Министерство образования Республики Беларусь
Учреждение образования
“Гомельский государственный университет
имени Франциска Скорины”

Ю.В. Малинковский

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА (ЧАСТЬ 2. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА)

Учебное пособие

Допущено Министерством образования Республики Беларусь в качестве учебного пособия для студентов математических специальностей учреждений, обеспечивающих получение высшего образования

Гомель 2004

УДК 519.2(075.8)
ББК 22.17 Я73
М 191

Рецензенты:
кафедра теории функций, функционального анализа, теории вероятностей и прикладной математики Гродненского государственного университета им. Янки Купалы,
д-р физ.-мат. наук, профессор Г.А.Медведев

Малинковский Ю.В.

Теория вероятностей и математическая статистика (часть 2. Математическая статистика): Учебное пособие / Ю.В.Малинковский. — Гомель: УО "ГГУ им. Ф.Скорины", 2004. — 146 с.

ISBN

В основу данного учебного пособия положены курсы лекций по теории вероятностей и математической статистике, которые читаются автором на математическом факультете Гомельского государственного университета им. Ф.Скорины начиная с 1973 года в рамках специальности "Математика". Во второй части изложены основы математической статистики.

Предназначено для студентов математических специальностей ВУЗов. Будет полезно для студентов других факультетов, изучающих теорию вероятностей и математическую статистику.

Библиогр.: 17 назв.

ПОСВЯЩАЕТСЯ 30–летию
Белорусской школы по
теории вероятностей и
математической статистике
и ее создателю Г.А.Медведеву

Предисловие

Вторая часть пособия посвящена математической статистике. Главное внимание уделяется основным идеям классической статистики, в основном на примере статистической структуры повторной выборки.

В первом параграфе главы 2 рассматривается понятие выборки, эмпирического распределения и выборочных моментов. Параграфы 2-4 посвящены точечным оценкам неизвестных параметров. В 5-ом параграфе изучается многомерное нормальное распределение, в 6-ом — интервальные оценки неизвестных параметров. Основы общей теории проверки статистических гипотез содержатся в параграфе 7. Параграфы 8 и 9 посвящены критериям значимости и весьма важному их частному случаю — критериям согласия. В 10-ом параграфе рассмотрены идеи однофакторного дисперсионного анализа. 11-ый параграф посвящен основам линейного регрессионного анализа. Здесь происходит знакомство с условными математическими ожиданиями (в основном для случая дискретных распределений). Автор считает, что давать общее определение условных ожиданий относительно σ-алгебр, основанное либо на теореме Радона-Никодима, либо на проекциях в гильбертовом пространстве, либо на функциональном подходе в общем курсе не стоит. Как показал многолетний опыт преподавания для неформального усвоения этого материала необходимо гораздо больше учебных часов, чем то количество, которое может быть выкроено из общего курса без ущерба для остальных разделов. В последнем параграфе можно ознакомиться с понятиями достаточных статистик, экспоненциальных семейств, минимальных и полных достаточных статистик, теоремой Дынкина и неравенством Колмогорова-Блекуэлла-Рао.

Если имеется ссылка на первую часть-главу, то впереди добавляется 1., например, (1.4.3) — формула (4.3) из главы 1, теорема 1.7.3 — теорема 7.3 из главы 1.

В заключение, автор выражает глубокую благодарность членам кафедры теории функций, функционального анализа, теории вероятностей и прикладной математики Гродненского государственного университета им. Янки Купалы, взявшим на себя нелегкий труд рецензирования данного пособия, и рецензенту Г.А.Медведеву за весьма полезные замечания.

г.Гомель, июнь 2004 г. *Ю.Малинковский*

ГЛАВА 2. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА.

§ 1. Выборка. Эмпирическое распределение. Выборочные моменты

В самых общих чертах можно определить статистику как науку, которая предназначена для создания и использования методов сбора, сокращения, хранения и обработки различной информации.

Теория вероятностей и математическая статистика — тесно связанные между собой разделы математики, в которых применяются близкие методы, основанные на понятии вероятности. Однако их цели и задачи во многом отличаются. Рассмотрим два примера.

1. Типичной задачей теории вероятностей является следующая. Заданы вероятности не более чем счетного набора событий $P(A_k), \quad k = 1, 2, \ldots$. Требуется найти вероятность события A, которое выражается через события набора с помощью теоретико-множественных операций. В теории вероятностей вероятности набора задают не задумываясь над тем, откуда они взялись. Но, как правило, их и невозможно задать в рамках теории вероятностей, а можно только постулировать их существование. Единственные случаи их возможного задания в указанных рамках — это те случаи, когда случайный эксперимент обладает симметрией возможных исходов, т.е. можно пользоваться классическим или геометрическим определением вероятности (попробуйте задать вероятность выпадения герба на монете, одна часть которой алюминиевая, а другая свинцовая). Математическая статистика позволяет по результатам проведенных экспериментов

строить оценки этих первоначальных вероятностей, проверять гипотезы о виде этих вероятностей и т.д.

2. Заданы совместное распределение случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, например, в виде плотности $p_{\xi_1,\xi_2,\ldots,\xi_n}(x_1, x_2, \ldots, x_n)$, и некоторая борелевская функция $f(x_1, x_2, \ldots, x_n)$. Требуется найти закон распределения случайной величины $\eta = f(\xi_1, \xi_2, \ldots, \xi_n)$. В теории вероятностей совместное распределение $\xi_1, \xi_2, \ldots, \xi_n$ задается и, как правило, не может быть найдено в ее рамках. В математической статистике ищутся оценки этого совместного распределения или его параметров, проверяются гипотезы об их виде и т.д.

Математическая статистика в основном занимается

а) оценкой неизвестных параметров (точечные и интервальные оценки);

б) проверкой статистических гипотез (параметрические и непараметрические гипотезы);

в) установлением статистических связей между случайными событиями, между случайными величинами и между случайными процессами (регрессионный и корреляционный анализ);

г) управлением случайными процессами.

Предположим, что проведено n экспериментов, в которых регистрируются значения, принимаемые некоторой случайной величиной ξ, называемой в дальнейшем **теоретической случайной величиной**. Ее функцию распределения $F(x)$ и распределение P_ξ также будем называть **теоретическими**. Пусть $x_1, x_2, \ldots, x_n$ — результаты экспериментов, т.е. n чисел. Предположим, что выполнены следующие предпосылки:

1) эксперименты проводятся в одинаковых условиях;

2) эксперименты проводятся независимо друг от дру-

га.

В этом случае набор чисел $x_1, x_2, \ldots, x_n$ называют **конкретной выборкой объема n из распределения** F. (Здесь, как это часто делается в теории вероятностей и математической статистике, проявляется небрежность в употреблении слов — некоторая функция, в данном случае функция распределения, однозначно определяющая распределение, сама именуется распределением).

Пусть теперь $\xi_1, \xi_2, \ldots, \xi_n$ — результаты предполагаемых n экспериментов (которые мы хотим произвести). Так как эксперименты случайны, то на результаты надо смотреть как на случайные величины. Из первой предпосылки и того, что измеряется одна и та же теоретическая случайная величина ξ, следует, что случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ имеют одно и то же распределение, совпадающее с теоретическим распределением F. Из второй предпосылки следует, что эти случайные величины независимы. Поэтому логично ввести следующее

Определение 1.1. Говорят, что случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ образуют **абстрактную выборку объема n из распределения** F, если они независимы и имеют одну и ту же функцию распределения F.

Фигурально выражаясь, абстрактная выборка — это n "независимых экземпляров" теоретической случайной величины ξ. Таким образом, конкретная выборка может рассматриваться как значение, которое принимает абстрактная выборка в результате составного случайного эксперимента, заключающегося в n измерениях теоретической случайной величины ξ.

Числовые характеристики теоретической случайной величины ξ называются **теоретическими** или **генеральными** характеристиками. Например, $M\xi = a$ — теоретическое или генеральное математическое ожидание, более коротко — **теоретическая** или **генеральная средняя**; $D\xi = \sigma^2$ — **теоретическая** или **генеральная дисперсия**.

Во многих руководствах по математической статистике сложилась своеобразная традиция, согласно которой различные понятия обозначаются одной и той же буквой. Например, абстрактная и конкретная выборки обозначаются одинаково как $x_1, x_2, \ldots, x_n$ и называются **выборкой объема** n **из распределения** F. Какая из этих выборок подразумевается, должно быть ясно из контекста рассматриваемого вопроса. В теоретических исследованиях выборка понимается как абстрактная, при применении конкретного статистического материала в приложениях она понимается как конкретная. В теории вероятностей функцию распределения и плотность вероятности первого элемента выборки мы обозначили бы соответственно $F_{\xi_1}(x)$ и $p_{\xi_1}(x)$. В статистике мы будем обозначать их $F(x_1)$ и $p(x_1)$. Здесь x_1 одновременно означает случайную величину — первый элемент абстрактной выборки, к которой относятся F и p, и действительный аргумент функции распределения и плотности вероятности. Например, в теории вероятностей и математической статистике одно и то же математическое ожидание первого элемента абстрактной выборки из абсолютно непрерывного распределения с плотностью p записывается в формах соответственно:

$$M\xi_1 = \int\limits_{-\infty}^{+\infty} x p_{\xi_1}(x) dx \qquad \text{и} \qquad M x_1 = \int\limits_{-\infty}^{+\infty} x_1 p(x_1) dx_1.$$

Итак, один и тот же символ x_k используется для обозначения числа — k-го элемента конкретной выборки, случайной величины — k-го элемента абстрактной выборки, переменной — аргумента закона распределения и переменной интегрирования. С одной стороны, это может внести некоторую путаницу для неискушенного читателя (правда, к таким обозначениям быстро привыкаешь и они не кажутся тогда странными), с другой стороны такие обозначения во многом сокращают запись и потому удобны.

Пусть $x_1, x_2, \ldots, x_n$ выборка из распределения F, где $F(x)$ — непрерывная функция распределения. В этом случае без потери общности можно считать, что все значения в выборке различны (ведь $P\{x_i = x_j\} = 0$ при $i \neq j$). Если никакой другой информации о теоретической случайной величине ξ кроме выборки у нас нет, то самое лучшее, что можно сделать, — это заменить ξ некоторой фиктивной случайной величиной η, которая принимает значения $x_1, x_2, \ldots, x_n$. Так как ни одному из выборочных значений нельзя отдать преимущество перед другими выборочными значениями, то естественно приписать всем значениям случайной величины η одинаковые вероятности. Итак, мы ожидаем, что фиктивная случайная дискретная величина с рядом распределения

η	x_1	x_2	......	x_n
P	$\frac{1}{n}$	$\frac{1}{n}$	......	$\frac{1}{n}$

,

называемая **эмпирической** или **выборочной случайной величиной**, в определенном смысле является величиной, напоминающей неизвестную нам теоретическую величину ξ. По крайней мере, ничего лучшего для того, чтобы хоть как-то представить ξ по выборке, мы сделать не можем.

Определение 1.2. Распределение $P_n^*(B) = P\{\eta \in B\}$, $B \in \mathcal{B}$, случайной величины η называется **эмпирическим распределением**, а ее функция распределения $F_n^*(x) = F_\eta(x) = P\{\eta < x\}$ называется **эмпирической функцией распределения**.

Очевидно, что

$$P_n^*(B) = \sum_{k\colon x_k \in B} P\{\eta = x_k\} = \sum_{k\colon x_k \in B} \frac{1}{n} = \frac{\mu(B)}{n},$$

где $\mu(B)$ — число выборочных значений, попадающих в B, а

$$F_n^*(x) = \sum_{k\colon x_k < x} P\{\eta = x_k\} = \sum_{k\colon x_k < x} \frac{1}{n} = \frac{\mu(x)}{n},$$

где $\mu(x)$ — число выборочных значений строго меньших x. Таким образом, $P_n^*(B)$ — относительная частота попадания теоретической случайной величины ξ в множество B, а $F_n^*(x)$ — относительная частота события $\xi < x$, если на выборку смотреть как на n независимых испытаний, в каждом из которых наблюдается теоретическая случайная величина ξ. Если за успех принять событие $\{\xi < x\}$, то имеем схему Бернулли с вероятностью успеха в отдельном испытании $p = P\{\xi < x\} = F(x)$ и вероятностью неудачи $q = 1 - F(x)$. При этом $\mu(x)$ есть число успехов в n испытаниях. Не давая пока формального определения оценки, а рассматривая при каждом фиксированном $x \in \mathbb{R}$ эмпирическую функцию распределения как оценку теоретической функции распределения, получим, что

1. $MF_n^*(x) = M\dfrac{\mu(x)}{n} = \dfrac{1}{n}M\mu(x) =$

$$= \frac{1}{n}np = \frac{1}{n}nF(x) = F(x),$$

т.е. среднее значение оценки $F_n^*(x)$ совпадает с оцениваемой величиной $F(x)$ (так называемая несмещенность оценки);

2. По закону больших чисел в форме Бернулли

$$F_n^*(x) = \frac{\mu(x)}{n} \xrightarrow{P} F(x) \quad \text{при} \quad n \to \infty$$

(так называемая состоятельность оценки).

На самом деле здесь имеет место сходимость почти наверное (по усиленному закону больших чисел в форме Бореля). Но и это утверждение можно усилить до равномерной сходимости почти наверное. Точнее, справедлива следующая

Теорема Гливенко-Кантелли. *Эмпирическая функция распределения почти наверное равномерно сходится к теоретической функции распределения*:

$$P\Big\{ \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| \to 0 \quad \text{при} \quad n \to \infty \Big\} = 1.$$

Доказательство этой теоремы для случая непрерывной функции $F(x)$ выносится в приложение.

Таким образом, для выборок большого объема эмпирическая функция распределения мало отличается от теоретической функции распределения, т.е. в этом случае эмпирическая случайная величина является достаточно хорошим образом теоретической случайной величины, построенным по выборке.

Расставим значения выборки в возрастающем порядке:

$$x_{(1)} < x_{(2)} < \ldots < x_{(n)}.$$

Тогда, в частности,

$$x_{(1)} = \min(x_1, x_2, \ldots, x_n), \qquad x_{(n)} = \max(x_1, x_2, \ldots, x_n).$$

При этом x_k называется **k-й порядковой статистикой**, $W_n = x_{(n)} - x_{(1)}$ — **размахом выборки**, а конечный ряд $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ называется **вариационным рядом выборки**.

Задача 1. Найдите функции распределения крайних членов вариационного ряда.

Задача 2. Предполагая теоретическую функцию распределения абсолютно непрерывной, найдите плотность распределения вероятностей k-й порядковой статистики и функцию распределения размаха выборки.

Очевидно, эмпирическая функция распределения может быть записана в виде

$$F_n^*(x) = \begin{cases} 0, & \text{если } x \le x_{(1)}, \\ \frac{1}{n}, & \text{если } x_{(1)} < x \le x_{(2)}, \\ \frac{2}{n}, & \text{если } x_{(2)} < x \le x_{(3)}, \\ \ldots & \ldots\ldots\ldots\ldots\ldots \\ 1, & \text{если } x > x_{(n)}. \end{cases}$$

Замечание 1. Если теоретическая функция распределения не является непрерывной (например, теоретическая случайная величина дискретна), то некоторые значения в выборке могут совпадать. Тогда элементы выборки можно записать в неубывающем порядке и обозначить k-е по величине значение через $x_{(k)}$. Эмпирическая случайная величина будет задаваться рядом распределения

$$\begin{array}{c||c|c|c|c||}
\eta & x_{(1)} & x_{(2)} & \ldots\ldots & x_{(l)} \\
\hline
P & \frac{n_1}{n} & \frac{n_2}{n} & \ldots\ldots & \frac{n_l}{n}
\end{array}\ ,$$

где l — число различных значений в выборке, n_k — число повторений значения $x_{(k)}$ в ней. При этом смысл $\mu(x)$ сохраняется, только каждое значение $x_{(k)} < x$ учитывается столько раз, сколько оно встречается в выборке. Эмпирическая функция распределения, как нетрудно понять, может быть записана в виде

$$F_n^*(x) = \begin{cases} 0, & \text{если } x \le x_{(1)}, \\ \frac{n_1}{n}, & \text{если } x_{(1)} < x \le x_{(2)}, \\ \frac{n_1+n_2}{n}, & \text{если } x_{(2)} < x \le x_{(3)}, \\ \ldots & \ldots\ldots\ldots\ldots\ldots \\ 1, & \text{если } x > x_{(l)}. \end{cases}$$

Все свойства эмпирической функции распределения как оценки теоретической функции распределения сохранятся.

Так как при больших n значения эмпирической и теоретической функций распределения неограниченно сближаются, то следует ожидать, что различные их числовые характеристики, например, моменты одного и того же порядка также будут близки друг к другу. Пусть $b_l = M\xi^l = \int\limits_{-\infty}^{+\infty} x^l dF(x)$ — теоретический начальный момент l-го порядка. В качестве его оценки возьмем выборочный начальный момент l-го порядка

$$\overline{x^l} = M\eta^l = \sum_{k=1}^{n} x_k^l\, \frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k^l.$$

В частности, при $l = 1$ получаем, что в качестве оценки теоретической средней $a = b_1 = M\xi$ берется выборочная средняя $\overline{x} = \frac{1}{n}\sum\limits_{k=1}^{n} x_k$.

Пусть теперь $c_k = M(\xi - M\xi)^l$ — теоретический центральный момент l-го порядка. В качестве его оценки возьмем выборочный центральный момент l-го порядка

$$\widetilde{S}^l = M(\eta - M\eta)^l = M(\eta - \overline{x})^l =$$

$$= \sum_{k=1}^{n}(x_k - \overline{x})^l \,\frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^l.$$

В частности, при $k = 2$ оценкой теоретической дисперсии $\sigma^2 = D\xi$ является выборочная дисперсия

$$\widetilde{S^2} = D\eta = M(\eta - M\eta)^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2.$$

В дальнейшем нам понадобится также другое выражение для выборочной дисперсии, получающееся применением иной формулы для дисперсии случайной величины,

$$\widetilde{S^2} = D\eta = M\eta^2 - (M\eta)^2 = M\eta^2 - \overline{x}^2 =$$

$$= \overline{x^2} - \overline{x}^2 = \frac{1}{n}\sum_{k=1}^{n}{x_k}^2 - \overline{x}^2,$$

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

Пусть $x_1, x_2, \dots, x_n$ — выборка, $x_{(1)}$ и $x_{(n)}$ — наименьшая и наибольшая порядковые статистики соответственно. Разделим отрезок $[x_{(1)}, x_{(n)}]$ на k равных частей длиной $l = \frac{x_{(n)} - x_{(1)}}{k}$ каждая. Обозначим эти части

$\Delta_1, \Delta_2, \dots, \Delta_k$ и пусть $\mu(\Delta_i)$ — число выборочных значений, попавших в Δ_i. На каждом Δ_i как на основании построим прямоугольники высотой $h_i = \frac{\mu(\Delta_i)}{nl}$. Полученная ступенчатая фигура называется **гистограммой относительных частот**. Ее площадь равна

$$S = \sum_{i=1}^{k} |\Delta_i| h_i = \frac{1}{n} \sum_{i=1}^{k} \mu(\Delta_i) = 1.$$

Предположим, что теоретическая случайная величина ξ абсолютно непрерывна с плотностью распределения $p(x)$. По теореме о среднем

$$P\{\xi \in \Delta_i\} = \int\limits_{\Delta_i} p(x)dx = p(c_i)|\Delta_i|.$$

По усиленному закону больших чисел в форме Бореля

$$\frac{\mu(\Delta_i)}{n} \overset{\text{п.н.}}{\to} P\{\xi \in \Delta_i\} = p(c_i)l,$$

откуда

$$h_i = \frac{\mu(\Delta_i)}{nl} \overset{\text{п.н.}}{\to} p(c_i).$$

Таким образом, в абсолютно непрерывном случае гистограмма дает качественное представление о плотности вероятности теоретической случайной величины. Иногда на основании проведенного рассуждения считают, что при большом числе наблюдений гистограмма стремится к плотности вероятности $p(x)$. Это неверно: для того чтобы это было так, нужно, чтобы выполнялось довольно сложное соотношение между n и k.

Заметим, что указанным недостатком не обладает эмпирическая функция распределения, поскольку $F_n(x) \overset{\text{п.н.}}{\to} F(x)$ равномерно по $x \in \mathbb{R}$ без всяких дополнительных ограничений.

§ 2. Точечные оценки неизвестных параметров

Начнем с примера. Предположим, что измеряется определенная физическая постоянная a. Результаты измерений

$$x_1 = a + \delta_1,$$

$$x_2 = a + \delta_2,$$

$$\dots\dots\dots$$

$$x_n = a + \delta_n$$

обладают некоторой случайной ошибкой δ. Если ошибка измерений складывается из большого числа независимых вкладов, каждый из которых оказывает равномерно малое влияние на ошибку, то по центральной предельной теореме ошибка будет приближенно иметь нормальное распределение. Предположим, что измерения лишены систематической ошибки, т.е. $M\delta_k = 0$. Тогда $\delta_k \sim N(0, \sigma^2)$. По лемме о нормальном распределении $x_k \sim N(a, \sigma^2)$. Таким образом, $x_1, x_2, \dots, x_n$ образуют выборку из нормального распределения с параметрами (a, σ^2). Возникает задача оценки по данной выборке параметра a, т.е. измеряемой физической постоянной.

Пусть $x_1, x_2, \dots, x_n$ — выборка из распределения $F(x/\theta)$, зависящего от неизвестного параметра θ.

Определение 2.1. **Оценкой** или **статистикой** неизвестного параметра θ называется любая борелевская функция

$$\widehat{\theta_n} = \widehat{\theta}(x_1, x_2, \dots, x_n)$$

от выборочных значений.

В этом определении требование измеримости по Борелю необходимо для того, чтобы при подстановке в оценку элементов абстрактной выборки получалась случайная величина. Это позволяет рассматривать, например, математическое ожидание оценки или другие числовые характеристики. Отметим, что определение оценки носит формальный характер и не отражает интуитивного понимания оценки как некоторой величины, близкой к оцениваемому параметру. Например, постоянная — борелевская функция. Поэтому в качестве оценки параметра θ можно взять $\widehat{\theta_n} = 0$, или $\widehat{\theta_n} = 10^{10}$, или $\widehat{\theta_n} = -10^{50}$, т.е. величину как угодно далекую от оцениваемого параметра.

Для того, чтобы оценка обладала действительно хорошими качествами, приближаясь в некотором смысле к оцениваемому параметру, на нее накладывают определенные дополнительные требования.

Определение 2.2. Оценка неизвестного параметра θ называется **несмещенной**, если ее среднее значение совпадает с оцениваемым параметром, т.е.

$$M\widehat{\theta_n} = \theta.$$

Для произвольной оценки с конечным математическим ожиданием **смещением** называется $\beta_n = M\widehat{\theta_n} - \theta$. Таким образом, несмещенная оценка — оценка со смещением, равным нулю.

Определение 2.3. Оценка неизвестного параметра θ называется **состоятельной**, если она сходится по вероятности к оцениваемому параметру, т.е.

$$\widehat{\theta_n} \xrightarrow{P} \theta.$$

Состоятельность означает, что с ростом объема выборки качество оценки должно улучшаться: чем больше n, тем ближе оценка к θ.

Как мы помним, в пространстве $L_2(\Omega, F, P)$ случайных величин, имеющих конечный начальный момент второго порядка ($M\xi^2 < \infty$), расстояние между случайными величинами ξ и η определяется как

$$\rho(\xi, \eta) = \sqrt{M|\xi - \eta|^2}.$$

Естественно пытаться построить такую оценку, которая была бы как можно ближе от оцениваемого параметра по этому расстоянию.

Определение 2.4. Оценка неизвестного параметра θ называется **эффективной**, если среди всех оценок с конечным моментом второго порядка она доставляет минимум $M|\widehat{\theta_n} - \theta|^2$.

Таким образом, если для любой оценки $\widehat{\theta_n}$ и некоторой оценки θ_n^* с конечными начальными моментами второго порядка выполняется неравенство

$$M|\theta_n^* - \theta|^2 \leq M|\widehat{\theta_n} - \theta|^2,$$

то оценка θ_n^* является эффективной.

Часто эффективную оценку ищут в классе несмещенных оценок. Так как в этом классе $M\widehat{\theta_n} = \theta$, то $M|\widehat{\theta_n} - \theta|^2$ превращается в дисперсию. Поэтому эффективной в классе несмещенных оценок будет такая несмещенная оценка, которая среди всех несмещенных оценок с конечной дисперсией имеет наименьшую дисперсию. Следовательно, она будет наиболее компактно разбрасывать свои значения вокруг оцениваемого параметра, когда для оценивания будут использоваться различные выборки объема n.

Теорема 2.5. *Если в классе несмещенных оценок существует эффективная оценка, то она единственна с точностью до эквивалентности (т.е., если $\widehat{\theta_n}$ и $\widehat{\psi_n}$ — две эффективные несмещенные оценки, то $P\{\widehat{\theta_n} = \widehat{\psi_n}\} =$ 1).*

Д о к а з а т е л ь с т в о. Пусть $\widehat{\theta_n}$ и $\widehat{\psi_n}$ — две эффективные несмещенные оценки, т.е.

$$M\widehat{\theta_n} = M\widehat{\psi_n} = \theta$$

и, если σ^2 — наименьшая из дисперсий несмещенных оценок, то

$$D\widehat{\theta_n} = D\widehat{\psi_n} = \sigma^2.$$

Рассмотрим оценку

$$\widehat{\tau}_n = \frac{\widehat{\theta_n} + \widehat{\psi_n}}{2}.$$

Она несмещенная:

$$M\widehat{\tau}_n = \frac{1}{2}M(\widehat{\theta_n} + \widehat{\psi_n}) = \theta.$$

Так как σ^2 — наименьшая из дисперсий несмещенных оценок, то

$$D\widehat{\tau}_n = \frac{1}{4}D(\widehat{\theta_n} + \widehat{\psi_n}) = \frac{1}{4}\{D\widehat{\theta_n} + D\widehat{\psi_n} + 2cov(\widehat{\theta_n}, \widehat{\psi_n})\} =$$

$$= \frac{1}{4}(\sigma^2 + \sigma^2 + 2r\sigma\sigma) = \frac{1}{2}\sigma^2(1 + r) \geq \sigma^2,$$

где r — коэффициент корреляции между $\widehat{\theta_n}$ и $\widehat{\psi_n}$. Из последнего неравенства следует, что $r \geq 1$. Так как коэффициент корреляции не может быть больше 1, то $r = 1$. По свойству коэффициента корреляции с вероятностью 1

$$\widehat{\psi_n} = A\widehat{\theta_n} + B, \qquad \text{где} \quad A > 0.$$

Поскольку второе слагаемое — постоянная, то слагаемые — независимые случайные величины. По свойствам дисперсии

$$D\widehat{\psi_n} = A^2 D\widehat{\theta_n}, \ \text{ т.е. } \ \sigma^2 = A^2\sigma^2,$$

откуда с учетом того, что $A > 0$, следует $A = 1$. Итак, с вероятностью 1

$$\widehat{\psi_n} = \widehat{\theta_n} + B.$$

Беря математическое ожидание от обеих частей, получим $\theta = \theta + B$, т.е. $B = 0$. Таким образом, $\widehat{\psi_n} = \widehat{\theta_n}$ с вероятностью 1. □

В качестве примера рассмотрим выборку $x_1, x_2, \ldots, x_n$ из некоторого распределения с теоретическими средней $a = M\xi$ и дисперсией $\sigma^2 = D\xi$. Как говорилось выше, их оценками являются соответственно выборочная средняя $\overline{x} = \frac{1}{n}\sum\limits_{k=1}^{n} x_k$ и выборочная дисперсия $\widetilde{S^2} = \frac{1}{n}\sum\limits_{k=1}^{n}(x_k - \overline{x})^2 = \frac{1}{n}\sum\limits_{k=1}^{n} {x_k}^2 - \overline{x}^2$. С формальной точки зрения — это оценки, поскольку они являются непрерывными функциями от $x_1, x_2, \ldots, x_n$, т.е. борелевскими.

Установим, какими качествами обладают эти оценки.

1. *Оценка $\overline{x}$ теоретической средней a.*

$$A.\ \ M\overline{x} = \frac{1}{n}\sum_{k=1}^{n} Mx_k = \frac{1}{n}(a + a + \ldots + a) = \frac{1}{n}na = a,$$

т.е. $\overline{x}$ — *несмещенная оценка a.*

Б. Так как $x_1, x_2, \ldots, x_n$ — независимые одинаково распределенные случайные величины с конечным математическим ожиданием $M\xi = Mx_k = a$, то по закону больших чисел в форме Хинчина

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k \xrightarrow{P} a,$$

т.е. $\overline{x}$ — *состоятельная оценка a.*

В. Вопрос об эффективности оценки $\overline{x}$ в классе несмещенных оценок решается более сложно и зависит от вида теоретического распределения. Позже мы убедимся, что если выборка взята из нормального распределения с параметрами (a, σ^2), то $\overline{x}$ является эффективной в классе несмещенных и так называемых регулярных оценок. Если выборка берется из равномерного распределения на отрезке $[a - h, a + h]$, то существуют более эффективные оценки a, чем $\overline{x}$.

