной совокупности.

Любая задача проверки непараметрических гипотез выглядит следующим образом. Из двух конкурирующих гипотез альтернатива всегда непараметрична, а нулевая гипотеза может быть либо простой, либо непараметрической. Так как одна из гипотез непараметрична, различие между гипотезами задается в некотором общем виде, не связанном с конкретным видом ф.р. Требуется предложить тест на основе выборки, результатом которого явилось бы решение об истинности одной из гипотез.

Отметим основные непараметрические задачи [28].

1. Задача согласия.

Пусть дано известное непрерывное распределение $F(x)$. Из неизвестного распределения $G(x)$, принадлежащего классу всех распределений, берется выборка $(x_1, x_2, \ldots, x_n)$.

Конкурирующие гипотезы:

нулевая гипотеза $H_0 : F = G$ – простая гипотеза.

Альтернатива:

а) $H_1^+ : F < G$;

б) $H_1^- : F > G$ – односторонние гипотезы;

в) $H_1 : F \neq G$ – двусторонняя гипотеза.

2. Задача сдвига (расположения).

Простейший случай. Известно, что альтернатива сводится только к сдвигу, то есть

$$F(x|\theta) = F_0(x - \theta).$$

Нулевая гипотеза $H_0 : \theta = 0$.

Альтернатива:

а) $H_1' : \theta > 0$;

б) $H_1'' : \theta < 0$ – односторонние гипотезы;

в) $H_1'' : \theta \neq 0$ – двусторонняя гипотеза.

Естественной мерой сдвига являются квантили. Тогда имеем задачу: нулевая гипотеза $H_0 : F^{-1}(p) = x_{-p}$

Альтернатива

а) $H_1^{'}: F^{-1}(p) > x_{-p}$;

б) $H_1^{''}: F^{-1}(p) < x_{-p}$ – односторонние гипотезы;

в) $H_1^{''}: F^{-1}(p) \neq x_{-p}$ – двусторонняя гипотеза.

Иногда известна симметричность распределения относительно медианы, тогда

$$H_0: F^{-1}(0{,}5) = x_0$$

и $F(x)$ симметрична относительно x_0:

а) $H_1^{'}: F^{-1}(0{,}5) = x_0,\ F(x)$ – симметрична;

б) $H_1^{''}: F^{-1}(0{,}5) \neq x_0,\ F(x)$ – несимметрична.

3. Задача расположения и симметрии (задача проверки симметричности распределения).

$H_1: F^{-1}(0{,}5) = x_0,\ F(x)$ – симметрична.

$H_1: F^{-1}(0{,}5) \neq x_0,\ F(x)$ – несимметрична.

4. Задача масштаба.

Если известно, что изменяется только масштаб, имеем альтернативу вида $F(x|\theta) = F_0(\theta x)$ и задачу: $H: \theta = 1$:

а) $H_1^{'}: \theta > 1$;

б) $H_1^{''}: \theta < 1$ – односторонние гипотезы;

в) $H_1^{''}: \theta \neq 1$ – двусторонняя гипотеза.

5. Задача независимости.

$H_0: F(x_1,...,x_n) = F(x_1)F(x_2)...F(x_n)$ для всех $x_1, x_2,..., x_n$;

$H_0: F(x_1,...,x_n) \neq F(x_1)F(x_2)...F(x_n)$ хотя бы при некоторых $x_1, x_2,..., x_n$.

6. Задача случайности.

Многие из результатов математической статистики получены в предположении, что выборка является чисто случайной, то есть состоит из одинаково распределенных и независимых с.в. Проверка этого факта:

$$H_0: F(x_1,...,x_n) = \prod_{i=1}^{n} F(x_i),$$

$$H_0 : F(x_1,...,x_n) = \prod_{i=1}^{n} F_i(x_i) \neq \prod_{i=1}^{n} F(x_i).$$

Отмеченные задачи допускают двухвыборочную формулировку (например, в смысле сравнения выборок или проверки их однородности (см. п.7 ниже)).

Проверка гипотез как в параметрическом, так и в непараметрическом вариантах, осуществляется с помощью тестовых статистик, которые являются статистическими оценками функционалов, выражающих степень различия между конкурирующими гипотезами.

В задаче согласия гипотетическое распределение $F(x)$ считается полностью заданным, и требуется обнаружить любое отличие G от F. Поэтому и функционал задается так, чтобы его значение изменялось при любых отклонениях G от F. Обычно рассматривают функционалы двух типов – интегральные и супремальные:

$$\rho_1(F,G) = \int \phi(F,G)dF, \;\; \rho_2(F,G) = \sup_x \phi(F,G).$$

В качестве примеров критериев согласия рассмотрим статистики, основанные на измерении «расстояния» между теоретической и эмпирической функциями распределения. Они обладают свойством непараметричности в том смысле, что при нулевой гипотезе распределения ρ_1 и ρ_2 не зависят от $F(x)$. Это свойство является следствием применения преобразования с помощью ф.р., а именно того факта, что если с.в. X имеет ф.р. $G(x)$, то с.в. $u = G(x)$ распределена равномерно на интервале (0,1) (см. теорему 12).

Таким образом, если дана выборка ($x_1, x_2,..., x_n$) из $G(x)$ и ее вариационный ряд ($x_{(1)}, x_{(2)},..., x_{(n)}$), то с.в. $u_1 = G(x_1)$,…, $u_n = G(x_n)$ имеют равномерное на интервале (0,1) распределение. Соответственно порядковые статистики равны $u_{(1)} = G(x_{(1)})$,…, $u_{(n)} = G(x_{(n)})$.

В таком виде задачу называют каноническим видом задачи о непараметрическом критерии согласия, и она сводится к проверке того, можно ли рассматривать наблюденные точки $u_1,...,u_n$ как равномерно распределенные на отрезке [0,1].

Односторонние и двусторонние статистики Колмогорова-Смирнова.

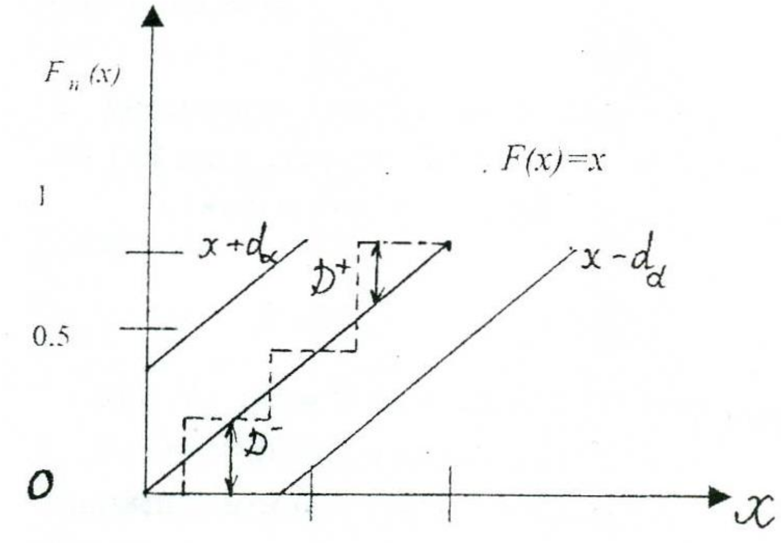

Рисунок 3 – Двусторонний критерий согласия Колмогорова – Смирнова.

– – – – – – – – – – – – эмпирическая ф.р.

———————— теоретическая ф.р.

$$D_n^+ = \sup_{-\infty<x<\infty}\{F_n(x)-F(x)\} = \max_{1\le i\le n}\left\{\frac{i}{n}F(x_{(i)})\right\},$$

$$D_n^- = \sup_{-\infty<x<\infty}\{F(x)-F_n(x)\} = \max_{1\le i\le n}\left\{F(x_{(i)})-\frac{i-1}{n}\right\},$$

$$D_n = \sup_{-\infty<x<\infty}\{F_n(x)-F\ (x)\} = \max_{1\le i\le n}\{D_n^+, D_n^-\}.$$

Пусть $d_{n,\alpha}$ означает такое число, что $P(D_n > d_{n,\alpha}) = \alpha$. Тогда следует отвергнуть нулевую гипотезу с уровнем значимости α, если наблюденное значение статистики D_n больше, чем $d_{n,\alpha}$.

Статистика D_n позволяет получить доверительные границы для неизвестной ф.р. с.в. X. Действительно, при данной эмпирической ф.р. нулевая гипотеза о $F(x)$ принимается с уровнем значимости α, согласно критерию D_n, всякий раз, когда для всех x

$$F_n(x) - d_{n,\alpha} \le F(x) \le F_n(x) + d_{n,\alpha}.$$

Поэтому область (см. рис. 3), которая ограничена сверху наименьшим из значений $F_n(x)+d_{n,\alpha}$ и 1, а снизу – наибольшим из значений $F_n(x)-d_{n,\alpha}$ и 0, является доверительной областью для $G(x)$ с коэффициентом доверия 1-α. Критерии D_n^+ и D_n^- являются несмещенными и состоятельными. D_n является смещенным при непрерывных конкурирующих гипотезах. Их распределения табулированы.

Критерий Крамера-Мизеса.

Его статистика имеет вид [17]:

$$\omega_n^2 = \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x) = \frac{1}{12n^2} + \frac{1}{n}\sum_{i=1}^{n} \left(F(x_{(i)}) - \frac{(2i-1)}{2n}\right)^2.$$

Приведем формальные преобразования, доказывающие это равенство:

$$\int_{-\infty}^{\infty}(F_n(x) - F(x))^2 dF(x) = \sum_{i=0}^{n} \int_{x_{(i)}}^{x_{(i+1)}} \left(\frac{i}{n} - F(x)\right)^2 dF(x) = \sum_{i=0}^{n} \Bigg(\int_{x_{(i)}}^{x_{(i+1)}} \left(\frac{i}{n}\right)^2 dF(x) -$$

$$\frac{2i}{2n}\int_{x_{(i)}}^{x_{(i+1)}} dF^2(x) + \frac{1}{3}\int_{x_{(i)}}^{x_{(i+1)}} dF^3(x) \Bigg) = \sum_{i=0}^{n} \Bigg(\left(\frac{i}{n}\right)^2 (F(x_{(i+1)}) - F(x_{(i)})) - \frac{i}{n}(F^2(x_{(i+1)}) - F^2(x_{(i)})) -$$

$$\frac{1}{3}(F^3(x_{(i+1)}) - F^3(x_{(i)})) \Bigg) = \sum_{i=0}^{n} \left(\frac{i}{n}\right)^2 F(x_{(i+1)}) - \sum_{j=-1}^{n-1} \left(\frac{j+1}{n}\right)^2 F(x_{(j+1)}) -$$

$$-\sum_{i=0}^{n} \frac{i}{n} F^2(x_{(i+1)}) + \sum_{j=-1}^{n-1} \frac{j+1}{n} F^2(x_{(j+1)}) + \frac{1}{3} = \sum_{i=0}^{n-1} \frac{F(x_{(i+1)})}{n^2}(i^2 - (i+1)^2) +$$

$$+\sum_{i=0}^{n-1} \frac{F^2(x_{(i+1)})}{n}(i - i + 1) + \frac{1}{3} = \sum_{i=1}^{n} \frac{F^2(x_{(i)})}{n} - \sum_{i=1}^{n} \frac{F(x_{(i)})}{n^2}(2i-1) + \frac{1}{3},$$

поскольку

$$F(x_{(0)}) = 0,\ F(x_{(n+1)}) = 1.$$

Учитывая далее, что

$$\frac{1}{n}\sum_{i=1}^{n} \left(\frac{(2i-1)}{2n}\right)^2 = \frac{1}{4n^3}\sum_{i=1}^{n} (4i^2 - 4i + 1) = \frac{1}{3} - \frac{1}{12n^2},$$

получим

$$\omega_n^2 = \sum_{i=1}^{n} \frac{F^2(x_{(i)})}{n} - \sum_{i=1}^{n} \frac{F(x_{(i)})}{n^2}(2i-1) + \frac{1}{3} - \frac{1}{12n^2} + \frac{1}{12n^2} =$$

$$= \sum_{i=1}^{n} \frac{F^2(x_{(i)})}{n} - \sum_{i=1}^{n} \frac{2F(x_{(i)})}{2n^2}(2i-1) + \frac{1}{n}\sum_{i=1}^{n} \left(\frac{(2i-1)}{2n}\right)^2 + \frac{1}{12n^2} =$$

$$= \frac{1}{12n^2} + \frac{1}{n}\sum_{i=1}^{n} \left(F(x_{(i)}) - \frac{(2i-1)}{2n}\right)^2,$$

что и требовалось доказать.

Распределение ω_n^2 при малой выборке известно частично. При n> 20 можно пользоваться асимптотическим распределением.

Критерий χ^2 – Пирсона.

За меру отклонения эмпирической ф.р. от теоретической принимается величина

$$\chi^2 = \sum_{i=1}^{r} \frac{n}{p_i}\left(\frac{m}{n} - p_i\right)^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i},$$

где p_i (*i*-я групповая вероятность) есть приращение гипотетической функции на *i*-м множестве $S_i,\ i = \overline{1,r}.$ точек некоторого разбиения области значений с.в. X; $\frac{n_i}{n}$ – приращение распределения $F_n(x)$ выборки $(x_1, x_2, ..., x_n)$ на том же множестве (*i*-я групповая частота).

Известна следующая теорема Пирсона, которая является следствием теоремы 29.

Теорема 33. Какова бы ни была ф.р. $F(x)$ с.в. X , при $n \to \infty$ распределение величины χ^2 стремится к $p_{r-1}(x)$, то есть

$$P(\chi^2 < x) \to \int_{-\infty}^{x} p_{r-1}(t)dt$$

в каждой точке x, где $p_{r-1}(x)$ – плотность χ^2 – распределения с $(r-1)$ степенями свободы.

Если n достаточно велико $P(\chi^2 < \chi_\alpha^2) = \alpha$, где χ_α^2 – решение интегрального уравнения $\int_{\chi_\alpha^2}^{\infty} p_{\kappa-1}(x)dx = \alpha$, которое находится по таблицам при заданных α и $r-1$.

Нулевую гипотезу следует отвергнуть с уровнем значимости α, если наблюденное значение χ^2 больше χ_α^2.

Критерий Пирсона удобен при группированной выборке и дает хорошие результаты во всех случаях, когда $np_i \geq 10, \quad i = 1,2,...,r.$

Замечание. Рассмотренный критерий χ^2 Пирсона следует отличать от аналогичного критерия, который используется в случае, когда по выборке оцениваются некоторые неизвестные параметры $\theta_1, ..., \theta_k$.

Нулевая гипотеза состоит в том, что $G(x) = F(x;\theta_1,...,\theta_k)$ при некоторых значениях параметров. Тогда

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - np_i(\theta_1,...,\theta_k))^2}{np_i(\theta_1,...,\theta_k)}.$$

Но $\theta_1,...\theta_k$ неизвестны; если же воспользоваться их оценками по выборке, то величины $\theta_1,...\theta_k$ сами станут случайными. Фишер доказал, что при весьма общих условиях при $n \to \infty$ предельным распределением для χ^2 является $\chi^2_{(r-k-1)}$, если неизвестные параметры оцениваются по выборке методом минимума χ^2 (или максимума правдоподобия).