Задача 1. Покажите, что если $x_1, x_2, \ldots, x_n$ — выборка из равномерного распределения на отрезке $[a - h, a + h]$, $\widehat{a_n} = \frac{x_{(1)} + x_{(n)}}{2}$, то $D\overline{x} = O(\frac{1}{n})$, в то время как $D\widehat{a_n} = O(\frac{1}{n^2})$, т.е. при больших n оценка $\widehat{a_n}$ эффективнее оценки $\overline{x}$.

2. *Оценка* $\widetilde{S^2}$ *теоретической дисперсии* σ^2.

А. По свойствам математического ожидания

$$M\widetilde{S^2} = \frac{1}{n}\sum_{k=1}^{n} M{x_k}^2 - M\overline{x}^2. \tag{2.1}$$

Для определения $M\widetilde{S^2}$ остается определить $M{x_k}^2$ и $M\overline{x}^2$. Имеем

$$\sigma^2 = D\xi = Dx_k = Mx_k^2 - (Mx_k)^2 = Mx_k^2 - a^2,$$

$$D\overline{x} = \frac{1}{n^2}\sum_{k=1}^{n} Dx_k = \frac{\sigma^2}{n} = M\overline{x}^2 - (M\overline{x})^2 = M\overline{x}^2 - a^2,$$

откуда

$$Mx_k^2 = \sigma^2 + a^2, \qquad M\overline{x}^2 = \frac{\sigma^2}{n} + a^2. \tag{2.2}$$

Подставляя (2.2) в (2.1), получим:

$$M\widetilde{S^2} = \sigma^2 + a^2 - \frac{\sigma^2}{n} - a^2 = \frac{n-1}{n}\sigma^2.$$

Итак, $\widetilde{S^2}$ *не является несмещенной оценкой* σ^2. Поэтому подправим ее, введя так называемую **исправленную выборочную дисперсию**

$$S^2 = \frac{n}{n-1}\widetilde{S^2} = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2.$$

Исправленная выборочная дисперсия является несмещенной оценкой теоретической дисперсии:

$$MS^2 = \frac{n}{n-1}M\widetilde{S^2} = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.$$

Поэтому лучше пользоваться исправленной выборочной дисперсией. Однако для выборок большого объема множитель $\frac{n-1}{n}$ мало отличается от 1 и разница от того, какой выборочной дисперсией пользоваться, не столь уж существенна.

В. Так как случайные величины $x_1, x_2, \ldots, x_n$ независимы и одинаково распределены, то и случайные величины $x_1^2, x_2^2, \ldots, x_n^2$ обладают этими же свойствами (обоснуйте почему). Если теоретическая случайная величина имеет конечную дисперсию (мы строим ее оценку, т.е. предполагается ее существование), то она имеет конечный начальный момент второго порядка $b_2 = M\xi^2$. Поэтому для последовательности случайных величин (ξ_n^2) выполнены условия закона больших чисел в форме Хинчина, по которому

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k^2 \xrightarrow{P} b_2 = M\xi^2 = \sigma^2 + a^2.$$

Так как $\overline{x} \xrightarrow{P} a$, то по свойствам сходимости по вероятности $\overline{x}^2 \xrightarrow{P} a^2$ и $\widetilde{S^2} = \frac{1}{n}\sum\limits_{k=1}^{n}\xi_k^2 - \overline{x}^2 \xrightarrow{P} (\sigma^2 + a^2) - a^2 = \sigma^2$. Таким образом, $\widetilde{S^2}$ — *состоятельная оценка* σ^2.

Так как $\frac{n}{n-1} \to 1$, то и подавно $\frac{n}{n-1} \xrightarrow{P} 1$. По свойствам сходимости по вероятности $S^2 = \frac{n}{n-1}\widetilde{S^2} \xrightarrow{P} 1 \cdot \sigma^2 = \sigma^2$. Поэтому S^2 также *является состоятельной оценкой* σ^2.

§ 3. Методы получения оценок

Мы рассмотрим только два наиболее популярных метода получения точечных оценок.

1. Метод моментов.

Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения $F(x/\theta_1, \theta_2, \ldots, \theta_m)$, зависящего от m неизвестных параметров $\theta_1, \theta_2, \ldots, \theta_m$. Ранее говорилось о том, что в силу близости теоретической и эмпирической функций распределения должны быть близки друг к другу соответствующие теоретические и выборочные моменты. Пусть

$$b_l = M\xi^l = \int\limits_{-\infty}^{+\infty} x^l dF(x/\theta_1, \theta_2, \ldots, \theta_m) = b_l(\theta_1, \theta_2, \ldots, \theta_m) \text{ —}$$

начальный теоретический момент l-го порядка, являющийся функцией от неизвестных параметров. Тогда он близок к соответствующему выборочному моменту $\overline{x^l} = \frac{1}{n}\sum\limits_{k=1}^{n} x_k^l$. Для первых m начальных моментов вместо приближенных запишем точные равенства, в которые вместо истинных значений параметров подставим их оценки:

$$\begin{gathered} b_1(\widehat{\theta_1}, \widehat{\theta_2}, \ldots, \widehat{\theta_m}) = \overline{x^1}, \\ b_2(\widehat{\theta_1}, \widehat{\theta_2}, \ldots, \widehat{\theta_m}) = \overline{x^2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ b_m(\widehat{\theta_1}, \widehat{\theta_2}, \ldots, \widehat{\theta_m}) = \overline{x^m}. \end{gathered} \tag{3.1}$$

Суть метода моментов состоит в нахождении оценок неизвестных параметров как решения полученной системы m уравнений с m неизвестными.

Замечание 1. Вместо начальных моментов можно в некоторых из уравнений использовать центральные моменты. При этом в каждом из уравнений слева и справа должны приравниваться одноименные моменты.

Пример 1. Пусть $x_1, x_2, \dots, x_n$ — выборка из нормального распределения с параметрами (a, σ^2). Приравнивая первые начальные и вторые центральные моменты, получим систему уравнений

$$\hat{a} = \overline{x},$$

$$\widehat{\sigma^2} = \widetilde{S^2},$$

которая уже разрешена относительно неизвестных оценок. Получили известные нам оценки $\overline{x}$ и $\widetilde{S^2}$ неизвестных параметров a и σ^2 соответственно. Если бы мы приравняли в обоих уравнениях начальные моменты, то второе уравнение системы пришлось бы решать. Результат, разумеется, получился бы таким же.

Пример 2. Рассмотрим схему из n независимых испытаний Бернулли с вероятностью успеха p. Результат x_k k-го испытания будем регистрировать как число успехов в этом испытании, т.е.

$$x_k = \begin{cases} 1, & \text{если в } k\text{-м испытании произошел успех,} \\ 0, & \text{если в } k\text{-м испытании произошла неудача.} \end{cases}$$

В данном случае выборка $x_1, x_2, \dots, x_n$ состоит только из нулей и единиц. Теоретическая случайная величина ξ — число успехов в отдельном испытании, ее распределение зависит от единственного параметра p. Приравнивая первые начальные моменты, получаем систему из одного уравнения для определения оценки $\widehat{p}_n$ неизвестного параметра p:

$$\widehat{p}_n = \overline{x} = \frac{1}{n}\sum_{k=1}^{n} x_k = \frac{\mu}{n},$$

где μ — число успехов в n испытаниях. Таким образом, оценкой вероятности успеха p является относительная частота успеха $\frac{\mu}{n}$.

Теорема 3.1. *Если теоретическая случайная величина имеет конечный момент m-го порядка, а система уравнений (3.1) имеет единственное решение*

$$\widehat{\theta}_k = b_k^{-1}(\overline{x^1}, \overline{x^2}, \dots, \overline{x^m}), \qquad k = \overline{1, m},$$

задаваемое непрерывными обратными функциями b_k^{-1}, то это решение дает состоятельные оценки параметров θ_k, $\quad k = \overline{1, m}$.

Д о к а з а т е л ь с т в о. Очевидно, что для $k = 1, 2, \dots, m$ случайные величины $x_1^k, x_2^k, \dots, x_n^k$ независимы, одинаково распределены и имеют конечное математическое ожидание $b_k(\theta_1, \theta_2, \dots, \theta_m)$. По закону больших чисел в форме Хинчина

$$\overline{x^k} = \frac{1}{n}\sum_{i=1}^{n} x_i^k \xrightarrow{P} b_k, \qquad k = \overline{1, m}.$$

По теореме 1.15.2 (обобщенной очевидным образом на функции n переменных)

$$\widehat{\theta}_k = b_k^{-1}(\overline{x^1}, \overline{x^2}, \dots, \overline{x^m}) \xrightarrow{P}$$

$$\xrightarrow{P} b_k^{-1}(b_1(\theta_1, \theta_2, \dots, \theta_m), \dots, b_m(\theta_1, \theta_2, \dots, \theta_m)) =$$

$$= \theta_k, \qquad k = \overline{1, m}. \qquad \square$$

Отметим, что метод моментов не учитывает конкретный вид теоретической функции распределения. Поэтому кроме состоятельности (при выполнении условий теоремы 1) он не гарантирует других хороших свойств оценок. При его применении полученные оценки необходимо дополнительно проверять, например, на несмещенность, эффективность и т.п. Поэтому рассмотрим еще один метод, который гарантирует хорошие асимптотические свойства оценок и позволяет пользоваться ими для выборок большого объема без дополнительной проверки их качества.

2. Метод максимального правдоподобия.

Пусть измеряется физическая постоянная величина a, причем результаты измерений

$$x_1 = a + \delta_1,$$

$$\dots\dots\dots\dots$$

$$x_n = a + \delta_n$$

лишены систематической ошибки и имеют нормальное распределение с параметрами (a, σ^2). Тогда $x_1, x_2, \dots, x_n$ — выборка из нормального распределения $N(a, \sigma^2)$. Для рассматриваемого случая еще Гаусс предложил выбирать оценку $\hat{a}$ параметра a таким образом, чтобы сумма квадратов ошибок всех измерений $\sum\limits_{i=1}^{n}(x_i - \hat{a})^2$ была минимальной. Пользуясь необходимым условием экстремума и приравнивая к нулю производную по $\hat{a}$ от этой суммы, получим, что $\hat{a} = \overline{x}$. Легко видеть, что это — точка минимума. Фишер усмотрел, что суть этого метода такова. Плотность совместного распределения случайных величин $x_1, x_2, \ldots, x_n$

$$L(x_1, x_2, \ldots, x_n/a, \sigma^2) = p(x_1)p(x_2)\ldots p(x_n) =$$

$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_1-a)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_2-a)^2}{2\sigma^2}} \ldots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_n-a)^2}{2\sigma^2}} =$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-a)^2}.$$

Она называется функцией правдоподобия. Очевидно, сумма квадратов ошибок $\sum\limits_{i=1}^{n}(x_i - \hat{a})^2$ минимальна в том и только том случае, когда функция правдоподобия $L(x_1, x_2, \ldots, x_n/a, \sigma^2)$ максимальна. Теперь можно забыть о нормальности теоретического распределения и, следуя Фишеру, искать оценки неизвестных параметров из того условия, чтобы функция правдоподобия принимала максимальное значение. Перейдем к более формальному описанию метода.

Пусть $x_1, x_2, \ldots, x_n$ — выборка из абсолютно непрерывного распределения с плотностью вероятности $p(x/\theta_1, \theta_2, \ldots, \theta_m)$, зависящей от m неизвестных параметров $\theta_1, \theta_2, \ldots, \theta_m$.

Определение 3.2. Совместная плотность распределения случайных величин $x_1, x_2, \ldots, x_n$

$$L(x_1, x_2, \ldots, x_n/\theta_1, \theta_2, \ldots, \theta_m) = L(\theta_1, \theta_2, \ldots, \theta_m) =$$

$$= p(x_1/\theta_1, \theta_2, \ldots, \theta_m)p(x_2/\theta_1, \theta_2, \ldots, \theta_m)\ldots$$

$$\ldots p(x_n/\theta_1, \theta_2, \ldots, \theta_m),$$

рассматриваемая как функция от $\theta_1, \theta_2, \ldots, \theta_m$, в которой x_i — случайные величины — элементы абстрактной выборки, называется **функцией правдоподобия**.

Введем понятие функции правдоподобия для выборки из дискретного распределения. Пусть $x_1, x_2, \ldots, x_n$ — выборка из дискретного распределения теоретической случайной величины ξ, принимающей значения u_k с вероятностями $P\{\xi = u_k\} = p(u_k/\theta_1, \theta_2, \ldots, \theta_m) \quad , k = 1, 2, \ldots,$ зависящими от m неизвестных параметров $\theta_1, \theta_2, \ldots, \theta_m$. Рассмотрим совместную вероятность принятия случайными величинами $\xi_1, \xi_2, \ldots, \xi_n$ — элементами абстрактной выборки соответственно значений $x_1, x_2, \ldots, x_n$:

$$P\{\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_n = x_n\} =$$

$$= P\{\xi_1 = x_1\}P\{\xi_2 = x_2\} \ldots P\{\xi_n = x_n\}.$$

Определение 3.3. Совместная вероятность значений случайных величин $x_1, x_2, \ldots, x_n$

$$L(x_1, x_2, \ldots, x_n/\theta_1, \theta_2, \ldots, \theta_m) = L(\theta_1, \theta_2, \ldots, \theta_m) =$$

$$= p(x_1/\theta_1, \theta_2, \ldots, \theta_m) p(x_2/\theta_1, \theta_2, \ldots, \theta_m) \ldots$$

$$\ldots p(x_n/\theta_1, \theta_2, \ldots, \theta_m),$$

рассматриваемая как функция от $\theta_1, \theta_2, \ldots, \theta_m$, в которой x_i — случайные величины — элементы абстрактной выборки, называется **функцией правдоподобия**.

Пример 3. Пусть $x_1, x_2, \dots, x_n$ — выборка из равномерного распределения на отрезке $[a, b]$. Тогда

$$p(x/a,b) = \begin{cases} \frac{1}{b-a}, & \text{если } a \le x \le b, \\ 0, & \text{если } x \notin [a,b]. \end{cases}$$

и функция правдоподобия принимает вид

$$L(x_1, x_2, \dots, x_n/a, b) = p(x_1/a,b)p(x_2/a,b)\dots p(x_n/a,b) =$$

$$= \begin{cases} \frac{1}{(b-a)^n}, & \text{если } x_{(1)} \ge a,\ x_{(n)} \le b, \\ 0 & \text{в иных случаях.} \end{cases}$$

Пример 4. Пусть $x_1, x_2, \dots, x_n$ — выборка из распределения Пуассона с параметром λ. Тогда

$$p(x/\lambda) = P\{\xi = x\} = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, \dots$$

и функция правдоподобия имеет форму

$$L(\lambda) = p(x_1/\lambda)p(x_2/\lambda)\dots p(x_n/\lambda) = \frac{\lambda^{x_1+x_2+\dots+x_n}}{x_1!x_2!\dots x_n!} e^{-n\lambda}.$$

Метод максимального правдоподобия состоит в том, что в качестве оценок выбираются такие значения $\theta_1, \theta_2, \dots, \theta_m$, при которых функция правдоподобия принимает максимальное значение. В силу необходимого условия экстремума надо приравнять нулю частные производные функции правдоподобия по каждому из параметров. Но находить частные производные от произведения n сомножителей не очень удобно. Лучше поступить следующим образом. Так как функция $\mathfrak{L} = \ln L$ возрастающая, то точка максимума функции правдоподобия

$L(\theta_1, \theta_2, \dots, \theta_m)$ совпадает с точкой максимума так называемой **логарифмической функции правдоподобия** $\mathfrak{L}(\theta_1, \theta_2, \dots, \theta_m) = \ln L(\theta_1, \theta_2, \dots, \theta_m)$. Приравняем нулю ее частные производные:

$$\frac{\partial \mathfrak{L}(\theta_1, \theta_2, \dots, \theta_m)}{\partial \theta_1} = \sum_{k=1}^{n} \frac{\partial \ln p(x_k/\theta_1, \theta_2, \dots \theta_n)}{\partial \theta_1} = 0,$$

$$\frac{\partial \mathfrak{L}(\theta_1, \theta_2, \dots, \theta_m)}{\partial \theta_2} = \sum_{k=1}^{n} \frac{\partial \ln p(x_k/\theta_1, \theta_2, \dots \theta_n)}{\partial \theta_2} = 0,$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\frac{\partial \mathfrak{L}(\theta_1, \theta_2, \dots, \theta_m)}{\partial \theta_m} = \sum_{k=1}^{n} \frac{\partial \ln p(x_k/\theta_1, \theta_2, \dots \theta_n)}{\partial \theta_m} = 0.$$

Из этой системы мы и найдем оценки неизвестных параметров $\widehat{\theta_1}, \widehat{\theta_2}, \dots, \widehat{\theta_m}$ (Здесь предполагается, что случай хороший и нет концевых экстремумов).

Пример 5. Рассмотрим выборку чисел успехов $x_1, x_2, \dots, x_n$ в схеме Бернулли из примера 2. Для теоретической случайной величины ξ — числа успехов в отдельном испытании —

$$P\{\xi = x\} = p^x (1-p)^{1-x}, \quad x = 1, 0.$$

Значит, $p(x_i/p) = p^{x_i}(1-p)^{1-x_i}$ и функция правдоподобия принимает вид

$$L(p) = L(x_1, x_2, \dots, x_n/p) = p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \dots$$

$$\dots p^{x_n}(1-p)^{1-x_n} = p^{\sum\limits_{i=1}^{n} x_i}(1-p)^{n - \sum\limits_{i=1}^{n} x_i} = p^{\mu}(1-p)^{n-\mu},$$

где μ — число успехов в n испытаниях. Логарифмическая функция правдоподобия равна

$$\mathfrak{L}(p) = \ln L(p) = \mu \ln p + (n - \mu) \ln (1 - p).$$

Приравнивая производную по p нулю и заменяя p на оценку $\hat{p}$, получим

$$\frac{\mu}{\hat{p}} - \frac{n - \mu}{1 - \hat{p}} = 0,$$

откуда $\hat{p} = \frac{\mu}{n}$, т.е. в качестве оценки вероятности успеха надо взять его относительную частоту. Очевидно, это несмещенная и, в силу закона больших чисел в форме Бернулли, состоятельная оценка. Позже мы убедимся, что она эффективна в классе несмещенных и так называемых регулярных оценок.

Пример 6. Из вида функции правдоподобия примера 3 заключаем, что оценками максимального правдоподобия для a и b в случае выборки из равномерного распределения на отрезке $[a, b]$ будут соответственно $\hat{a} = x_{(1)}$ и $\hat{b} = x_{(n)}$.

Пример 7. Для выборки из распределения Пуассона примера 4 логарифмическая функция правдоподобия имеет форму

$$\mathfrak{L}(\lambda) = (x_1 + x_2 + \ldots + x_n) \ln \lambda - \ln (x_1! x_2! \ldots x_n!) - n\lambda.$$

Приравнивая производную по λ нулю и заменяя λ на оценку $\hat{\lambda}$, получим

$$\frac{x_1 + x_2 + \ldots + x_n}{\hat{\lambda}} - n = 0,$$

откуда $\hat{\lambda} = \overline{x}$.

Рассмотрим случай, когда плотность теоретического распределения $p(x/\theta)$ зависит от одного параметра θ. Установим асимптотические свойства оценок максимального правдоподобия. Предположим выполненными следующие условия.

1. Параметр θ и его истинное значение θ_0 лежат внутри интервала (θ_1, θ_2), в котором существуют производные $\frac{\partial p}{\partial \theta}$, $\frac{\partial^2 p}{\partial \theta^2}$, $\frac{\partial^3 p}{\partial \theta^3}$.

2. Допустимо двукратное дифференцирование под знаком интеграла $\int\limits_{-\infty}^{+\infty} p(x/\theta)dx = 1$.

3. $i(\theta_0) = M\left(\frac{\partial \ln p}{\partial \theta}\right)^2\Big|_{\theta=\theta_0} > 0$, $\left|\frac{\partial^3 p}{\partial \theta^3}\right| \leq H(x)$, $MH(\xi) \leq M$, где M не зависит от θ.

Теорема 3.4. *Если выполнены условия 1,2,3, то уравнение правдоподобия*

$$\frac{\partial L(x/\theta)}{\partial \theta} = 0$$

имеет решение $\hat{\theta}$, являющееся состоятельной оценкой θ_0. Эта оценка максимального правдоподобия асимптотически нормальна с параметрами $(\theta_0, \frac{1}{n \cdot i(\theta_0)})$ и асимптотически эффективна.

Доказательство теоремы 3.4 выносится в приложение.

Замечание 1. В дальнейшем мы увидим, что $i(\theta_0)$ может трактоваться как информация о параметре θ_0 в отдельном наблюдении (т.е. в одном выборочном значении), а величина $I(\theta_0) = ni(\theta_0)$ — как информация о параметре θ_0 в выборке. Асимптотически нормальная оценка называется асимптотически эффективной, если ее дисперсия равна $\frac{1}{I(\theta_0)}$.

§ 4. Информация в наблюдении и выборке. Неравенство Рао-Крамера

Пусть $x_1, x_2, \dots, x_n$ — выборка из абсолютно непрерывного распределения с плотностью вероятности $p(x/\theta)$, зависящей от неизвестного параметра θ. Запишем функцию и логарифмическую функцию правдоподобия

$$L(\theta) = L(x_1, x_2, \dots, x_n/\theta) = p(x_1/\theta)p(x_2/\theta)\dots p(x_n/\theta)$$

и

$$\mathfrak{L}(\theta) = \mathfrak{L}(x_1, x_2, \dots, x_n/\theta) = \ln L(x_1, x_2, \dots, x_n/\theta).$$

Определение 4.1. Величина

$$I(\theta) = M\left|\frac{\partial \mathfrak{L}(x_1, x_2, \dots, x_n/\theta)}{\partial \theta}\right|^2$$

называется **информацией по Фишеру о параметре θ в выборке** $x_1, x_2, \dots, x_n$. Если выборка состоит из одного наблюдения x_1, то информация в такой выборке

$$i(\theta) = M\left|\frac{\partial \ln p(x_1/\theta)}{\partial \theta}\right|^2$$

называется **информацией по Фишеру о параметре θ в наблюдении** x_1.

§ 4. Информация в наблюдении и выборке. Неравенство Рао-Крамера

Предположим, что законно дифференцирование по параметру θ под знаком интеграла $\int\limits_{\mathbb{R}} p(x/\theta)dx$, так что

$$\int\limits_{\mathbb{R}} \frac{\partial p(x/\theta)}{\partial\theta}dx = \frac{\partial \int\limits_{\mathbb{R}} p(x/\theta)dx}{\partial\theta} = \frac{\partial 1}{\partial\theta} = 0.$$

Так как случайные величины $x_1, x_2, \ldots, x_n$ в абстрактной выборке одинаково распределены, то информации в наблюдениях $x_1, x_2, \ldots, x_n$ одинаковы и равны $i(\theta)$. Поэтому

$$I(\theta) = M\Big|\frac{\partial \ln\left[p(x_1/\theta)p(x_2/\theta)\ldots p(x_n/\theta)\right]}{\partial\theta}\Big|^2 =$$

$$= M\Big[\sum_{i=1}^{n}\frac{\partial \ln p(x_i/\theta)}{\partial\theta}\Big]^2 = \sum_{i=1}^{n} M\Big|\frac{\partial \ln p(x_i/\theta)}{\partial\theta}\Big|^2 +$$

$$+\sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} M\Big[\frac{\partial \ln p(x_i/\theta)}{\partial\theta}\frac{\partial \ln p(x_j/\theta)}{\partial\theta}\Big] = \sum_{i=1}^{n} i(\theta) +$$

$$+\sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} \iint\limits_{\mathbb{R}^2} \frac{\partial \ln p(x_i/\theta)}{\partial\theta}\frac{\partial \ln p(x_j/\theta)}{\partial\theta} p(x_i/\theta)p(x_j/\theta)\cdot$$

$$\cdot dx_i dx_j = ni(\theta) +$$

$$+\sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} \int\limits_{\mathbb{R}} \frac{\partial p(x_i/\theta)}{\partial\theta}dx_i \int\limits_{\mathbb{R}} \frac{\partial p(x_j/\theta)}{\partial\theta}dx_j = ni(\theta).$$

Тот факт, что информации в наблюдениях сложились, обусловлен независимостью этих наблюдений. Итак, $I(\theta) = ni(\theta)$.

Теорема 4.2. *Пусть*

1) область значений, где $L(x_1, x_2, \dots, x_n/\theta) = 0$, *не зависит от параметра* θ*;*

2) допустимо дифференцирование под знаком следующих двух интегралов по параметру θ*:*

$$\int\limits_{\mathbb{R}^n} \cdots \int \widehat{\theta}(x_1, x_2, \dots, x_n) L(x_1, x_2, \dots, x_n/\theta) dx_1 dx_2 \dots dx_n$$

$$\text{и} \qquad \int\limits_{\mathbb{R}^n} \cdots \int L(x_1, x_2, \dots, x_n/\theta) dx_1 dx_2 \dots dx_n,$$

где $\widehat{\theta_n} = \widehat{\theta}(x_1, x_2, \dots, x_n)$ *— некоторая оценка параметра* θ*;*

3) $I(\theta) \neq 0$.

Тогда выполняется **неравенство Рао-Крамера***:*

$$M(\widehat{\theta_n} - \theta)^2 \geq \frac{|1 + b'(\theta)|^2}{I(\theta)},$$

где $b(\theta) = M\widehat{\theta_n} - \theta$ *— смещение оценки* $\widehat{\theta}_n$.

Д о к а з а т е л ь с т в о проведем для случая, когда $L(x_1, x_2, \dots, x_n/\theta) > 0$, поскольку во всех далее рассматриваемых примерах это выполняется. Модификация доказательства на случай условия 1) теоремы не представляет большой сложности. По условию теоремы

$$M\widehat{\theta_n} = \theta + b(\theta) =$$

$$= \int\limits_{\mathbb{R}^n} \cdots \int \widehat{\theta}(x_1, x_2, \dots, x_n) L(x_1, x_2, \dots, x_n/\theta) dx_1 dx_2 \dots dx_n.$$

По свойству плотности распределения вероятностей случайного вектора

$$1 = \int\limits_{\mathbb{R}^n}\cdots\int L(x_1, x_2, \dots, x_n/\theta)dx_1dx_2\dots dx_n.$$

Дифференцируя последние два равенства по θ и используя условие 2), получим

$$1 + b'(\theta) = \int\limits_{\mathbb{R}^n}\cdots\int \widehat{\theta}(x_1, x_2, \dots, x_n)\frac{\partial L(x_1, x_2, \dots, x_n/\theta)}{\partial\theta}\cdot$$

$$\cdot dx_1dx_2\dots dx_n, \tag{4.1}$$

$$0 = \int\limits_{\mathbb{R}^n}\cdots\int \frac{\partial L(x_1, x_2, \dots, x_n/\theta)}{\partial\theta}dx_1dx_2\dots dx_n. \tag{4.2}$$

Вычитая из (4.1) равенство (4.2), умноженное на θ, и возводя обе части полученного равенства в квадрат, приходим к соотношению

$$|1 + b'(\theta)|^2 = \Big|\int\limits_{\mathbb{R}^n}\cdots\int (\widehat{\theta}(x_1, x_2, \dots, x_n) - \theta)\cdot$$

$$\cdot\frac{\partial L(x_1, x_2, \dots, x_n/\theta)}{\partial\theta}dx_1dx_2\dots dx_n\Big|^2.$$

Так как $\frac{\partial L}{\partial\theta} = L\frac{\partial\mathfrak{L}}{\partial\theta}$, то применяя неравенство Коши-Буняковского-Шварца, получим

$$|1 + b'(\theta)|^2 = \Big|\int\limits_{\mathbb{R}^n}\cdots\int (\widehat{\theta}(x_1, x_2, \dots, x_n) - \theta)\cdot$$

$$\cdot\sqrt{L(x_1, x_2, \dots, x_n/\theta)}\cdot\frac{\partial\mathfrak{L}(x_1, x_2, \dots, x_n/\theta)}{\partial\theta}\cdot$$

$$\cdot\sqrt{L(x_1, x_2, \dots, x_n/\theta)}dx_1dx_2\dots dx_n\Big|^2 \leq$$

$$\leq \int\limits_{\mathbb{R}^n}\cdots\int(\widehat{\theta}(x_1, x_2, \dots, x_n) - \theta)^2 L(x_1, x_2, \dots, x_n/\theta)dx_1\cdot$$

$$\cdot dx_2\dots dx_n\cdot\int\limits_{\mathbb{R}^n}\cdots\int\Big|\frac{\partial\mathfrak{L}(x_1, x_2, \dots, x_n/\theta)}{\partial\theta}\Big|^2\cdot$$

$$\cdot L(x_1, x_2, \dots, x_n/\theta)dx_1dx_2\dots dx_n =$$

$$= M(\widehat{\theta_n}-\theta)^2\cdot M\Big|\frac{\partial\mathfrak{L}(x_1, x_2, \dots, x_n/\theta)}{\partial\theta}\Big|^2 = M(\widehat{\theta_n}-\theta)^2\cdot I(\theta).$$

Из последнего неравенства следует неравенство Рао-Крамера. □

В частности, если оценка несмещенная, то смещение $b(\theta) = M\widehat{\theta_n} - \theta = 0$ и $M(\widehat{\theta_n} - \theta)^2 = D\widehat{\theta_n}$. Поэтому для несмещенных оценок неравенство Рао-Крамера принимает следующий вид:

$$D\widehat{\theta_n} \geq \frac{1}{I(\theta)} = \frac{1}{ni(\theta)}.$$

Отсюда видно, что если дисперсия оценки маленькая (оценка информативная), то информация в выборке о параметре большая, и наоборот, если информация о параметре маленькая, то дисперсия оценки большая. Этот факт объясняет естественность названия "информация" для $I(\theta)$.