Типичный критерий для гипотезы сдвига – критерий знаков.

Пусть $(x_1,...,x_k)$ – выборка наблюдений над с.в. X с непрерывной ф.р. Гипотеза – медиана п.в. с.в. X равна $x_0 : F^{-1}(0{,}5) = x_0$.

Пусть r из величин $(x_1 - x_0),...,(x_n - x_0)$ отрицательны, а s – положительны, где $r + s \le n$. Условное распределение с.в. r при фиксированном $r+s$ является биномиальным $Bi(r+s, p=0{,}5)$. Испытуемая гипотеза эквивалентна гипотезе о том, что $p = 0{,}5$, если известно, что в схеме испытаний Бернулли в $r+s$ испытаниях осуществлялось r успехов. Можно использовать точный критерий, основанный на таблицах биномиального распределения, либо приближенный, так как при больших $r+s$ распределение с.в. r нормально.

7. Двухвыборочные критерии однородности.

Критерий χ^2.

Пусть $(x_1,...,x_m)$ – простая выборка объема m из непрерывного распределения F_1 с.в. X, аналогично $(y_1,...,y_n)$ – выборка обьема n из непрерывного распределения F_2 с.в. Y. Мы хотим проверить параметрическую нулевую гипотезу $F_1 = F_2$ или иначе – проверить, что выборки взяты из одной совокупности, если ничего неизвестно о гипотетических групповых вероятностях p_i.

Введем обозначения наблюденных частот в двух выборках (см. Таблицу 1)

Таблица 1

Группы / Выборка	1	2	...	k	$\sum$
Первая	n_{11}	n_{12}	$\dots$	n_{1k}	$n_{1.}$
Вторая	n_{21}	n_{22}	$\dots$	n_{2k}	$n_{2.}$
Сумма $(n_{.i})$	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.k}$	$n_{..}$
$q_i = \dfrac{n_1(i)}{n(.i)}$ $i = 1,\dots,k.$	q_1	q_2	$\dots$	q_k	q

Пусть $p_{11},\dots,p_{1k};\ p_{21},\dots,p_{2k}$ – групповые вероятности в двух совокупностях. Гипотеза, которую требуется проверить, состоит в том, что $p_{1i} = p_{2j}, i = 1,2,\dots k.$

Таким образом, можно использовать критерий χ^2 в модификации, отмеченной в замечании предыдущего раздела.

Так как $p_{1i} = \dfrac{n_{1i}}{n_1};\ p_{2i} = \dfrac{n_{2i}}{n_2};\ p_i = \dfrac{n_i}{n_{1..}}$,

то $\chi^2 = \sum\limits_{j=1}^{2}\sum\limits_{i=1}^{k}\dfrac{\left(n_{ji} - n_{j.}p_{ji}\right)^2}{n_{j.}p_{ji}} = \dfrac{1}{q(1-q)}\left(\sum\limits_{i=1}^{k} n_{1i}q_i - n_{1.}q\right)$ имеет $k-1$ степеней свободы, ибо проверяемая гипотеза состоит в равенстве $k-1$ независимых вероятностных исходов.

Критерий типа Колмогорова-Смирнова.

Обозначим $F_{1,m}$ и $F_{2,n}$ эмпирические ф.р. выборок $(x_1',\dots,x_m')$, $(x_1^2,\dots,x_n^2)$ соответственно, тогда

$$K^+ = \left(\frac{mn}{m+n}\right)^{1/2} \max_{-\infty<x<\infty}\left(F_{2,n}(x) - F_{1,m}(x)\right),$$

$$K^\pm = \left(\frac{mn}{m+n}\right)^{1/2} \max_{-\infty<x<\infty}\left|F_{2,n}(x) - F_{1,m}(x)\right|.$$

Как K^+, так и $K^\pm$ называют статистиками Колмогорова-Смирнова.

Пусть

$$C_i = \begin{cases} 1, & i = 1,\ldots,m, \\ 0, & i = m+1,\ldots,m+n. \end{cases}$$

Антиранг $D_k = j$ тогда и только тогда, когда k – я порядковая статистика $x_{(k)}$ есть x_j. Иначе, $C_{D_k} = 1$, если k – я порядковая статистика $x_{(k)}$ принадлежит второй выборке. $C_{D_k} = 0$, если $x_{(k)}$ принадлежит второй выборке. Тогда

$$K^+ = \left(\frac{m+n}{mn}\right)^{1/2} \max_{-\infty < x < \infty}\left(K\frac{m}{m+n} - C_{D_1} - \ldots - C_{D_k}\right),$$

$$K^\pm = \left(\frac{m+n}{mn}\right)^{1/2} \max_{-\infty < x < \infty}\left|K\frac{m}{m+n} - C_{D_1} - \ldots - C_{D_1}\right|.$$

Критерии χ^2, K^+ и $K^\pm$ полезны при анализе произвольных отклонений от $H_0 : F_1 = F_2$, если же необходимы только средние значения отклонений, то применяют следующие более простые критерии:

Критерий серий Вальда-Вольфовитца (блочности структур).

Расположим $m+n$ наблюдений смешанных выборок в возрастающем порядке.

Каждое наблюдение первой выборки заменим на 0, а второй на – 1, получим последовательность 0 и 1. Если $F_1 = F_2$, то все $\frac{(m+n)!}{n!m!}$ возможных расположений m нулей n и единиц равновероятны и можно точно определить распределение любой статистики, основанной на последовательности нулей и единиц. Одной из таких статистик является число серий нулей и единиц. Наблюденное число серий считается значимым, если оно больше заранее выбранного согласно уровню значимости числа.

Критерий Уилкоксона.

Расположим $m+n$ наблюдений в порядке возрастания и перенумеруем их заново от 1 до $m+n$. Пусть номера второй выборки будут $s_1,...,s_n$. Критическим значением является сумма этих номеров.

Если верна нулевая гипотеза, то вероятность любого выбора номеров равна $\frac{n!m!}{(m+n)!}$ и, следовательно, можно получить распределение их суммы. Большое значение статистики (суммы номеров) может указать на отклонение от нулевой гипотезы.

Глава 5. СТАТИСТИЧЕСКИЕ РЕШАЮЩИЕ ПРАВИЛА И ОПТИМАЛЬНЫЕ ВЫВОДЫ

5.1. Постановка статистической задачи. Статистическая структура. Функция риска. Упорядочение решающих правил

Предметом математической статистики является статистическая структура, т.е. тройка объектов <S, *B* (S) , *P* >, в которой S – множество, имеющее смысл пространства наблюдений, *B*(S) – σ-алгебра соответствующих измеримых (борелевских) множеств, иначе говоря,

<S, *B* (S) >, – выборочное измеримое пространство; *P* – семейство распределений на выборочном пространстве, причем обычно эти распределения параметризуются: $P = \langle P_\theta, \theta \in \Theta \rangle$, где Θ – пространство параметров.[1] Если это семейство состоит из одного элемента, то статистическая структура превращается в вероятностное пространство.

Исходным материалом при изучении случайных явлений статистическими методами обычно служит совокупность наблюдений (выборка) $\tau_n = (x_1, ..., x_n)$ (кратко τ_n обозначают через $x \in$ S, где S = R^n - $n-$мерное евклидово пространство) значений некоторой с.в., имеющей ф.р. $F(x;\theta)$. Предположим, что $P_\theta, \theta \in \Theta$ на { R^n, *B* (R^n) } имеет вид

[1] Более тонкие построения, связанные с понятием статистической структуры, см. в [1, 21]. Обобщение понятия статистической структуры связано с построением конкретного плана П – правила проведения эксперимента, испытания и сбора уже существующих данных [32]. Данные могут быть представлены в виде таблиц, списков и т.п., содержащих выборки как вещественных чисел, так и значений величин, не имеющих числовой интерпретации. Непараметрические статистические структуры обычно строятся на базе использования распределения Дирихле.

$$\mathrm{P}_{\theta}(A) = \int\limits_{A}\!...\!\int dF(x_1;\theta)...dF(x_n;\theta),$$

где $F(x;\theta)$ – некоторая ф.р. в R^1, с.в. X. В этом случае говорят, что распределение P_{θ} является n-й степенью F,

$$d\mathrm{P}_{\theta} = dF(x_1;\theta)...dF(x_n;\theta).$$

Статистические выводы делаются по выборке τ_n, например, об истинном значении θ. Говорят, что по τ_n выносится решение $d \in D$, кратко обозначая $\tau_n \leftrightarrow d$; D – множество возможных решений, в частности, $D = \Theta$. Отображение $\delta : S \to D$ называется решающим правилом (решающей процедурой, решающей функцией). Решающая функция $\delta \equiv \delta(x) \equiv \delta(\tau_n)$ – это правило, определяющее конкретное решение, которое должно быть принято для заданного $x = \tau_n$.

Задача математической статистики состоит в отыскании решающей процедуры в некотором классе $\Delta = \{\delta\}$ решающих правил.

Возникает необходимость иметь некоторые критерии, позволяющие упорядочивать решающие правила. С этой целью следующим образом вводится понятие функции риска.

Функция $\mathrm{L}(d/\theta)$, определенная на $\Theta \times D$, называется функцией потерь, если она определяет потери, обусловленные ошибочным оцениванием. Иными словами, это мера погрешности при использовании решения *d*, например, дисперсия оценивания, функционалы Колмогорова и ω_n^2.

Пусть $\mathrm{L}\ (d/\theta)$ – потери при использовании решения d, если истинное значение параметра есть θ. Пусть $\delta = \delta(x)$ – некоторое решающее правило, $\tau_n \leftrightarrow d$. Тогда можно определить средние потери $R(\delta/\theta)$ при использовании δ, если истинное значение параметра равно θ, следующими равенствами:

$$R(\delta/\theta) = M_{\theta}\mathrm{L}\,(\delta(x)/\theta) = \int\limits_{S} \mathrm{L}\,(\delta(x)/\theta)dP_{\theta}(x).$$

Функция $R(\delta,\theta)$ называется функцией риска, она позволяет частично упорядочить решающие правила.

Будем считать, что δ_1 предпочтительнее $\delta_2 (\delta_1 \prec \delta_2)$, если $R(\delta_1/\theta) \leq R(\delta_2/\theta)$ для любого фиксированного θ. Однако не всякие два решающие правила сравнимы между собой по всем значениям θ. Это видно, в частности, из рис. 4.

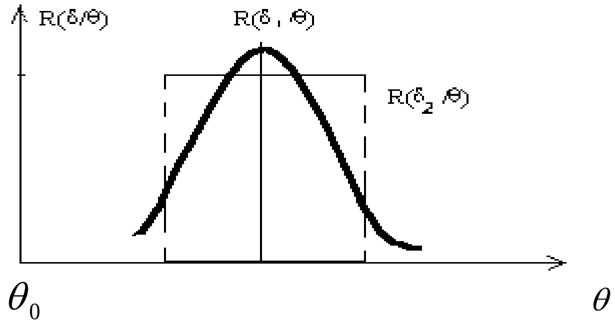

Рисунок 4

Иными словами, не во всех точках θ равномерно будет выполняться упомянутое неравенство. В этом состоит основная трудность постановки статистической задачи.

Решающее правило δ^*, для которого выполняется $\delta^* \prec \delta$ для всех δ, называется решающим правилом с равномерно наименьшим риском (см., например, РНМК в п.4.1).

Для упорядочения решающих правил используются минимаксный и байесовский подходы.

1. Минимаксный подход.

Пусть $R(\delta) = \sup\limits_{\theta} R(\delta/\theta)$. Будем считать, что $\delta_1 \prec \delta_2$, если $R(\delta_1) \leq R(\delta_2)$ для $\forall \delta_2$. В этих условиях δ_1 является минимаксным решающим правилом. Минимаксный подход дает самую большую гарантию от больших потерь.

Из рис.4 легко усмотреть недостаток минимаксного подхода: максимум кривой 1 меньше максимума кривой 2 лишь в ближайшей окрестности точки θ_O.

2. Байесовский подход (см. также п. 3.2.1).

Если известно распределение $\mu(\theta)$ параметра θ, то можно определить $R(\delta)$ путем интегрирования $R(\delta/\theta)$ по этому распределению:

$$R(\delta) = \int\limits_{\Theta} R(\delta/\theta)\, d\mu(\theta).$$

Тогда $\delta_1 \prec \delta_2$, если $R(\delta_1) \le R(\delta_2)$ для $\forall \delta_2$. Таким образом, определенное δ_1 называется байесовским решающим правилом (б.р.п.). Достоинством байесовского подхода является то, что всегда существует б.р.п. δ^* такое, что $\delta^* \prec \delta$, если $R(\delta^*) \le R(\delta)$ при $\forall \delta$.

Основной недостаток б.р.п. состоит в том, что параметр обьявляется с.в. и априорная информация о виде $\mu(\theta)$ обычно отсутствует.

5.2. Теория точечных оценок, основанная на понятии полной статистики

Пусть пространство D возможных решений является выпуклым множеством векторного пространства (например, многомерного евклидова пространства). Пусть функция потерь $\mathrm{L}(d/\theta)$ при $\forall \theta$ выпукла вниз по d. Пусть задана функция $\tau(\theta)$ со значениями в D, то есть задано отображение $\tau : \Theta \to D$ при $\forall \theta$. Решающее правило δ называется несмещенным относительно функции τ, если $M_\theta(\delta(x)) = \tau(\theta)$ (иногда решающее правило будем называть оценкой). Пусть $\Delta = \{\delta\}$ – класс несмещенных решающих правил. Тогда справедлива следующая теорема.

Теорема 34 (Блэкуэлла-Колмогорова).

Для решающих правил из Δ, среди которых $\delta^*(x) = h(T(x))$, где $T(x)$ – любая достаточная статистика ($h(T) = M_\theta(\delta(x)/T(x) = T)$ не зависит от θ, так как $T(x)$ – достаточная статистика), средние потери $R(\delta/\theta) \ge R(\delta^*/\theta)$ при $\forall \delta \in \Delta, \forall \theta \in \Theta$.

Доказательство следует из неравенства Иенсена (п.1.4):

$$R(\delta/\theta) = M_\theta[\ \mathrm{L}(\delta(x)/\theta] = M_\theta(M_\theta\ \mathrm{L}\ (\delta(x)/\theta)) \ge M_\theta(\ \mathrm{L}(M_\theta(\delta(x)/\theta))) =$$
$$= M_\theta \mathrm{L}(h(T)/\theta)) = R(\delta^*/\theta).$$

Определение 1. Скажем, что некоторое свойство, зависящее от выборки, выполнено Р – почти всюду в P, если оно выполнено в P почти всюду для всех θ (для всех распределений P_θ из заданного семейства).

Теорема 35. Если функция потерь $\mathrm{L}(d/\theta)$ при $\forall\theta$ строго выпукла по d, то решающее правило с равномерно наименьшим риском (РНР) Р – почти всюду в *P* единственно.

Доказательство этого утверждения основано на доказательстве предыдущей теоремы и очевидно.

Определение 2 [1]. Говорят, что структура <S,*B*(S),*P*> или семейство *P* доминируемы, если существует положительная $\sigma-$конечная мера μ на <S,*B*(S) > такая, что выполнено одно из двух эквивалентных утверждений:

а) каждое распределение из *P* абсолютно непрерывно;

б) каждое распределение из *P* имеет плотность вероятностей $p(x/\theta)\equiv p(x;\theta)=\dfrac{dP_\theta(x)}{d\mu(x)}$ относительно μ.

Замечание. Напомним, что мера μ называется $\sigma-$конечной на <S,*B* (S)>, если S представимо в виде счетного объединения измеримых множеств $\sigma-$конечной меры μ.

Замечание. Если ф.р. $F(x;\theta)$ задается п.в. $p(x;\theta)$ по мере Лебега, то семейство $\{P_\theta\}$ будет доминироваться мерой Лебега μ в R^n, то есть

$$\frac{dP_\theta}{d\mu}=p(x;\theta)=\prod_1^n p(x_i;\theta).$$

Замечание. Если пространство S или семейство *P* не более, чем счетно, то структура <S,*B*(S),*P*> также доминируема [1].

Определение 3 *[7,1]*. Статистика $T(x)$, заданная на доминируемой структуре <S,*B*(S),*P*> называется ограниченно полной, если из $M_\theta\phi(T(x))=0$ при $\forall\theta\in\Theta$ следует $\phi(T(x))=0$, Р – почти всюду в <S,*B*(S),*P*> при ограниченности функции $\phi(T(x))$.

Пример 1. Пусть $X\in Br(\theta),T(x)=\sum_{i=0}^n x_i,\phi_n(T(x))=T(x)$, тогда, если $M_\theta\phi(T(x))=\sum_{i=0}^n T(x_i)C_n^i\theta^i(1-\theta)^{n-i}=0,(0\le\theta\le1)$, то $T(x_0)=T(x_1)=T(x_2)=...=$ $=T(x_n)=0$ с вероятностью 1, поскольку рассмотренная сумма – тождественно равный нулю полином степени n. Отсюда вытекает полнота $T(x)$.

Определение 3′. Семейство $P = \{p(x;\theta), \theta \in \Theta\}$ называется ограниченно полным, если из $M_\theta \phi(x) = 0$ следует $\phi(x) = 0$, Р – почти всюду в P для $\forall \theta \in \Theta$ при ограниченности ϕ.

Пример 2. Рассмотрим семейство нормальных распределений $N^2(a,\sigma^2)$, $-\infty < a < \infty$, $0 < \sigma < \infty$. Пусть $\phi(x)$ интегрируема и $\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi(x) \exp\left\{-\frac{1}{2\sigma^2}(x-a)^2\right\} dx = 0$ или $\int_{-\infty}^{\infty} \exp\left\{\frac{a}{\sigma^2}x\right\} \phi(x) \exp\left\{-\frac{1}{2\sigma^2}x^2\right\} dx = 0$, откуда следует, что двустороннее преобразование Лапласа функции $\phi(x)\exp\left\{-\frac{1}{2\sigma^2}x^2\right\}$ равно нулю для всех $\tau = \frac{a}{\sigma^2}, -\infty < \tau < \infty$. Отсюда следует, что $\phi(x)\exp\left\{-\frac{1}{2\sigma^2}x^2\right\} = 0$ почти всюду в P, т.е. семейство P полно.

Примеры неполных статистик и семейств см. [1, с. 186]. Свойство полноты гарантирует единственность решающих правил, основанных на использовании достаточных статистик, в частности, наилучших несмещенных оценок и критериев.

Достаточные условия существования полной статистики устанавливаются следующей теоремой, доказанной Леманом.

Теорема 36. Предположим, что для заданного, так называемого «экспоненциального», семейства распределений $P = \{P_\theta, \theta \in \Theta\}$ выполнено

$$dP_\theta = \exp\left\{\sum_{j=1}^{r} A_j(\theta) B_j(x) + C(x) + D(\theta)\right\} d\mu(x),$$

где $\theta = (\theta_1, ..., \theta_r) \in \Theta \subset R^r, r < \infty$, μ – некоторая мера на выборочном пространстве, $A(.), B(.), C(.), D(.)$ – известные функции. Предположим, что Θ содержит некоторую точку вместе с ее окрестностью. Тогда статистика $T(x) = \{\sum_{i=1}^{n} B_1(x_i), ..., \sum_{i=1}^{n} B_r(x_i)\}$ является полной и достаточной для оценки θ.

Для доказательства полноты $T(x)$ покажем, что P^T таково, что в условиях теоремы для любой измеримой функции ϕ на R^r (R^r – выборочное пространство статистики Т), для которой справедливо равенство

$$\int \phi(t_1, ..., t_r) \mathrm{P}_\theta^T(dt_1, ..., dt_r) = 0$$

(Р^T – вероятностная мера, соответствующая статистике Т), для всех $\theta \in \Theta$ имеем $\phi(t_1,...,t_r) = 0$ почти всюду в P^T. Определим меру m как умеренно растущую в том смысле, что существуют два вероятностных распределения P_1 и P_2 на $\{R^r, \text{В}_{R^r}\}$, две неотрицательные постоянные A и B и вектор $\theta \in \Theta \subset R^r$, для которых $dm = Ae^{<\theta,t>}dP_1 - Be^{<\theta,t>}dP_2$, знак $<...>$ – знак скалярного произведения. Нетрудно заметить, что тогда справедливо представление

$$e^{<\theta,t>}dm = P_\theta(dt_1,...,dt_r)$$

для рассматриваемого экспоненциального семейства, где m – умеренно растущая функция. Но если ϕ – функция со свойством

$$\int \phi\, e^{<\theta,t>}dm = 0, \forall \theta \in \Theta$$

и θ_O – точка в Θ, то мера m' вида $dm' = \phi e^{<\theta_O,t>}dm$ также умеренно возрастает и согласно свойствам преобразования Лапласа [1, с.180] имеет вид:

$$\theta_O - \text{Re}(z) \in \Theta \Rightarrow L_{m'}(\text{Re}(z)) = 0,$$

где $L_{m'}(z)$ – преобразование Лапласа меры m'. Известно, что если функция $L_{m'}(z)$ тождественно равна нулю при вещественных значениях z (*Re(z))* из некоторого открытого подмножества $D_{m'} = D_{(m')^+} \cup D_{(m')^-}$, $(m')^+$ и $(m')^-$ – соответственно положительная и отрицательная вариация меры $m' = (m')^+ - (m')^-$, то m' равна 0. Отсюда $\phi = 0$ m' – почти всюду, а значит и P^T – почти всюду.

Доказательство достаточности статистики следует из теоремы факторизации (см. п.2.3.1) с очевидностью. Теорема доказана.

Пусть $\theta \in \Theta \subseteq R^1$, рассмотрим примеры использования т.36.

Пример 3. Пусть X – число успехов в схеме испытаний Бернулли, θ – вероятность успеха в единичном испытании, $0 \le \theta \le 1$,

$$p_\theta(x) = p_\theta(\tau_n) = \theta^x(1-\theta)^{n-x}, x = \overline{1,n}.$$

Таким образом, мы имеем экспоненциальное семейство $B(x) = x$, $A(\theta) = \ln\dfrac{\theta}{1-\theta}$, $D(\theta) = \ln(1-\theta)$. Поэтому статистика $T(x) = (\sum_1^n x_i)/n$ является полной и достаточной для оценки параметра θ.

Пример 4. Рассмотрим нормальное семейство распределений с известной дисперсией и неизвестным средним θ. Тогда $\frac{dP_\theta(x)}{d\mu(x)} = p(x/\theta) = \frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{1}{2\sigma^2}(x-\theta)^2\}$, где $x = \sum_{i=1}^{n} x_i$ и, таким образом,

$A(\theta) = \frac{\theta}{\sigma^2}$, $B(x) = x$, $D(\theta) = \frac{\theta^2}{2\sigma^2}$, $C(x) = -\frac{X^2}{2\sigma^2} - \ln\sqrt{2\pi}\sigma$. Поэтому $T(x) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ является полной и достаточной статистикой для θ (математического ожидания).

Дармуа доказал необходимость упомянутого свойства экспоненциальности для существования достаточной статистики. Приведем это доказательство, следуя [7].

Действительно, при $r=1$, если $T(x)$ достаточна для θ в выборке из n независимых наблюдений, то из представления ((2.9) п.2.3.1):

$$L(x/\theta) = \phi(T(x)/\theta)h(x) \tag{5.1}$$

следует

$$\frac{\partial \ln L}{\partial\theta} = \sum_{i=1}^{n}\frac{\partial \ln p(x_i/\theta)}{\partial\theta} = \Pi\,(T,\theta), \tag{5.2}$$

где Π есть некоторая функция от T и $p(x_i/\theta)$ – п.в. с.в. X_i в непрерывном (абсолютно) случае (сравни с п.2.4.2). Рассматривая (5.2) как уравнение относительно T, мы видим, что оно справедливо для любого значения θ.

Отсюда T можно представить $T = H\{\sum_{i=1}^{n}\pi(x_i)\}$, где H и π – произвольные функции. Если $w = \sum\pi(x_i)$, то π будет функцией (обозначим ее $\Phi(\theta,w)$) только от θ и w. Поэтому из (5.2) вытекает, если существуют написанные ниже производные, что

$$\frac{\partial^2 \ln L}{\partial\theta\,\partial x} = \frac{\partial\Phi}{\partial w}\frac{\partial w}{\partial x_i}. \tag{5.3}$$

Левая часть (5.3) зависит от θ и x_i, а $\partial w/\partial x_i$ только от x_i. Отсюда следует, что $\partial\Phi$ зависит только от θ и x_i. Но производная $\partial\Phi/\partial w$ симметрична [7] относительно x_i $(i=\overline{1,n})$ и, следовательно, зависит лишь от θ. Поэтому, интегрируя ее по w, получим

$$\Phi(\theta,w)=\omega_p(\theta)+g(\theta).$$

Таким образом, (5.2) принимает вид

$$\frac{\partial}{\partial\theta}\ln L=\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\ln f(x_i/\theta)=p(\theta)\sum\pi(x_i)+g(\theta),$$

где p,q – произвольные функции от θ, откуда

$$\frac{\partial}{\partial\theta}\ln p(x/\theta)=p(\theta)\pi(x)+q(\theta)/n,$$

что дает необходимое условие существования достаточной статистики:

$$p(x/\theta)=\exp\{A(\theta)B(x)+C(x)+D(\theta)\}.$$

Доказательство при $r>1$ аналогично. Утверждение Дармуа доказано.

Замечание. Внимательный читатель наверняка обратил внимание, что в приведенном утверждении Дармуа (с использованием логарифма функции правдоподобия) имеются дополнительные ограничения в виде регулярности экспоненциального семейства – непрерывной дифференцируемости по x и существования (конечного или нет) интервала $M\subset R^1$ такого, что

$$p(x/\theta)>0,\ \ x\in M,\ \ \theta\in\Theta,\ \ p(x/\theta),\ \ x\bar{\in}M,\theta\in\Theta.$$

Имеются другие доказательства утверждения Дармуа, в которых из условий регулярности используется (неявно) лишь положительность $p(x/\theta)$.

Можно показать, что для с.в. с экспоненциальной плотностью равенство размерностей $\dim(T(x))=\dim(\theta)$ является необходимым и достаточным условием для полноты статистики T [9, с. 41].

Определение 4. Говорят, что статистика T: <S,B> $\rightarrow$ <V,$\Re$> служит минимальной достаточной статистикой для семейства

$P = \{P_{\theta}\}$, если T является функцией от всякой достаточной статистики u:<S,B> $\to$ <W, Σ>, точнее $T^{-1}(\Re) \subset u^{-1}(\Sigma)$, где $T^{-1}(\Re) \subset$ B и $u^{-1}(\Sigma) \subset$ B – соответствующие полные прообразы.

Замечание. Резюмируя вышеизложенное, можно сказать, что если достаточная статистика ограниченно полна, то она – минимальная достаточная статистика, обратное неверно.

Минимальную достаточную статистику желательно использовать всюду, где это возможно [9].

Рассмотрим статистику

$$T : x \to g_x(\theta) = \log \frac{p(x/\theta)}{p(x/\theta_O)}, \tag{5.4}$$

где $p(x/\theta) > 0$ при всех $x \in S$ и $\theta \in \Theta$, θ_θ – произвольная точка в Θ. Статистика $T(x)$ отображает пространство <S,B> в пространство W $= \{g_x(\theta); x \in S\}$ функций на Θ. В W выделим σ-алгебру Σ, порожденную цилиндрическими множествами вида $\{\gamma \in W; (\gamma(\theta_1),...,\gamma(\theta_r)) \in B_r\}$, где $\theta_1,...,\theta_r$ – точки из Θ, B_r — r-мерное борелевское множество.

Таким образом, статистика (5.4) представляет собой измеримое отображение <S,B> в <W, Σ>.

Теорема 37 [6]. Если $\dfrac{dP_\theta}{d\mu} = p(x/\theta) > 0,$ то статистика (5.4) является минимальной достаточной для семейства $P = \{P_\theta\}$.

Доказательство. Поскольку $p(x/\theta) = \tilde{\varphi}(T/\theta)\tilde{h}(x),$ где $\tilde{h}(x) = p(x/\theta_0),$ а $\tilde{\varphi}(T/\theta) = \exp g_x(\theta)$ при $\forall \theta \in \Theta$ Σ– измерима, то из теоремы факторизации ((1) п.2.3.1) заключаем, что $T(x)$ – достаточная статистика. Далее, если $K(x)$ – произвольная достаточная статистика, то $p(x/\theta) = \tilde{\varphi}(K(x)/\theta)h(x)$ и $g_x(\theta) = \dfrac{\tilde{\varphi}(K(x)/\theta)}{\tilde{\varphi}(K(x)/\theta_0)}$, так что, действительно, T является функцией достаточной статистики. Теорема доказана.

Статистика (5.4) используется в методе отношения правдоподобия (см. гл. 4).

Достаточные условия существования решающего правила с равномерно наименьшим риском устанавливаются следующей теоремой.

Теорема 38. Пусть: 1) существует полная достаточная статистика *T(x)* (*x*- выборка), 2) функция потерь $L(d/\theta)$ выпукла на $d \in D$, 3) класс несмещенных решающих правил $\Delta \neq \varnothing$, тогда:

а) существует P – почти всюду в P единственное решающее правило $\delta \in \Delta$, для которого $R(\delta^*/\theta) \leq R(\delta/\theta)$ для любых θ и $\delta \in \Delta$;

b) $\delta^*(x) = h(T(x))$, где $h(T(x)) = M_\theta(\delta(x)/(T(x) = T)$, где δ – произвольное решающее правило, таким образом, δ^* зависит от x через достаточную статистику $T(x)$;

c) δ^* не зависит от вида функции L.

Доказательство. Достаточно показать, что δ^* не зависит от выбора $\delta \in \Delta$. Это следует из полноты статистики $T(x)$. Остальные утверждения доказаны ранее.

Несмещенные оценки с минимальной дисперсией

Пусть D – вещественная прямая $R^1, \tau : \Theta \to R^1$ – отображение. Рассмотрим функцию потерь специального вида $L(d/\theta) = [d - \tau(\theta)]^2$. Она строго выпукла по d при $\forall\theta$.