Если для оценки выполняются условия теоремы 1, при которых справедливо неравенство Рао-Крамера, то такая оценка называется **регулярной**.

Пример 1. Пусть $x_1, x_2, \ldots, x_n$ — выборка из нормального распределения с параметрами (a, σ^2). Неизвестным параметром будем считать a. Так как

$$p(x/a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad \ln p(x/a) = -\ln\sqrt{2\pi}\sigma - \frac{(x-a)^2}{2\sigma^2},$$

$$\frac{\partial \ln p(x/a)}{\partial a} = \frac{x-a}{\sigma^2},$$

то информация о параметре a в наблюдении

$$i(a) = M\left|\frac{\partial \ln p(x_1/a)}{\partial a}\right|^2 = M\left|\frac{x_1 - a}{\sigma^2}\right|^2 = \frac{1}{\sigma^4} M(x_1 - a)^2 =$$

$$= \frac{1}{\sigma^4} D x_1 = \frac{1}{\sigma^4}\sigma^2 = \frac{1}{\sigma^2}.$$

В качестве оценки теоретической средней a, как и ранее, рассмотрим выборочную среднюю, которая является несмещенной и состоятельной оценкой a. Найдем ее дисперсию. Так как случайные величины $x_1, x_2, \ldots, x_n$ независимы, то по свойствам дисперсии

$$D\overline{x} = \frac{1}{n^2}\sum_{i=1}^{n} D x_i = \frac{\sigma^2}{n}.$$

В силу неравенства Рао-Крамера для любой несмещенной регулярной оценки $\widehat{a_n}$ параметра a

$$D\widehat{a_n} \geq \frac{1}{I(a)} = \frac{1}{n i(a)} = \frac{\sigma^2}{n} = D\overline{x}.$$

Итак, среди всех несмещенных регулярных оценок параметра a оценка $\overline{x}$ имеет наименьшую дисперсию. Таким образом, она эффективна в классе несмещенных регулярных оценок.

Пример 2. Пусть $x_1, x_2, \dots, x_n$ — выборка из распределения Пуассона с параметром λ. Для теоретической случайной величины

$$p(x/\lambda) = P\{\xi = x\} = \frac{\lambda^x}{x!}e^{-\lambda}, \quad x = 0, 1, \dots .$$

Поэтому

$$\ln p(x/\lambda) = x\ln\lambda - \ln x! - \lambda, \quad \frac{\partial \ln p(x/\lambda)}{\partial\lambda} = \frac{x}{\lambda} - 1.$$

Следовательно, информация о параметре λ в наблюдении

$$i(\lambda) = M\left|\frac{\partial \ln p(x_1/\lambda)}{\partial\lambda}\right|^2 = M\left|\frac{x_1}{\lambda} - 1\right|^2 = \frac{1}{\lambda^2}M(x_1 - \lambda)^2 =$$

$$= \frac{1}{\lambda^2}Dx_1 = \frac{1}{\lambda^2}\lambda = \frac{1}{\lambda}.$$

Так как $M\xi = \lambda$, то в качестве оценки λ снова возьмем выборочную среднюю $\overline{x}$. Ее дисперсия

$$D\overline{x} = \frac{1}{n^2}\sum_{i=1}^{n} Dx_i = \frac{\lambda}{n}.$$

Если $\widehat{\lambda}_n$ — любая несмещенная регулярная оценка λ, то в силу неравенства Рао-Крамера

$$D\widehat{\lambda}_n \geq \frac{1}{ni(\lambda)} = \frac{\lambda}{n} = D\overline{x}.$$

Поэтому $\overline{x}$ — эффективная оценка параметра λ в классе несмещенных регулярных оценок.

Замечание 1. Теорема 4.2 сформулирована и доказана для абсолютно непрерывного случая, а в примере 2 мы применили ее к дискретному случаю. На самом деле применен следующий аналог этой теоремы для выборок из дискретного распределения.

Теорема 4.3. *Пусть*

1) область значений, где $L(x_1, x_2, \dots, x_n/\theta) = 0$*, не зависит от параметра* θ*;*

2) допустимо дифференцирование под знаком следующих двух сумм по параметру θ*:*

$$\sum_{x_1}\sum_{x_2}\dots\sum_{x_n}\widehat{\theta}(x_1, x_2, \dots, x_n)L(x_1, x_2, \dots, x_n/\theta) \qquad \text{и}$$

$$\sum_{x_1}\sum_{x_2}\dots\sum_{x_n}L(x_1, x_2, \dots, x_n/\theta),$$

где $\widehat{\theta_n} = \widehat{\theta}(x_1, x_2, \dots, x_n)$ *— некоторая оценка параметра* θ*;*

3) $I(\theta) \neq 0$.

Тогда выполняется **неравенство Рао-Крамера***:*

$$M(\widehat{\theta_n} - \theta)^2 \geq \frac{|1 + b'(\theta)|^2}{I(\theta)},$$

где $b(\theta) = M\widehat{\theta_n} - \theta$ *— смещение оценки* $\widehat{\theta_n}$*.*

Для выборок из дискретного распределения оценки, удовлетворяющие условиям этой теоремы, также называются **регулярными**. Доказательство теоремы 4.3 почти дословно повторяет доказательство теоремы 4.2, только вместо неравенства Коши-Буняковского-Шварца для интегралов надо применять одноименное неравенство для сумм.

Задача 1. Докажите теорему 4.3.

Пример 3. Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения Бернулли с параметром p (т.е. ξ — число успехов в отдельном испытании в схеме n независимых испытаний Бернулли с вероятностью успеха p). Ранее была найдена логарифмическая функция правдоподобия

$$\mathfrak{L}(p) = \mu \ln p + (n - \mu) \ln (1 - p).$$

Поэтому информация по Фишеру в выборке о параметре p равна

$$I(p) = M\left|\frac{\partial \mathfrak{L}(p)}{\partial p}\right|^2 = M\left|\frac{\mu}{p} - \frac{n-\mu}{1-p}\right|^2 =$$

$$= \frac{1}{p^2(1-p)^2} M(\mu - np)^2 = \frac{1}{p^2(1-p)^2} D\mu = \frac{n}{p(1-p)}.$$

В качестве оценки вероятности успеха p как и ранее возьмем относительную частоту $\frac{\mu}{n}$. Ранее установлено, что она несмещенная и состоятельная. Ее дисперсия

$$D\frac{\mu}{n} = \frac{1}{n} D\mu = \frac{pq}{n}, \quad \text{где} \quad q - 1 - p.$$

Если $\widehat{p}$ — любая несмещенная регулярная оценка p, то в силу неравенства Рао-Крамера

$$D\widehat{p} \geq \frac{1}{I(p)} = \frac{pq}{n} = D\frac{\mu}{n}.$$

Таким образом, $\frac{\mu}{n}$ — эффективная оценка p в классе несмещенных регулярных оценок.

§ 5. Многомерное нормальное распределение. Лемма Фишера

В этом разделе будем использовать матричные обозначения, в частности, буква T будет обозначать операцию транспонирования матрицы, вектора будут рассматриваться как матрицы. Так, если $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, то $\mathbf{x}$ — вектор-столбец, а $\mathbf{x}^T = (x_1, x_2, \dots, x_n)$ — вектор-строка. Если $A = (a_{ij})$ — квадратная матрица порядка n, то

$$\mathbf{x}^T A \mathbf{x} = \sum_{i,j=1}^{n} a_{ij} x_i x_j \quad —$$

квадратичная форма, матрица которой есть A.

Пусть $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ — случайный вектор; тогда вектор $\mathbf{a}_\xi = M\xi = (M\xi_1, M\xi_2, \dots, M\xi_n)^T$ называется **математическим ожиданием вектора** ξ, а матрица $\mathbf{C}_\xi = (cov(\xi_i, \xi_j))$ — **ковариационной матрицей** или **матрицей ковариаций вектора** ξ. Очевидно, $\mathbf{C}_\xi$ — симметрическая матрица.

Лемма 5.1. *Если* $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ *— неслучайный вектор, а* $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$ *— случайный вектор, то*

$$D(\mathbf{x}^T \xi) = \mathbf{x}^T \mathbf{C}_\xi \mathbf{x}.$$

Д о к а з а т е л ь с т в о. Согласно свойствам дисперсии

$$D(\mathbf{x}^T \xi) = D\Big(\sum_{i=1}^{n} x_i \xi_i\Big) = \sum_{i,j=1}^{n} x_i x_j cov(\xi_i, \xi_j) = \mathbf{x}^T \mathbf{C}_\xi \mathbf{x}.$$

□

Так как $\mathbf{x}^T\mathbf{C}_\xi\mathbf{x} = D(\mathbf{x}^T\xi) \geq 0$ для любых $x_1, x_2, \dots, x_n$, то из леммы 5.1 следует, что $\mathbf{C}_\xi$ — неотрицательно определенная матрица. Итак, ковариационная матрица любого случайного вектора симметрическая и неотрицательно определенная.

Определение 5.2. Говорят, что **распределение случайного вектора ξ сосредоточено на поверхности S**, если

$$P\{\xi \in S\} = 1.$$

Лемма 5.3. *Для того, чтобы распределение случайного вектора ξ было сосредоточено в некоторой гиперплоскости, необходимо и достаточно, чтобы его ковариационная матрица $\mathbf{C}_\xi$ была вырождена.*

Д о к а з а т е л ь с т в о. Любая гиперплоскость имеет вид

$$L = \{\mathbf{x}: \ \mathbf{a}^T\mathbf{x} = q\},$$

где $\mathbf{a}$ — ненулевой вектор, ортогональный L.

Необходимость. Пусть распределение ξ сосредоточено в L. Тогда $P\{\mathbf{a}^T\xi = q\} = 1$, т.е. с вероятностью 1 случайная величина $\mathbf{a}^T\xi$ постоянна. По свойствам дисперсии и лемме 5.1 $D(\mathbf{a}^T\xi) = \mathbf{a}^T\mathbf{C}_\xi\mathbf{a} = 0$. Так как $\mathbf{C}_\xi$ симметрична и неотрицательно определена, то в некотором базисе она приводится к диагональному виду, причем на диагонали стоят числа $\lambda_1 \geq 0, \lambda_2 \geq 0, \dots, \lambda_n \geq 0$. В этом базисе квадратичная форма $\mathbf{a}^T\mathbf{C}_\xi\mathbf{a}$ приводится к виду ${\lambda_1 a_1'}^2 + \lambda_2 {a_2'}^2 + \ldots + \lambda_n {a_n'}^2 = 0$. Следовательно, существует $\lambda_i = 0$ (в противном случае, в силу того, что $\mathbf{a} \neq \mathbf{0}$, было бы $\lambda_1 {a_1'}^2 + \lambda_2 {a_2'}^2 + \ldots + \lambda_n {a_n'}^2 > 0$). Значит $|\mathbf{C}_\xi| = 0$ и $\mathbf{C}_\xi$ вырождена.

Достаточность. Пусть $\mathbf{C}_\xi$ вырождена; тогда в некотором базисе она диагональна с элементами $\lambda_1 \geq 0, \lambda_2 \geq 0, \dots, \lambda_n \geq 0$, причем хотя бы одно $\lambda_j = 0$. Выбирая в качестве $\mathbf{a}$ единичный вектор, j-я координата которого равна 1, получим $\mathbf{a}^T\mathbf{C}_\xi\mathbf{a} = \lambda_1 a_1^2 + \lambda_2 a_2^2 + \ldots + \lambda_n a_n^2 = 0$. По лемме 5.1 $D(\mathbf{a}^T\xi) = 0$, т.е. существует постоянная q такая, что $P\{\mathbf{a}^T\xi = q\} = 1$. Значит распределение ξ сосредоточено в гиперплоскости $L = \{\mathbf{x} : \ \mathbf{a}^T\xi = q\}$. □

Лемма 5.4. *Пусть $\eta = A\xi + \mathbf{a}$, где A — неслучайная матрица, ξ — случайный вектор, $\mathbf{a}$ — неслучайный вектор. Тогда*

$$\mathbf{C}_\eta = A\mathbf{C}_\xi A^T.$$

Д о к а з а т е л ь с т в о. В силу леммы 5.1

$$\mathbf{x}^T\mathbf{C}_\eta\mathbf{x} = D(\mathbf{x}^T\eta) = D(\mathbf{x}^T(A\xi + \mathbf{a})) = D(\mathbf{x}^T A\xi) =$$

$$= D((A^T\mathbf{x})^T\xi) = (A^T\mathbf{x})^T\mathbf{C}_\xi A^T\mathbf{x} = \mathbf{x}^T(A\mathbf{C}_\xi A)\mathbf{x},$$

т.е. две квадратичные формы совпадают при любых значениях $x_1, x_2, \dots, x_n$. Поэтому равны между собой матрицы этих квадратичных форм: $\mathbf{C}_\eta = A\mathbf{C}_\xi A^T$. □

Определение 5.5. **Стандартным нормальным распределением в** $\mathbb{R}^n$ называется распределение случайного вектора $\xi = (\xi_1, \xi_2, \dots, \xi_n)^T$, координаты которого — независимые случайные величины, имеющие стандартное нормальное распределение.

Очевидно, что

$$p_\xi(\mathbf{x}) = \prod_{i=1}^n p_{\xi_i}(x_i) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}\right) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum\limits_{i=1}^n x_i^2} =$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}.$$

Так как для стандартного нормального распределения $M\xi_i = 0$, $cov(\xi_i, \xi_j) = 0$ при $i \neq j$ и $cov(\xi_i, \xi_i) = D\xi_i = 1$, то $\mathbf{a}_\xi = M\xi = \mathbf{0}$, $\mathbf{C}_\xi = E$ — единичная матрица.

Определение 5.6. **Нормальным распределением в** $\mathbb{R}^n$ называется распределение случайного вектора

$$\eta = A\xi + \mathbf{a},$$

где A — неслучайная ненулевая матрица, $\mathbf{a}$ — неслучайный вектор, ξ — случайный вектор, имеющий стандартное нормальное распределение в $\mathbb{R}^n$.

Очевидно, $\mathbf{a}_\eta = M\eta = A \cdot M\xi + \mathbf{a} = A \cdot \mathbf{a}_\xi + \mathbf{a} = \mathbf{a}$, $\mathbf{C}_\eta = A\mathbf{C}_\xi A^T = AA^T$. Символически тот факт, что случайный вектор η имеет n-мерное нормальное распределение с математическим ожиданием $\mathbf{a}_\eta$ и ковариационной матрицей $\mathbf{C}_\eta$, будем обозначать как $\eta \sim N_n(\mathbf{a}_\eta,\ \mathbf{C}_\eta)$. В частности, если ξ имеет стандартное нормальное распределение в $\mathbb{R}^n$, то эта фраза может быть записана как $\xi \sim N_n(\mathbf{0},\ E)$.

Так как $|\mathbf{C}_\eta| = |A|^2$, то матрицы $\mathbf{C}_\eta$ и A вырождены или невырождены одновременно. Если матрица A вырождена, то по лемме 5.3 распределение вектора η сосредоточено в некоторой гиперплоскости и, следовательно, случайный вектор η не имеет плотности вероятности в $\mathbb{R}^n$.

Если же A невырождена, то $\mathbf{C}_\eta$ невырождена и имеет место

Теорема 5.7. *Если матрица A невырождена, то распределение случайного вектора $\eta = A\xi + \mathbf{a}$, где $\xi \sim N_n(\mathbf{0},\ E)$, абсолютно непрерывно, а соответствующая плотность распределения вероятностей задается формулой*

$$p_\eta(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n\sqrt{|\mathbf{C}_\eta|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{a})^T \mathbf{C}_\eta{}^{-1}(\mathbf{x}-\mathbf{a})}. \qquad (5.1)$$

Д о к а з а т е л ь с т в о. Соответствующее преобразованию $\eta = A\xi + \mathbf{a}$ случайных векторов преобразование переменных $\mathbf{y} = A\mathbf{x} + \mathbf{a}$ является гладким с отличным от нуля якобианом $|A|$. При этом обратное преобразование имеет вид $\mathbf{x} = A^{-1}(\mathbf{y} - \mathbf{a})$. По теореме 1.13.1 (формула (1.13.1)) случайный вектор η абсолютно непрерывен, а плотность его распределения вероятностей равна

$$p_\eta(\mathbf{y}) = p_\xi(A^{-1}(\mathbf{y}-\mathbf{a}))\cdot\frac{1}{|A|} =$$

$$= \frac{1}{(\sqrt{2\pi})^n|A|} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{a})^T(AA^T)^{-1}(\mathbf{y}-\mathbf{a})} =$$

$$= \frac{1}{(\sqrt{2\pi})^n\sqrt{|\mathbf{C}_\eta|}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{a})^T \mathbf{C}_\eta{}^{-1}(\mathbf{y}-\mathbf{a})}. \qquad \square$$

Рассмотрим частный случай двумерного случайного вектора $\eta = (\eta_1, \eta_2)^T$. Обозначим через $\sigma_1^2 = D\xi_1$, $\sigma_2^2 = D\xi_2$, $r = \frac{cov(\eta_1,\, \eta_2)}{\sigma_1\sigma_2}$ — коэффициент корреляции между η_1 и η_2. Легко видеть, что $|\mathbf{C}_\eta| = (1-r^2)\sigma_1^2\sigma_2^2$,

$$\mathbf{C}_\eta{}^{-1} = \frac{1}{1-r^2}\cdot\begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{r}{\sigma_1\sigma_2} \\ -\frac{r}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

Поэтому

$$p_{\eta_1,\eta_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\cdot$$

$$\cdot e^{-\frac{1}{2(1-r^2)}\left[\frac{(x_1-a_1)^2}{\sigma_1^2}-2r\cdot\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2}+\frac{(x_2-a_2)^2}{\sigma_2^2}\right]}. \qquad (5.2)$$

В пространстве Ox_1x_2z эта плотность имеет вид "холма" с вершиной в точке $(a_1, a_2, \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}})$, который лежит выше плоскости Ox_1x_2 и асимптотически спускается к ней при $x_1 \to \infty$, $x_1 \to \infty$. В сечении графика плотности плоскостями, параллельными оси Oz, получаются кривые, подобные одномерным плотностям нормального распределения, а в сечении плоскостями, перпендикулярными оси Oz, получаются элипсы (см. Рис.6)

$$\frac{(x_1-a_1)^2}{\sigma_1^2}-2r\cdot\frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2}+\frac{(x_2-a_2)^2}{\sigma_2^2}=C$$

(поскольку квадратичная форма, стоящая в левой части последнего равенства, положительно определена).

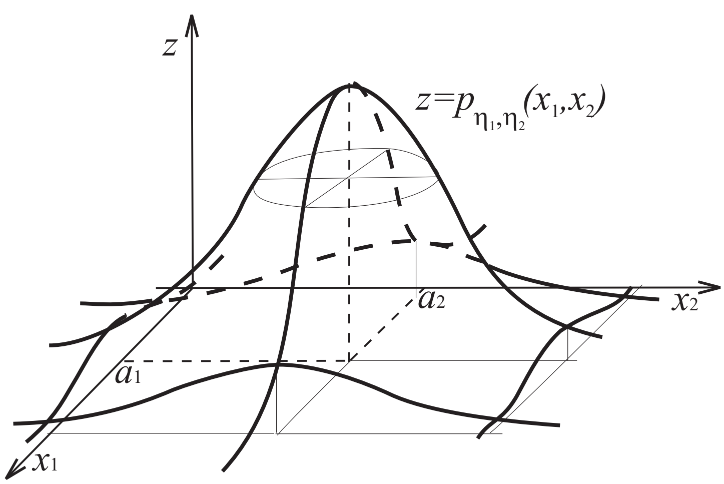

Рис.6

Следствие 5.8. *Если случайный вектор* $\eta = (\eta_1, \eta_2)^T$ *имеет нормальное распределение и его координаты* η_1 *и* η_2 *— некоррелированные случайные величины, то* η_1 *и* η_2 *независимы и каждая из них имеет нормальное распределение.*

Д о к а з а т е л ь с т в о. Если η_1 и η_2 — некоррелированные случайные величины, то коэффициент корреляции между ними $r = 0$. Из (5.2) следует, что

$$p_{\eta_1,\eta_2}(x_1, x_2) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-a_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-a_2)^2}{2\sigma_2^2}}.$$

В силу следствия 1.12.4 это означает, что η_1 и η_2 независимы и $\eta_1 \sim N(a_1, \sigma_1^2)$, $\quad \eta_2 \sim N(a_2, \sigma_2^2)$. $\square$

Заметим, что это утверждение очевидным образом обобщается на n-мерный случай.

Теорема 5.9. *Для того, чтобы* $\eta \sim N_n(\mathbf{a}, D)$*, где* $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T$, D *— диагональная матрица с элементами* $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ *на диагонали, необходимо и достаточно, чтобы случайные величины* $\eta_1, \eta_2, \ldots, \eta_n$ *были независимыми и* $\eta_i \sim N(a_i,\ \sigma_i^2)$.

Д о к а з а т е л ь с т в о. Достаточность очевидна. Для доказательства необходимости заметим, что с учетом вида матрицы D формула (5.1) примет вид

$$p_\eta(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n \sigma_1 \sigma_2 \ldots \sigma_n} e^{-\frac{1}{2}\sum\limits_{i=1}^{n} \frac{(x_i-a_i)^2}{\sigma_i^2}} =$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-a_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-a_2)^2}{2\sigma_2^2}} \ldots \frac{1}{\sqrt{2\pi}\sigma_n} e^{-\frac{(x_n-a_n)^2}{2\sigma_n^2}}.$$

Далее остается применить рассуждения, использованные при доказательстве следствия 5.8. □

Таким образом, если координаты n-мерного нормального вектора попарно некоррелированы, то они независимы и каждая из них имеет нормальное распределение.

Теорема 5.10. *Пусть* $\eta = (\eta_1, \eta_2, \ldots, \eta_n)^T \sim N_n(\mathbf{0},\ \sigma^2 E)$, *$C$ — ортогональная матрица. Тогда* $\zeta = C\eta \sim N_n(\mathbf{0},\ \sigma^2 E)$, *т.е. ортогональное преобразование сохраняет независимость координат нормального случайного вектора.*

Д о к а з а т е л ь с т в о. Соответствующее преобразованию $\zeta = C\eta$ случайных векторов преобразование переменных $\mathbf{y} = C\mathbf{x}$ — гладкое с отличным от нуля якобианом. Обратное преобразование есть $\mathbf{x} = C^{-1}\mathbf{y} = C^T\mathbf{y}$, причем $|C^T| = |C| = 1$. По теореме 1.13.1 (формула (1.13.1)) случайный вектор ζ абсолютно непрерывен, а плотность его распределения вероятностей равна

$$p_\zeta(\mathbf{y}) = p_\eta(C^T\mathbf{y}) \cdot 1 = \frac{1}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2\sigma^2}(C^T\mathbf{y})^T C^T\mathbf{y}} =$$

$$= \frac{1}{(\sqrt{2\pi})^n \sigma^n} e^{-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{y}}.$$

□

Напомним, что случайная величина $\chi^2_n = \xi_1^2 + \xi_2^2 + \ldots + \xi_n^2$ называется случайной величиной **хи-квадрат с n степенями свободы**, если случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ независимы, а каждая из них имеет стандартное нормальное распределение.

Лемма Фишера. *Пусть случайные величины* $\xi_1, \xi_2, \dots, \xi_n$ *независимы, каждая из них имеет нормальное распределение с параметрами* $(0, \sigma^2)$ *и для* $p < n$ *случайные величины* $\eta_1, \eta_2, \dots, \eta_p$ *являются линейными комбинациями от* $\xi_1, \xi_2, \dots, \xi_n$ *с матрицей коэффициентов, строки которой ортонормированы (т.е. ортогональны и длина каждой вектор-строки равна 1). Тогда квадратичная форма* $Q(\xi_1, \xi_2, \dots, \xi_n) = \sum\limits_{k=1}^{n} \xi_k^2 - \eta_1^2 - \eta_2^2 - \dots - \eta_p^2$ *не зависит от* $\eta_1, \eta_2, \dots, \eta_p$ *и* $\frac{Q}{\sigma^2} = \chi^2_{n-p}$.

Доказательство. Пусть

$$\eta_1 = c_{11}\xi_1 + c_{12}\xi_2 + \dots + c_{1n}\xi_n,$$

$$\eta_2 = c_{21}\xi_1 + c_{22}\xi_2 + \dots + c_{2n}\xi_n,$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\eta_p = c_{p1}\xi_1 + c_{p2}\xi_2 + \dots + c_{pn}\xi_n$$

— заданные в условии леммы линейные комбинации, причем c_{ij} удовлетворяют условию ортонормированности

$$\sum_{j=1}^{n} c_{ij}c_{kj} = \delta_{ik}, \quad i, k = \overline{1, p}.$$

Дополним матрицу коэффициентов $n - p$ строками так, чтобы полученная матрица C была ортогональной (любую систему ортонормированных векторов можно дополнить до ортонормированного базиса в n-мерном линейном пространстве), и введем случайные величины

$$\eta_{p+1} = c_{p+1,1}\xi_1 + c_{p+1,2}\xi_2 + \dots + c_{p+1,n}\xi_n,$$

$$\eta_{p+2} = c_{p+2,1}\xi_1 + c_{p+2,2}\xi_2 + \ldots + c_{p+2,n}\xi_n,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\eta_p = c_{n1}\xi_1 + c_{n2}\xi_2 + \ldots + c_{nn}\xi_n.$$

Тогда случайные вектора $\eta = (\eta_1, \eta_2, \ldots, \eta_n)^T$ и $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^T$ связаны ортогональным преобразованием $\eta = C\xi$. По теореме 5.10 случайные величины $\eta_1, \eta_2, \ldots, \eta_n$ независимы, а каждая из них имеет нормальное распределение с параметрами $(0, \sigma^2)$. Очевидно, $\xi = C^{-1}\eta = C^T\eta$. Поэтому

$$Q(\xi_1, \xi_2, \ldots, \xi_n) = \xi^T\xi - \eta_1^2 - \eta_2^2 - \ldots - \eta_p^2 = (C^T\eta)^T C^T\eta -$$

$$-\eta_1^2 - \eta_2^2 - \ldots - \eta_p^2 = \eta^T\eta - \eta_1^2 - \eta_2^2 - \ldots - \eta_p^2 =$$

$$= \eta_{p+1}^2 + \eta_{p+2}^2 + \ldots + \eta_n^2 =$$

$$= \sigma^2\left[\left(\frac{\eta_{p+1}}{\sigma}\right)^2 + \left(\frac{\eta_{p+2}}{\sigma}\right)^2 + \ldots + \left(\frac{\eta_n}{\sigma}\right)^2\right] = \sigma^2\chi^2_{n-p}.$$

Так как $Q(\xi_1, \xi_2, \ldots, \xi_n) = \eta_{p+1}^2 + \eta_{p+2}^2 + \ldots + \eta_n^2$ — борелевская функция от $\eta_{p+1}, \eta_{p+2}, \ldots, \eta_n$, которые не зависят от $\eta_1, \eta_2, \ldots, \eta_p$, то и Q не зависит от них. □

Следствие из леммы Фишера. *Пусть $x_1, x_2, \ldots, x_n$ — выборка из нормального распределения с параметрами (a, σ^2). Тогда случайные величины $\overline{x}$ и S^2 независимы и*

$$\frac{n\widetilde{S^2}}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} = \chi^2_{n-1}.$$

Д о к а з а т е л ь с т в о. Не ограничивая общности, можно считать $a = 0$ (иначе надо перейти к случайным величинам $x_1 - a, x_2 - a, \dots, x_n - a$, которые образуют выборку из $N(0, \sigma^2)$ с теми же $\widetilde{S^2}$ и S^2). Тогда

$$Q(x_1, x_2, \dots, x_n) = n\widetilde{S^2} = (n-1)S^2 =$$

$$= \sum_{i=1}^{n} x_i^2 - n\overline{x}^2 = \sum_{i=1}^{n} x_i^2 - \eta_1^2,$$

где $\eta_1 = \sqrt{n}\overline{x} = \frac{1}{\sqrt{n}}x_1 + \frac{1}{\sqrt{n}}x_2 + \dots + \frac{1}{\sqrt{n}}x_n = c_{11}x_1 + c_{12}x_2 + + \dots + c_{1n}x_n$ и $\sum\limits_{i=1}^{n} c_{1i}^2 = 1$. По лемме Фишера $n\widetilde{S^2}$ не зависит от $\eta_1^2 = n\overline{x}^2$ и

$$\frac{n\widetilde{S^2}}{\sigma^2} = \chi^2_{n-1}. \qquad \square$$

§ 6. Интервальные оценки неизвестных параметров

В схеме Бернулли, как мы помним, несмещенной, состоятельной и эффективной в классе несмещенных регулярных оценок оценкой вероятности успеха p является относительная частота $\hat{p} = \frac{\mu}{n}$. Выясним, насколько она близка к оцениваемому параметру p. По следствию из интегральной теоремы Муавра-Лаплпса при больших n

$$P\{|\hat{p} - p| < \varepsilon\} = P\left\{\left|\frac{\mu}{n} - p\right| < \varepsilon\right\} \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right).$$

Обозначая $\varepsilon = t\sqrt{\frac{pq}{n}}$, получим

$$P\left\{|\hat{p} - p| < t\sqrt{\frac{pq}{n}}\right\} \approx 2\Phi(t).$$

Неравенство, стоящее под знаком вероятности, эквивалентно неравенству

$$\left(1+\frac{t^2}{n}\right)p^2-2\left(\hat{p}+\frac{t^2}{n}\right)p+\hat{p}^2\le 0,$$

решение которого имеет вид

$$\frac{\hat{p}+\frac{t^2}{2n}-t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\left(\frac{t}{2n}\right)^2}}{1+\frac{t^2}{n}}\le p\le\frac{\hat{p}+\frac{t^2}{2n}+t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\left(\frac{t}{2n}\right)^2}}{1+\frac{t^2}{n}}. \tag{6.1}$$

Зададим **надежность (доверительную вероятность, коэффициент доверия)** P, где $0.5<P<1$. Чаще всего берут $P=0.95$ или $P=0.99$. Находя t по таблицам функции Лапласа из уравнения $2\Phi(t)=P$, получаем доверительный интервал (отрезок) для неизвестного параметра p, а именно (6.1). Считаем, что p всегда находится в нем, при этом в среднем ошибаемся приблизительно в $100\cdot(1-P)$ случаях из 100. Если $P=0.99$, то ошибка происходит в среднем приблизительно в 1 случае из 100. Число $\alpha=1-P$ называется **уровнем значимости**.

Отметим, что случайными являются концы отрезка (интервала), а p хоть и неизвестно, но является числом, а не случайной величиной. Поэтому правильнее говорить, что доверительный интервал

$$\left(\frac{\hat{p}+\frac{t^2}{2n}-t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\left(\frac{t}{2n}\right)^2}}{1+\frac{t^2}{n}},\ \frac{\hat{p}+\frac{t^2}{2n}+t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}+\left(\frac{t}{2n}\right)^2}}{1+\frac{t^2}{n}}\right)$$

покрывает с надежностью P неизвестный параметр p, но не утверждать, что с надежностью P параметр p попадает в этот доверительный интервал.

Если n велико, то пренебрегая в (6.1) величинами более высокого порядка, чем $\frac{1}{\sqrt{n}}$, получим более простой доверительный интервал

$$\hat{p} - t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Так как $\hat{p}$ — точечная оценка p, то доверительный интервал в этом случае оказывается симметричным относительно этой точечной оценки. Половина длины этого интервала, т.е. величина $\delta = t\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, на которую может по абсолютной величине отклониться $\hat{p}$ от p, называется **точностью оценки**.

Перейдем к систематическому изложению данной темы. Пусть $x_1, x_2, \dots, x_n$ — выборка из распределения $F(x/\theta)$, зависящего от неизвестного параметра θ. Под статистикой будем понимать любую борелевскую функцию от $x_1, x_2, \dots, x_n$, так что при подстановке в нее случайных величин — элементов абстрактной выборки — получается случайная величина.

Определение 6.1. Статистика $\theta_{\text{в}} = \theta_{\text{в}}(x_1, x_2, \dots, x_n)$ называется **односторонним верхним доверительным пределом для неизвестного параметра θ с надежностью (доверительной вероятностью, коэффициентом доверия)** P (0.5<P<1) или **с уровнем значимости** $\alpha = 1 - P$, если

$$P\{\theta < \theta_{\text{в}}\} = P.$$

Статистика $\theta_{\text{н}} = \theta_{\text{н}}(x_1, x_2, \dots, x_n)$ называется **односторонним нижним доверительным пределом для неизвестного параметра θ с надежностью (доверительной вероятностью, коэффициентом доверия)**

P (0.5<P<1) или **с уровнем значимости** $\alpha = 1 - P$, если

$$P\{\theta > \theta_{\text{н}}\} = P.$$

При этом интервалы $(-\infty, \theta_{\text{в}})$ и $(\theta_{\text{н}}, +\infty)$ называются соответственно **верхним** и **нижним односторонними (левосторонним** и **правосторонним) доверительными интервалами для неизвестного параметра** θ **с надежностью (доверительной вероятностью, коэффициентом доверия)** P или **с уровнем значимости** $\alpha = 1 - P$.

Определение 6.2. Статистики $\theta_{\text{н}} = \theta_{\text{н}}(x_1, x_2, \dots, x_n)$ и $\theta_{\text{в}} = \theta_{\text{в}}(x_1, x_2, \dots, x_n)$ называются соответственно **двусторонними нижним и верхним доверительными пределами** соответственно **для неизвестного параметра** θ **с надежностью (доверительной вероятностью, коэффициентом доверия)** P или **с уровнем значимости** $\alpha = 1 - P$, если

$$P\{\theta_{\text{н}} < \theta < \theta_{\text{в}}\} = P.$$

При этом интервал $(\theta_{\text{н}}, \theta_{\text{в}})$ называется **двусторонним доверительным интервалом для неизвестного параметра** θ **с надежностью (доверительной вероятностью, коэффициентом доверия)** P или **с уровнем значимости** $\alpha = 1 - P$.

Замечание 1. Отметим, что в этих определениях θ является хоть и неизвестным, но числом, а случайными являются $\theta_{\text{н}}$ и $\theta_{\text{в}}$.

Замечание 2. Если $\widehat{\theta_n}$ — точечная оценка параметра θ, а двусторонний доверительный интервал для неизвестного параметра θ симметричен относительно оценки $\widehat{\theta_n}$:

$$\widehat{\theta_n} - \delta < \theta < \widehat{\theta_n} + \delta,$$

то δ называется **точностью интервальной оценки**.

Лемма (о построении двусторонних доверительных интервалов по односторонним). *Если $\theta_н$ и $\theta_в$ — соответственно нижний и верхний односторонние пределы для неизвестного параметра θ с надежностью $P \quad (0.5 < P < 1)$, то $\theta_н$ и $\theta_в$ являются двусторонними доверительными пределами для θ с надежностью* $\gamma = 2P - 1$.

Д о к а з а т е л ь с т в о. По условию $P\{\theta > \theta_{\text{н}}\} = P, \quad P\{\theta < \theta_{\text{в}}\} = P$. Следовательно,

$$P\{\theta_{\text{н}} < \theta < \theta_{\text{в}}\} = P\{\theta < \theta_{\text{в}}\} - P\{\theta \le \theta_{\text{н}}\} =$$

$$= P - (1 - P) = 2P - 1 = \gamma.$$

Согласно определению 6.2 это означает, что $\theta_{\text{н}}$ и $\theta_{\text{в}}$ — двусторонние доверительные пределы для θ с надежностью γ. □

Лемма (о равномерном распределении). *Если функция распределения $F(x)$ случайной величины ξ непрерывна, то случайная величина $\eta = F(\xi)$ равномерно распределена на отрезке* $[0, 1]$.

Д о к а з а т е л ь с т в о. Так как $0 \le F(x) \le 1 \quad \forall x$, то $0 \le F(\xi) \le 1$. Поэтому для $x \le 0 \quad F_\eta(x) = P\{F(\xi) < x\} = 0$, а для $x > 1 \quad F_\eta(x) = P\{F(\xi) < x\} = 1$. Если же $0 < x \le 1$, то

$$F_\eta(x) = P\{F(\xi) < x\} = P\{\xi < F^{-1}(x)\} = F(F^{-1}(x)) = x.$$

Итак,

$$F_\eta(x) = \begin{cases} 0, & \text{если } x \le 0, \\ x, & \text{если } 0 < x \le 1, \\ 1, & \text{если } x > 1. \end{cases}$$

Это означает, что η равномерно распределена на отрезке $[0, 1]$. $\square$

Основная теорема интервального оценивания. *Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения $F(x/\theta)$, зависящего от неизвестного параметра θ и $\zeta = \zeta(x_1, x_2, \ldots, x_n/\theta)$ — некоторая борелевская функция от $x_1, x_2, \ldots, x_n$, также зависящая от θ. Кроме того, пусть $G(x/\theta)$ — зависящая от θ функция распределения случайной величины ζ, которая непрерывна по x.*

1. Если $G(x/\theta)$ при фиксированном x не возрастает по θ, то односторонние доверительные пределы для θ с надежностью P находятся из уравнений

$$G(\zeta/\theta_{н}) = P, \quad G(\zeta/\theta_{в}) = 1 - P, \quad 0.5 < P < 1, \tag{6.2}$$

а двусторонние пределы с надежностью γ — из уравнений

$$G(\zeta/\theta_{н}) = \frac{1+\gamma}{2}, \quad G(\zeta/\theta_{в}) = \frac{1-\gamma}{2}. \tag{6.3}$$

2. Если $G(x/\theta)$ при фиксированном x не убывает по θ, то односторонние доверительные пределы для θ с надежностью P находятся из уравнений

$$G(\zeta/\theta_{н}) = 1 - P, \quad G(\zeta/\theta_{в}) = P, \quad 0.5 < P < 1, \tag{6.4}$$

а двусторонние пределы с надежностью γ — из уравнений

$$G(\zeta/\theta_{\text{н}}) = \frac{1-\gamma}{2}, \quad G(\zeta/\theta_{\text{в}}) = \frac{1+\gamma}{2}. \tag{6.5}$$

Доказательство. 1. Пусть $G(x/\theta)$ при фиксированном x не возрастает по θ, а $\theta_{\text{н}}$ и $\theta_{\text{в}}$ — корни уравнений (6.2). Так как по лемме о равномерном распределении $\eta = G(\zeta/\theta)$ равномерно распределена на отрезке $[0, 1]$, то

$$P\{\theta > \theta_{\text{н}}\} = P\{G(\zeta/\theta) \leq G(\zeta/\theta_{\text{н}})\} = P\{G(\zeta/\theta) \leq P\} =$$

$$= P\{\eta \leq P\} = P.$$

По лемме о нахождении двусторонних пределов по односторонним $\theta_{\text{н}}$ и $\theta_{\text{в}}$ — двусторонние доверительные пределы с надежностью $\gamma = 2P - 1$. Значит, они найдутся из тех же уравнений (6.2), которые принимают форму (6.3) при замене P на $\frac{\gamma+1}{2}$.

2. Случай неубывающей по θ функции $G(x/\theta)$ рассматривается подобным способом. $\square$

Задача 1. Докажите пункт 2 теоремы.

Плотность распределения случайной величины хи-квадрат с n степенями свободы есть

$$k_n(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{если } x > 0, \\ 0, & \text{если } x \leq 0. \end{cases}$$

Соответствующую функцию распределения обозначим $K_n(x) = P\{\chi_n^2 < x\} = \int\limits_{-\infty}^{x} k_n(t)dt$. Очевидно, $K_n(x) = 0$ для $x \leq 0$ и $K_n(x) = \int\limits_{0}^{x} k_n(t)dt$ для $x > 0$. Зададим число $\alpha \in (0, 1)$.

Число $\chi^2_{\alpha,\, n} = \chi^2(100\alpha\%;\ n)$, называется **критической точкой распределения хи-квадрат**, соответствующей уровню значимости α и числу степеней свободы n или 100α-**процентной точкой распределения хи-квадрат** с n степенями свободы, если

$$P\{\chi_n^2 \geq \chi^2(100\alpha\%;\ n)\} = \int\limits_{\chi^2(100\alpha\%;\ n)}^{\infty} k_n(x)dx = \alpha.$$

Имеются таблицы процентных точек распределения хи-квадрат.

Напомним, что отношение

$$t_n = \frac{\zeta}{\sqrt{\frac{\chi^2}{n}}},$$

в котором числитель и знаменатель — независимые случайные величины, $\zeta \sim N(0,1)$, а χ^2 — случайная величина хи-квадрат с n степенями свободы, называется **отношением Стьюдента с** n **степенями свободы**, а распределение этого отношения — **распределением Стьюдента с** n **степенями свободы**. Плотность распределения Стьюдента имеет вид

$$s_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+x^2\right)^{-\frac{n+1}{2}}.$$

Обозначим соответствующую функцию распределения $S_n(x) = P\{t_n < x\} = \int\limits_{-\infty}^{x} s_n(t)dt$. Так как $s_n(x)$ — четная функция, то $S_n(x)+S_n(-x) = S_n(x)+\int\limits_{-\infty}^{-x} s_n(t)dt = S_n(x) + \int\limits_{x}^{+\infty} s_n(t)dt = S_n(x) + P\{t_n \geq x\} = S_n(x) +$

$1 - S_n(x) = 1$. Итак, функция распределения отношения Стьюдента удовлетворяет свойству

$$S_n(x) + S_n(-x) = 1.$$

Как и для распределения хи-квадрат, для распределения Стьюдента имеются таблицы процентных точек. Число $t_{\alpha,\, n} = t(100\alpha\%;\ n)$, называется **критической точкой распределения Стьюдента**, соответствующей уровню значимости α и числу степеней свободы n или 100α-**процентной точкой распределения Стьюдента** с n степенями свободы, если

$$P\{t_n \geq t(100\alpha\%;\ n)\} = \int\limits_{t(100\alpha\%;\, n)}^{\infty} s_n(x)dx = \alpha.$$

Займемся оценкой параметров нормального распределения. Пусть $x_1, x_2, \dots, x_n$ — выборка из нормального распределения $N(a, \sigma^2)$. Требуется построить интервальные оценки для параметров a и σ^2. Возможны ситуации, когда один из параметров точно известен и необходимо оценить другой и когда неизвестны оба параметра и необходимо оценить каждый из них. Рассмотрим 4 случая.

1. *Пусть σ^2 известна, требуется оценить a.*

Так как $x_1, x_2, \dots, x_n$ независимы и каждая из этих случайных величин имеет нормальное распределение, то и $\sum\limits_{i=1}^{n} x_i$ имеет нормальное распределение. По лемме о нормальном распределении выборочная средняя $\overline{x} =$

$\frac{1}{n}\sum\limits_{i=1}^{n} x_i$ также имеет нормальное распределение. Поскольку $M\overline{x} = a$, $D\overline{x} = \frac{\sigma^2}{n}$, то $\overline{x} \sim N(a,\ \frac{\sigma^2}{n})$. По лемме о нормальном распределении $\zeta = \frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$. Подставим ζ в ее функцию распределения $N(x)$. При увеличении a величина $\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}}$ уменьшается. Так как $N(x)$ — возрастающая функция, то при этом $N\Big(\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}}\Big)$ также уменьшится. Таким образом, $N\Big(\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}}\Big)$ не возрастает по a. По основной теореме интервального оценивания нижний и верхний двусторонние доверительные пределы $a_{\text{н}}$ и $a_{\text{в}}$ с надежностью γ найдутся из уравнений

$$N\Big(\frac{\overline{x}-a_{\text{н}}}{\sigma}\sqrt{n}\Big) = \frac{1+\gamma}{2}, \quad N\Big(\frac{\overline{x}-a_{\text{в}}}{\sigma}\sqrt{n}\Big) = \frac{1-\gamma}{2}.$$

Обозначим через $t = t_\gamma$ корень уравнения $\Phi(t) = \frac{\gamma}{2}$. Поскольку $N(x) = \frac{1}{2} + \Phi(x)$ и $\Phi(x)$ — нечетная функция, то

$$\frac{\overline{x}-a_{\text{н}}}{\sigma}\sqrt{n} = t, \quad \frac{\overline{x}-a_{\text{в}}}{\sigma}\sqrt{n} = -t,$$

откуда искомый доверительный интервал имеет вид

$$\overline{x} - \frac{\sigma}{\sqrt{n}}t_\gamma < a < \overline{x} + \frac{\sigma}{\sqrt{n}}t_\gamma,$$

где t_γ — корень уравнения $\Phi(t_\gamma) = \frac{\gamma}{2}$, который находится из таблиц функции Лапласа.

2. *Пусть σ^2 неизвестна, требуется оценить a.*

Пользоваться построенным выше интервалом мы не можем, так как в его концы входит неизвестная величина σ. В предыдущем пункте в качестве ζ, фигурирующем в основной теореме интервального оценивания, мы брали

$\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}}$. Чтобы исключить отсюда неизвестную σ, выберем в качестве ζ случайную величину

$$\zeta = t_{n-1} = \frac{\overline{x}-a}{\frac{S}{\sqrt{n}}} = \frac{\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}}}{\frac{S}{\sigma}},$$

которая представляет собой отношение Стьюдента с $n-1$ степенями свободы. В самом деле, числитель $\frac{\overline{x}-a}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ и знаменатель $\frac{S}{\sigma}$ — независимые случайные величины как борелевские функции от независимых случайных величин $\overline{x}$ и S^2, а по следствию из леммы Фишера $\frac{(n-1)S^2}{\sigma^2} = \chi^2_{n-1}$, откуда $\frac{S}{\sigma} = \sqrt{\frac{\chi^2_{n-1}}{n-1}}$.

Подставим случайную величину ζ в ее же функцию распределения $S_{n-1}(t)$, которая не убывает. При увеличении a аргумент $\frac{\overline{x}-a}{\frac{S}{\sqrt{n}}}$ уменьшается, следовательно, $S_{n-1}\Big(\frac{\overline{x}-a}{\frac{S}{\sqrt{n}}}\Big)$ не увеличивается. Поэтому функция $S_{n-1}\Big(\frac{\overline{x}-a}{\frac{S}{\sqrt{n}}}\Big)$ не возрастает по a и по основной теореме интервального оценивания нижний и верхний двусторонние доверительные пределы с надежностью γ найдутся из уравнений

$$S_{n-1}\Big(\frac{\overline{x}-a_{\text{н}}}{\frac{S}{\sqrt{n}}}\Big) = \frac{1+\gamma}{2}, \quad S_{n-1}\Big(\frac{\overline{x}-a_{\text{в}}}{\frac{S}{\sqrt{n}}}\Big) = \frac{1-\gamma}{2}.$$

Обозначим $\frac{\overline{x}-a_{\text{н}}}{\frac{S}{\sqrt{n}}} = t$. Тогда $S_{n-1}(t) = \frac{1+\gamma}{2}$. В силу свойства функции распределения Стьюдента $S_{n-1}(-t) = 1 - S_{n-1}(t) = 1 - \frac{1+\gamma}{2} = \frac{1-\gamma}{2}$, откуда $\frac{\overline{x}-a_{\text{в}}}{\frac{S}{\sqrt{n}}} = -t$. Итак, $a_{\text{н}} = \overline{x} - \frac{S}{\sqrt{n}}t, \quad a_{\text{в}} = \overline{x} + \frac{S}{\sqrt{n}}t$. Далее,

$$P\{t_{n-1} \geq t\} = 1 - S_{n-1}(t) = \frac{1-\gamma}{2}.$$

Таким образом, искомый доверительный интервал

$$\overline{x} - \frac{S}{\sqrt{n}}t < a < \overline{x} + \frac{S}{\sqrt{n}}t,$$

где $t = t(100\frac{1-\gamma}{2}\%;\ n-1)$.

Разумеется, что этим доверительным интервалом можно пользоваться и в случае, когда σ^2 известна, но при одной и той же надежности интервал, построенный в 1., будет входить в данный интервал, т.е. будет более информативным. Это связано с тем, что при построении доверительного интервала в 2. не использовалась информация о точном значении σ^2.

На практике часто возникает задача: каково число наблюдений n, обеспечивающее с данной надежностью γ нужную точность Δ_a интервальной оценки. Для построенной нами оценки точность $\delta_a = \frac{S}{\sqrt{n}}t$, т.е.

$$P\{\overline{x} - \delta_a < a < \overline{x} - \delta_a\} = \gamma.$$

Нам надо выбрать n из условия

$$P\{\overline{x} - \Delta_a < a < \overline{x} - \Delta_a\} \geq \gamma,$$

откуда $\Delta_a \geq \delta_a = \frac{S}{\sqrt{n}}t$. Таким образом,

$$n \geq t^2 \frac{S^2}{\Delta_a^2},$$

где $t = t(100\frac{1-\gamma}{2}\%;\ n-1)$.

3. *Пусть a известна, требуется оценить* σ^2.

Ранее были исследованы две точечные оценки $\widetilde{S^2}$ и S^2 теоретической дисперсии σ^2. Но в случае, когда значение параметра a точно известно, лучше использовать точечную оценку

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2,$$

в которую входит это значение. Так как случайные величины $\frac{x_1-a}{\sigma}$, $\frac{x_2-a}{\sigma}$, $\ldots$ $\frac{x_n-a}{\sigma}$ независимы, а по лемме о нормальном распределении каждая из них имеет стандартное нормальное распределение, то случайная величина

$$\zeta = \frac{nS^{*2}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{x_i - a}{\sigma}\right)^2$$

является случайной величиной хи-квадрат с n степенями свободы. Подставим ζ в ее же функцию распределения $K_n(t)$. Так как при увеличении σ^2 величина $\frac{nS^{*2}}{\sigma^2}$ уменьшается, $K_n(t)$ — неубывающая функция, то $K_n\left(\frac{nS^{*2}}{\sigma^2}\right)$ не возрастает по σ^2. По основной теореме интервального оценивания двусторонние доверительные пределы для σ^2 с надежностью γ найдутся из уравнений

$$K_n\left(\frac{nS^{*2}}{\sigma_{\text{н}}^2}\right) = \frac{1+\gamma}{2}, \quad K_n\left(\frac{nS^{*2}}{\sigma_{\text{в}}^2}\right) = \frac{1-\gamma}{2}.$$

Обозначим $\frac{nS^{*2}}{\sigma_{\text{н}}^2} = \chi^2_{1,\gamma}$, $\frac{nS^{*2}}{\sigma_{\text{в}}^2} = \chi^2_{2,\gamma}$. Тогда $\sigma_{\text{н}}^2 = \frac{nS^{*2}}{\chi^2_{1,\gamma}}$, $\sigma_{\text{в}}^2 = \frac{nS^{*2}}{\chi^2_{2,\gamma}}$. Далее,

$$P\{\chi_n^2 \geq \chi^2_{1,\gamma}\} = 1 - K_n(\chi^2_{1,\gamma}) = 1 - K_n\left(\frac{nS^{*2}}{\sigma_{\text{н}}^2}\right) =$$

$$= 1 - \frac{1+\gamma}{2} = \frac{1-\gamma}{2},$$

$$P\{\chi_n^2 \geq \chi_{2,\gamma}^2\} = 1 - K_n(\chi_{2,\gamma}^2) = 1 - K_n\Big(\frac{nS^{*2}}{\sigma_{\text{в}}^2}\Big) =$$

$$= 1 - \frac{1-\gamma}{2} = \frac{1+\gamma}{2}.$$

Таким образом, искомый доверительный интервал

$$\frac{nS^{*2}}{\chi_{1,\gamma}^2} < \sigma^2 < \frac{nS^{*2}}{\chi_{2,\gamma}^2},$$

где

$$\chi_{1,\gamma}^2 = \chi^2\Big(100\frac{1-\gamma}{2}\%;\ n\Big), \quad \chi_{2,\gamma}^2 = \chi^2\Big(100\frac{1+\gamma}{2}\%;\ n\Big).$$

4. *Пусть a неизвестна, требуется оценить* σ^2.

По следствию из леммы Фишера

$$\frac{(n-1)S^2}{\sigma^2} = \chi_{n-1}^2.$$

Повторяя рассуждения предыдущего пункта с заменой n на $n-1$, $\quad S^{*2}$ на S^2, получаем, что искомый доверительный интервал имеет вид

$$\frac{(n-1)S^2}{\chi_{1,\gamma}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{2,\gamma}^2},$$

где

$$\chi_{1,\gamma}^2 = \chi^2\Big(100\frac{1-\gamma}{2}\%;\ n-1\Big), \quad \chi_{2,\gamma}^2 = \chi^2\Big(100\frac{1+\gamma}{2}\%;\ n-1\Big).$$

§ 7. Основания теории проверки статистических гипотез. Лемма Неймана-Пирсона. Равномерно наиболее мощные критерии

Пусть $x_1, x_2, \ldots, x_n$ — выборка из абсолютно непрерывного распределения с плотностью $p(x)$, которая нам неизвестна. Вектор $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ назовем **вектором наблюдений**, а множество его возможных значений X — **выборочным пространством**. Мы хотим проверить **основную** или как ее еще называют **нулевую гипотезу** H_0, состоящую в том, что $p(x) = p_0(x)$, где $p_0(x)$ — заданная плотность распределения вероятностей (т.е. $p_0(x) \geq 0$, $\int\limits_{-\infty}^{+\infty} p_0(x)dx = 1$). Нулевую гипотезу можно сформулировать по другому. Так как наблюдения в выборке независимы, то плотность распределения случайного вектора $\mathbf{x}$ равна $L(\mathbf{x}) = p(x_1)p(x_2)\ldots p(x_n)$. Поэтому нулевая гипотеза состоит в том, что в выборочном пространстве X вектор наблюдений $\mathbf{x}$ распределен с плотностью $L(\mathbf{x})(= L(x_1, x_2, \ldots, x_n)) = L_0(\mathbf{x})$, где $L_0(\mathbf{x}) = p_0(x_1)p_0(x_2)\ldots p_0(x_n)$. В соответствие со сказанным, символически нулевую гипотезу будем записывать либо в виде

$$H_0: \quad p(x) = p_0(x), \quad \text{либо в виде} \quad H_0: \quad L(\mathbf{x}) = L_0(\mathbf{x}).$$

Рассматриваемая нулевая гипотеза является **простой**, ибо выражает единственное распределение $p_0(x)$. Если нулевая гипотеза выражает принадлежность $p(x)$ к некоторому классу распределений, состоящему более чем из одного распределения, то она называется **сложной**. Например, гипотеза, состоящая в том, что выборка взята из нормального распределения, является сложной, ибо она состоит в принадлежности $p(x)$ классу нормальных плотностей $\{p(x/a, \sigma^2), \quad a \in R, \quad \sigma > 0\}$.

Предположим, что мы хотим проверить простую гипотезу H_0 против **конкурирующей гипотезы** или **альтернативы** H_1. Предполагается, что верна одна и только одна из гипотез H_0 или H_1. Сначала рассмотрим наиболее простой случай, когда проверяется простая гипотеза $H_0 \ : \quad L(\mathbf{x}) = L_0(\mathbf{x})$ против простой альтернативы $H_1 \ : \quad L(\mathbf{x}) = L_1(\mathbf{x}) \qquad (p(x) = p_1(x))$. Здесь $p_1(x)$ —некоторая плотность вероятности, отличная от $p_0(x), \quad L_1(\mathbf{x}) = p_1(x_1)p_1(x_2)\ldots p_1(x_n)$.