Пусть $\delta \in \Delta$, т.е. $R(d/\theta) = M_\theta[\delta(x) - \tau(\theta)]^2 = \operatorname*{var}_\theta \delta_{(x)}$. В этом случае δ^* называется несмещенной оценкой с минимальной дисперсией.

Замечание. При довольно общих условиях несмещенная оценка (имеющая вид $h(T(x))$, где T – полная достаточная статистика) с минимальной дисперсией является несмещенной оценкой с РНР для всякой выпуклой функции потерь [6]. Эти условия, как правило, выделяют оценку однозначно, если такая оценка существует (см. теорему 38). Существование же несмещенной оценки с минимальной дисперсией имеет место далеко не всегда (см. теорему 36).

Пример 5. Пусть X – число успешных испытаний в серии из n независимых испытаний с постоянной вероятностью θ успеха, $p(x/\theta) = C_n^x \theta^x (1-\theta)^{n-x}, x = 0,1,...,n$, тогда x/n – несмещенная оценка с минимальной дисперсией для θ. Это следует из того, что x/n – несмещенная оценка для θ, а $T(x) = x$ – полная и достаточная (см. пример 3) статистика.

Пример 6. Пусть $x_1,...,x_n$ – независимые с.в. с нормальным распределением $N(\mu,\sigma^2), \theta = (\mu,\sigma^2), \sigma^2 > 0$. В данном случае статистика $T(x) = (\bar{x}, s^2)$ – полная и достаточная (см. пример 4), где $\bar{x} = \frac{1}{n}(x_1 + .. + x_n), s^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2$. Отсюда $\bar{x}$ является несмещенной оценкой с минимальной дисперсией для $\mu = \tau_1(\theta)$, а s^2 аналогичная оценка для $\sigma^2 = \tau_2(\theta)$, так как эти оценки являются несмещенными и зависят от x через достаточную и полную статистику.

5.3. Теория несмещенных оценок с минимальной дисперсией по Рао-Крамеру (теория эффективных оценок)

Теорема 39 (неравенство Рао-Крамера). Предположим, что множество $\{x / L(x/\theta) > 0\}$ не зависит от θ:$L(x/\theta)$ дифференцируема по θ, $t(x)$ – несмещенная оценка для $\tau(\theta)$, и предположим**,** что существует $\tau'(\theta)$**.** Тогда

$$\operatorname{var}_\theta(t) \geq \frac{[\tau'(\theta)]^2}{M_\theta[\frac{\partial}{\partial\theta}\ln L]^2}, \tag{5.5}$$

причем равенство выполняется тогда и только тогда, когда справедлива линейная зависимость $\frac{\partial}{\partial\theta}\ln L = a(\theta)[t(x) - \tau(\theta)]$, при этом $\operatorname{var}_\theta t = \frac{\tau'(\theta)}{a(\theta)}$.

Доказательство.

$$\int L(x/\theta)dx = 1, \tag{5.6}$$

где функция правдоподобия $L(x/\theta) = f(x_1/\theta)f(x_2/\theta)...f(x_n/\theta)$ ввиду несмещенности $t(x)$

$$\int t(x)L(x/\theta)dx = \tau(\theta). \tag{5.7}$$

Продифференцируем (5.6) и (5.7) по θ (поскольку $\{x / L(x/\theta) > 0\}$ не зависит от θ, то можно изменять порядок интегрирования и дифференцирования)

$$O=\int \frac{\partial \ln L}{\partial \theta} L dx = M_{\theta}\left(\frac{\partial \ln L}{\partial \theta}\right), \tag{5.8}$$

$$\tau'(\theta)=\int t(x)\frac{\partial \ln L}{\partial \theta} L dx . \tag{5.9}$$

Из (5.8) и (5.9) можно получить (предварительно умножив (5.8) на τ):

$$\tau'(\theta)=M_{\theta}[(t-\tau)\frac{\partial \ln L}{\partial \theta}].$$

Применим неравенство Шварца к последнему равенству, тогда получим $[\tau'(\theta)]^2=\int [t(x)-\tau(\theta)]^2 L dx \cdot \int [\frac{\partial \ln L}{\partial \theta}]^2 L dx$, откуда непосредственно следует (5.5).

Равенство в (5.5) следует из неравенства Коши-Буняковского. Действительно, неравенство Коши-Буняковского для с.в. X, Y

$$M(XY)^2 \le M(X^2)M(Y^2)$$

вытекает из неравенства

$$M((aX+Y)^2)\ge 0, \forall a,$$

которое обращается в равенство тогда и только тогда, когда $aX+Y=0$. Равенство $\mathrm{var}_{\theta}(t)=\dfrac{\tau'(\theta)}{a(\theta)^2}$ в этом случае очевидно. Теорема доказана.

Определение 1. Оценка называется эффективной, если для нее неравенство Рао-Крамера (5.5) обращается в равенство.

Пример 1. Пусть $x_1,\dots,x_n$ – независимы и являются элементами нормально распределенной генеральной совокупности

$$x_k \in N(\theta,\sigma^2), \quad \frac{\partial \ln L}{\partial \theta}=\frac{n}{\sigma^2}(\bar{x}-\theta), \text{ т.е. } A(\theta)=\frac{n}{\sigma^2},\ t(x)-\tau(\theta)=\bar{x}-\theta .$$

Тогда несмещенная оценка $\bar{x}$ – эффективная оценка для $\tau(\theta)=\theta$. При этом вариация $\mathrm{var}_{\theta}\,\bar{x}=\dfrac{\sigma^2}{n}$ (см. п.2.3.2, а).

Замечание 1. Из теоремы 39 следует, что ни для какой функции $\tau(\theta)$, отличной от линейной, не существует эффективной оценки.

Замечание 2. В силу неравенства Рао-Крамера любая эффективная оценка является несмещенной с минимальной дисперсией. Обратное утверждение в общем случае неверно. Действительно, эффективная оценка является достаточной, так как

$$\frac{\partial \ln L}{\partial \theta} = a(\theta)[t(x) - \tau(\theta)] \tag{5.10}$$

следует из факторизации равенства (критерия) (см. п.2.3.1) для достаточной статистики:

$$L(x/\theta) = \varphi(t(x)/\theta)h(x). \tag{5.11}$$

Чтобы в этом убедиться, нужно лишь заметить, что (5.10) есть частный случай следующего эквивалентного (5.11) утверждения

$$\frac{\partial \ln L(x/\theta)}{\partial \theta} = \frac{\partial \ln \varphi(t(x)/\theta)}{\partial \theta}.$$

Ясно также, что для экспоненциального семейства распределений (см. теорему 36) эффективная оценка для $\tau(\theta)$ существует, она есть $t(x) = \frac{1}{n}\sum_{1}^{n} B(x_i)$, и ее дисперсия

$$D_\theta t = \frac{D'(\theta)A''(\theta) - D''(\theta)A'(\theta)}{n[A'(\theta)]} = \frac{\tau'(\theta)}{nA'(\theta)}. \tag{5.12}$$

Действительно, в экспоненциальном семействе, задаваемом плотностью $p(x/\theta) = \exp\{A(\theta)B(x) + C(x) + D(\theta)\}$,
выполняется

$$\frac{\partial \ln L}{\partial \theta} = A'(\theta)\sum_{1}^{n} B(x_i) + nD'(\theta) = nA'(\theta)\left[\frac{1}{n}\sum_{1}^{n} B(x_i) + \frac{D'(\theta)}{A'(\theta)}\right] \tag{5.13}$$

И правая часть (5.13) имеет вид (5.10), если

$$t(x) = \frac{1}{n}\sum_{1}^{n} B(x_i);\ \tau(\theta) = \frac{D'(\theta)}{A'(\theta)};\ a(\theta) = nA'(\theta).$$

Но мы знаем (см. теорему 36), что для экспоненциального семейства достаточная статистика удовлетворяет условиям (ограниченной) полноты, а значит и несмещенности, то есть условиям теоремы 38.

Улучшение неравенства Рао-Крамера. Неравенство Баттачария

В тех случаях, когда эффективная несмещенная оценка не существует, можно найти лучшую (т.е. большую) нижнюю границу для дисперсии оценки, чем та, которую дает неравенство Рао-Крамера [7].

В неравенстве Рао-Крамера $t(x)-\tau(\theta)\sim\frac{L'(\theta)}{L}$, где $\sim$ – знак пропорциональности. Идея Баттачария состоит в том, чтобы сделать $t(x)-\tau(\theta)\sim(\frac{L'_\theta}{L},...,\frac{L^{(s)}_\theta}{L})$. Выдвигается предположение: $t(x)$ – несмещенная оценка для $\tau(\theta)$.

Теорема 40. Пусть $\mathrm{var}_\theta\, t=\sum_{i=1}^{n}a_{ij}c_ic_j$; $a_{ij}=M_\theta(\frac{L^{(i)}}{L}\cdot\frac{L^{(j)}}{L})$, где $L^{(i)}=\frac{\partial^i}{\partial\theta^i}L(x/\theta)$; $\sum_j a_{ij}c_j=\tau^{(i)}$; $A^{-1}=\{a^{ij}\}$; $\tau^{(i)}=\frac{d^i\tau(\theta)}{d\theta^i}$.

Тогда
$$\mathrm{var}\, t\geq\sum\tau^{(i)}a^{ij}\tau^{(j)}, \tag{5.14}$$
и (5.14) обращается в равенство тогда и только тогда, когда $t(x)-\tau(\theta)=\frac{1}{L}\sum_{i=1}^{s}c_iL^{(i)}$ для некоторых $c_i=c_i(\theta)$.

Доказательство. Пусть $x_1,...,x_s$ – элементы гильбертова пространства. Модуль вектора больше своей проекции $\|x\|\geq\left\|\sum_1^s c_ix_i\right\|$. Если $\sum_j(x_i,x_j)\,c_j=(x,x_j)$ для $\forall i$, то $x=\sum_{i=1}^{s}c_ix_i$, то есть $\|x\|=\left\|\sum_1^s c_ix_i\right\|$.

Рассмотрим $\forall\theta\in\Theta$ (гильбертово пространство), тогда $(x,y)=M_\theta(x,y)$. Возьмем в качестве $x=t-r$, а в качестве $x_i=\frac{L^{(i)}}{L}$, тогда $\|x\|^2=\mathrm{var}_\theta\, t$, $(x_i,x_j)=a_{ij}=a_{ji}$. Соответственно $(x_i,x_j)=\tau^{(i)}$.

Допустим, что A обратима, $\bar\tau=(\tau^{(1)},...,\tau^{(s)})$, $c=(c_1,...,c_s)$, тогда $Ac=\tau$; $(Ac,c)=\bar\tau,A^{-1}\bar\tau=\sum_1^s\tau^{(i)}a^{(ij)}\tau^{(j)}$, откуда непосредственно следует, что поскольку $\int t(x)Ldx=\tau(\theta)$ (несимметричность $t(x)$), а также $\int Ldx=1$, то $\int(t\frac{L^{(i)}}{L})Ldx=\tau^{(i)}\tau(\theta)\int\frac{L^{(i)}}{L}Ldx=0$. Теорема доказана.

5.4. Регрессия, несмещенные и достаточные статистики. Теорема Рао-Блекуэлла

Теорема 41 [23]. Пусть X и Y – случайные величины и Y имеет среднее μ и положительную дисперсию σ_Y^2. Пусть $M(Y/x)=\varphi(x)$. Тогда $M[\varphi(x)]=\mu$ и $\sigma_{\varphi(x)}^2 \le \sigma_Y^2$.

Доказательство (проведем для непрерывных с.в.). Пусть $p(x,y), p_1(x), p_2(y)$ и $p(y/x)$ – соответственно совместная п.в. X и Y, две маргинальные п.в. и условная п.в. с.в. Y при условии $X=x$. Тогда для регрессии среднего имеем

$$M(Y/x)=\int_{-\infty}^{\infty} yp(y/x)dx=\frac{\int_{-\infty}^{\infty} yp(x,y)dy}{p_1(x)}=\varphi(x),$$

откуда $\int_{-\infty}^{\infty} yp(x,y)dy=\varphi(x)p_1(x)$. Имеем также

$$M[\varphi(x)]=\int_{-\infty}^{\infty}\varphi(x)p_1(x)dx=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} yp(x,y)dy\right]dx=\int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} p(x,y)dx\right]dy=\int_{-\infty}^{\infty} yp_2(y)dy=\mu.$$

Первая часть теоремы доказана. Запишем далее

$$\sigma_Y^2=M[(Y-\mu)^2]=M\{[(Y-\varphi(x))+(\varphi(x)-\mu)]^2\}=$$
$$=M\{[Y-\varphi(x)]^2\}+M\{[\varphi(x)-\mu]^2\}+2M\{[Y-\varphi(x)][\varphi(x)-\mu]\}. \qquad (5.15)$$

Покажем, что последнее слагаемое в правой части (5.15) равно нулю. Имеем

$$M\{[Y-\varphi(x)][\varphi(x)-\mu]\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y-\varphi(x)][\varphi(x)-\mu]p(x,y)dxdy=$$
$$=\int_{-\infty}^{\infty}[\varphi(x)-\mu]\{\int_{-\infty}^{\infty}[y-\varphi(x)]p(y/x)dy\}p_1(x)dx.$$

Но $\varphi(x)$ – есть среднее для условной п.в. $p\,(y/x)$, следовательно,

$$\int_{-\infty}^{\infty}[y-\varphi(x)]p(y/x)dy=0$$

и соответственно $M\{[Y-\varphi(x)][\varphi(x)-\mu]\}=0$. Поскольку $\sigma_{\varphi(x)}^2=M\{[\varphi(x)-\mu]^2\}$ и $M\{[Y-\varphi(x)]^2\}\ge 0$, то $\sigma_Y^2 \ge \sigma_{\varphi(x)}^2$. Доказательство в дискретном случае аналогично. Теорема доказана.

Пример 1. Пусть X и Y имеют двумерное нормальное распределение со средними μ_1 и μ_2, положительными дисперсиями σ_1^2 и σ_2^2 и коэффициентом корреляции ρ. Тогда (см. п.1.2)

$$M(Y/x) = \varphi(x) = \mu^2 + \rho\frac{\sigma_1}{\sigma_2}(x - \mu_1),$$

$$\varphi(x) = \mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1)$$

и теорема 40 утверждает, что $M[\varphi(x)] = \mu_2$. Далее $\sigma^2_{\varphi(x)} = M\{[\varphi(x) - \mu_2]^2\} = M\{[\rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)]^2\} = \rho^2\sigma_2^2$. При $-1 < \rho < 1$ имеем строгое неравенство $\sigma^2_{\varphi(x)} > \sigma_2^2$. Теорему Рао-Блекуэлла можно использовать при построении наилучшей оценки параметра. Пусть $(x_1,...,x_n)$ – выборка из распределения с п.в. $p(x/\theta)$, $\theta \in \Theta$, и пусть известно, что $Y_1 = T_1(x_1,...,x_n)$ является достаточной статистикой для θ. Пусть $Y_2 = T(x_1,...,x_n)$ – другая статистика (не являющаяся явной функцией от Y_1) есть несмещенная оценка θ. Рассмотрим $M(Y_1/y_1^1) = \varphi(y_1)$. Поскольку Y_1 – достаточная статистика для θ, то условная п.в. Y_2 при условии $Y_1 = y_1$ не зависит θ. Следовательно, $\varphi(Y_1)$ является достаточной статистикой для θ. Таким образом, имеем следующее следствие из теоремы 41.