Критерий или **решающее правило** для проверки нулевой гипотезы будем строить на основании выбора в выборочном пространстве X критической области K такой, что если вектор наблюдений $\mathbf{x} \in K$, то нулевая гипотеза H_0 отвергается (т.е. принимается альтернатива H_1), а если $\mathbf{x} \in \bar{K} = X \setminus K$, то нулевая гипотеза H_0 принимается (т.е. отвергается H_1). Если критическая область имеет вид $K = \{\mathbf{x} \in X : \ \varphi(\mathbf{x}) \geq C\}$, где $\varphi : \ \mathbb{R}^n \longrightarrow \mathbb{R}$ — некоторая борелевская функция, C — некоторая постоянная, то $\varphi(\mathbf{x})$ называется **статистикой критерия**.

Принимая или отвергая нулевую гипотезу, мы можем совершить ошибки двух родов.

1. Ошибка 1-го рода происходит, если нулевая гипотеза H_0 отвергается, хотя, на самом деле, она верна. **Вероятность ошибки 1-го рода** равна

$$\alpha = P\{\mathbf{x} \in K/H_0 \quad \text{верна}\} = \int\limits_K L_0(\mathbf{x})d\mathbf{x}.$$

По другому вероятность ошибки 1-го рода α называют **уровнем значимости критерия**.

2. Ошибка 2-го рода происходит, если нулевая гипотеза H_0 принимается, хотя она неверна. **Вероятность ошибки 2-го рода** равна

$$\beta = P\{\mathbf{x} \in \bar{K}/H_1 \quad \text{верна}\} = \int\limits_{\bar{K}} L_1(\mathbf{x})d\mathbf{x}.$$

Здесь рассматриваемые интегралы являются n-кратными, $d\mathbf{x} = dx_1 dx_2 \ldots dx_n$.

Критерий проверки являлся бы идеальным, если бы вероятности ошибок обоих родов α и β равнялись нулю. К сожалению, мы не можем этого сделать, поскольку находимся в условиях неопределенности, вызванной случайностью проводимых измерений. Уменьшая вероятность ошибки 1-го рода, мы тем самым увеличиваем вероятность ошибки 2-го рода и наоборот, увеличивая вероятность ошибки 1-го рода, мы уменьшаем вероятность ошибки 2-го рода. Действительно, сужая критическую область K, мы уменьшаем первый из интегралов и увеличиваем второй интеграл и наоборот. Поэтому необходимо найти некоторый компромисс, при котором вероятности ошибок обоих родов были бы незначительными. При классическом подходе предлагается выбирать критическую область K из того условия, чтобы вероятность ошибки 2-го рода была минимальна при условии, что вероятность ошибки 1-го рода не превосходит некоторый критический уровень (уровень значимости) α_0. Символически эту задачу оптимизации записывают в виде

$$\text{Найти} \quad K: \quad \beta \to \min \quad \text{при условии, что } \alpha \le \alpha_0.$$

Так как при уменьшении α величина β растет, то данная оптимизационная задача эквивалентна следующей:

$$\text{Найти} \quad K: \quad \beta \to \min \quad \text{при условии, что } \alpha = \alpha_0.$$

Определение 7.1. Число

$$1-\beta = 1-P\{\mathbf{x} \in \bar{K}/H_1 \text{ верна}\} = P\{\mathbf{x} \in K/H_1 \text{ верна}\} = \int\limits_K p_1(\mathbf{x})d\mathbf{x}$$

называется **мощностью критерия**. Критерий, для которого мощность максимальна, называется **наиболее мощным**.

Минимизировать ошибку 2-го рода — это то же самое, что максимизировать мощность критерия. Поэтому сформулированная выше задача оптимизации может быть записана в форме

$$\text{Найти} \quad K: \quad 1-\beta \to \max \quad \text{при условии, что } \alpha = \alpha_0.$$

Следующий результат решает поставленную задачу.

Лемма Неймана-Пирсона. *Из всех подмножеств* $K^* \subset X$ *с фиксированной вероятностью ошибки 1-го рода*

$$\int\limits_{K^*} L_0(\mathbf{x})d\mathbf{x} = \alpha_0$$

критическая область специального вида $K = \{\mathbf{x} : L_1(\mathbf{x}) \geq zL_0(\mathbf{x})\}$ *дает наиболее мощный критерий (дает наименьшую ошибку 2-го рода). Оптимальное значение* z *может быть найдено из уравнения*

$$\int\limits_K L_0(\mathbf{x})d\mathbf{x} = \alpha_0.$$

Д о к а з а т е л ь с т в о. По условию

$$\int\limits_K L_0(\mathbf{x})d\mathbf{x} = \int\limits_{K^*} L_0(\mathbf{x})d\mathbf{x} = \alpha_0.$$

Поэтому

$$(1-\beta)-(1-\beta^*) = \int\limits_K L_1(\mathbf{x})d\mathbf{x} - \int\limits_{K^*} L_1(\mathbf{x})d\mathbf{x} =$$

$$= \int\limits_{K\setminus(K\cap K^*)} L_1(\mathbf{x})d\mathbf{x} - \int\limits_{K^*\setminus(K\cap K^*)} L_1(\mathbf{x})d\mathbf{x} \geq$$

$$\geq z \int\limits_{K\setminus(K\cap K^*)} L_0(\mathbf{x})d\mathbf{x} - z \int\limits_{K^*\setminus(K\cap K^*)} L_0(\mathbf{x})d\mathbf{x} =$$

$$= z\Big(\int\limits_K L_0(\mathbf{x})d\mathbf{x} - \int\limits_{K^*} L_0(\mathbf{x})d\mathbf{x}\Big) = z(\alpha_0-\alpha_0) = 0,$$

откуда $1-\beta \geq 1-\beta^*$. $\square$

Предположим теперь, что вектор наблюдений $\mathbf{x}$ имеет плотность распределения $L(\mathbf{x})$, принадлежащую семейству распределений $\{L(\mathbf{x}/\theta)\}$, непрерывно зависящему от некоторого параметра $\theta \in \Theta$. Мы хотим проверить простую гипотезу $H_0: \quad \theta = \theta_0$ против сложной альтернативы $H_1: \quad \theta \in \Theta \setminus \{\theta_0\}$. В терминах функции правдоподобия нулевая гипотеза и альтернатива могут быть записаны как

$$H_0: \ \mathbf{x} \ \text{в выборочном пространстве распределен}$$

$$\text{с плотностью} \ \ L(\mathbf{x}/\theta_0),$$

$H_1: \;\; \mathbf{x}$ в выборочном пространстве распределен

с плотностью $L(\mathbf{x}/\theta)$,

где $\theta \in \Theta \setminus \{\theta_0\}$ — любое действительное число.

Функцию от параметра θ, определенную равенством $f(\theta) = P\{\mathbf{x} \in K/$истинное значение параметра есть $\theta\}$, назовем **функцией мощности**. Если зафиксировать $\theta = \theta_1 \neq \theta_0$, то по лемме Неймана-Пирсона оптимальная критическая область для проверки простой гипотезы $H_0 : \theta = \theta_0$ против простой альтернативы $H_1 : \quad \theta = \theta_1$ имеет вид $K_{\theta_1} = \{\mathbf{x} : \quad L(\mathbf{x}/\theta_1) \geq z_1 L(\mathbf{x}/\theta_0)\}$. Если зафиксировать другое значение параметра θ_2, то оптимальное значение z_2 может оказаться, вообще говоря, другим. Итак, вид оптимальной критической области $K_\theta = \{\mathbf{x} : \;\; L(\mathbf{x}/\theta) \geq z(\theta) L(\mathbf{x}/\theta_0)\}$ зависит от параметра θ.

Критерий проверки простой гипотезы $H_0 : \quad \theta = \theta_0$ против сложной альтернативы $H_1 : \quad \theta \in \Theta \setminus \{\theta_0\}$ называется **равномерно наиболее мощным**, если вид соответствующей ему оптимальной критической области не зависит от параметра θ. Отметим, что равномерно наиболее мощный критерий существует далеко не всегда.

Пример 1. Пусть $x_1, x_2, \ldots, x_n$ — выборка из нормального распределения $N(a, 1)$. Функция правдоподобия имеет форму

$$L(\mathbf{x}/a) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum\limits_{i=1}^{n}(x_i - a)^2}.$$

Проверим простую гипотезу $H_0 : \;\; a = 0$ против сложной альтернативы $H_1 : \;\; a > 0$. Вид оптимальной критической области находится из условия $L(\mathbf{x}/a) \geq z(a)L(\mathbf{x}/0)$, т.е.

$$e^{-\frac{1}{2}\sum\limits_{i=1}^{n}(x_i-a)^2+\frac{1}{2}\sum\limits_{i=1}^{n}x_i^2} \geq z(a),$$

что эквивалентно

$$a\sum_{i=1}^{n} x_i \geq \frac{na^2}{2} + \ln z(a).$$

Таким образом, критическая область имеет вид

$$K = \{\mathbf{x}: \ \overline{x} \geq C\},$$

который не зависит от параметра a. Значит существует равномерно наиболее мощный критерий. Для определения постоянной C выберем ее из того условия, что вероятность ошибки 1-го рода равна заданному уровню значимости α_0:

$$P\{\mathbf{x} \in K/H_0\} = P\{\overline{x} \geq C/a = 0\} = \alpha_0.$$

Так как при $a = 0$ статистика критерия $\overline{x} \sim N\left(0, \frac{1}{n}\right)$, т.е. $\sqrt{n}\overline{x} \sim N(0,1)$, то

$$\alpha_0 = P\{\sqrt{n}\overline{x} \geq C\sqrt{n}/a = 0\} = 1 - N(C\sqrt{n}) = \frac{1}{2} - \Phi(C\sqrt{n}).$$

Таким образом, постоянная C находится из уравнения

$$\Phi(C\sqrt{n}) = \frac{1}{2} - \alpha_0.$$

Критерий для проверки H_0 против H_1 следующий:

если $\overline{x} \geq C$, то считаем, что $a > 0$;

если $\overline{x} < C$, то считаем, что $a = 0$.

Заметим, что если бы сложная гипотеза H_1 имела вид $a \neq 0$, то в зависимости от знака a вид критической области области определялся бы неравенствами противоположного знака, т.е. вид критической области зависел бы от a. Поэтому для такого случая равномерно наиболее мощного критерия не существует.

§ 8. Критерии значимости. Критерий χ^2 Пирсона

Если имеется только нулевая гипотеза и нет никакой альтернативы или альтернатива слишком сложная, то про вероятность ошибки 2-го рода забывают и строят критерии, учитывающие только вероятность ошибки 1-го рода. Вероятность ошибки 1-го рода тогда называют **уровнем значимости критерия**. В этом случае проверяемую гипотезу называют **гипотезой значимости**, а критерии для ее проверки — **критериями значимости**. Поскольку не учитывается ошибка 2-го рода, то такие критерии менее надежны. Задавая уровень значимости α (обычно берут $\alpha = 0.05$ или $\alpha = 0.01$), находят критическую область K из условия

$$P\{\mathbf{x} \in K/H_0\} = \alpha,$$

где H_0 — проверяемая гипотеза. Если по имеющимся статистическим данным оказывается, что $\mathbf{x} \in K$, то предполагая, что гипотеза H_0 верна, мы получаем, что в результате проведенных случайных экспериментов произошло событие, имеющее ничтожно малую вероятность α. Доверять этому вряд ли имеет смысл, поэтому считаем, что выдвинуто неправильное предположение, и гипотезу H_0 надо отвергнуть. Если же окажется, что

$\mathbf{x} \in \bar{K}$, то предположив, что наша гипотеза выполняется, мы получаем, что произошло событие, имеющее большую вероятность $1 - \alpha$. Поэтому можно сделать вывод о том, что статистические данные не противоречат выдвинутой гипотезе. Это вовсе не означает, что гипотезу надо принять. На самом деле, для того, чтобы принять с определенной уверенностью проверяемую гипотезу, надо ее проверить на достаточно большом количестве выборок. Если во всех случаях гипотеза не противоречит имеющимся данным, то только тогда ее можно принять. В этом основной недостаток игнорирования вероятности ошибки 2-го рода.

А. *Построение критериев значимости по интервальным оценкам.*

Одним из способов построения критериев значимости является использование доверительных интервалов. Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения $F(x/\theta)$, зависящего от параметра θ, и для этого параметра построен двусторонний доверительный интервал:

$$P\{\theta_{\text{н}} < \theta_0 < \theta_{\text{в}}\} = \gamma,$$

где θ_0 — истинное значение параметра θ (заметим, что при построении интервальных оценок истинное значение параметра обозначалось θ, поскольку другие значения параметра там не использовались). Необходимо проверить гипотезу $H_0: \quad \theta = \theta_0$. В качестве критической области возьмем

$$K_\theta = \{\mathbf{x}: \ \ \theta \le \theta_{\text{н}} \ \text{ или } \ \theta \ge \theta_{\text{в}}\}.$$

Тогда вероятность ошибки первого рода

$$\alpha = P\{\mathbf{x} \in K_\theta / \theta = \theta_0\} = P\{\mathbf{x} \in K_{\theta_0}\} =$$

$$= 1 - P\{\theta_{\text{н}} < \theta_0 < \theta_{\text{в}}\} = 1 - \gamma.$$

Решающее правило (критерий) для проверки гипотезы H_0:

H_0 отвергается, если $\theta_0 \le \theta_{\text{н}}$ или $\theta_0 \ge \theta_{\text{в}}$;

H_0 не противоречит экспериментальным данным, если $\theta_{\text{н}} < \theta_0 < \theta_{\text{в}}$.

Заметим, что при построении критерия уровень значимости критерия α получается вычитанием из 1 надежности соответствующего доверительного интервала γ. А критическая область является дополнением к доверительному интервалу в выборочном пространстве.

Предположим, например, что $x_1, x_2, \dots, x_n$ — выборка из нормального распределения $N(a, \sigma^2)$.

* Пусть параметр σ^2 известен. Мы хотим проверить гипотезу $H_0: \quad a = a_0$. При построении интервальных оценок мы получили, что

$$P\Big\{\overline{x} - \frac{\sigma}{\sqrt{n}} t_\gamma < a_0 < \overline{x} + \frac{\sigma}{\sqrt{n}} t_\gamma\Big\} = \gamma,$$

где t_γ — корень уравнения $\Phi(t_\gamma) = \frac{\gamma}{2}$. Зададим уровнь значимости $\alpha = 1 - \gamma$ и выберем критическую область как

$$K_a = \Big\{\mathbf{x}: \ a \le \overline{x} - \frac{\sigma}{\sqrt{n}} t_\gamma \quad \text{или} \quad a \ge \overline{x} + \frac{\sigma}{\sqrt{n}} t_\gamma\Big\}.$$

Тогда $P\{\mathbf{x} \in K_a / H_0\} = P\{\mathbf{x} \in K_{a_0}\} = \alpha$. Получаем следующий критерий проверки гипотезы H_0:

если $a_0 \le \overline{x} - \frac{\sigma}{\sqrt{n}} t_\gamma$ или $a_0 \ge \overline{x} + \frac{\sigma}{\sqrt{n}} t_\gamma$, то H_0 отвергается, т.е. считаем, что $a \ne a_0$;

если $\overline{x} - \frac{\sigma}{\sqrt{n}} t_\gamma < a_0 < \overline{x} + \frac{\sigma}{\sqrt{n}} t_\gamma$, то считаем, что H_0 не противоречит экспериментальным данным.

При этом, по заданному уровню значимости α число t_γ найдется как корень уравнения $\Phi(t_\gamma) = \frac{1-\alpha}{2}$.

** Пусть теперь параметр σ^2 неизвестен и надо проверить ту же гипотезу $H_0 : \quad a = a_0$. При построении интервальных оценок мы получили, что

$$P\Big\{\overline{x} - \frac{S}{\sqrt{n}}t < a_0 < \overline{x} + \frac{S}{\sqrt{n}}t\Big\} = \gamma,$$

где $t = t(100\frac{1-\gamma}{2}\%;\ n-1)$.

Критерий проверки нашей гипотезы выглядит следующим образом:

если $a_0 \leq \overline{x} - \frac{S}{\sqrt{n}}t$ или $a_0 \geq \overline{x} + \frac{S}{\sqrt{n}}t$, то H_0 отвергается, т.е. считаем, что $a \neq a_0$;

если $\overline{x} - \frac{S}{\sqrt{n}}t < a_0 < \overline{x} + \frac{S}{\sqrt{n}}t$, то считаем, что H_0 не противоречит экспериментальным данным.

При этом, при заданном уровне значимости α число $t = t(100\frac{\alpha}{2}\%;\ n-1)$.

*** Допустим, что параметр a известен. Требуется проверить гипотезу $H_0 : \quad \sigma^2 = \sigma_0^2$. Ранее было установлено, что

$$P\Big\{\frac{nS^{*2}}{\chi^2_{1,\gamma}} < \sigma_0^2 < \frac{nS^{*2}}{\chi^2_{2,\gamma}}\Big\} = \gamma,$$

где $\chi^2_{1,\gamma} = \chi^2\Big(100\frac{1-\gamma}{2}\%;\ n\Big), \quad \chi^2_{2,\gamma} = \chi^2\Big(100\frac{1+\gamma}{2}\%;\ n\Big)$.

Поэтому критерий прверки гипотезы $\sigma^2 = \sigma_0^2$ выглядит следующим образом:

если $\sigma_0^2 \leq \frac{nS^{*2}}{\chi^2_{1,\gamma}}$ или $\sigma_0^2 \geq \frac{nS^{*2}}{\chi^2_{2,\gamma}}$, то H_0 отвергается, т.е. считаем, что $\sigma^2 \neq \sigma_0^2$;

если $\frac{nS^{*2}}{\chi^2_{1,\gamma}} < \sigma_0^2 < \frac{nS^{*2}}{\chi^2_{2,\gamma}}$, то считаем, что H_0 не противоречит экспериментальным данным.

При этом $\chi^2_{1,\gamma} = \chi^2\Big(100\frac{\alpha}{2}\%;\quad n\Big)$, $\chi^2_{2,\gamma} = \chi^2\Big(100\Big(1-\frac{\alpha}{2}\Big)\%;\ n\Big)$.

**** Пусть параметр a неизвестен и требуется проверить гипотезу $H_0:\ \ \sigma^2 = \sigma_0^2$. Ранее было установлено, что

$$P\Big\{\frac{(n-1)S^2}{\chi^2_{1,\gamma}} < \sigma_0^2 < \frac{(n-1)S^2}{\chi^2_{2,\gamma}}\Big\} = \gamma,$$

где

$$\chi^2_{1,\gamma} = \chi^2\Big(100\frac{1-\gamma}{2}\%;\ n-1\Big),$$

$$\chi^2_{2,\gamma} = \chi^2\Big(100\frac{1+\gamma}{2}\%;\ n-1\Big).$$

Поэтому критерий прверки гипотезы $\sigma^2 = \sigma_0^2$ выглядит следующим образом:

если $\sigma_0^2 \leq \frac{(n-1)S^2}{\chi^2_{1,\gamma}}$ или $\sigma_0^2 \geq \frac{(n-1)S^2}{\chi^2_{2,\gamma}}$, то H_0 отвергается, т.е. считаем, что $\sigma^2 \neq \sigma_0^2$;

если $\frac{(n-1)S^2}{\chi^2_{1,\gamma}} < \sigma_0^2 < \frac{(n-1)S^2}{\chi^2_{2,\gamma}}$, то считаем, что H_0 не противоречит экспериментальным данным.

При этом $\chi^2_{1,\gamma} = \chi^2\Big(100\frac{\alpha}{2}\%;\quad n-1\Big)$, $\chi^2_{2,\gamma} = \chi^2\Big(100\Big(1-\frac{\alpha}{2}\Big)\%;\ n-1\Big)$.

Отметим, что если мы хотим построить критерий значимости с односторонней критической областью, то можно использовать односторонние доверительные интервалы. Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения $F(x/\theta)$, зависящего от параметра θ, и для этого параметра построен правосторонний доверительный интервал:

$$P\{\theta_0 > \theta_{\text{н}}\} = P,$$

где θ_0 — истинное значение параметра θ. Для проверки гипотезы $H_0: \ \theta = \theta_0$ надо взять левостороннюю критическую область

$$K_\theta = \{\mathbf{x}: \ \theta \le \theta_{\text{н}}\}.$$

Очевидно, что критерий проверки гипотезы H_0 имеет следующий вид:

H_0 отвергается, если $\theta_0 \le \theta_{\text{н}}$;

H_0 не противоречит экспериментальным данным, если $\theta_0 > \theta_{\text{н}}$.

При этом уровень значимости критерия α получается вычитанием из 1 надежности соответствующего доверительного интервала P.

Б. *Критерии сравнения параметров двух выборок.*

Пусть имеются две независимые выборки

$x_1, x_2, \ldots, x_{n_1}$ из $N(a_1, \sigma_1^2)$ и

$y_1, y_2, \ldots, y_{n_2}$ из $N(a_2, \sigma_2^2)$

(выборки называются **независимыми**, если случайные величины $x_1, x_2, \ldots, x_{n_1}, y_1, y_2, \ldots, y_{n_2}$ независимы).

* Пусть σ_1^2 и σ_2^2 известны. Требуется проверить гипотезу $H_0: \ a_1 = a_2$.

Так как $\overline{x} \sim N\Big(a_1, \frac{\sigma_1^2}{n_1}\Big)$, $\quad \overline{y} \sim N\Big(a_2, \frac{\sigma_2^2}{n_2}\Big)$ независимы, то $\overline{y} - \overline{x}$ имеет нормальное распределение. Его параметры $M(\overline{y} - \overline{x}) = a_2 - a_1$, $\quad D(\overline{y} - \overline{x}) = D\overline{y} + D\overline{x} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$, т.е. $\overline{y} - \overline{x} \sim N\Big(a_2 - a_1, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\Big)$. По лемме о нормальном распределении

$$\zeta = \frac{\overline{y} - \overline{x} - (a_2 - a_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$

Предположим, что гипотеза H_0 верна, тогда

$$\zeta_0 = \zeta\Big|_{H_0} = \frac{\overline{y} - \overline{x}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

Понятно, что чем ближе друг к другу a_1 и a_2, тем меньше по абсолютной величине статистика ζ. Поэтому в качестве критической области естественно взять

$$K = \{\mathbf{x} : \ |\zeta| \geq C\}.$$

Назначая уровень значимости α, выберем постоянную C из условия

$$P\{\mathbf{x} \in K/H_0\} = P\{|\zeta| \geq C/H_0\} = P\{|\zeta_0| \geq C\} =$$

$$= 1 - P\{|\zeta_0| < C\} = 1 - 2\varPhi(C) = \alpha,$$

откуда C находится по таблицам функции Лапласа как корень уравнения $\varPhi(C) = \frac{1-\alpha}{2}$.

Критерий проверки гипотезы $a_1 = a_2$ имеет следующий вид:

если $|\zeta_0| \geq C$, то гипотеза отвергается;

если $|\zeta_0| < C$, то считаем, что гипотеза не противоречит экспериментальным данным.

** Пусть $\sigma_1^2 = \sigma_2^2 = \sigma^2$ неизвестна. Требуется проверить ту же гипотезу $H_0 : \ a_1 = a_2$. В силу определения случайной величины хи-квадрат для выборочных дисперсий $\widetilde{S_x^2} = \frac{1}{n_1}\sum\limits_{i=1}^{n_1}(x_i - \overline{x})^2$ и $\widetilde{S_y^2} = \frac{1}{n_2}\sum\limits_{i=1}^{n_2}(y_i - \overline{y})^2$ выполняются равенства

$$\frac{n_1\widetilde{S_x^2}}{\sigma^2} = \chi^2_{n_1-1}; \quad \frac{n_2\widetilde{S_y^2}}{\sigma^2} = \chi^2_{n_2-1}; \quad \frac{n_1\widetilde{S_x^2} + n_2\widetilde{S_y^2}}{\sigma^2} = \chi^2_{n_1+n_2-2}. \tag{8.1}$$

Поэтому

$$\frac{\sqrt{n_1\widetilde{S_x^2}+n_2\widetilde{S_y^2}}}{\sigma\sqrt{n_1+n_2-2}}=\sqrt{\frac{\chi^2_{n_1+n_2-2}}{n_1+n_2-2}}.$$

Как было установлено выше,

$$\frac{\overline{y}-\overline{x}-(a_2-a_1)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1). \tag{8.2}$$

В силу леммы Фишера и того, что рассматриваются независимые выборки, случайные величины (8.1) и (8.2) независимы, следовательно,

$$\zeta=\frac{\frac{\overline{y}-\overline{x}-(a_2-a_1)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\frac{\sqrt{n_1\widetilde{S_x^2}+n_2\widetilde{S_y^2}}}{\sigma\sqrt{n_1+n_2-2}}}=\frac{\overline{y}-\overline{x}-(a_2-a_1)}{\sqrt{n_1\widetilde{S_x^2}+n_2\widetilde{S_y^2}}}\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}$$

является отношением Стьюдента с n_1+n_2-1 степенями свободы.

Предположим, что гипотеза H_0 верна, тогда

$$\zeta_0=\frac{\overline{y}-\overline{x}}{\sqrt{n_1\widetilde{S_x^2}+n_2\widetilde{S_y^2}}}\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}$$

имеет распределение Стьюдента с n_1+n_2-1 степенями свободы. Понятно, что чем ближе друг к другу a_1 и a_2, тем меньше по абсолютной величине статистика ζ. Поэтому в качестве критической области естественно взять

$$K=\{\mathbf{x}:\ \ |\zeta|\geq C\}.$$

Назначая уровень значимости α, выберем постоянную C из условия

$$P\{\mathbf{x} \in K/H_0\} = P\{|\zeta| \geq C/H_0\} = P\{|\zeta_0| \geq C\} =$$
$$= 2P\{\zeta_0 \geq C\} = \alpha,$$

откуда $C = t\Big(100\frac{\alpha}{2}\%;\ \ n_1 + n_2 - 1\Big)$.

Критерий проверки гипотезы $a_1 = a_2$ имеет следующий вид:

если $|\zeta_0| \geq C$, то гипотеза отвергается;

если $|\zeta_0| < C$, то считаем, что гипотеза не противоречит экспериментальным данным.

Для рассмотрения следующего случая введем

Определение 8.1. Отношение

$$F_{n_1,n_2} = \frac{\frac{\chi^2_{n_1}}{n_1}}{\frac{\chi^2_{n_2}}{n_2}},$$

в котором числитель и знаменатель являются независимыми случайными величинами, называется F**- отношением** или **отношением Фишера с n_1 и n_2 степенями свободы**, а распределение этого отношения — F**-распределением** или **распределением Фишера с n_1 и n_2 степенями свободы**. Число $F(100\alpha\%;\ n_1,\ n_2)$, удовлетворяющее уравнению

$$P\{F_{n_1,n_2} \geq F(100\alpha\%;\ n_1,\ n_2)\} = \alpha,$$

называется 100α**-процентной точкой F-распределения с n_1 и n_2 степенями свободы**.

Имеются таблицы процентных точек F-распределения.

Задача 1. Покажите, что плотность распределения случайной величины F_{n_1,n_2} имеет вид

$$g_{n_1,n_2}(x) = \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} B^{-1}\left(\frac{n_1}{2},\ \frac{n_2}{2}\right) x^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}},$$

где $B(a,b) = \int\limits_0^1 x^{a-1}(1-x)^{b-1}dx$ — интеграл Эйлера.

*** Предположим, что a_1 и a_2 неизвестны; требуется проверить гипотезу $H_0 : \quad \sigma_1^2 = \sigma_2^2$. Для исправленных выборочных дисперсий $S_x^2 = \frac{1}{n_1-1}\sum\limits_{i=1}^{n_1}(x_i - \overline{x})^2$ и $S_y^2 = \frac{1}{n_2-1}\sum\limits_{i=1}^{n_2}(y_i - \overline{y})^2$ в силу следствия из леммы Фишера

$$\frac{S_x^2}{\sigma_1^2} = \frac{\chi^2_{n_1-1}}{n_1-1}, \quad \frac{S_y^2}{\sigma_2^2} = \frac{\chi^2_{n_2-1}}{n_2-1}.$$

Эти две случайные величины независимы, поскольку относятся к независимым выборкам. Поэтому при выполнении гипотезы H_0 отношение

$$\frac{S_x^2}{S_y^2} = \frac{\frac{\chi^2_{n_1-1}}{n_1-1}}{\frac{\chi^2_{n_2-1}}{n_2-1}}$$

имеет распределение Фишера с n_1-1 и n_2-1 степенями свободы, а отношение

$$\frac{S_y^2}{S_x^2} = \frac{\frac{\chi^2_{n_2-1}}{n_2-1}}{\frac{\chi^2_{n_1-1}}{n_1-1}}$$

имеет распределение Фишера с n_2-1 и n_1-1 степенями свободы. Принято выбирать в качестве статистики критерия то из этих двух отношений, у которого числитель больше знаменателя. Пусть $S_1^2 = \max(S_x^2, S_y^2)$, $S_2^2 = \min(S_x^2, S_y^2)$. Тогда

$$\zeta = \frac{S_1^2}{S_2^2},$$

причем при выполнении гипотезы H_0

$$\zeta_0 = \zeta\Big|_{H_0} = F_{k_1,k_2},$$

где k_1 и k_2 легко выписываются по n_1 и n_2, и меньше одного из них на 1 (т.е., если $S_x^2 > S_y^2$, то $k_1 = n_1 - 1$, $k_2 = n_2 - 1$, а если $S_x^2 < S_y^2$, то $k_1 = n_2 - 1$, $k_2 = n_1 - 1$). Пусть числа F_1 и F_2 выбраны из условия

$$P\{F_{k_1,k_2} \le F_1\} = P\{F_{k_1,k_2} \ge F_2\} = \frac{\alpha}{2},$$

где α — уровень значимости критерия. Очевидно,

$$P\{F_{k_1,k_2} \le F_1\} = P\{F_{k_2,k_1} \ge \frac{1}{F_1}\} = \frac{\alpha}{2}.$$

Таким образом,

$$F_2 = F\Big(100\frac{\alpha}{2}\%;\ \ k_1,\ k_2\Big); \quad \frac{1}{F_1} = F\Big(100\frac{\alpha}{2}\%;\ \ k_2,\ k_1\Big).$$

Выберем критическую область как

$$K_{F_1,F_2} = \{\mathbf{x}:\ \ \zeta \le F_1 \quad \text{или} \quad \zeta \ge F_2\}.$$

Тогда

$$P\{\mathbf{x} \in K/H_0\} = P\{\zeta \le F_1 \text{ или } \zeta \ge F_2/H_0\} =$$

$$= P\{F_{k_1,k_2} \le F_1 \text{ или } F_{k_1,k_2} \ge F_2\} =$$

$$= P\{F_{k_1,k_2} \le F_1\} + P\{F_{k_1,k_2} \ge F_2\} = \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha.$$

Критерий проверки гипотезы $\sigma_1^2 = \sigma_2^2$ имеет следующий вид:

если $\zeta_0 \le F_1$ или $\zeta_0 \ge F_2$, то гипотеза H_0 отвергается;

если $F_1 < \zeta_0 < F_2$, то гипотеза H_0 согласуется с экспериментальными данными.

Отметим, что точно таким же способом, которым строились критерии значимости по доверительным интервалам, можно решать обратную задачу, а именно, по критериям значимости можно строить доверительные интервалы. Например, для критерия сравнения * в случае известных дисперсий σ_1^2 и σ_2^2 для $a_2 - a_1$ получается доверительный интервал

$$\overline{y} - \overline{x} - C\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < a_2 - a_1 < \overline{y} - \overline{x} + C\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

где C — корень уравнения $\Phi(C) = \frac{1-\alpha}{2}$.

В. *Критерий χ^2 Пирсона как критерий соответствия вероятностей в полиномиальной схеме.*

Рассмотрим n независимых испытаний, в каждом из которых может произойти один из k несовместимых исходов

$$u_1, \quad , u_2, \quad \dots, \quad u_k,$$

причем вероятности этих исходов в отдельном испытании равны

$$p(u_1) = P(\{u_1\}) = p_1, \quad p(u_2) = P(\{u_2\}) = p_2, \quad \dots,$$

$$p(u_k) = P(\{u_k\}) = p_k \qquad (\sum_{i=1}^{k} p_i = 1).$$

Обозначим через μ_i число появлений исхода u_i в этих n испытаниях ($i = \overline{1,k}$, $\sum\limits_{i=1}^{k} \mu_i = n$). Составим следующую статистику:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\mu_i - np_i)^2}{np_i}.$$

Примем без доказательства *результат К.Пирсона*:

$$\lim_{n\to\infty} P\{\chi^2 < t\} = P\{\chi^2_{k-1} < t\},$$

где χ^2_{k-1} *— случайная величина хи-квадрат с* $k-1$ *степенями свободы.*

Следовательно, можно считать, что при большом объеме выборки n случайная величина χ^2 ведет себя приближенно как случайная величина хи-квадрат с $k-1$ степенями свободы.

Предположим теперь, что по результатам n независимых испытаний надо проверить гипотезу

$$H_0: \quad p_1 = p_1^0, \;\; p_2 = p_2^0, \;\; \ldots, \;\; p_k = p_k^0,$$

где p_1^0, p_2^0, $\ldots$, p_k^0 — фиксированные числа такие, что $p_1^0 + p_2^0 + \ldots + p_k^0 = 1$. Так как при большом n по закону больших чисел $\frac{\mu_i}{n} \approx p_i$, т.е. $\mu_i - np_i \approx 0$, то значения статистики χ^2 не должны быть очень большими. Поэтому критическую область целесообразно выбрать таким образом, чтобы при превышении ею некоторого порогового значения гипотеза отвергалась. Поэтому возьмем

$$K_p = \{\mathbf{x}: \;\; \chi^2 \geq \chi^2_{\text{крит.}}\}.$$

Зададим уровень значимости α. Пусть

$$\chi_0^2 = \chi^2\Big|_{H_0} = \sum_{i=1}^{k} \frac{(\mu_i - np_i^0)^2}{np_i^0}.$$

Выберем $\chi^2_{\text{крит.}}$ так, чтобы

$$P\{\mathbf{x} \in K_p/H_0\} = P\{\chi^2 \geq \chi^2_{\text{крит.}}/H_0\} = P\{\chi^2_0 \geq \chi^2_{\text{крит.}}\} = \alpha.$$

В силу результата Пирсона это равенство можно заменить приближенным равенством

$$P\{\chi^2_{k-1} \geq \chi^2_{\text{крит.}}\} \approx \alpha.$$

Поэтому приближенно $\chi^2_{\text{крит.}} = \chi^2(100\alpha\%;\ k-1)$.

Критерий проверки гипотезы H_0 имеет следующую форму:

если $\chi^2_0 \geq \chi^2(100\alpha\%;\ k-1)$, то гипотеза H_0 отвергается;

если $\chi^2_0 < \chi^2(100\alpha\%;\ k-1)$, то гипотеза H_0 согласуется с экспериментальными данными.

Подчеркнем, что данное решающее правило можно применять только для выборок достаточно большого объема.

§ 9. Критерии согласия

Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения $F(x)$, причем функция распределения F неизвестна. Требуется проверить гипотезу $H_0: \quad F(x) = F^0(x)$, где $F^0(x)$ — заданная функция распределения (т.е. F^0 не убывает, непрерывна слева, $F^0(-\infty) = 0,\ F^0(+\infty) = 1$). Гипотезу такого вида называют **гипотезой согласия**, а критерии для ее проверки — **критериями согласия**.

Обычно критерии согласия строят так. Как мы знаем, теоретическая функция распределения F и эмпирическая функция распределения F_n^* мало отличаются при больших n. Выберем некоторую меру отклонения $\nu_n =$

$\nu_n(F_n^*, F) = \nu_n(x_1, x_2, \dots, x_n)$ функции F_n^* от функции F. Эта мера выбирается неоднозначно, и, в зависимости от способа ее выбора, получаются разные критерии согласия. Предположим, что нам удалось разыскать предельное распределение случайной величины ν_n при $n \to \infty$. Пусть случайная величина ν обладает этим распределением. Тогда критическую область выбираем как

$$K_F = \{\mathbf{x} : \ \nu \geq \nu_{\text{крит.}}\}.$$

Далее по заданному уровню значимости α находим $\nu_{\text{крит.}}$ из уравнения

$$P\{\mathbf{x} \in K_F / H_0\} = P\{\nu \geq \nu_{\text{крит.}} / H_0\} = P\{\nu_0 \geq \nu_{\text{крит.}}\} = \alpha,$$

где $\nu_0 = \nu\Big|_{H_0}$ — значение предельной меры ν в предположении, что верна гипотеза H_0. И, наконец, критерий проверки гипотезы H_0 строится стандартным способом:

если $\nu_0 \geq \nu_{\text{крит.}}$, то H_0 отвергается;

если $\nu_0 < \nu_{\text{крит.}}$, то считаем гипотезу H_0 согласующейся с экспериментальными данными.

Отметим, что критерии согласия являются частными видами критериев значимости, так как не учитывают вероятность ошибки 2-го рода.

1. Критерий χ^2 Пирсона как критерий согласия.

Пусть $x_1, x_2, \dots, x_n$ — выборка из распределения $F(x)$, которое нам неизвестно. Проверяется гипотеза $H_0 :$ $F(x) = F^0(x)$, где F^0 — заданная функция распределения. Разобъем числовую ось на k непересекающихся частей $\Delta_1 = (z_0 = -\infty, z_1)$, $\Delta_2 = [z_1, z_2)$, $\Delta_3 =$

$[z_2, z_3)$, ... , $\Delta_k = [z_{k-1}, z_k = +\infty)$. Обозначим через u_i событие, состоящее в том, что теоретическая случайная величина ξ попадет в Δ_i, тогда $p_i = p(u_i) = P\{\xi \in \Delta_i\} = F(z_i) - F(z_{i-1})$ $(i = \overline{1,k})$. В силу свойств функции распределения это выполняется и для Δ_1, Δ_k. Пусть μ_i — число выборочных значений попавших в Δ_i, т.е. число появлений исхода u_i в n испытаниях.

Предположим, что гипотеза H_0 верна, т.е. $F = F^0$. Тогда тем более верна гипотеза H_0' : $p_1 = p_1^0$, $p_2 = p_2^0$, ... , $p_k = p_k^0$, где $p_1^0 = F^0(z_1) - F^0(z_0) = F^0(z_1)$, $p_2^0 = F^0(z_2) - F^0(z_1)$, ... , $p_k^0 = F^0(z_k) - F^0(z_{k-1}) = 1 - F^0(z_{k-1})$.

Таким образом, задача свелась к проверке гипотезы соответствия вероятностей в полиномиальной схеме. Составим статистику

$$\chi^2 = \sum_{i=1}^{k} \frac{(\mu_i - np_i)^2}{np_i},$$

предельное распределение которой при $n \to \infty$ в силу результата Пирсона совпадает с распределением χ^2 с $k-1$ степенями свободы. Вычисляем $p_i^0 = F^0(z_i) - F^0(z_{i-1})$, $i = \overline{1,k}$, затем вычисляем значение статистики χ^2 в предположении, что гипотеза H_0 верна:

$$\chi_0^2 = \chi^2\Big|_{H_0} = \sum_{i=1}^{k} \frac{(\mu_i - np_i^0)^2}{np_i^0}.$$

По заданному уровню значимости α находим 100α-процентную точку распределения χ^2, соответствующую числу $k-1$ степеней свободы.

Критерий проверки гипотезы H_0 имеет следующую форму:

если $\chi_0^2 \geq \chi^2(100\alpha\%;\ k-1)$, то гипотеза H_0 отвергается;

если $\chi_0^2 < \chi^2(100\alpha\%;\ k-1)$, то гипотеза H_0 согласуется с экспериментальными данными.

Замечание 1. В данном случае в роли меры отклонения F_n^* от F служит

$$\chi^2 = \nu(F_n^*,\ F) = \sum_{i=1}^{k} \frac{(\mu_i - np_i)^2}{np_i}.$$

Замечание 2. На сколько частей и каким образом разбивать числовую прямую на Δ_i? На этот счет существует много различных мнений. Кендалл и Стьюарт предлагают довольно сложную процедуру, в которой части разбиения выбираются таким образом, чтобы вероятности попадания в них были одинаковы: $p_1^0 = p_2^0 = \ldots = p_k^0 = \frac{1}{k}$. Обычно поступают проще. Разбивают полуотрезок $[x_{(1)}, x_{(n)})$ на достаточно большое число k равных частей, затем внутренние части берут за $\Delta_2, \Delta_3, \ldots, \Delta_{k-1}$. В качестве Δ_1 берут объединение крайней левой части с $(-\infty, x_{(1)})$, а в качестве Δ_n берут объединение крайней правой части с $[x_{(n)}, +\infty)$.

Рекомендуется, чтобы для каждого Δ_i выполнялось неравенство $np_i^0 \geq 5$. Поэтому должна быть разработана процедура изменения Δ_i для тех частей, в которых это условие нарушено. Само разбиение числовой прямой и эта процедура должны быть выработаны до того, как мы увидели выборку. Если они подгоняются под выборку, то концы z_i окажутся функциями от $x_1, x_2, \ldots, x_n$, т.е. случайными величинами, а предельный результат Пирсона доказан только для случая частей с фиксированными концами. Примером процедуры, позволяющей удовлетворить условие $np_i^0 \geq 5$, является следующая: если это условие выполняется в Δ_1, то оставляем его; если нет, то в качестве нового Δ_1 берем объединение старых Δ_1 иΔ_2. Если и в новом Δ_1 условие не выполняется, то объединяем Δ_1, Δ_2, Δ_3 в новое Δ_1 и т.д. до тех пор, пока выполнится условие. Выбрав Δ_1, берем следующий интервал и объединяем его с последующими до тех пор, когда условие выполнится и т.д. При этом число частей разбиения уменьшится или останется таким же.

Отметим также, что критерием Пирсона можно пользоваться только для выборок большого объема. Наиболее осторожные авторы рекомендуют применять его для $n \geq 150$, есть рекомендации более смелые, например, предлагается, чтобы $n \geq 30$.

Замечание 3. Мы подозреваем, что теоретическая случайная величина имеет нормальное распределение, но параметры неизвестны. Тогда естественно проверять гипотезу $H_0: \ \xi \sim N(\overline{x},\ S^2)$. Можно ли пользоваться критерием χ^2? Доказано, что в этом случае вся процедура остается в силе с единственным различием, что надо брать процентную точку с числом степеней свободы $k-3$, а не $k-1$. И вообще, если гипотетически выборка берется из распределения $F^0(x/\theta_1, \theta_2, \ldots, \theta_m)$, зависящего от m неизвестных параметров, то сначала находятся их оценки $\widehat{\theta_1},\ \widehat{\theta_2},\ \ldots,\ \widehat{\theta_m}$ (строго говоря с помощью так называемого метода минимума χ^2 или асимптотически эквивалентного ему метода максимального правдоподобия). Затем проверяется гипотеза $H_0: \quad F(x) = F^0(x/\widehat{\theta_1},\ \widehat{\theta_2},\ \ldots,\ \widehat{\theta_m})$. Тогда результат Пирсона остается в силе, только предельным распределением статистики

$$\chi^2 = \sum_{i=1}^{k} \frac{(\mu_i - np_i)^2}{np_i}$$

будет хи-квадрат распределение с $k-m-1$ степенями свободы. Поэтому критерий Пирсона будет применим, только число степеней свободы у процентной точки изменится с $k-1$ на $k-m-1$. Естественно, что число частей разбиения k должно браться большим, чем m.

2. Критерий согласия ω^2.

При применении критерия согласия χ^2 разбиение числовой прямой на части в определенной степени произвольно. Результат применения критерия Пирсона зависит от способа разбиения, что связано с потерей информации. Этого недостатка лишен рассматриваемый ниже критерий.

Г.Крамер, Р.Мизес и Н.В.Смирнов впервые рассмотрели меры отклонения эмпирической функции распределения от теоретической функции распределения вида

$$\nu(F_n^*,\ F) = \int\limits_{-\infty}^{+\infty} |F_n^*(x) - F(x)|^2 dK(x),$$

где $K(x)$ — в достаточной мере произвольная неубывающая функция. Н.В.Смирнов предложил в качестве $K(x)$ взять $F(x)$ и рассмотрел меру

$$\omega^2 = \int\limits_{-\infty}^{+\infty} |F_n^*(x) - F(x)|^2 dF(x).$$

Если подставить в этот интеграл

$$F_n^*(x) = \begin{cases} 0, & \text{если } x \le x_{(1)}, \\ \frac{1}{n}, & \text{если } x_{(1)} < x \le x_{(2)}, \\ \frac{2}{n}, & \text{если } x_{(2)} < x \le x_{(3)}, \\ \dots & \dots\dots\dots\dots\dots \\ 1, & \text{если } x > x_{(n)} \end{cases}$$

и посчитать, то получится

$$\omega^2 = \frac{1}{12n^2} + \frac{1}{n}\sum_{k=1}^{n}\Big[F(x_{(k)}) - \frac{2k-1}{2n}\Big]^2.$$

Задача 1. Посчитайте этот интеграл.

Смирнов доказал, что *если $F(x)$ непрерывна, то для $W_n = n\omega^2$ существует*

$$\lim_{n\to\infty} P\{W_n < x\} = P\{W < x\},$$

не зависящий от F.

Имеются таблицы критических точек распределения предельной случайной величины W, позволяющие по уровню значимости α находить критическую точку W_α, которая является корнем уравнения $P\{W \geq W_\alpha\} = \alpha$.

Зададим уровень значимости α и найдем по нему W_α. Далее подсчитываем

$$W_{n0} = n\omega_0 = n\omega\Big|_{H_0} = n\Big\{\frac{1}{12n^2}+\frac{1}{n}\sum_{k=1}^{n}\Big[F^0(x_{(k)})-\frac{2k-1}{2n}\Big]^2\Big\}.$$

Критерий проверки гипотезы H_0 такой:

если $W_{n0} \geq W_\alpha$, то гипотеза $F(x) = F^0(x)$ отвергается,

если $W_{n0} < W_\alpha$, то гипотеза $F(x) = F^0(x)$ не противоречит экспериментальным данным.

3. Критерий согласия Колмогорова.

А.Н.Колмогоров предложил в качестве меры отклонения эмпирической функции распределения от теоретической функции распределения рассматривать

$$D_n = \sup_{x\in R} |F_n^*(x) - F(x)|.$$

Он доказал, что *если $F(x)$ непрерывна, то для случайной величины $K_n = \sqrt{n}D_n$ существует не зависящий от F предел*:

$$\lim_{n\to\infty} P\{K_n < x\} = P\{K < x\} = 1 - 2\sum_{r=1}^{\infty}(-1)^{r-1}e^{-2r^2x^2}.$$

Имеются таблицы критических точек распределения предельной случайной величины K, позволяющие по уровню значимости α находить критическую точку K_α, которая является корнем уравнения $P\{K \geq K_\alpha\} = \alpha$.

Зададим уровень значимости α и найдем по нему K_α. Далее подсчитываем

$$K_{n0} = \sqrt{n}D_{n0} = \sqrt{n}D_n\Big|_{H_0} = \sqrt{n}\sup|F_n^*(x) - F^0(x)|.$$

Критерий проверки гипотезы H_0 такой:

если $K_{n0} \geq K_\alpha$, то гипотеза $F(x) = F^0(x)$ отвергается,

если $K_{n0} < K_\alpha$, то гипотеза $F(x) = F^0(x)$ не противоречит экспериментальным данным.

§ 10. Однофакторный дисперсионный анализ

Предположим, что наблюдается m независимых нормально распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_m$, имеющих одну и ту же дисперсию σ^2, и пусть $M\xi_1 = a_1$, $M\xi_2 = a_2$, $\dots$, $M\xi_m = a_m$. Пусть каждое ξ_i измеряется n раз:

$$\xi_1: \quad x_{11},\ x_{12},\ \dots,\ x_{1n}$$

$$\xi_2: \quad x_{21},\ x_{22},\ \dots,\ x_{2n}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\xi_i: \quad x_{i1},\ x_{i2},\ \dots,\ x_{in}$$

$$\dots\dots\dots\dots\dots\dots\dots\dots$$

$$\xi_m: \quad x_{m1},\ x_{m2},\ \dots,\ x_{mn}$$

По сравнению с тем, что рассматривалось ранее, у нас имеется не одна теоретическая случайная величина, а m независимых теоретических случайных величин, и каждая измеряется n раз. При этом все mn случайных величин x_{ij} независимы между собой (когда выборки рассматриваются как абстрактные).

Мы хотим проверить нулевую гипотезу

$$H_0: \quad a_1 = a_2 = \ldots = a_m.$$

Результаты эксперимента удобно записать в виде таблицы

Здесь

$$\overline{x_{1.}} = \frac{1}{n}\sum_{j=1}^{n} x_{1j}, \dots, \overline{x_{i.}} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}, \dots, \overline{x_{m.}} = \frac{1}{n}\sum_{j=1}^{n} x_{mj} \text{ —}$$

Номер случайной величины (группы)	Номер измерения					Средние
	1	2	3	...	n	
1	x_{11}	x_{12}	x_{13}	...	x_{1n}	$\overline{x_{1.}}$
2	x_{21}	x_{22}	x_{23}	...	x_{2n}	$\overline{x_{2.}}$
3	x_{31}	x_{32}	x_{33}	...	x_{3n}	$\overline{x_{3.}}$
....						
m	x_{m1}	x_{m2}	x_{m3}	...	x_{mn}	$\overline{x_{m.}}$

средние арифметические результатов измерений внутри групп. Пусть

$$\overline{x} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = \frac{1}{m}\sum_{i=1}^{m}\overline{x_{i.}} \quad —$$

среднее арифметическое всех измерений.

Разложим полную сумму квадратов отклонений выборочных значений от выборочной средней по всем mn измерениям:

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{x})^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{x_{i.}}+\overline{x_{i.}}-\overline{x})^2 =$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{x_{i.}})^2 + \sum_{i=1}^{m}\sum_{j=1}^{n}(\overline{x_{i.}}-\overline{x})^2 +$$

$$+2\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{x_{i.}})(\overline{x_{i.}}-\overline{x}).$$

Но последняя сумма

$$S = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{x_{i.}})(\overline{x_{i.}}-\overline{x}) = \sum_{i=1}^{m}(\overline{x_{i.}}-\overline{x})\sum_{j=1}^{n}(x_{ij}-\overline{x_{i.}}) = 0,$$

так как

$$\sum_{j=1}^{n}(x_{ij} - \overline{x_{i.}}) = \sum_{j=1}^{n} x_{ij} - n\overline{x_{i.}} = n\overline{x_{i.}} - n\overline{x_{i.}} = 0.$$

Таким образом,

$$Q^2 = Q_1^2 + Q_2^2,$$

где

$$Q^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \overline{x})^2$$ — полная сумма квадратов,

$$Q_1^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(\overline{x_{i.}} - \overline{x})^2 = n\sum_{i=1}^{m}(\overline{x_{i.}} - \overline{x})^2$$ — межгрупповая

сумма квадратов (характеризует степень расхождения

в систематических погрешностях групп),

$$Q_2^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij}-\overline{x_{i.}})^2$$ — внутригрупповая сумма квад-

ратов (характеризует степень расхождения внутри групп).

Предположим, что гипотеза H_0 верна. Тогда все mn наблюдений можно рассматривать как выборку из одного и того же нормального распределения $N(a_1,\ \sigma^2)$. Поэтому по определению случайной величины хи-квадрат

$$\frac{Q^2}{mn-1} = \sigma^2 \cdot \frac{\chi^2_{mn-1}}{mn-1}.$$

Очевидно, $\overline{x_{1.}}$, $\overline{x_{2.}}$, ..., $\overline{x_{m.}}$ независимы, $D\overline{x_{i.}} = \frac{\sigma^2}{n}$. Поэтому

$$\frac{1}{m-1}\sum_{i=1}^{m}(\overline{x_{i.}} - \overline{x})^2 = \frac{Q_1^2}{n(m-1)} = \frac{\sigma^2}{n}\frac{\chi^2_{m-1}}{m-1}.$$

Далее,

$$\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \overline{x_{i.}})^2 = \sigma^2\frac{\chi^2_{n-1}}{n-1},$$

откуда

$$\sum_{j=1}^{n}(x_{ij} - \overline{x_{i.}})^2 = \sigma^2\chi^2_{n-1}.$$

Так как при различных $i = \overline{1,m}$ эти случайные величины независимы, то

$$Q_2^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}(x_{ij} - \overline{x_{i.}})^2 = \sigma^2\chi^2_{m(n-1)},$$

т.е.

$$\frac{Q_2^2}{m(n-1)} = \frac{\sigma^2\chi^2_{m(n-1)}}{m(n-1)}.$$

Поэтому, в предположении, что гипотеза H_0 верна, статистика

$$F = \frac{\frac{1}{m-1}Q_1^2}{\frac{1}{m(n-1)}Q_2^2} = \frac{\frac{\chi^2_{m-1}}{m-1}}{\frac{\chi^2_{m(n-1)}}{m(n-1)}}$$

есть отношение Фишера с $m-1$ и $m(n-1)$ степенями свободы. Пусть $F_\alpha = F(100\alpha\%;\ m-1,\ m(n-1))$ — процентная точка F-распределения, т.е. корень уравнения

$$P\{F_{m-1,m(n-1)} > F_\alpha\} = \alpha.$$

По заданному уровню значимости α по таблицам процентных точек F-распределения находим F_α.

Критерий проверки гипотезы H_0 выглядит так:

если $F \geq F_\alpha$, то гипотеза H_0 отвергается;

если $F < F_\alpha$, то гипотеза не противоречит экспериментальным данным.

Схему однофакторного дисперсионного анализа удобно представить в виде таблицы (см. табл. 1).

При этом для вычислений удобно пользоваться следующими формулами:

$$Q^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}^2 - mn\overline{x}^2; \quad Q_1^2 = n\Big[\sum_{i=1}^{m} \overline{x_{i.}}^2 - m\overline{x}^2\Big];$$

$$Q_2^2 = Q^2 - Q_1^2.$$

Рассмотрим случай, когда числа наблюдений в группах могут отличаться. Если n_i — число наблюдений в i-й группе, $N = \sum\limits_{i=1}^{m} n_i$, то рассмотренная схема останется справедливой со следующими изменениями:

$$\overline{x_{i.}} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, \quad \overline{x} = \frac{1}{N}\sum_{i=1}^{m}\sum_{j=1}^{n_i} x_{ij} = \frac{1}{N}\sum_{i=1}^{m} n_i\overline{x_{i.}},$$

$$Q^2 = Q_1^2 + Q_2^2,$$

где теперь

$$Q^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij} - \overline{x})^2; \quad Q_1^2 = \sum_{i=1}^{m} n_i(\overline{x_{i.}} - \overline{x})^2;$$

$$Q_2^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(x_{ij} - \overline{x_{i.}})^2.$$

Таблица 1.

Компонента дисперсии	Сумма квадратов	Число степеней свободы	Средний квадрат
Между группами (факторная)	$\sum\limits_{i,j}(\overline{x_{i.}}-\overline{x})^2$	$m-1$	$\frac{1}{m-1}\sum\limits_{i,j}(\overline{x_{i.}}-\overline{x})^2$
Внутри групп (остаточная)	$\sum\limits_{i,j}(x_{ij}-\overline{x_{i.}})^2$	$mn-m$	$\frac{1}{mn-m}\sum\limits_{i,j}(x_{ij}-\overline{x_{i.}})^2$
Общая	$\sum\limits_{i,j}(x_{ij}-\overline{x})^2$	$mn-1$	$\frac{1}{mn-1}\sum\limits_{i,j}(x_{ij}-\overline{x})^2$

Табличная схема изменится следующим образом:

Таблица 2.

Компонента дисперсии	Сумма квадратов	Число степеней свободы	Средний квадрат
Между группами (факторная)	$\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i}(x_{ij}-\overline{x})^2$	$m-1$	$\frac{1}{m-1}\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i}(x_{ij}-\overline{x})^2$
Внутри групп (остаточная)	$\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i}(x_{ij}-\overline{x_{i.}})^2$	$N-m$	$\frac{1}{N-m}\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i}(x_{ij}-\overline{x_{i.}})^2$
Общая	$\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i}(x_{ij}-\overline{x})^2$	$N-1$	$\frac{1}{N-1}\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n_i}(x_{ij}-\overline{x})^2$

§ 11. Основы линейного регрессионного анализа

Регрессионный анализ является мощным оружием математической статистики и используется для установления эмпирических зависимостей между случайными событиями, между случайными величинами и между случайными процессами. Мы рассмотрим начала этой теории применительно к случайным величинам. Центральным понятием в ней является понятие условного математического ожидания, общее определение которого строится с помощью известной теоремы Радона-Никодима из теории меры. В общем курсе ввести такое определение представляется невозможным, поэтому мы введем определение для частных случаев дискретных и абсолютно непрерывных случайных величин, причем более подробно рассмотрим дискретный случай, для которого рельефно просматриваются идейные особенности вводимого понятия.