Следствие. Пусть $(x_1,...,x_n)$ – выборка объема n из распределения с п.в. $p(x;\theta), \theta \in \Theta$. Пусть $Y_1 = T_1(x_1,...,x_n)$ – достаточная статистика параметра θ, а $Y_2 = T(x_1,...,x_n)$ – не связанная функционально с Y_1 несмещенная оценка параметра θ. Тогда $M(Y_2/y_1^1) = \varphi(y_1)$ является статистикой $\varphi(Y_1)$. Эта статистика $\varphi(Y_1)$ есть функция от достаточной статистики для θ, является несмещенной оценкой для θ и ее дисперсия меньше, чем дисперсия Y_2.

ГЛАВА 6. РЕШЕНИЕ ЗАДАЧ

Для иллюстрации большинства рассмотренных выше понятий рассмотрим следующие 4 родственные задачи, в процессе решения которых нами будут установлены довольно интересные факты.

Задача 1. Оценка свертки п.в. по гистограммам компонент.

Пусть непрерывные с.в. X_1 *и* X_2 независимы, причем $X_1 \in p_1(x)$, $X_2 \in p_2(x)$. По группированной выборке обьемов n_1 *и* n_2 соответственно (не имея исходных выборок $\tau_{n1}=\{x_1^1,...,x_n^1\}$ и $\tau_{n2}=\{x_1^2,...,x_n^2\}$) оценить п.в. $p_y(y)$ суммы $X_1+X_2=У$.

Решение. Известно (п.1.3) , что $p_y(y)=p_{x1} * p_{x2}(y)$. Очевидно также, что если $p_{x1}(x) \in L_2$, $p_{x2} \in L_2$, то $p_y(y) \in L_2$. Поэтому точность оценивания будем изучать в метрике L_2.

Заметим при этом, что всякая п.в. принадлежит также и L_1, поскольку $\int\limits_{-\infty}^{\infty} p(x)dx = 1.$

Введем некоторые обозначения.

Пусть

$$\tilde{p}_i(x) = \begin{cases} \dfrac{g_i(k)}{h}, & a_i + (k-1)h < x < a_i + kh, \\ 0\,, & x \notin [a_i, \quad a_i + s_i h], \end{cases} \tag{6.1}$$

где

$$g_i(k) = \int\limits_{a_i+(k-1)h}^{a_i+kh} p_i(x)dx, \quad k = \overline{1, s_i}, \quad i = 1{,}2,$$

$$\tilde{p}_{(2)}(x) = \begin{cases} \dfrac{g_{(2)}(k)}{h}, & b + (k-1)h \le x < b + kh, \\ 0\,, & x \notin [b, b + dh], \end{cases}$$

$$g_{(2)}(k) = \int\limits_{b+(k-1)h}^{b+kh} \int\limits_{-\infty}^{\infty} \tilde{p}_1(t)\tilde{p}_2(x-t)dtdx, \quad k = \overline{1, d}; \; b = \sum_{i=1}^{2} a_i; \; d = \sum_{i=1}^{2} s_i \, ,$$

$$\widehat{p}_i(x) = \begin{cases} \dfrac{\widehat{g}(k)}{h}, & a_i + (k-1)h < x < a_i + kh, \\ 0, & x \notin [a_i, \quad a_i + s_i h], \end{cases} \tag{6.2}$$

где $g_i(k)$ – частоты попадания элементов выборки с.в. X_i в интервал $[a_i+(k-1)h,\ a_i+kh],\ k=\overline{1,s_i},\ i=1,2.$

В этих обозначениях $\widetilde{p}_i(x)$ и $p_i(x)$ являются соответственно аппроксимацией и оценкой плотности $p_{x_i}(x)$. Нетрудно проверить, что справедлива точная формула

$$g_2(k) = \frac{1}{2}\sum_{i+j=k}[g_2(i) + g_2(i+1)]g_1(j) = \frac{1}{2}\Big(\sum_{i+j=k} g_2(i)g_1(j) + \sum_{i+j=k+1} g_2(i)g_1(j)\Big), \tag{6.3}$$

где $i=\overline{0,s_i},\ j=\overline{0,s_2},\quad k=\overline{0,d},\quad g_1(0)=g_2(0)=0$.

Поэтому имеет смысл оценивать $p_Y(y)$ статистиками

$$g_{(2)}(k) = \frac{1}{2}\sum_{i+j=k}[g_2(i) + g_2(i+1)]g_1(j), \quad k=\overline{0,d}. \tag{6.4}$$

Оценки (6.4) являются несмещенными и состоятельными для величин $g_2(k),\ k=\overline{0,d}$ (докажите самостоятельно, учитывая, что с.в. $g_2(i), g_1(j)$ статистически независимы и состоятельны).

Так как частоты

$$g_i(k) = \frac{1}{n_i}\sum_{r=1}^{n_i}\chi_k(x_r), \tag{6.5}$$

где x_r^i , $r=\overline{1,n_i},\ k=\overline{0,d}$ – элемент выборки, $\chi_k(x)$ – индикатор κ-го интервала разбиения в (6.2), имеют мультиномиальное распределение, то и $g_2(k),\ k=\overline{0,d}$ также будут иметь мультиномиальное распределение, которое, как известно (см. теорему 26), асимптотически нормально и при $n_1\to\infty$ и $n_2\to\infty$ (точнее – при $\min(n_1, n_2)\to\infty$) переходит в d-мерное нормальное распределение.

Поскольку

$$\left\|p_{(2)}(x) - g_{(2)}(x)\right\|_{L_2} \le \left\|p_{(2)}(x) - \widetilde{p}_{(2)}(x)\right\|_{L_2} + \left\|g_{(2)}(x) - \widetilde{p}_{(2)}(x)\right\|_{L_2}, \text{ то}$$

$$\left\|p_{(2)}(x) - g_{(2)}(x)\right\|_{L_2} \le \left\|p_{(2)}(x) - \widetilde{p}_{(2)}(x)\right\|_{L_2} + \Big\{\frac{1}{h}\sum_{k=1}^{d}[g_{(2)}(k) - g_{(2)}(k)]^2\Big\}^{1/2}, \tag{6.6}$$

предположим, что существуют ограниченные производные $\left|p_i'(x)\right| \le K_i$, $i = 1,2.$

Тогда, учитывая, что $\left|p_1(x-t) - \tilde{p}_1(x-t)\right| \le K_1 h$, $\left|p_2(t) - \tilde{p}_2(t)\right| \le K_2 h$, $\int\limits_{-\infty}^{\infty} p_2(t)dt = \int\limits_{-\infty}^{\infty} p_1(t)dt = 1$, получим:

$$\left| p_{(2)}(x) - \frac{g_{(2)}(k)}{h} \right| = \frac{1}{h} \int\limits_{b(k-1)h}^{b+kh} \int\limits_{-\infty}^{\infty} [p_2(\tau - t)p_1(t) - \tilde{p}_2(\tau - t)p_1(t)]dtd\tau =$$

$$= \frac{1}{h} \int\limits_{b(k-1)h}^{b+kh} \int\limits_{-\infty}^{\infty} [p_2(\tau - t) - \tilde{p}_2(\tau - t))p_1(t) + (p_1(t) - \tilde{p}_1(t))\tilde{p}_2(\tau - t)]dtd\tau \le (K_1 + K_2)h\,.$$

Иначе говоря, для точности аппроксимации $\tilde{p}_2(x)$ в (6.6) имеем

$$\left\| p_{(2)} - \tilde{p}_{(2)} \right\|_{L_2} \le \sqrt{dh}\, h(K_1 + K_2). \tag{6.7}$$

Для оценки второго члена в правой части (6.6) воспользуемся методами интервального оценивания (см., в частности, пп. *3.2.3*, *3.2.4*). Поскольку $g_{(2)}(k),\ k = \overline{1,d}$, асимптотически нормальны, то упомянутый член имеет распределение хи-квадрат и асимптотически

$$P\Big(\frac{N}{S}\sum_{k=1}^{d}[\hat{g}_{(2)}(k) - g_2(k)]^2 < x\Big) \ge \chi_{d-2}(x), \tag{6.8}$$

где $N = \min\{n_i\},\ i = \overline{1,2},\ s = \max\{s_i\},\ i = \overline{1,2},\ \chi_{d-2}(x)$ – значение функции распределения хи-квадрат с *d-2* степенями свободы. Действительно, ввиду (6.4) справедливо неравенство

$$[g_{(2)}(k) - \hat{g}_{(2)}(k)]^2 \le \frac{1}{4}\{\sum_{i+j=k}[g_1(i) + g_1(i+1)]^2[\hat{g}_2(j) - g_2(j)]^2 +$$

$$+ [\hat{g}_2(j) + \hat{g}_2(j+1)]^2[\hat{g}_1(i) - g_1(i)]^2\}, \text{ то}$$

$$\sum_{k=1}^{d} N[g_{(2)}(k) - g_{(2)}(k)]^2 \le \frac{N}{4}\{2n\sum_{j=1}^{s_2}[g_2(j) - g_2(j)]^2 +$$

$$+ 2n\sum_{i=1}^{s_1}[g_1(j) - g_1(j)]^2\} \le \frac{n}{2}\{\sum_{j=1}^{s_2}\frac{n_2}{g_2(j)}[g_2(j) - g_2(j)]^2 +$$

$$+ \sum_{i=1}^{s_1}\frac{n_1}{g_1(i)}[g_1(i) - g_1(i)]^2\},$$

где две суммы в фигурных скобках представляют собой независимые статистики критериев χ^2 Пирсона. (см. теорему *29*) с s_1-1, s_2-1 степенями свободы соответственно, а $s_1+s_2 \le d$, откуда следует (6.8).

Итак, пусть t_γ – γ – квантиль распределения $\chi^2_{d-2}(x)$, тогда из (6.6)–(6.8) получаем окончательно

$$\left\| p_{(2)}(x) - g_2(x) \right\| \le \sqrt{dh}h(K_1 + K_2) + \sqrt{\frac{t_{\gamma n} n}{hN}}. \qquad (6.9)$$

Мы построили несмещенную, состоятельную оценку (6.4) свертки п.в. с.в. X_1 и X_2 в условиях задачи и указали точность (6.9) и надежность γ этой оценки. В приложении I приведена программа для решения задачи 1.

Задача 2. Оценка компоненты свертки. Пусть непрерывные с.в. X_1 и X_2 независимы, причем $X_1 \subseteq p_1(x)$, $X_2 \subseteq p_2(x)$. По гистограмме $g_2(k),\ k=\overline{1,d}$, их суммы $Y= X_1+X_2$ (построенной с шагом h по выборке объема M на интервале $[a,b]$, $a=\min\beta_i$, $b=\max\beta_i$ по $1\le i\le M$, причем выборка $(\beta_1, ..., \beta_M)$ не сохраняется) и по аналогичной гистограмме (с шагом h на интервале $[a',b']$ по выборке $(x_1^1,...x_n^1)$ обьема n_1) $g_i(k),\ k=\overline{1,s_1}$, одной из них (X_1) оценить гистограммой с тем же шагом h плотность $p_2(x)$ другой с.в. (X_2) (см. аналогичные обозначения в задаче 1).

Решение. Очевидно, данная задача является обратной к задаче 1. Иначе говоря, она состоит в оценке компоненты свертки п.в.

$$p_{(2)}(x) = \int_{-\infty}^{\infty} p_1(x-t)p_2(t)dt,$$

то есть разрешение относительно $\hat{p}_2(x)$ следующего уравнения:

$$p_{(2)}(x) = \int_{-\infty}^{\infty} p_1(x-t)p_2(t)dt. \qquad (6.10)$$

Обозначим $z=(g_2(1),..,g_2(s_2),0,...,0)^T$, $u=(g_{(2)}(1),...,g_2(d))^T$ (Т – знак транспонирования, тогда, учитывая (6.4) запишем (6.10) в матричной форме:

$$u = Az, \tag{6.11}$$

где A – оператор, соответствующий (6.4) (нижняя треугольная матрица размера $d \times d$, составленная из функций от $g_1(k),\ k = \overline{1,s},$ все элементы главной диагонали которой равны $\frac{1}{2}(g_1(1))$.

Обоснуем корректность разрешения уравнения (6.11) относительно z.

Поскольку объем n_1 выборки отличен от нуля, то $g_1(1) > 0$ (по крайней мере один элемент выборки $(x_1^1,...x_n^1)$ попадает в интервал $[a,a+h]$, то есть $g_1(1) \geq \frac{1}{n_1}$). Следовательно, $\left|A_d^{-1}\right| \leq \frac{n_1}{2} < \infty,$ то есть обратная задача (6.11) разрешима.

Поскольку преобразование (6.11) билинейно, а оценки u и $g_1(k),\ k = \overline{1,s}$ состоятельны и несмещенные для $\tilde{u} = (g_{(2)}(1),...,\tilde{g}_{(2)}(d))$ и $\tilde{g}_1(k),\ k = \overline{1,s},$ то и z — состоятельная и несмещенная оценка для $\tilde{z} = (\tilde{g}_2(1),...,\tilde{g}_2(s_2),0,...,0),\ s_1 = d - s_1$.

Следовательно, погрешность оценивания $p_2(x)$ задается по аналогии с (6.9):

$$\left\| p_2(x) - p_2(x) \right\|_{L_2} \leq \sqrt{s_2 h} h K_2 + \sqrt{\frac{t_\gamma s_2}{hN}},$$

$p_2'(x) \leq K_2$ (вопрос об оценке K_2 как функции от условий задачи оставим открытым).

$$N = min\ \{n_1, M\},$$

t_{γ} – γ – квантиль хи-квадрат распределения с $(d\text{-}s_1)$ степенями свободы.

Замечания

1. По существу рассмотренная задача дает метод оценки компонента свертки любых финитных п.в. (метод оценки решения уравнений типа свертки от п.в.).

2. Частным случаем данной задачи является задание известной п.в. $p_1(x)$; данную задачу можно рассматривать как задачу оценки априорной п.в. параметра t сдвига в (6.10). В приложении II приведена программа решения задачи *2* на персональном компьютере.

Задача 3. Оценка решения уравнения Линдли. Уравнение Линдли используется в теории массового обслуживания для получения предельного (по m) распределения $W(y)$ времени ожидания $\omega_{m+1} = max(0, \omega_m + \upsilon_m)$ при $m \to \infty$:

$$W(y) = \begin{cases} \int\limits_{-\infty}^{y} W(y-u)dC(u), \; y \ge 0, \\ 0\,, \;\; y < 0, \end{cases} \qquad (6.12)$$

где $C_m(y)=C(y)$ – ф.р. с.в. $u_m = t_m - x_m$, то есть разности интервала времени t_m между проявлениями m-го и $(m+1)$ –требований и времени обслуживания x_m m-го требования. Имеется следующая процедура решения дискретного аналога уравнения Линдли (Клейнрок. Вычислительные системы с очередями. М.: Мир, 1979).