11.1. Условное математическое ожидание

Пусть (ξ, η) — дискретный случайный вектор, принимающий значения (x_i, y_j) из некоторого не более чем счетного множества возможных значений с вероятностями

$$p_{ij} = P\{\xi = x_i,\, \eta = y_j\}.$$

Как мы знаем, частные распределения компонент этого вектора определяются соотношениями

$$p_{i.} = P\{\xi = x_i\} = \sum_j p_{ij}; \quad p_{.j} = P\{\eta = y_j\} = \sum_i p_{ij}. \tag{11.1}$$

Предположим, что случайная величина η имеет математическое ожидание

$$M\eta = \sum_j y_j\, P\{\eta = y_j\} = \sum_j y_j p_{.j} \tag{11.2}$$

(т.е. ряд (11.2) абсолютно сходится). Мы хотим определить условное математическое ожидание случайной величины η при условии, что произошло событие $\{\xi = x_i\}$. Для этого естественно в (11.2) вероятность $p_{.j} = P\{\eta = y_j\}$ заменить на условную вероятность

$$q_{j/i} = P\{\eta = y_j/\xi = x_i\} = \frac{P\{\xi = x_i,\ \eta = y_j\}}{P\{\xi = x_i\}} =$$

$$= \frac{p_{ij}}{p_{i.}} = \frac{p_{ij}}{\sum\limits_j p_{ij}}.$$

Поэтому даем следующее

Определение 11.1. Пусть случайная величина η имеет математическое ожидание. **Условным математическим ожиданием случайной величины η при условии, что случайная величина ξ приняла значение** x_i, называется число

$$M(\eta/\xi = x_i) = \sum_j y_j\, q_{j/i} = \sum_j y_j\, P\{\eta = y_j/\xi = x_i\} =$$

$$= \frac{1}{p_{i.}} \sum_j y_j\, p_{ij}. \tag{11.3}$$

Так как $p_{ij} \leq p_{.j}$, то абсолютная сходимость ряда (11.3) вытекает из абсолютной сходимости ряда (11.2).

Аналогично определяется условное математическое ожидание случайной величины ξ при условии, что произошло событие $\{\eta = y_j\}$:

$$M(\xi/\eta = y_j) = \sum_i x_i\, p_{i/j} = \sum_i x_i\, P\{\xi = x_i/\eta = y_j\} =$$

$$= \frac{1}{p_{.j}} \sum_i x_i\, p_{ij}. \qquad (11.4)$$

в предположении существования $M\xi$, где

$$p_{i/j} = P\{\xi = x_i/\eta = y_j\} = \frac{P\{\xi = x_i,\ \eta = y_j\}}{P\{\eta = y_j\}} =$$

$$= \frac{p_{ij}}{p_{.j}} = \frac{p_{ij}}{\sum\limits_i p_{ij}}.$$

Поскольку для различных значений i числа x_i различны, то различны также события $\{\xi = x_i\}$. Это означает, что для различных i условные математические ожидания $M(\eta/\xi = x_i)$, вообще говоря, отличаются. Таким образом, $M(\eta/\xi = x_i)$ является функцией от x_i, которую обозначим через $f(x_i)$:

$$f(x_i) = M(\eta/\xi = x_i). \qquad (11.5)$$

Определение 11.2. Пусть случайная величина η имеет математическое ожидание. **Условным математическим ожиданием случайной величины** η **при условии** ξ или **относительно** ξ называется $f(\xi)$, где функция f определяется равенством (11.5).

Так как $\xi = \xi(\omega)$ является функцией от элементарного исхода $\omega \in \Omega$, то $M(\eta/\xi)$ является сложной функцией $f(\xi(\omega))$, которая на множествах $\{\omega : \xi(\omega) = x_i\}$ принимает значения $f(x_i) = M(\eta/\xi = x_i)$. Таким образом, $M(\eta/\xi)$ — дискретная случайная величина, принимающая значения $M(\eta/\xi = x_i)$ с вероятностями $p_{i.} = P(\{\omega : \xi(\omega) = x_i\}) = P\{\xi = x_i\}$.

Свойства условного математического ожидания.

1. Если $\eta(\omega) = C$ — постоянная (неслучайная) величина, то

$$M(\eta/\xi) = M(C/\xi) = C.$$

Действительно, на всех множествах $\{\omega : \xi(\omega) = x_i\}$, образующих разбиение Ω, случайная величина $M(\eta/\xi)$ принимает одно и то же значение C в силу того, что

$$M(C/\xi = x_i) = C\,P\{\eta = C/\xi = x_i\} = C\,P\{\Omega/\xi = x_i\} = C$$

(условная вероятность достоверного события равна единице).

2. $M(\eta_1 + \eta_2/\xi) = M(\eta_1/\xi) + M(\eta_2/\xi)$.

Доказательство этого свойства аналогично доказательству соответствующего свойства для безусловных (обычных) математических ожиданий.

3. Имеет место формула **полного математического ожидания**:

$$MM(\eta/\xi) = M\eta.$$

В самом деле, так как $M(\eta/\xi)$ — дискретная случайная величина, принимающая значения $M(\eta/\xi = x_i)$ с вероятностями $p_{i.}$, то с учетом формул (11.1)-(11.3) ее математическое ожидание равно

$$MM(\eta/\xi) = \sum_i M(\eta/\xi = x_i)\, p_{i.} = \sum_i \sum_j \frac{1}{p_{i.}} y_j\, p_{ij}\, p_{i.} =$$

$$= \sum_i \sum_j y_j\, p_{ij} = \sum_j \sum_i y_j\, p_{ij} = \sum_j y_j \sum_i p_{ij} =$$

$$= \sum_j y_j\, p_{.j} = M\eta.$$

Возможность изменения порядка суммирования следует из абсолютной сходимости двойного ряда.

Замечание 1. Рассмотрим частный случай, когда η является индикатором события A, т.е.

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

Предположим, что события H_1, H_2, $\ldots$ H_n образуют полную группу, т.е. $H_1 + H_2 + \ldots + H_n = \Omega$. Пусть случайная величина ξ принимает на множествах H_1, H_2, $\ldots$ H_n соответственно значения $1, 2, \ldots, n$. На множестве $H_i = \{\omega : \xi(\omega) = i\}$ случайная величина $M(I_A/\xi)$ принимает значение $M(I_A/\xi = i) = 1 \cdot P\{I_A = 1/\xi = i\} + 0 \cdot P\{I_A = 0/\xi = i\} = P\{A/\xi = i\} = P(A/H_i)$. Поэтому ее математическое ожидание равно $MM(I_A/\xi) = \sum\limits_i P(A/H_i)P(H_i)$. С другой стороны, $MI_A = 1 \cdot P(I_A = 1) + 0 \cdot P(I_A =$

$0) = P(A)$. Таким образом, свойство 3 в рассматриваемом частном случае превращается в обычную формулу полной вероятности:

$$P(A) = \sum_i P(H_i) P(A/H_i).$$

4. Если $a(x)$ — борелевская функция, то

$$M(a(\xi)\eta/\xi) = a(\xi)M(\eta/\xi).$$

Сначала поясним смысл на интуитивном уровне. При фиксированнном значении ξ, равном x_i, множитель $a(\xi)$ превращается в постоянную (неслучайную, т.е. не зависящую от ω) величину. Постоянную можно выносить за знак математического ожидания, следовательно

$$M(a(\xi)\,\eta/\xi = x_i) = M(a(x_i)\,\eta/\xi = x_i) = a(x_i)M(\eta/\xi = x_i),$$

откуда и следует доказываемое свойство.

Теперь дадим строгое доказательство. Занумеруем произвольным образом все возможные значения случайной величины $a(\xi)\,\eta$ с помощью индекса s так, что когда ξ и η пробегают множества своих значений x_k и y_j соответственно, она пробегает значения z_s из не более чем счетного множества. Изменение порядка суммирования и группировка членов ряда, производимые далее, законны в силу его абсолютной сходимости. Имеем:

$$M(a(\xi)\,\eta/\xi = x_i) = \sum_s z_s\, P(a(\xi)\,\eta = z_s/\xi = x_i) =$$

$$= \sum_s \; z_s \sum_{(k,j):\, z_s = a(x_k)\, y_j} P(\xi = x_k,\, \eta = y_j/\xi = x_i) =$$

$$= \sum_{(k,j)} \sum_{s:\, z_s = a(x_k)\, y_j} z_s\, P(\xi = x_k,\, \eta = y_j/\xi = x_i) =$$

$$= \sum_k \sum_j a(x_k)\, y_j\, P(\xi = x_k,\, \eta = y_j/\xi = x_i) =$$

$$= \sum_j a(x_i)\, y_j\, P(\eta = y_j/\xi = x_i) = a(x_i)\, M(\eta/\xi = x_i).$$

Здесь учитывалось, что при фиксированных x_k и y_j существует единственное s, для которого $z_s = a(x_k)\, y_j$, и тот факт, что

$$P(\xi = x_k,\, \eta = y_j/\xi = x_i) = \frac{P(\xi = x_k,\, \eta = y_j,\, \xi = x_i)}{P(\xi = x_i)} =$$

$$= \delta_{ki}\, \frac{P(\eta = y_j,\, \xi = x_i)}{P(\xi = x_i)} = \delta_{ki}\, P(\eta = y_j/\xi = x_i),$$

где δ_{ki} — символ Кронекера. Значит на множествах $\{\omega : \xi(\omega) = x_i\}$ случайные величины $M(a(\xi)\, \eta/\xi)$ и $a(\xi)\, M(\eta/\xi)$ совпадают. Так как эти множества образуют разбиение пространства элементарных исходов Ω, то указанные случайные величины совпадают на всем Ω.

5. Если ξ и η — независимые случайные величины, то $M(\eta/\xi) = M\eta$.

Действительно,

$$M(\eta/\xi = x_i) = \sum_j y_j\, P(\eta = y_j/\xi = x_i) =$$

$$= \sum_j y_j\, P(\eta = y_j) = M\eta.$$

Рассмотрим теперь абсолютно непрерывный случайный вектор (ξ, η) с плотностью распределения вероятностей $p_{\xi,\eta}(x,y)$. Частные плотности распределения вероятностей компонент этого вектора, как мы знаем, определяются соотношениями

$$p_\xi(x) = \int\limits_{-\infty}^{+\infty} p_{\xi,\eta}(x,y)dy; \quad p_\eta(y) = \int\limits_{-\infty}^{+\infty} p_{\xi,\eta}(x,y)dx.$$

Определение 11.3. Пусть случайная величина η имеет математическое ожидание. **Условным математическим ожиданием случайной величины η при условии, что случайная величина ξ приняла значение** x, называется число

$$M(\eta/\xi = x) = \int\limits_{-\infty}^{+\infty} y\, p_{\eta/\xi}(y/x)dy,$$

где

$$p_{\eta/\xi}(y/x) = \frac{p_{\xi,\eta}(x,y)}{p_\xi(x)} \ —$$

так называемая **условная плотность распределения вероятностей** случайной величины η при условии, что случайная величина ξ приняла значение x.

Абсолютная сходимость интеграла, участвующего в определении, следует из абсолютной сходимости интеграла $M|\eta|$.

При фиксированном значении x условное математическое ожидание является числом. Так как для различных x эти числа, вообще говоря, различны, то

$M(\eta/\xi = x)$ является функцией от x. Если в эту функцию вместо x подставить саму случайную величину ξ, то получится случайная величина, которую называют условным математическим ожиданием η относительно ξ.

Определение 11.4. Пусть случайная величина η имеет математическое ожидание. **Условным математическим ожиданием случайной величины** η **при условии** ξ или **относительно** ξ называется $f(\xi)$, где функция f определяется равенством $f(x) = M(\eta/\xi = x)$.

Свойства условных математических ожиданий, доказанные для случая дискретного случайного вектора, остаются в силе и для случая абсолютно непрерывного вектора.

11.2. Регрессия. Основная теорема регрессионного анализа. Уравнения линейной регрессии

Определение 11.5. Уравнение $\eta = f(\xi) = M(\eta/\xi)$ называется **уравнением регрессии** η **на** ξ, а уравнение $\xi = g(\eta) = M(\xi/\eta)$ — **уравнением регрессии** ξ **на** η.

Регрессия называется **параболической**, если $f(x)$ — многочлен не менее второй степени, и **линейной**, если $f(x) = a_0 + a_1 x$, где $a_1 \neq 0$. Регрессия η на ξ показывает как в среднем меняется величина η при изменении величины ξ.

Основная теорема регрессионного анализа (принцип наименьших квадратов).

Наилучшим прогнозом случайной величины η по случайной величине ξ в среднем квадратическом смысле является условное математическое ожидание $M(\eta/\xi)$ (наилучший в этом смысле прогноз дается уравнением регрессии).

Поясним более подробно формулировку теоремы. Если в пространстве случайных величин ввести среднюю квадратическую метрику $\rho(\xi, \eta) = (M(\xi - \eta)^2)^{\frac{1}{2}}$ (метрику L_2) и называть прогнозом η по ξ любую функцию $h(\xi) \in L_2$, то в теореме утверждается, что из всех возможных прогнозов η по ξ условное математическое ожидание $M(\eta/\xi)$ является самым близким к прогнозируемой величине η по этому расстоянию: $M(\eta - f(\xi))^2 \leq M(\eta - h(\xi))^2$, где $f(\xi) = M(\eta/\xi)$.

Доказательство. По свойству линейности математического ожидания

$$M(\eta - h(\xi))^2 = M(\eta - f(\xi) + f(\xi) - h(\xi))^2 =$$

$$= M(\eta - f(\xi))^2 + M[(\eta - f(\xi))(f(\xi) - h(\xi))] + M(f(\xi) - h(\xi))^2.$$

По свойствам 4 и 1 условного математического ожидания

$$M[(\eta - f(\xi))(f(\xi) - h(\xi))/\xi] = [f(\xi) - h(\xi)]\, M[(\eta - f(\xi))/\xi] =$$

$$= [f(\xi) - h(\xi)]\, [M(\eta/\xi) - f(\xi)\, M(1/\xi)] =$$

$$= [f(\xi) - h(\xi)]\, [M(\eta/\xi) - f(\xi)] = 0,$$

По формуле полного математического ожидания (свойство 3)

$$M[(\eta - f(\xi))(f(\xi) - h(\xi))] = 0.$$

Следовательно,

$$M(\eta - h(\xi))^2 = M(\eta - f(\xi))^2 + M(f(\xi) - h(\xi))^2 \geq M(\eta - f(\xi))^2.$$

Теорема доказана. $\square$

Рассмотрим случай линейной регрессии η на ξ:

$$f(\xi) = M(\eta/\xi) = a_0 + a_1\,\xi. \qquad (11.6)$$

Введем обозначения:

$$a = M\xi;\;\; b = M\eta;\;\; \sigma_\xi = \sqrt{D\xi};\;\; \sigma_\eta = \sqrt{D\eta};\;\; r = \frac{cov(\xi,\, \eta)}{\sigma_\xi\,\sigma_\eta}.$$

По свойству 3 условных математических ожиданий

$$b = M\eta = MM(\eta/\xi) = a_0 + a_1\,M\xi = a_0 + a_1\,a. \qquad (11.7)$$

Вычитая (11.7) из (11.6) и используя свойство 2, получим:

$$M(\eta - b/\xi) = a_1\,(\xi - a). \qquad (11.8)$$

Умножая (11.8) на $\xi - a$ и применяя свойство 4, имеем

$$M[(\xi - a)\,(\eta - b)/\xi] = a_1\,(\xi - a)^2. \qquad (11.9)$$

Беря математическое ожидание от обеих частей (11.9) и используя свойство 3, получим:

$$M[(\xi - a)(\eta - b)] = a_1\,M(\xi - a)^2$$

или, что то же самое, $r\,\sigma_\xi\,\sigma_\eta = a_1\,\sigma_\xi^2$, откуда

$$a_1 = r\,\frac{\sigma_\eta}{\sigma_\xi}.$$

Таким образом,

$$M(\eta - b/\xi) = r\,\frac{\sigma_\eta}{\sigma_\xi}\,(\xi - a).$$

Поэтому уравнение линейной регрессии η на ξ имеет вид

$$\eta - b = r\,\frac{\sigma_\eta}{\sigma_\xi}\,(\xi - a). \tag{11.10}$$

Аналогично, уравнение линейной регрессии ξ на η будет иметь форму

$$\xi - a = r\,\frac{\sigma_\xi}{\sigma_\eta}\,(\eta - b). \tag{11.11}$$

Основная теорема регрессионного анализа показывает, что в случае линейной регрессии η на ξ уравнение (11.10) дает наилучший в среднем квадратическом смысле прогноз случайной величины η по случайной величине ξ. Ошибка такого приближения равна

$$\delta_{\eta/\xi} = \eta - b - r\frac{\sigma_\eta}{\sigma_\xi}(\xi - a).$$

Дисперсия этой ошибки —

$$\sigma^2_{\eta/\xi} = M(\delta_{\eta/\xi} - M\delta_{\eta/\xi})^2 = M\delta^2_{\eta/\xi} = M[\eta - b - r\frac{\sigma_\eta}{\sigma_\xi}(\xi - a)]^2 =$$

$$= M(\eta - b)^2 - 2r\frac{\sigma_\eta}{\sigma_\xi}M[(\xi - a)(\eta - b)] + r^2\frac{\sigma^2_\eta}{\sigma^2_\xi}M(\xi - a)^2 =$$

$$= \sigma^2_\eta - 2r\frac{\sigma_\eta}{\sigma_\xi}r\sigma_\xi\sigma_\eta + r^2\frac{\sigma^2_\eta}{\sigma^2_\xi}\sigma^2_\xi = \sigma^2_\eta(1 - r^2).$$

Итак, среднеквадратическое отклонение ошибки приближения в уравнении регрессии η на ξ равно

$$\sigma_{\eta/\xi} = \sigma_\eta\sqrt{1 - r^2}. \tag{11.12}$$

Аналогично, среднеквадратическое отклонение ошибки приближения в уравнении регрессии ξ на η равно

$$\sigma_{\xi/\eta} = \sigma_\xi \sqrt{1 - r^2}. \tag{11.13}$$

Из уравнений линейной регрессии видно, что обе прямые регрессии проходят через центр совместного распределения $(M\xi,\, M\eta) = (a,\, b)$. Угловые коэффициенты прямых регрессии равны

$$tg\, \alpha = r \frac{\sigma_\eta}{\sigma_\xi}; \quad tg\, \beta = \frac{1}{r} \frac{\sigma_\eta}{\sigma_\xi}.$$

Оба угловых коэффициента одного знака с r, следовательно, либо у обеих прямых угол наклона к оси OX острый, либо у обеих — тупой. При $r = 0$ из (11.10) видно, что среднеквадратическое отклонение ошибки приближения η на ξ совпадает со среднеквадратическим отклонением случайной величины η. Это означает, что прогноз с помощью уравнения регрессии (11.10) такой же неопределенный как сама случайная величина η, т.е. уравнение фактически не прогнозирует значения случайной величины η. Уравнения регрессии превращаются в

$$\eta = b, \quad \xi = a.$$

т.е. линии регрессии (11.10), (11.11) перпендикулярны. При этом прогноз η по ξ выражается как математическое ожидание η и не использует никакой информации о случайной величине ξ. Из (11.12), (11.13) видно, что при увеличении r^2 от 0 до 1 среднеквадратические ошибки приближения уравнениями регрессии убывают до 0, превращаясь в него при $r^2 = 1$. Угол между линиями при этом убывает от $90^o C$ до 0^o. Заметим, что прямые регрессии совпадают тогда и только тогда, когда $r^2 = 1$. В этом случае, как мы знаем, η и ξ с вероятностью 1 связаны линейной зависимостью, которая как раз выражается совпадающими между собой уравнениями регрессии. Таким образом, уравнение регрессии выражает в этом случае точный прогноз. Чем ближе r^2 к 1, тем лучше качество прогноза уравнениями регрессии, чем ближе r к 0, тем качество прогноза хуже.

Предположим теперь, что регрессия η на ξ не является линейной:

$$M(\eta/\xi) \neq a_0 + a_1\,\xi;$$

тем не менее мы хотим построить наилучший в смысле среднего квадратического отклонения линейный прогноз для величины η. Другими словами, требуется подобрать a_0 и a_1 таким образом, чтобы $M(\eta - a_0 - a_1\,\xi)^2$ принимало наименьшее значение. Имеем

$$M(\eta - a_0 - a_1\,\xi)^2 = M[(\eta - b) - a_1(\xi - a) + (b - a_1\,a - a_0)]^2 =$$
$$= \sigma_\eta^2 + a_1^2\sigma_\xi^2 + (b - a_1\,a - a_0)^2 - 2a_1 r\sigma_\xi\sigma_\eta.$$

Для того, чтобы это выражение давало минимум, необходимо, чтобы частные производные по a_0 и a_1 обращались в 0:

$$b - a_1 a - a_0 = 0;$$
$$2a_1\sigma_\xi^2 - 2a_1 r\sigma_\xi\sigma_\eta = 0;$$

откуда

$$a_1 = r\frac{\sigma_\eta}{\sigma_\xi}; \quad a_0 = b - r\frac{\sigma_\eta}{\sigma_\xi}a.$$

Легко проверяется, что найденная точка — точка минимума. Поэтому наилучшее линейное приближение для η по ξ дается с помощью

$$\eta = a_0 + a_1\,\xi = b - r\frac{\sigma_\eta}{\sigma_\xi}a + r\frac{\sigma_\eta}{\sigma_\xi}\,\xi,$$

что совпадает с уравнением линейной регрессии (11.10) величины η на ξ.

Таким образом, уравнения линейной регрессии дают наилучший прогноз среди всех линейных прогнозов даже в том случае, когда регрессия не является линейной.

11.3. Выборочные уравнения линейной регрессии

Пусть имеется выборка (x_1, y_1), (x_2, y_2), ... , $(x_n, y_n,)$ неэависимых наблюдений случайного вектора (ξ, η). На практике числовые характеристики $a = M\xi$, $b = M\eta$, $\sigma_\xi = \sqrt{D\xi}$, $\sigma_\eta = \sqrt{D\eta}$, $r = cov(\xi, \eta)/(\sigma_\xi \sigma_\eta)$, как правило, неизвестны, а мы можем построить только их оценки

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

$$S_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}; \quad S_y = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2};$$

$$r_{\text{в}} = \frac{\widehat{cov(\xi, \eta)}}{S_x S_y}.$$

Здесь

$$\widehat{cov(\xi, \eta)} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) -$$

оценка $cov(\xi, \eta)$. Поэтому уравнения линейной регрессии построить невозможно. Но мы можем подставить в эти уравнения вместо параметров указанные оценки, а вместо ξ и η записать x и y. Получатся уравнения, которые в определенном смысле являются приближенными уравнениями линейной регрессии и называются **выборочными уравнениями линейной регрессии**.

Итак, выборочные уравнения линейной регрессии η на ξ имеют вид

$$y - \overline{y} = r_{\text{в}} \frac{S_y}{S_x} (x - \overline{x}),$$

а выборочные уравнения линейной регрессии ξ на η — вид

$$x - \overline{x} = r_{\text{в}} \frac{S_x}{S_y} (y - \overline{y}).$$

Замечание 2. При записи выборочных уравнений регрессии можно пользоваться смещенными оценками дисперсий

$$\tilde{S}_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2; \quad \tilde{S}_y^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y})^2.$$

Но тогда для оценки квадрата ковариации также надо воспользоваться смещенной оценкой

$$\widehat{cov_1(\xi, \eta)} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}) (y_i - \overline{y}).$$

Так как угловые коэффициенты в уравнениях регрессии при этом сохраняются, то и выборочные уравнения регрессии не изменятся.

Замечание 3. Предположим, что случайный вектор (ξ, η) имеет двумерное нормальное распределение. В этом случае говорят о так называемой нормальной корреляции. Рассмотрим случайный вектор

$$\theta = \eta - b - r\,\frac{\sigma_\eta}{\sigma_\xi}\,(\xi - a).$$

Имеем

$$M[\theta(\xi - a)] = M[(\xi - a)(\eta - b)] - r\,\sigma_\eta\sigma_\xi = 0.$$

Так как вектор $(\theta,\ \xi - a)$ имеет двумерное нормальное распределение и его компоненты некоррелированы, то они независимы. Следовательно, независимы также θ и ξ, и по свойству 5 условных математических ожиданий $M(\theta/\xi) = M\theta = 0$, т.е.

$$M(\theta/\xi) = M(\eta/\xi) - b - r\,\frac{\sigma_\eta}{\sigma_\xi}\,(\xi - a) = 0,$$

откуда

$$M(\eta/\xi) = b + r\,\frac{\sigma_\eta}{\sigma_\xi}\,(\xi - a).$$

Таким образом, при нормальной корреляции регрессия является обязательно линейной.

Можно показать, что при нормальной корреляции в случае большого объема выборки n с надежностью γ для коэффициента корреляции r и коэффициента регрессии $k_{\eta/\xi} = r\,\frac{\sigma_\eta}{\sigma_\xi}$ получаются следующие доверительные интервалы:

$$r_\text{в} - \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big)\,\frac{1 - r_\text{в}^2}{\sqrt{n}} < r < r_\text{в} + \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big)\,\frac{1 - r_\text{в}^2}{\sqrt{n}};$$

$$k^\text{в}_{y/x} - \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big)\,\frac{1 - r_\text{в}^2}{\sqrt{n}}\,\frac{S_y}{S_x} < k_{\eta/\xi} <$$

$$< k^{\text{в}}_{y/x} + \Phi^{-1}\Big(\frac{1+\gamma}{2}\Big)\,\frac{1-r_{\text{в}}^2}{\sqrt{n}}\,\frac{S_y}{S_x},$$

где $k^{\text{в}}_{y/x} = r_{\text{в}}\,(S_y/S_x)$ — выборочный коэффициент регрессии η на ξ, являющийся оценкой коэффициента регрессии $k_{\eta/\xi}$, $\Phi(x)$ — функция Лапласа.

Задача 1. Пользуясь приведенными интервальными оценками, построить критерии значимости для проверки гипотез $H_0:\ r = r_0$ и $H_0:\ k_{\eta/\xi} = k_0$, где $0 \le r_0 \le 1$.

Укажем еще один способ получения выборочных уравнений линейной регрессии. Поскольку уравнения линейной регрессии выражают наилучший линейный прогноз даже в случае нелинейной регрессии, то можно надеяться, что похожая ситуация будет сохраняться и для выборочных уравнений линейной регрессии. Еще Гаусс для эмпирического определения наилучшего линейного прогноза η по ξ по результатам наблюдений предложил так называемый **метод наименьших квадратов**. При линейном прогнозе

$$\tilde{\eta} = A\xi + B$$

в i-м наблюдении мы делаем ошибку $\varepsilon_i = y_i - Ax_i - B$. Коэффициенты A и B находятся из условия, чтобы сумма квадратов ошибок была минимальной:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - Ax_i - B)^2 \quad \longrightarrow \quad \min.$$

Необходимым условием минимума функции двух переменных A и B является равенство ее частных производных по A и B нулю:

$$\sum_{i=1}^{n} x_i\,(y_i - Ax_i - B) = 0;$$

$$\sum_{i=1}^{n} (y_i - Ax_i - B) = 0.$$

Для определения неизвестных коэффициентов A и B получается система так называемых **нормальных уравнений**:

$$(\sum_{i=1}^{n} x_i^2)\,A + (\sum_{i=1}^{n} x_i)\,B = \sum_{i=1}^{n} x_i\,y_i;$$

$$(\sum_{i=1}^{n} x_i)\,A + n\,B = \sum_{i=1}^{n} y_i.$$

Если найти A и B, то окажется, что уравнение $y = Ax+B$ совпадет с выборочным уравнением линейной регрессии η на ξ.

Задача 2. Докажите этот результат.

На практике для установления вида регрессионной зависимости изображают точки (x_1, y_1), (x_2, y_2), … , (x_n, y_n) и по графику прикидывают, вокруг какой кривой (по форме) они группируются. Неизвестные коэффициенты находят методом наименьших квадратов.

Пусть, например, регрессионная зависимость имеет вид $y = C\,e^{Ax} + B$. Сумма квадратов ошибок равна

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - C\,e^{Ax_i} - B)^2.$$

Для того, чтобы она была минимальна, необходимо, чтобы ее частные производные по B, C и A обращались в ноль:

$$\sum_{i=1}^{n} (y_i - C\, e^{Ax_i} - B) = 0;$$

$$\sum_{i=1}^{n} (y_i - C\, e^{Ax_i} - B)\, e^{Ax_i} = 0;$$

$$\sum_{i=1}^{n} (y_i - C\, e^{Ax_i} - B)\, e^{Ax_i}\, x_i = 0.$$

Остается только решить эту систему трансцендентных уравнений.