Пусть $c\,(k) = p\,(u_m=kh), k=0, \pm 1, ..., \forall m$;

$p\,(m_k) = p\,(\omega_m=kh), k=0,1,...$

I шаг: вычислить дискретную свертку (см. п.1.3) $p_{m^*}\,c(k)$.

II шаг: применить к результату свертки оператор π: вероятности из распределения на отрицательной полуоси переносятся в начало координат. Получив таким образом $p_{m+1}(k)$, приступим к новому циклу. При этом сначала находим

$$c(k) = q_1 \bar{*}\, q_2(k) = \sum_{i=-\infty}^{\infty} q_1(k+i)q_2(i), \qquad (6.13)$$

где $q_1(k) = P(t_m = k),\; q_2(k) = P(x_m = k),\; \forall\; m = 0,1,...$.

Пусть с.в. x_m, t_m – непрерывны и имеют финитные (регулярные) п.в. $q_1(x)$ и $q_2(x)$. Зададим гистограммы $q_1(x)$ и $q_2(x)$, построенные с одинаковым шагом h. Требуется оценить решение уравнение (6.12) по ним.

Решение. Поскольку непрерывную свертку финитных п.в. всегда (в том числе и для разности с.в.) можно точно записать с помощью равенств типа (6.3), то $C(x)$ оценивается как в 1-й задаче и отличие непрерывного варианта от дискретного не принципиально. Вернемся поэтому к дискретному случаю и будем считать, что нам известны функции $c_j(k),\; k = \overline{k_1, k_2},\; k_1 < 0,\; k_2 > k_1,\; k_2 - k_1 = n$ (собственные распределения *W(y)* существуют лишь при *Mu<0*), полученные по выборкам объема *N* значений одинаково распределенных и независимых дискретных с.в. u_j, *j=0,1,…,m-1,* как оценки (частоты) соответствующих вероятностей $c_j(k)$. Обозначим $W_m(y)=P(\omega_m<y)$,

$$W_m(y) = \Pi(C_{m-1} * \ldots * \Pi(C_1 * \Pi(C_0 * W_0))\ldots)(y)\,, \quad C_j(y) = C_j(k) = \sum_{i<[y]} C_j(i),\ k = [y],$$

где [.] – целая часть.

$$W_0(x) = W_0(x),\ \ W_0(x) = 0,\ \ x < 0;\ \ W_0(x) = 1,\ \ x \geq 0,$$

Π: $F(x) \to \{F(x),\ x \geq 0,\ F(x) = 0,\ x < 0\}$ (т.е. оператор Π «срезки» ф.р. в нуле равносилен оператору π для п.в.). Для любого исхода (элементарного события $\omega\in\Omega$)

$$\|C_m - C_m\|_{L_2} = \sqrt{\sum_{k=k_1}^{k_2} (C_m(k) - C_m(k))^2} \leq \sqrt{(n+1)\sum_{k=k_1}^{k_2} \frac{(C_m(k) - C_m(k))^2}{C(k)}},$$

в чем предлагается самостоятельно убедиться читателю. Кроме того,

$$\|W_m - W_m\|_{L_2} \leq \sum_{i=1}^{m} \|C_i - C_i\|_{L_2}\,, \tag{6.14}$$

поскольку для $\forall\, m<\infty$ неравенство $\|W_m - W_m\|_{L_2} \leq \sum_{i=1}^{m-1} \|C - C\|_{L_2}$ следует из того, что для любых ф.р. F_i, G_i независимых с.в. η_i , ε_i

$$i=1,2,\ \left\|F_{(\varepsilon_1+\eta_1)+} - F_{(\varepsilon_2+\eta_2)+}\right\|_{L_2} \leq \left\|F_{\varepsilon_1} - F_{\varepsilon_2}\right\|_{L_2} + \left\|G_{\eta_1} - G_{\eta_2}\right\|_{L_2},$$

$x^+=max(0,x)$, что в свою очередь следует из неравенств: $\|F_1 * F_2 - G_1 * G_2\|_{L_2} \leq \|F_1 - G_1\|_{L_2} + \|F_2 - G_2\|_{L_2}$ и $\left\|F_{\varepsilon_1^+} - F_{\varepsilon_2^+}\right\|_{L_2} \leq \left\|F_{\varepsilon_1} - F_{\varepsilon_2}\right\|_{L_2}$. Из (6.14) и неравенства треугольника следует также, что

$$\|W_{m+1} - W_m\|_{L_2} \leq \|W_{m+1} - W_{m+1}\|_{L_2} + \|W_m - W_m\|_{L_2} + \|W_{m+!} - W_m\|_{L_2} \leq$$
$$\leq 2\sum_{i=0}^{m-1} \|C_i - C_i\|_{L_2} + \|C_{m-1} - C_m\|_{L_2} + \|W_{m+1} - W_m\|_{L_2}\,.$$

Заметим далее, что $W_m(y),\ k \leq y \leq k+1,\ k > k_1$ является состоятельной и асимптотически нормальной оценкой $W_m(y)$, $k \leq y < k+1,\ k > k_1$, поскольку получена линейным (рекуррентным) преобразованием из оценок $c_0(k), c_1(k),...,c_{m-1}(k),\ k=\overline{k_1,k_2}$, которые являются состоятельными и асимптотически нормальными оценками одинаковых (по условию) распределений $c_0(k),...,c_{m-1}(k),\ k=\overline{k_1,k_2}$.

Поэтому, учитывая монотонность убывания (сходимости), $W_m(y)$ для $\forall\, y \geq 0$ к собственному распределению $W(y)$ при $Mu < 0$, задавшись $\varepsilon > 0$ и $0< \gamma <1$, найдем m, при котором при заданных N и n (а в непрерывном случае и h таких, что $n/Nh=0(1)$)

$\sqrt{\dfrac{t_\gamma(n+1)}{N}} < \dfrac{\varepsilon}{2m+1}$, где t_γ – квантиль распределения $\chi^2_{(n-1)}$.

И, наоборот, задавшись ε, γ и m, можно найти N, n такие, что

$\sqrt{\dfrac{t_\gamma(n+1)}{N}} < \dfrac{\varepsilon}{2m+1}$. Тогда $W_m(y)$, вычисленное по оценкам $c(\kappa),\ k=k_1,k_2,\ m=0,...,m-1,$ с доверительной вероятностью, близкой к γ, с точностью ε можем считать искомым значением ф.р. $W(y)$. Причем $\varepsilon = \varepsilon_1 + 2\|W_m - W\|_{L_2}$.

Здесь мы вплотную подошли к проблематике теории устойчивости стохастических моделей, с которой можно познакомиться по работам В.М. Золотарева, А.А. Боровкова и В.В. Калашникова, хотя статическим аспектам оценки решения уравнения Линдли в ней пока не уделялось внимания.

В приложении III приводиться пример программы, использующей уравнение Линдли для получения предельного распределения времени ожидания обслуживания требования в системе массового обслуживания G/G/1/. Машинный алгоритм решения данной задачи разработан с учетом его реализации на персональном компьютере. Программа написана на языке *«C»*. Текст программы находится в файле *stat4*.

Задача 4. О мощности критерия согласия для свертки функций распределения. Пусть $\xi \in N(\mu_1,\sigma^2),\ \eta \in N(\mu_2,\sigma^2)$ – независимые с.в. Рассмотрим простые гипотезы:

$$H_0 : \mu_1 = 0,\ \mu_2 = 0,$$
$$H_1 : \mu_1 = b,\ \mu_2 = c,\ b = const > 0,\ \ c = const > 0.$$

Пусть $(x_1,\dots,x_n)$, $(y_1,\dots,y_n)$ – выборки объема n для с.в. ξ и η соответственно. Для с.в. $\zeta=\xi+\eta$ рассмотрим 2 выборки

$$Z_i=x_i+y_i,\ i=\overline{1,n};\ \overline{1,n}; \tag{6.15}$$

$$Z_{ij}=x_i+y_j,\ i,j=\overline{1,n}.\overline{1,n}. \tag{6.16}$$

Требуется сравнить (по мощности) наиболее мощные критерии согласия уровня γ ($0<\gamma=\mathrm{const}<<1$) для п.в. с.в. ζ по выборкам (6.15) и (6.16) между собой.

Решение. Заметим прежде всего, что случайные величины z_i в выборке (6.15) независимы, в то время как выборка (6.16) – зависимая.

Согласно лемме Неймана-Пирсона, для любого уровня значимости (размера) критическая область отношения правдоподобия является наилучшей критической областью (см. п. *4.1.2*, а также [9, с. 107]).

Найдем критическую область отношения правдоподобия размера γ сначала для случая (6.15). В этом случае очевидно отношение правдоподобия

$$\Lambda_n=\prod_{i=1}^{n}\frac{p_1(z_i)}{p_0(z_i)}=\exp\left\{\frac{b+c}{2\sigma^2}\sum_{i=1}^{n}(2z_i-(b+c))\right\}.$$

Иначе

$$\ln\Lambda_n=A\sum_{i=1}^{n}z_i+B,\ \ A=\frac{b+c}{2\sigma^2},\ \ B=-\frac{(b+c)^2}{2\sigma^2}. \tag{6.17}$$

Поскольку $z_i\in N(\mu_1+\mu_2,2\sigma^2),\ \ i=\overline{1,n}$, то $A\sum\limits_{i=1}^{n}z_i\in N(An\mu,\ A^2n2\sigma^2)$, где $\mu=\mu_1+\mu_2$.

Отсюда, если верна H_0, то $A\sum\limits_{i=1}^{n}z_i\in N(0\,,2A^2n\sigma^2)$ $(\mu_1=0,\mu_2=0)$.

Обозначим соответствующую ф.р. с.в. $A\sum\limits_{i=1}^{n}z_i$ через F_0. Если верна альтернатива H_1, то $A\sum\limits_{i=1}^{n}z_i\in N(A_n(b+c),2A^2n\sigma^2)\,(\mu_1=b,\mu_2=c)$, т.е. F_1. Из (6.17) (см. также [8]), следует, что ф.р. с.в. Λ_n равна соответственно

$\Phi_0(x)=F_0(x\text{-}B)$ и $\Phi_1(x)=F_1(x\text{-}B)$. Обозначим далее $\bar{\mu}=An(b+c),\ \bar{\sigma}=2A^2n\sigma^2$, тогда γ – квантиль t_γ найдем как решение уравнения

$$\gamma=\Phi_0(t_\gamma)=F_0(t_\gamma-B)=\Phi(\frac{t_\gamma-B}{\sigma})=\Phi(K_\gamma),$$

где $\Phi(x)$ – нормальное распределение $N(0,1)$, K_γ – γ-квантиль распределения $N(0,1)$. Тогда мощность соответствующего критерия для случая (6.15)

$$\beta=\Phi_1(t_\gamma)=F_1(t_\gamma-B)=F_1(K_\gamma\bar{\sigma})=\Phi\left(\frac{K_\gamma\bar{\sigma}-\bar{\mu}}{\bar{\sigma}}\right)=\Phi(K_\gamma-\frac{\bar{\mu}}{\bar{\sigma}})=\Phi(K_\gamma-q) \quad (6.18)$$

где $q=\dfrac{\bar{\mu}}{\bar{\sigma}}$.

Замечание. Нахождение величины критической зоны при $c>0$, $b>0$, очевидно, равносильно нахождению соответствующего квантиля.

Перейдем к случаю (6.16).

Для с.в. $\zeta=\xi+\eta$ образуем зависимую выборку $(z_{11},\dots,z_{1n},\dots,z_{nn})$, где $z_{ij}=x_i+y_j$, $i,j=\overline{1,n}$. С.в. $z_{ij}\in N(\mu,2\sigma^2)$, где $\mu=\mu_1+\mu_2$.

Рассмотрим гипотезы:

$$H_0:\ \mu=0$$

$$H_1:\ \mu=bc$$

Для того, чтобы аналогично случаю (6.15) записать отношение правдоподобия, найдем явный вид совместного распределения вектора $z_{11},\dots,z_{nn}$. Естественно, что оно будет нормальным и при этом нетрудно показать, что $Mz_{ij}=\mu=\mu_1+\mu_2$, $Dz_{ij}=2\sigma^2$, а матрица ковариаций

$$\|\sigma_n\|=\begin{pmatrix} G_n & K_n & \dots & K_n \\ K_n & G_n & \dots & K_n \\ \dots & \dots & \dots & \dots \\ K_n & K_n & \dots & G_n \end{pmatrix},$$

где $G_n = \sigma^2 \begin{pmatrix} 21...1 \\ 12...1 \\ .... \\ 1...12 \end{pmatrix}_{n*n}$, $K_n = \sigma^2 I_n$.

(I – единичная матрица размерности $n*n$).

Ясно, что $det\|\sigma_n\| = 0$ для любого $n=2,3,...,$ то есть матрица $\|\sigma_n\|$ является вырожденной.

Поэтому для получения совместного распределения вектора $(z_{11},...,z_{nn})$ имеет смысл воспользоваться каким-либо методом псевдо-обращения по Муру-Пенроузу матрицы $\|\sigma_n\|$, например, методом Кэли-Гамильтона [13].

Так, в частном случае при $n=2$ получаем

$$\left|\sigma_2^+\right|^{1/2} = \frac{1}{\sqrt{\det\|\sigma^2\|}} = \frac{s\sqrt{21}}{2^s \sigma^4},$$

где “+” – знак псевдообращения.

Отсюда по аналогии с (6.17) при $n=2$ нетрудно записать логарифм отношения правдоподобия для (6.16):

$$\ln \Lambda_2' = \partial\{A(z_{11} + z_{12}) + B(z_{21} + z_{22}) + c\}, \tag{6.19}$$

где $A=6b+c;\ B=6c+b;\ c= -6(b^2+c^2)-bc;$

$$\kappa = \frac{1}{2^4 \sigma^2}.$$

Учитывая (6.16), (6.19) можно переписать в виде

$$\ln \Lambda_2' = \partial(a_1 x_1 + a_2 x_2 + a_3 y_1 + a_4 y_1) + \tau = \varepsilon + \tau,$$

где $a_1=2(6b+c),\ a_2=2(6c+b),\ a_3=a_4=7(b+c),\ \tau=c\partial.$

Введем обозначения: $\overline{\mu} = \partial(a_1 + a_2)\mu_1 + 2\partial a_3 \mu_2;\ \overline{\overline{\sigma}}^2 = \partial^2 \sigma^2 (a_1^2 + a_2^2 + 2).$

Тогда $\xi \in N(\overline{\overline{\mu}}, \overline{\overline{\sigma}}^2),$ поскольку $x_i \in N(\mu_1, \sigma^2),\ y_i \in N(\mu_2, \sigma^2).$

Действуя и далее по аналогии со случаем (6.15), получаем в случае (6.16) для мощности критерия размера γ равенства (при $n=2$)

$$\beta' = \Phi(K_\gamma - q'),$$

где $q' = \dfrac{\overline{\overline{\mu}}}{\overline{\overline{\sigma}}} = \dfrac{14(b+c)^2}{\sigma\sqrt{6x7^2(b+c)^2 - 48(b+c) - 8x37bc}}.$

Возвращаясь к (6.18), заметим, что в случае (2) при $n=2$ $q=(b+c)/\sigma$. Поскольку ф.р. $\Phi(x)$ – не убывающая, и если $q < q'$, то критерий, соответствующий случаю (6.16), мощнее критерия, построенного по (6.15).

Обобщить рассмотренное решение для $n>2$, а также сделать окончательные выводы по задаче 4 предлагается любознательному читателю.

Замечание. Частичное решение задачи 4 можно представить более компактно. Критерии проверки простой гипотезы о математическом ожидании с.в. ξ по выборкам (6.15) и (6.16) имеют одинаковую мощность.

Действительно, образовав базис из элементов z_i, $i = \overline{1,n}$;

$z_{i+1,i}$, $i = \overline{1,n-1}$, видим, что любой элемент (6.16) можно выразить через элементы базиса рекуррентно, например, с помощью равенств:

$$z_{k,m} = z_{k-1,m} + z_{k,k-1} - z_{k-1,k-1},\ k>m,$$

при этом $rank\left\|z_{i,j}\right\| \le 2n-1$. Если перейти к новому базису

$$\{z_{i,i}\ ,\ i=\overline{1,n}\,;\ z_{i+1,i} - z_{i,i+1}\ ,\ i=\overline{1,n-1}\},$$

учитывая, что $z_{i+1,i} = \dfrac{1}{2}[z_{i+1,i+1} + z_{i,i} + (z_{i+1,i} - z_{i,i+1})]$, то убеждаемся, что $E(z_{i+1,i} - z_{i,i+1}) = 0$ для каждой из гипотез.

В заключение заметим, что [4, 10, 14, 17] в приведенном ниже списке литературы могут быть использованы в качестве задачников и справочников по рассмотренным разделам математической статистики. Кроме того, читателю предлагается в качестве упражнения решить данные задачи с помощью статистических пакетов программ SPSS v10.0.5, Statistica v5.5a, MathCAD 2000 Pro, STADIA, STATGRAPHICS, подробное описание которых можно найти в [29, 30].

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Барра Ж.Р. Основные понятия математической статистики / Пер.с франц. М.: Мир, 1974.
2. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1965.
3. Боровков А.А. Математическая статистика. Новосибирск, 1997.
4. Володин Б.Г., Свешников А.А. [и др.]. Сборник задач по теории вероятностей, математической статистике и теории случайных функций / Под общ. Ред. А.А. Свешникова. М.: Наука, 1965.
5. Гнеденко Б.В. Курс теории вероятностей. 4-е изд. М.: Наука, 1965.
6. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. М.: Наука, 1972.
7. Кендалл М., Стюарт А. Статистические выводы и связи / Пер. с англ. М.: Наука, 1973.
8. Коваленко И.Н., Филиппова А.А. Теория вероятностей и математическая статистика. М.: Высшая школа, 1973.
9. Кокс Д., Хинкли Д. Задачи по теоретической статистике с решениями / Пер. с англ. М.: Мир, 1978.
10. Кокс Д., Хинкли Д. Теоретическая статистика / Пер. с англ. М.: Мир, 1981.
11. Королюк В.С. и др. Справочник по теории вероятностей и математической статистике / Под редакцией В.С. Королюка. Киев: Наукова думка, 1978.
12. Крамер Г. Математические методы статистики. Изд. 2-е, стереотипное / Пер. с англ. М.: Мир, 1975.
13. Ли Ц., Джадд Д., Зельнер А. Оценивание параметров Марковских моделей по агрегированным временным рядам / Пер. с англ. М.: Статистика, 1977.
14. Митропольский А.К. Техника статистических вычислений. 2-е изд. М.: 1971.

15. Михок Г., Урсяну В. Выборочный метод и статистическое оценивание / Пер. с румын. М.: Финансы и статистика, 1982.
16. Рао С.Р. Линейные статистические методы и их применение / Пер. с англ. М.: Наука, 1968.
17. Севостьянов Б.А., Чистяков В.П., Зубков А.М. Сборник задач по теории вероятностей. М.: Наука, 1980.
18. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики. 3-е изд. М.: Наука, 1969.
19. Уилкс С. Математическая статистика / Пер. с англ. М.: Наука, 1967.
20. Феллер В. Введение в теорию вероятностей и ее приложения. 2-е изд / Пер. с англ. М.: Мир, 1964.
21. Хеннекен П.А., Тортра А. Теория вероятностей и некоторые ее приложения / Пер. с франц. М.: Наука, 1974.
22. Шметтерер Л. Введение в математическую статистику / Пер. с нем. М.: Мир, 1976.
23. Харин Ю.С., Степанова М.Д. Практикум на ЭВМ по математической статистике. М.: Мир, 1987.
24. Беляев В.К, Чепурин Е.В. Основы математической статистики. Ч.I. М.: Изд. МГУ, 1982.
25. Беляев В.К, Чепурин Е.В., Основы математической статистики. Ч.II-III. М.: Изд. МГУ, 1982.
26. Ивченко Г.И. Математическая статистика: учебное пособие для вузов. М.: Высшая школа, 1984.
27. Леман Э. Проверка статистических гипотез / Пер. с англ. М.: Наука, 1982.
28. Тарасенко Ф.П. Непараметрическая статистика. Томск: Изд. ТГУ, 1976
29. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. М.: ИНФРА-М, 1998.
30. Радзинский Р.Н. Справочник по вероятностным распределениям. СПб.: Наука, 2001.
31. Закс Ш. Теория статистических выводов/ Пер. с англ. М.: Мир, 1975.

32. Беляев Ю.К., Носко В.П. Основные понятия и задачи математической статистики. М.: ЧеРо, изд.-во МГУ, 1998.
33. Климов Г.П. Теория вероятностей и математическая статистика. М.: Изд. МГУ, 1983.
34. Хампель Ф., Рончетти Э. Робастность в статистике/ Пер. с англ. М.: Мир, 1989.

ПРИЛОЖЕНИЕ I

Здесь приводится пример программы, использующейся для вычисления оценки значения критерия согласия о близости свертки и оценки свертки, через значения критерия согласия о близости компонент свертки и их оценок. Программа наглядно демонстрирует задачу 1.

1. Характеристики программы.

   Машинный алгоритм решения данной задачи разработан с учетом его реализации на ПЭВМ.

   Программа решается в течение менее 1 минуты в зависимости от длины входного массива.

2. Сведения о форме представления программы.

   Программа написана на языке ФОРТРАН с использованием подпрограммы для вычисления гистограммы и функции.

   Имя программы rik2.
   Имя подпрограммы SM.
   Имя функции RAN.
   Текст программы находится в файле rik2.for.
   Обращение к подпрограмме:
   CALL SM(a,b,d,u,g,m)

3. Входные и выходные данные.

   В качестве входной информации используются:
   h - шаг разбиения,
   p - плотность вероятности для с.в. β_1
   q - плотность вероятности для с.в. β_2
   nl - объем выборки для первой компоненты,
   n2 - объем выборки для второй компоненты,
   $[a_i,b_i]$, i=1,3 - интервалы разбиения для β_1, β_2, β сооответственно.
   В результате работы программы выводятся на печать:
   P1- гистограмма для первой компоненты,
   Q1- гистограмма для второй компоненты,
   g - гистограмма для свертки,
   dis - дисперсия,
   u - значение статистики.
   Выходные данные записаны в файл QQ.QQQ.

4. Таблица соответствия переменных.

Алгоритм	программа	содержание
p	P	$p_1(x)$
q	q	$p_2(x)$
p_1	p_1	$\hat{g}_1(k)$
q_1	q_1	$\hat{g}_2(k)$
v	u	v

```
   SUBROUTINE SM(a,b,d,u1,g3,m)
   INTEGER d.
   DIMENSION u1(m),g3(d)
   h=(b-a)/d
   DO 10 j=1,d
10 g3(j)=0
   x=a
   DO 11 j=1,d
   DO 12 i=1,m
   IF (u1(i).ge.x+h) GOTO 12
   IF (u1(i).lt.x) GOTO 12
   g3(j)=g3(j)+1
12 CONTINUE
   s3=m
   g3(j)=g3(j)/s3
   x=x+h
11 CONTINUE
   RETURN
   END
   DIMENSION x1(1000),x2(2000),p(10),q(10),p1(10),
   *q1(10),g(20),g1(20), gr(20),dis(20),u(50),z(20)
   DIMENSION fred (12), ubo(3)
   n1=10
   n2=20
   n=20
   a1=0
   b1=10
   a2=0
   b2=20
   m1=10
   m2=10
   h=(b1-a1)/m1
   DO 20 i=1,10
20 p(i)=0.1
   DO 30 i=1,10
30 q(i)=0.1
   WRITE(1.240)(p(i),i=1,10)
240 FORMAT (' ПЛОТНОСТЬ p'/6f10.3/)
   WRITE(1,250)(q(i),i=1,10)
250 FORMAT (' ПЛОТНОСТЬ q'/6f10.3/) .
   dis(1)=p(1)**2*q(1)*(1-q(1))/n2+
   *q(1)**2*p(1)*(1-p(1))/n1
   DO 90 i=2,10
```

```
      ih=i-1
      ir=i+1
      r1=0
      r2=0
      r3=0
      r4=0
      DO 100 j=1,ih
      D0  200 j1=1,ih
   200 r1=p(j)*p(j1)*q(i-j)*q(i-j1)*(1./n1+1./n2)+r1
   100 r1=r1-p(j)**2*q(i-j)**2*(1./n1+1./n2)
      DO 300 j=1,i
      DO 400 j1=1,i
   400 r2=r2+p(j)*p(j1)*q(ir-j)*q(ir-j1)*(1.n1+1./n2)
   300 r2=r2-p(j)**2*q(ir-j)**2*(1./n1+1./n2)
      DO 120 j=1,ih
   120 r3=r3+q(j)**2*p(i-j)*(1-p(i-j))/n1+
       *p(j)**2*q(i-j)*(1-q(i-j))/n2
       DO 130 j=1,i
   130 r4=r4+q(j)**2*p(ir-j)*(1-p(ir-j))/n1+
      *p(j)**2**q(i-j)*(1-q(i-j))/n2
    90 dis(i)=(r1+r2+r3+r4)/2
      DO 31 i=1,10
      j=21-i
   31 dis(j)=dis(i)
      WRITE(1,650)(dis(i),i=1,20)
 650 FORMAT (' ',10x,'ДИСПЕРСИЯ'/(3f15.10/))
      Вычисление свертки и оценка свертки
      ix=1
      D0 290 m=1,50
      DO 40 j=1,1000
      ayfl=ran(ix,iy)
      ix=iy
   40 x1(j)=ayfl*10.
      DO 50 j=1,2000
      ayfl=ran(ix,iy)
      ix=iy50 x2(j)=ayfl*10.
         CALL SM(a1,b1,m1,x1,p1,n1 )
         CALL SM(a2,b2,m2,x2,q1,n2)
         g(1)=p(1)*q(1)/2
         g1(1)=p1(1)*q(1)/2
         gr(1)=p1(1)*q1(1)
         DO 60 i=2,10
         ih=i-1
         ir=i+1
```

```
      s1-0.
      s2=0.
      t1=0.
      t2=0.
      DO 70 j=1,ih
      s1=s1+p(j)*q(i-j)
  70 t1=t1+p1(j)*q1(i-j)
       DO 80 j=1,i
       s2=s2+p(j)*q(ir-j)
 80 t2=t2+p1(j)*q1(ir-j)
        g(i)=(s1+s2)/2
        g1(i)=(t1+t2)/2
60 gr(i)=g(i)-g1(i)
        DO 61 iv=1,9
        i=20-iv
        ir=iv+1
        s1=0.
        s2=0.
        t1=0.
        t2=0.
        DO 71 j=1,ir
        jr=11-j
        s1=s1+p(jr)*q(i-jr)
  71 t1=t1+p1(jr)*q1(i-jr)
        DO 81 j=1,iv
        jr=11-j
        s2=s2+p(jr)*q(ir-jr)
    81 t2=t2+p1(jr)*q1(ir-jr)
        g(i)=(s1+s2)/2
      g1(i)=(t1+t2)/2
  61 gr(i)=g(i)-g1(i)
        g(20)=p(10)*q(10)/2
        g1 (20)=p1(10)*q1(10)/2
        gr(20)=g(20)=g(20)-g1(20)
        вычисление статистики
      u(m)=0
      DO 140 i=1,20
      u(m)=u(m)+gr(i)**2/dis(i)
  140 CONTINUE
       WRITE(1,270)u(m)
  270 FORMAT(' ', 'статистика u=',f11 .5)
  290 CONTINUE
         WRITE (1,103)
 103 FORMAT (' ' ,20x, 'выборка для первой компоненты свертки')
```

```
      WRITE(1,150)(x1(i),i=1,500)
150 FORMAT (' ', 6f10.5/)
      WRITE(1,104)
104 FORMAT (' ',20x, 'выборка для второй компоненты свертки' )
      WRITE(1,320)(x2(i),i=1,500)
 32  FORMAT(' ',6f10.5/)
      WRITE(1,600)(g1(i),i=1,20)
600 FORMAT (' ', 15x,'свертка g',(6f10.3/))
     WRITE (1,510) (gr(i),i=1,20)
510 FORMAT (' ' ,10x, гистограмма для g1 ' , (6f10.3/))
      WRITE(1,601)(gr(i),i=1,20)
601 FORMAT (' ',151, 'значение g-g1',(6f10.3/))
      WRITE(1,500)(pi(i),i=1,10)
500 FORMAT (' гистограмма для q1',6f10.3/)
       WRITE (1,501)(q19i), i-1, 10)
501 FORMAT(' гистограмма для q1 ',6f10.3/)
     a3=a1+a2
     b3=30.
     m3=10
     n=50
     CALL SM (a3, b3, m3, u, z,n)
     WRITE (1, 410)
410 FORMAT(' ' ,20x, 'результат')
      WRITE (1,280) (z(i),i=1,10)
280 FORMAT (' гистограмма для u',6f10.3/)
     STOP
     END
     FUNCTION RAN (n1, n)
     N=8003*n1
     IF(n) 3,4,4
  3 n=n+32767+1
  4 n1=n
    RAN=n/32768.
```

ПРИЛОЖЕНИЕ II

Здесь приводится пример программы построения оценки компоненты свертки плотностей вероятностей, демонстрирующей применение задачи 2 на практике.

1. Характеристики программы.

Машинный алгоритм решения данной задачи разработан с учетом его реализации на ПЭВМ.

Программа решается в течение меньше 1 минуты в зависимости от длины входного массива.

2. Сведения о форме представления программы.

Программа написана на языке PYTHON с использованием подпрограммы для вычисления гистограммы.

Имя программы task_2.

Имя подпрограммы sm.

Текст программы находится в файле task_2.py.

Обращение к подпрограмме;

SM(a,b,d,u,g,m)

3. Входные и выходные данные.

В качестве входной информации используются;

h - шаг разбиения,
m - объем выборки для свертки,
n1 - объем выборки для известной компоненты,
tg - квантиль χ^2- распределения с 9 степенями свободы,
[a,b] - интервал разбиения для случайной величины,
[c,f] - интервал разбиения для случайной величины,
g3 - гистограмма для свертки,
g1 - гистограмма для компоненты.

В результате работы программы выдаются на печать:

g2 - гистограмма для компоненты,
E1 - точность оценивания.

Входные данные считываются из файлов input_t.csv и input_u.csv
Выходные данные выводятся в консоль.
Запуск программы осуществляется командой run из редактора.

Таблица соответствия переменных.

Алгоритм	Программа	Содержание
$\hat{g}_3(x)$	*g3*	$\hat{g}_{(2)}(k)$
$\hat{g}_1(x)$	*g1*	$\hat{g}_1(k)$
$\hat{g}_2$	*g2*	$\hat{g}_2(k)$
$u(i)$	*u(i)*	X_i^1
$T(i)$	*t(i)*	Y_i

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from math import sqrt

#Функция (подпрограмма)для расчета гистограммы
def SM(a, b, u1, d, m):

    g3=[0]*200
    g = [0 for i in range(d)]
    h = (b-a)/d
    x = a
    for j in range(0, d):

        for i in range(0, m):

            if int(u1[i]) < x + h and int(u1[i]) >= x :

                g[j] = g[j]+1

        x = x+h
        g[j] = g[j]/m
        print(g[j])
    return g

input_t = pd.read_csv('D:/Calendar/input_t.csv')
u = pd.read_csv('D:/Calendar/input_u.csv')

print('Параметры исходных гистограмм')
a = 0
b = 20
h = 1
```

```
c = 0
f = 10
m = 200
n1 = 20
K1 = 1
tg = 14.68
d = int((b - a)/h)
d1=d
print('a= ', a)
print('b= ', b)
print('h= ', h)
print('c= ', c)
print('f= ', f)

u1=np.array(u)

print('Гистограмма для g3')
g3=SM(a, b, u1, d, m)
u.hist('u', bins=int(d))

a=c
b=f
d=int((b-a)/h)
s1=(f-c)/h
print('d= ', d)
print('Гистограмма для g1')
g1=SM(a, b, u1, d, m)
input_t.hist('t', bins=int(d))
plt.show()

g2=[0]*10
s2=int(d1-s1)
if g1[0] != 0:
    g2[0]=2*g3[0]/g1[0]
for j in range(1,s2):
    s=0
    j1=j-1
    for i in range(0,s2-1):
        s=s+(g1[i]+g1[i+1])*g2[j-1]
    if g1[0] !=0:
        g2[j]=(2*g3[j]-s)/g1[0]
if n1 >= m:
    n=m
else:
```

```
    n=n1
sq=tg/n
sg=K1
e1=sqrt(s2*h)*h*sg+sqrt(sg)*sqrt(sq)

print('Гистограмма для g2')
print(g3)
print()
print('Точность оцениевания e1= ', e1)
print('Надежность оценивания 0.9')
```

ПРИЛОЖЕНИЕ III

Здесь приводится пример программы, использующей уравнение Линдли для получения предельного распределения времени между появлениями m-го и (m+1) требований и временем обслуживания m-го требования.

1. Характеристики программы

Машинный алгоритм решения данной задачи разработан с учетом его реализации на ПЭВМ

2. Сведения о форме представления программы.

Программа написана на языке "СИ". Текст программы находится в файле stat4.

Входные параметры

```
#include <stdio.h>
#include <process.h>
#include <malloc.h>
#include <math.h>
#include <conio.h>
#define NOV       fprint( fvyvod,"\n");
#define DL_STRlNG 32L /*длина строки журнала регистрации*/
#define P1 (v,a,n) a=v,a+=(n);
#define P2(v,a,m,n,k1) P1(v,a(m)*(k1)+(n))
#define P3(v,a,m,l,n,k1,k2)
P1 (v,a,(m)*(k1)*(k2)+(1)*(k2)+(n))
void gsvert (float * , float * , int , int , int );
float mako(int , float * );
float frw( float * , int , float * , int , float * );
void message( short , short , short * , char * *, int );
void gists( );
void steps( );
void initline(),line();
void wait1();

extern char * zagwait [ ] ;
extern char * zagstep;
extern char * step[];
extern short zentr[2];
extern short colors [ ];
extern short colors _graph[];
char * mess_vyr[] = {"Вырожденная выборка.", NULL };

char * mess_ne_st[] = {Стационарное время ожидания не
```

```
существует.",
NULL};
char * mess_0[ ] = {"Распределение времени ожидания полностью сосредо-
точено в
точке 0.",
NULL};

static int k1,k2, ng=20;
static float hg;
extern FILE * fvyvod;
static float * t_t00, * t_t0;
static float min,max;
static float * c , * c0;
static float * w , * wO;
static float * www, * wwwO;
static float s, x, y, ww, toch , stime ;

static int k3,jj;
float st;
extern int dll;
int kline, klinel;
float amn,stmax,stshag;

/*Формирование гистограммы */

void
stat4(t_t , nelem , m , n , gamma )
float * t_t ;
float * gamma ;
int nelem , m, n ;
{ register int i , j;
 int k;
 steps(1,zentr[1],zentr[0],colors,zagstep,step,1);
initline();
rprintf( fvyvod,"\n\n  Результаты расчетов :\n");
fot( i=0,min=max=(*t_t),t_tO=t_t;
i<nelem;
i++,t_t0++ )
{ if( min>(*t_t0) ) min=(*t_t0);
if( max<(*t_t0) ) max=(*t_t0);
}
if( (hg=(max-min)/ng)==0.)
{fprintf (fvyvod,"\nВырожденна выборка.");
steps(-1,entr[1],zentr[0],colors,zagstep,step,1);
```

```
wait1(5,18,zagwait,0,-1);
mesage(zentr[1],zentr[0],colors,mess_vyr,1);
goto to_return;
}
k1 =(int)floor(min/hg) -1;
k2 =(int)ceil(max/hg) +1;
fprintf( fvyvod,"\nШаг гистограммы =%f",hg);
fprintf( fvyvod,”\nГистограмма на интервале %d, %d”, k1, k2,);
c = (float *)calloc(k3=k2-k1+1 , sizeof (float) );

  stmax=nelem+3; st=1.; stshag=stmax/dll; kline1=0;
  if( (kline=(int)floor(st++/stshag))>kline1 )
  { line(kline-kline1); kline1=kline; }
  for( k=0,t_t0=t;k<nelem;k++,t_t0++,st++ )
  {for( j=0,i=k1,c0=c;j<k3;j++,i++,c0++ )
  if( ((*t_t0)>-(i-1)*hg) && ((*t_t0)<i*hg) )
  (*c0)+=1./nelem);
  if( (kline=(int)floor(st/stshag))>kline1 )
  { line(kline-kline1); kline1=kline; }
  }
  free( (void *)t_t );
/*  fprintf( fvyvod,”\n Сформированная гистограмма:\n);
  for( i=1, c0=c;i<k3;i++,c0++ )
  { fprintf( fvyvod,”%f “,*c0);
  if( i%5 == 0 ) NOV
  }
*/
  for( min=0,i=0,j=k1=1,c0=c;
      i<k3;
      i++,j++,c0++ )
  min+=(j*(*c0));
  if( (kline=(int)floor(st++/stshag))>kline1 )
   { line(kline-kline1); kline1=kline; )
  if( min>=0. )
    { fprintf( fvyvod,”\n\nСтационарное время ожидания не существует.”);
  steps(-1,zentr[1],azentr[0],colors,zagstep,step,1);
  wait1(5,18,zagwait,0,-1);
  message( zentr[1],zentr[0],colors,mess_ne_st,1);
  free( (void *)c );
  goto to_return;
   }
  if( k2<=0 )
   {fprintf(fvyvod,”\n\nРаспределение времени ожидания полностью сосре-
доточено в точке 0.”);
```

```
  steps(-1,zentr[1],zentr[0],colors,zagstep,step,1);
  wait(5,18,zagwait,0,-1);
  message(zentr{1},zentr[0],colors,mess_0,1 );
  free( (void *)c );
  goto to_return;
 }
   if( (kline=(int)floor(st++/stshag))>kline1 )
  { line(kline-kline1); kline1=kline; }
/* Основной программный блок "/
  steps(2,zentr[1],zentr[0],colors,zafstep,step,1);
  w=(float )calloc( ( jj=(m*k2+1))*2, sizeof(float) );
  www=(float *)calloc( jj, sizeof (float) );
  initline();
  gsvert( w , c , k1 , k2 , m );
  steps(3,zentr[1],zentr[0],colors,zagstep,step,1);
  P2(w,w0,1,0,jj);
  for( i=0,s=0.;i<jj;i++,w0++)
  s+=(*w0);
  initline();
  stmax=jj+jj+3; st=1.; stshag=stmax/dll;kline1=0;
  if( (kline=(int)floor(st++/stshag))>kline1 )
    {line(kline-kline1); kline1=kline;  }
  P2(w,w0,1,0,jj)
  if( s>=1. ) (*w0)+=(1.-s);
  else (*w0)-=(1.-s);
/* fprintf( fvyvod,"\n Результат свертки гистограммы:\n);
  for( i=1;i<+jj;i++,w0++ )
    {fprintf( fvyvod,"%f ,*w0);
  if( i%8 == 0 ) NOV
    }
/*
  fprintf( fvyvod,"\n Оценка функции распределения функции ожида-
ния:\n");
  fprintf( fvyvod," (шаг = %f начиная с 0. )\n",hg);
  for( i=1,x=0.,www0=www;i<=jj;i++,x+=hg,www0++,st++ )
   { fprintf( fvyvod,"%f ",(*www0)=frw(w,v,&hg,k2,&x));
     if( i%8 == 0 ) NOV
  if( (kline=(int)floor(st/stshag))>kline1 )
   { line(kline-kline1); kline1=kline; }
 }
  for( i=1,min=(*c),c0=c+1;i<k3;i++,c0++ )
   { if( min>(*c0) ) min=(*c0);
  (*c0)/=hg;
 }
```

```
  (*c)/hg;
  if( (kline=(int)floor(st++/stshag))>kline1 )
   { line(kline-kline1); kline1=kline; }
  if( (toch =m*( mako(nelem,gamma) +min*hg))>1.)toch=1.;
  fprintf( fvyvod,”\n Точность по формуле 1=%f”,toch);
  if( (toch=m*mako(nelem,gamma)+min*hg)>1. ) toch=1.,
  fprintf( fvyvod,”\n Точность по формуле 2=%f”, toch);
  fprintf( fvyvod,”\n Статистическая погрешность =%f”, toch –m*hg*min );
  fprintf( fvyvod,”\n Статистическая погрешность =%f”, hg*min );
  for( i=0,x=0.,y=hg,stime=0.;i<jj;i++,x+=hg,y+=hg,st++ )
   { stime+=(x*(frw(w,m,&hg,k2,&y)-frw(w,m,&hg,k2,&x) ));
  if( (kline=(int)floor(st/stshag))>kline1 )
   {line(kline-kline1); kline=kline1; }
 }
  fprintf( fvyvod,”\n\n Среднее время ожидания .=%f”,stime);
  if( (kline=(int)floor(st++/stshag))>kline1 )
   {line(kline-kline1); kline1=kline; }
  gists(1,jj,&www,0.,hg1.,
  “Оценка функции распределения времени ожидания:”,
  colors_graph );
  free( (void *)c );
  free( (void *)w );
  free( (void *)www );
  steps(-1,zentr[1],zentr[0],colors,zagstep,step,1);
  wait(5,18,zagwait,0,-1);
  to_return:


 return;
 }

/*   Вычисление сверток гистограмм кратности до m, заданных на интер-
вале (k1,k2).
*/

 #include <memory.h>

 static float * ew0 , * ew00 , * cc0;

 void
 dsvert( ew , cc , k11 , k22 , mn )
  float * ew , * cc;
  int k11 , k22 , mn;
  { register int i , j;
```

```
    int 1 , j1 , k;
    stmax=mn; st=1.; stshag=stmax/dll; kline1=0;
    j=(mn*k22+1)*2;
    (*ew0)=1.;
    for( i=1,ew0=ew+1;i<j;i++,ew0++ )
     (*ew0)=0.;
    j1=j/2;
    for( l=0;l<mn;l++,st++ )
      { P2(ew,ew0,1,0,j1)
        (*ew0)=0.;
        for( k=k11;k<=0;k++ )
          for( j=0,ew00=ew;j<=l*k22;j++,ew00++ )
            for( i=k11,cc0=cc;i<k22;i++,cc0++ )
             if( i+j == k )
                  (*ew0)+=(0.5*((cc0)+(*(cc0+1)))*(*ew00));
       P2(ew,ew0,1,1,j1)
       for( k=1;k<=(l+1)*k22;k++,ew00++ )
        { (*ew0)=0.;
          for( j=0,ew00=ew;j<=l*k22;j++,ew00++ )
           for( i=k11,cc0=cc;i<k22;i++,cc0++ )
             if( i+j == k )
                  (*ew0)+=(0.5*((cc0)+(*(cc0+1)))*(*ew00));
  }
P2(ew,ew0,1,0,j1)
Memcpy( (void *)ew , (void )ew0 , j1sizeof(float) );
If( (kline=(int)floor(st/stshag))>kline1 )
   { line(kline-kline1); kline1=kline; }
  }
}

/* Преобразование гистограммы в эмпирическую ф.р. */

static float * ew000;

float
frw( ew , mn , hgg , k22 , xx )
float * ew;
float * hgg , * xx ;
int mn , k22;
{ float ret , s=0. ,ss=0.;
 register int i , j1;
 if( (*xx)<0. ) ret=0.;
 else { if( (*xx))>mn*k22*(*hgg) ) ter=1.;
      else { j1 =mn*k22+1;
```

```
            P2(ew,ew000,1,0,j1)
            if( (*xx)==0. ) ret=(*ew000);
            else { for( i=0;i<j1;i++,ew000++ )
                      { ss+=(*hgg);
                      if( s>(*xx) )break;
                     }
                     ret=ss;
                    }
             }
     }
return( ret );

/* Вычисление квантилей распределений Масси-Колмогорова для уровней
    доверия gamma= 0.8 , 0.9 , 0.95 , 0.98 , 0.99.
/*include <math.h>*/
#include “quanmako.h”
float
mako( n_ ,gamm )
   int n_ ;
   float * gamm ;
   { register int j;
     float ret;
     if( (*gamm)<=0.81 ) j=0;
     else { if( (*gamm)<=0.91 ) j=1;
          else { if( (*gamm)<=0.951 ) j=2;
               else { if( (*gamm)<=0.981 ) j=3;
                      else j=4;
                     }
            }
     }
if( n_>100 ) ret =quantils[100][j]/(float)
            sqrt((double)n_ );
       else ret =quantils[n_-1][j];
       return (ret );
   }
```

Учебное издание

Солодянников Юрий Васильевич

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Редакционно-издательская обработка Л.Р. Дмитриенко

Подписано в печать 10.04.2023. Формат 60х84 1/16.
Бумага офсетная. Печ. л. 9,5.
Тираж 45 экз. Заказ № .

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)
443086, Самара, Московское шоссе, 34.

Издательство Самарского университета.
443086, Самара, Московское шоссе.