§ 12. Достаточные статистики. Необходимая и полная достаточные статистики. Экспоненциальные семейства. Теорема Колмогорова-Блекуэлла-Рао

При больших объемах статистической информации важное значение имеет сокращение имеющихся данных. Желательно, чтобы такое сокращение не приводило к потере информации. Оказывается, что во многих случаях это можно сделать с помощью так называемых достаточных статистик, количество информации по Фишеру в которых совпадает с количеством информации в выборке.

Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения, зависящего от неизвестного параметра θ (возможно векторного, т.е. $\theta = (\theta_1, \theta_2, \ldots, \theta_m)$). Пусть $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ — вектор наблюдений, а

$$\mathbf{T}(\mathbf{x}) = (T_1(x_1, x_2, \ldots, x_n), T_2(x_1, x_2, \ldots, x_n), \ldots$$

$$\ldots, T_r(x_1, x_2, \ldots, x_n)) \quad —$$

векторная статистика, т.е. $T_i(\mathbf{x}) : \ \mathbb{R}^n \to \mathbb{R}$ — борелевские функции.

Определение 12.1. Статистика $\mathbf{T}(\mathbf{x}) = \mathbf{T}(x_1, x_2, \ldots, x_n)$ называется **достаточной**, если при заданном значении статистики $\mathbf{T}$ распределение вектора наблюдений $\mathbf{x}$ не зависит от θ, т.е. для любого $B \in \mathcal{B}$ условная вероятность

$$P\{\mathbf{x} \in B/\mathbf{T}(\mathbf{x}) = \mathbf{T}\}$$

не зависит от θ (и, значит, не несет никакой информации относительно θ).

Теорема факторизации Неймана. *Векторная статистика $T(\mathbf{x}) = \mathbf{T}(x_1, x_2, \ldots, x_n)$ является достаточной статистикой тогда и только тогда, когда существуют борелевские функции $h(\mathbf{x})$ и $g_\theta(\mathbf{y})$ такие, что функция правдоподобия представляется в виде*

$$L(\mathbf{x}/\theta) = g_\theta(\mathbf{T}(\mathbf{x}))h(\mathbf{x}). \tag{12.1}$$

Подчеркнем, что здесь функция $h(\mathbf{x})$ не зависит от θ.

Д о к а з а т е л ь с т в о проведем только для случая, когда теоретическая случайная величина дискретна (выборка взята из дискретного распределения).

Необходимость. Пусть $\mathbf{T}$ — достаточная статистика, $\mathbf{x} \in \mathbb{R}$ и $\mathbf{T}(\mathbf{x}) = \mathbf{t}$. Тогда, если $\vec{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ — теоретический вектор наблюдений, то

$$L(\theta) = L(\mathbf{x}/\theta) = P\{\vec{\xi} = \mathbf{x}\} = P\{\vec{\xi} = \mathbf{x}, \ \mathbf{T}(\mathbf{x}) = \mathbf{t}\} =$$

$$= P\{\vec{\xi} = \mathbf{x}, \ \mathbf{T}(\vec{\xi}) = \mathbf{t}\} = P\{\mathbf{T}(\vec{\xi}) = \mathbf{t}\}P\{\vec{\xi} = \mathbf{x}/\mathbf{T}(\vec{\xi}) = \mathbf{t}\} =$$

$$= g_\theta(\mathbf{t})h(\mathbf{x}) = g_\theta(\mathbf{T}(\mathbf{x}))h(\mathbf{x}),$$

где $h(\mathbf{x}) = P\{\vec{\xi} = \mathbf{x}/\mathbf{T}(\vec{\xi}) = \mathbf{t}\}$ не зависит от θ, так как $\mathbf{T}$ — достаточная статистика.

Достаточность. Пусть $L(\mathbf{x}/\theta) = g_\theta(\mathbf{T}(\mathbf{x}))h(\mathbf{x})$. Если $\mathbf{T}(\mathbf{x}) = \mathbf{t}$ и $P\{\mathbf{T}(\vec{\xi}) = \mathbf{t}\} \neq 0$, то

$$P\{\vec{\xi} = \mathbf{x}/\mathbf{T}(\vec{\xi}) = \mathbf{t}\} = \frac{P\{\vec{\xi} = \mathbf{x}, \mathbf{T}(\vec{\xi}) = \mathbf{t}\}}{P\{\mathbf{T}(\vec{\xi}) = \mathbf{t}\}} =$$

$$= \frac{P\{\vec{\xi} = \mathbf{x}, \mathbf{T}(\mathbf{x}) = \mathbf{t}\}}{P\{\mathbf{T}(\vec{\xi}) = \mathbf{t}\}} = \frac{P\{\vec{\xi} = \mathbf{x}\}}{P\{\mathbf{T}(\vec{\xi}) = \mathbf{t}\}} =$$

$$= \frac{P\{\vec{\xi} = \mathbf{x}\}}{\sum\limits_{\mathbf{z}:\ \mathbf{T}(\mathbf{z})=\mathbf{t}} P\{\vec{\xi} = \mathbf{z}\}} = \frac{g_\theta(\mathbf{T}(\mathbf{x}))h(\mathbf{x})}{\sum\limits_{\mathbf{z}:\ \mathbf{T}(\mathbf{z})=\mathbf{t}} g_\theta(\mathbf{T}(\mathbf{z}))h(\mathbf{z})} =$$

$$= \frac{g_\theta(\mathbf{t})h(\mathbf{x})}{\sum\limits_{\mathbf{z}:\ \mathbf{T}(\mathbf{z})=\mathbf{t}} g_\theta(\mathbf{t})h(\mathbf{z})} = \frac{h(\mathbf{x})}{\sum\limits_{\mathbf{z}:\ \mathbf{T}(\mathbf{z})=\mathbf{t}} h(\mathbf{z})}$$

не зависит от θ. □

Замечание 1. Для выборки из абсолютно непрерывного распределения функцией правдоподобия мы называли плотность случайного вектора наблюдений $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, а для выборки из дискретного распределения — вероятность принятия координатами этого вектора своих значений (с подстановкой в них вместо переменных случайных величин $x_1, x_2, \ldots, x_n$ — элементов абстрактной выборки). Точно так же можно ввести **функцию правдоподобия для достаточной статистики $\mathbf{T}$** — $L_\mathbf{T}(\mathbf{T}(\mathbf{x})/\theta)$ — плотность случайного вектора $\mathbf{T}(\mathbf{x})$ в абсолютно непрерывном случае и вероятность принятия этим вектором своих значений в дискретном случае. Пусть θ — скалярный параметр. Продолжая эту аналогию, назовем **информацией по Фишеру** в достаточной статистике $\mathbf{T}$ число

$$I_{\mathbf{T}}(\theta) = M\left|\frac{\partial \ln L_{\mathbf{T}}(\mathbf{T}(\mathbf{x})/\theta)}{\partial\theta}\right|^2.$$

Как видно из доказательства необходимости теоремы Неймана $g_\theta(\mathbf{t}) = L_{\mathbf{T}}(\mathbf{t}/\theta)$. Поэтому

$$I_{\mathbf{T}}(\theta) = M\left|\frac{\partial \ln g_\theta(\mathbf{T}(\mathbf{x}))}{\partial\theta}\right|^2.$$

В силу (12.1)

$$\ln L(\mathbf{x}/\theta) = \ln g_\theta(\mathbf{T}(\mathbf{x})) + \ln h(\mathbf{x}),$$

а так как $h(\mathbf{x})$ не зависит от θ, то

$$\frac{\partial \ln L(\mathbf{x}/\theta)}{\partial\theta} = \frac{\partial \ln g_\theta(\mathbf{T}(\mathbf{x}))}{\partial\theta}.$$

Поэтому

$$I_{\mathbf{T}}(\theta) = M\left|\frac{\partial \ln g_\theta(\mathbf{T}(\mathbf{x}))}{\partial\theta}\right|^2 = M\left|\frac{\partial \ln L(\mathbf{x}/\theta)}{\partial\theta}\right|^2 = I(\theta),$$

т.е. информация о параметре θ в достаточной статистике $\mathbf{T}(\mathbf{x})$ совпадает с информацией в выборке.

Замечание 2. Уравнение правдоподобия

$$\frac{\partial \ln L(\mathbf{x}/\theta)}{\partial\theta_i} = 0, \quad i = \overline{1,m},$$

эквивалентно уравнению

$$\frac{\partial \ln g_\theta(\mathbf{T}(\mathbf{x}))}{\partial\theta_i} = 0, \quad i = \overline{1,m}.$$

Поэтому оценки максимального правдоподобия, являющиеся его решением, зависят от выборочных значений только через достаточную статистику.

Пример 1. Пусть $x_1, x_2, \dots, x_n$ — выборка из равномерного распределения на отрезке $[0, 1]$. Если $\mathbf{T} = (x_{(1)}, x_{(n)})$, где $x_{(1)} = \min(x_1, x_2, \dots, x_n)$, $x_{(n)} = \max(x_1, x_2, \dots, x_n)$ — крайние порядковые статистики, то

$$L(\mathbf{x}/a, b) = \begin{cases} \frac{1}{(b-a)^n}, & \text{если } x_{(1)} \geq a,\ x_{(n)} \leq b, \\ 0, & \text{в иных случаях} \end{cases} =$$

$$= g_{a,b}(\mathbf{T}) \cdot h(\mathbf{x}),$$

где $h(\mathbf{x}) = 1$. Поэтому $\mathbf{T}$ — достаточная статистика.

Пример 2. Пусть $x_1, x_2, \dots, x_n$ — выборка из нормального распределения $N(a, \sigma^2)$.

1. Пусть a неизвестна, σ^2 известна. Тогда

$$L(\mathbf{x}/a) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} (x_i - a)^2} = g_a(\overline{x}) \cdot h(\mathbf{x}),$$

где

$$g_a(\overline{x}) = e^{\frac{na}{\sigma^2}(\overline{x} - \frac{a}{2})}, \quad h(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} x_i^2}.$$

Следовательно, $\overline{x}$ — достаточная статистика.

2. Пусть σ^2 неизвестна, a известна. Тогда

$$L(\mathbf{x}/\sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} (x_i - a)^2} = g_{\sigma^2}(S^{*2}) \cdot h(\mathbf{x}),$$

где

$$S^{*^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2, \quad g_{\sigma^2}(S^{*^2}) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{n}{2\sigma^2}S^{*^2}},$$

$$h(\mathbf{x}) = 1.$$

По теореме Неймана S^{*^2} — достаточная статистика.

3. Пусть a и σ^2 неизвестны. Тогда

$$L(\mathbf{x}/a, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i - \overline{x} + \overline{x} - a)^2} =$$

$$= \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\left[\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - a)^2\right]} = g_{a,\sigma^2}(\overline{x}, S^2) \cdot h(\mathbf{x}),$$

где

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

$$g_{a,\sigma^2}(\overline{x}, S^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{1}{2\sigma^2}\left[(n-1)S^2 + n(\overline{x} - a)^2\right]}, \quad h(\mathbf{x}) = 1.$$

По теореме факторизации $(\overline{x},\ S^2)$ — достаточная статистика.

Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения $F(x/\theta)$, зависящего от неизвестного параметра θ. Мы называли векторной статистикой $\mathbf{T} = \mathbf{T}(x_1, x_2, \ldots, x_n) = (T_1(x_1, x_2, \ldots, x_n), \ldots, T_r(x_1, x_2, \ldots, x_n))$, где T_i — борелевские функции от выборочных значений. При этом она не зависела от неизвестного параметра θ. В этом разделе временно разрешим ей зависеть от θ : $\mathbf{T} = \mathbf{T}(\mathbf{x}/\theta)$ и будем называть ее θ-статистикой.

Определение 12.2. θ-статистика $\mathbf{T}(\mathbf{x}/\theta)$ называется **достаточной**, если

$$L(\mathbf{x}/\theta) = g_\theta(\mathbf{T}(\mathbf{x}/\theta)) \cdot h(\mathbf{x}).$$

В силу факторизационной теоремы Неймана это определение согласуется с тем, которое мы вводили для статистики (т.е. когда $\mathbf{T}(\mathbf{x}/\theta)$ не зависело от θ).

Определение 12.3. θ-статистика $\mathbf{T}^*(\mathbf{x}/\theta)$ называется **необходимой** или **минимальной**, если она определяется значениями любой достаточной статистики.

Смысл этого определения таков. Если $\mathbf{T}$ — достаточная статистика, то для любой статистики $\mathbf{t}$ статистика $(\mathbf{T},\ \mathbf{t})$ тоже является достаточной. Предпочтительнее иметь дело только с $\mathbf{T}$, а не с $(\mathbf{T},\ \mathbf{t})$, поскольку наша цель — представить данные в сжатом виде. Для необходимой статистики характерно то, что при сохранении свойства достаточности невозможно дальнейшее сокращение данных по сравнению с $\mathbf{T}^*$.

Рассмотрим вопрос о существовании необходимой и достаточной статистики. Для θ-статистик этот вопрос решается тривиально.

Теорема 12.4. *Если $L(\mathbf{x}/\theta) > 0$, то θ-статистика*

$$\mathbf{T}^*(\mathbf{x}/\theta) = \ln \frac{L(\mathbf{x}/\theta)}{L(\mathbf{x}/\theta_0)},$$

называемая **логарифмическим отношением правдоподобия**, *является необходимой и достаточной.*

Д о к а з а т е л ь с т в о. По определению $\mathbf{T}^*(\mathbf{x}/\theta)$

$$L(\mathbf{x}/\theta) = L(\mathbf{x}/\theta_0)e^{\mathbf{T}^*(\mathbf{x}/\theta)} = h(\mathbf{x})g_\theta(\mathbf{T}^*(\mathbf{x}/\theta)),$$

где $h(\mathbf{x}) = L(\mathbf{x}/\theta_0)$, $\quad g_\theta(\mathbf{T}^*(\mathbf{x}/\theta)) = e^{\mathbf{T}^*(\mathbf{x}/\theta)}$. По факторизационной теореме Неймана это означает, что $\mathbf{T}^*(\mathbf{x}/\theta)$ — достаточная θ-статистика.

Пусть $\mathbf{T}(\mathbf{x}/\theta)$ — любая достаточная θ-статистика. Используя опять факторизационную теорему, получим

$$\mathbf{T}^*(\mathbf{x}/\theta) = \ln\frac{L(\mathbf{x}/\theta)}{L(\mathbf{x}/\theta_0)} = \ln\frac{h(\mathbf{x})g_\theta(\mathbf{T}(\mathbf{x}/\theta))}{h(\mathbf{x})g_\theta(\mathbf{T}(\mathbf{x}/\theta_0))} =$$

$$= \ln\frac{g_\theta(\mathbf{T}(\mathbf{x}/\theta))}{g_\theta(\mathbf{T}(\mathbf{x}/\theta_0))},$$

т.е. $\mathbf{T}^*(\mathbf{x}/\theta)$ выражается через любую достаточную θ-статистику $\mathbf{T}(\mathbf{x}/\theta)$. Значит, $\mathbf{T}^*(\mathbf{x}/\theta)$ — необходимая θ-статистика. $\square$

Это не решение вопроса, так как в отличие от θ-статистики обычная статистика не зависит от θ. Как избавиться от θ?

Теорема Дынкина. *Пусть $\mathfrak{L}$ — наименьшее линейное пространство, содержащее*

$$\Big\{const, \;\; \mathbf{T}^*(\mathbf{x}/\theta) = \ln\frac{L(\mathbf{x}/\theta)}{L(\mathbf{x}/\theta_0)}, \quad \theta \in \Theta\Big\}$$

и $dim\mathfrak{L} = m + 1$ (случай $m = \infty$ не исключается).

1. Если $n \le m$, то необходимая и достаточная статистика тривиальна (т.е. совпадает с самой выборкой).

2. Если $n > m$, то существует необходимая и достаточная статистика размерности m. Для любого базиса

1, $\mathbf{T_1}(\mathbf{x})$, ... , $\mathbf{T_m}(\mathbf{x})$ *в* $\mathfrak{L}$ *такой статистикой может служить* $(\mathbf{T_1}(\mathbf{x}), \ \dots, \ \mathbf{T_m}(\mathbf{x}))$.

Доказательство. Пусть 1, $\mathbf{T_1}(\mathbf{x})$, ... , $\mathbf{T_m}(\mathbf{x})$ — базис в $\mathfrak{L}$. По определению $\mathfrak{L}$ $\mathbf{T}^*(\mathbf{x}/\theta) \in \mathfrak{L}$ при каждом $\theta \in \Theta$. Поэтому

$$\mathbf{T}^*(\mathbf{x}/\theta) = \gamma_0(\theta) + \sum_{k=1}^{m} \gamma_k(\theta)\mathbf{T_k}(\mathbf{x}),$$

где $\gamma_k(\theta)$, $k = \overline{0,m}$, — коэффициенты разложения по базису. По теореме 1 $\mathbf{T}^*(\mathbf{x}/\theta)$ — необходимая и достаточная статистика. Отсюда следует, что $(\mathbf{T_1}(\mathbf{x}), \ \dots, \ \mathbf{T_m}(\mathbf{x}))$ также является необходимой и достаточной статистикой. Действительно, так как $\mathbf{T}^*(\mathbf{x}/\theta)$ достаточна, то по определению достаточной θ-статистики

$$L(\mathbf{x}/\theta) = g_\theta(\mathbf{T}^*(\mathbf{x}/\theta)) \cdot h(\mathbf{x}) = g_\theta(\gamma_0(\theta) +$$

$$+ \sum_{k=1}^{m} \gamma_k(\theta)\mathbf{T_k}(\mathbf{x})) \cdot h(\mathbf{x}) = \widetilde{g_\theta}(\mathbf{T_1}(\mathbf{x}), \dots, \mathbf{T_m}(\mathbf{x})) \cdot h(\mathbf{x}).$$

По факторизационной теореме Неймана $(\mathbf{T_1}(\mathbf{x}), \ \dots, \ \mathbf{T_m}(\mathbf{x}))$ — достаточная статистика.

Так как $\mathbf{T}^*(\mathbf{x}/\theta)$ — необходимая θ-статистика, то $\mathbf{T}^*(\mathbf{x}/\theta) = f(\mathbf{T}(\mathbf{x}/\theta))$, где $\mathbf{T}(\mathbf{x}/\theta)$ — любая θ-статистика. Пусть $\mathbf{T}(\mathbf{x})$ — любая статистика, т.е. и θ-статистика, откуда $\mathbf{T}^*(\mathbf{x}/\theta) = f(\mathbf{T}(\mathbf{x}))$, т.е.

$$\gamma_0(\theta) + \sum_{k=1}^{m} \gamma_k(\theta)\mathbf{T_k}(\mathbf{x}) = f(\mathbf{T}(\mathbf{x})).$$

Если Θ таково, что найдутся $\theta_1, \dots, \theta_m \in \Theta$ такие, что система уравнений

$$\gamma_0(\theta_i) + \sum_{k=1}^{m} \gamma_k(\theta_i)\mathbf{T_k}(\mathbf{x}) = f(\mathbf{T}(\mathbf{x})), \quad i = \overline{1,m},$$

однозначно разрешима, то

$$\mathbf{T_k}(\mathbf{x}) = f_k(\mathbf{T}(\mathbf{x})), \quad k = \overline{1, m},$$

т.е. $(\mathbf{T_1}(\mathbf{x}), \ \ldots, \ \mathbf{T_m}(\mathbf{x}))$ выражается через любую достаточную статистику. Значит, $(\mathbf{T_1}(\mathbf{x}), \ \ldots, \ \mathbf{T_m}(\mathbf{x}))$ — необходимая статистика.

Теперь,

1) если $n \leq m$, то размерность этой статистики больше размерности вектора наблюдений. Следовательно, само наблюдение $\mathbf{x} = (x_1, \ldots, x_n)$, являющееся достаточной статистикой, "лучше" $(\mathbf{T_1}(\mathbf{x}), \ \ldots, \ \mathbf{T_m}(\mathbf{x}))$;

2) если $n > m$, то $(\mathbf{T_1}(\mathbf{x}), \ \ldots, \ \mathbf{T_m}(\mathbf{x}))$ — необходимая и достаточная статистика. □

Замечание 3. Чтобы не усложнять формулировку теоремы Дынкина, в ее условии не указано, что должна быть однозначно разрешима система уравнений

$$\gamma_0(\theta_i) + \sum_{k=1}^{m} \gamma_k(\theta_i)\mathbf{T_k}(\mathbf{x}) = f(\mathbf{T}(\mathbf{x})), \quad i = \overline{1, m}.$$

Из теоремы Дынкина следует, что если существует нетривиальная необходимая и достаточная статистика, то $m < n$ и

$$L(\mathbf{x}/\theta) = L(\mathbf{x}/\theta_0)e^{\gamma_0(\theta) + \sum\limits_{k=1}^{m} \gamma_k(\theta)\mathbf{T_k}(\mathbf{x})} = h(\mathbf{x})c(\theta)e^{\sum\limits_{k=1}^{m} \gamma_k(\theta)\mathbf{T_k}(\mathbf{x})}. \tag{12.2}$$

Определение 12.5. Семейство распределений $\{L(\mathbf{x}/\theta), \quad \theta \in \Theta\}$ вида (12.2) называется **экспоненциальным**.

Итак, нами получено

Следствие 12.6. *Если существует нетривиальная необходимая и достаточная статистика, то семейство распределений* $\{L(\mathbf{x}/\theta), \quad \theta \in \Theta\}$ *является экспоненциальным.*

Задача 1. Пусть $x_1, x_2, \dots, x_n$ — выборка из $N(a, \sigma^2)$, где $a \in \mathbb{R}, \ \ \sigma > 0$. Покажите, что

а) если a известна, а σ^2 неизвестна, то скалярная статистика $T_0(\mathbf{x}) = \sum\limits_{i=1}^{n} (x_i - a)^2$ является необходимой и достаточной;

б) если a неизвестна, а σ^2 известна, то скалярная статистика $T_1(\mathbf{x}) = \sum\limits_{i=1}^{n} x_i$ является необходимой и достаточной;

в) если a и σ^2 неизвестны, то векторная статистика $\mathbf{T}(\mathbf{x}) = (T_1(\mathbf{x}), T_2(\mathbf{x}))$, где $T_1(\mathbf{x}) = \sum\limits_{i=1}^{n} x_i, \ \ T_2(\mathbf{x}) = \sum\limits_{i=1}^{n} x_i^2$, является необходимой и достаточной.

Теорема Колмогорова-Блекуэлла-Рао. *Пусть* $\mathbf{T}$ *— достаточная статистика,* $\hat{\theta} = \hat{\theta}(\mathbf{x})$ *— несмещенная оценка параметра* θ*. Тогда*

1) $\hat{\theta}^* = M(\hat{\theta}/\mathbf{T})$ *— несмещенная оценка* θ*;*

2) $\hat{\theta}^*$ *зависит лишь от* $\mathbf{T}(\mathbf{x})$*;*

3) $D\hat{\theta}^* \le D\hat{\theta}$*, причем равенство возможно лишь в случае, когда* $\hat{\theta}^* = \hat{\theta}$ *с вероятностью 1.*

Д о к а з а т е л ь с т в о. То, что $\hat{\theta}^*$ — статистика, а не θ-статистика, т.е. не зависит от θ, следует из того, что $\mathbf{T}$ — достаточная статистика, т.е. при фиксированном $\mathbf{T}$ распределение $\mathbf{x}$ не зависит от θ, а значит $M(\hat{\theta}(\mathbf{x})/\mathbf{T})$ не зависит от θ. Далее, по свойствам условных математических ожиданий

$$M\hat{\theta}^* = MM(\hat{\theta}/\mathbf{T}) = M\hat{\theta},$$

т.е. $\hat{\theta}^*$ — несмещенная оценка. Пункт 1) доказан.

По определению условного математического ожидания $\hat{\theta}^* = M(\hat{\theta}/\mathbf{T})$ — функция только от $\mathbf{T}(\mathbf{x})$, т.е. доказан пункт 2) теоремы.

Далее,

$$D\hat{\theta} = M(\hat{\theta} - \theta)^2 = M[(\hat{\theta} - \hat{\theta}^*) + (\hat{\theta}^* - \theta)]^2 =$$

$$= M(\hat{\theta} - \hat{\theta}^*)^2 + 2M[(\hat{\theta} - \hat{\theta}^*)(\hat{\theta}^* - \theta)] + M(\hat{\theta}^* - \theta)^2.$$

По свойствам условных математических ожиданий с использованием того, что $\hat{\theta}^* - \theta$ — борелевская функция только от $\mathbf{T}$ (свойство 4 из п.11), получим

$$M[(\hat{\theta} - \hat{\theta}^*)(\hat{\theta}^* - \theta)] = MM[(\hat{\theta} - \hat{\theta}^*)(\hat{\theta}^* - \theta)/\mathbf{T}] =$$

$$= M\{(\hat{\theta}^* - \theta)M[(\hat{\theta} - \hat{\theta}^*)/\mathbf{T}]\} = M\{(\hat{\theta}^* - \theta)[M(\hat{\theta}/\mathbf{T}) - \hat{\theta}^*]\} =$$

$$= M0 = 0.$$

Поэтому

$$D\hat{\theta} = M(\hat{\theta} - \hat{\theta}^*)^2 + D\hat{\theta}^* \geq D\hat{\theta}^*,$$

причем равенство возможно в том и только том случае, если $M(\hat{\theta} - \hat{\theta}^*)^2 = 0$, т.е. $\hat{\theta} = \hat{\theta}^*$ с вероятностью 1. □

Определение 12.7. Статистика $\mathbf{T}$ называется **полной**, если для любой борелевской функции f

$$\text{из} \quad Mf(\mathbf{T}(\mathbf{x})) = 0 \ \ \forall \theta \in \Theta \quad \text{следует} \quad f(\mathbf{T}(\mathbf{x})) = 0$$

с вероятностью 1.

Пример 3. Пусть $x_1, x_2, \ldots, x_n$ — выборка из распределения Бернулли с параметром p. Помним, что x_i может рассматриваться как число успехов в i-м испытании, p — вероятность успеха. В качестве возможных значений параметра p рассмотрим множество $\mathfrak{P} = (0,1)$. Функция правдоподобия есть

$$L(\mathbf{x}/p) = p^{\mu}(1-p)^{n-\mu}.$$

По факторизационной теореме Неймана $\mu = \sum\limits_{i=1}^{n} x_i$ — достаточная статистика. Когда p пробегает $\mathfrak{P} = (0,1)$, переменная $\gamma = \frac{p}{1-p}$ пробегает $(0, \infty)$. Если

$$Mf(\mu) = \sum_{k=0}^{n} f(k) C_n^k p^k (1-p)^{n-k} =$$

$$= (1-p)^n \sum_{k=0}^{n} f(k) C_n^k \gamma^k = 0,$$

то для всех $\gamma \in (0, \infty)$ многочлен $\sum\limits_{k=0}^{n} f(k) C_n^k \gamma^k$ обращается в ноль. Поэтому все его коэффициенты равны нулю, откуда $f(k) = 0, \quad k = \overline{0,n}$. Значит, $\forall\, p \in (0,1)$

$$P\{f(\mu) = 0\} = \sum_{k\colon f(k)=0} P\{\mu = k\} = \sum_{k=0}^{n} P\{\mu = k\} = 1.$$

Следовательно, μ — полная достаточная статистика.

Теорема 12.8. *Если достаточная статистика* $\mathbf{T}$ *— полная, то любая несмещенная оценка произвольной функции* $b(\theta)$, *зависящая только от* $\mathbf{T}$, *единственна с вероятностью 1.*

Д о к а з а т е л ь с т в о. Пусть $f_1(\mathbf{T})$ и $f_2(\mathbf{T})$ — две несмещенные оценки $b(\theta)$, т.е.

$$Mf_1(\mathbf{T}) = b(\theta), \quad Mf_2(\mathbf{T}) = b(\theta).$$

Тогда $M[f_1(\mathbf{T}) - f_2(\mathbf{T})] = 0$, откуда, в силу полноты $\mathbf{T}$, следует, что $f_1(\mathbf{T}) - f_2(\mathbf{T}) = 0$ с вероятностью 1. □

Теорема 12.9. *Если* $\mathbf{T}$ *— полная достаточная статистика,* $\hat{\theta}$ *— несмещенная оценка параметра* θ, *то* $\widehat{\theta^*} = M(\hat{\theta}/\mathbf{T})$ *— единственная эффективная оценка параметра* θ.

Д о к а з а т е л ь с т в о. Пусть $\tilde{\theta}$ — любая другая несмещенная оценка θ, $\quad \widetilde{\theta^*} = M(\tilde{\theta}/\mathbf{T})$. По пунктам 1),2) теоремы Колмогорова-Блекуэлла-Рао $\widehat{\theta^*}$ и $\widetilde{\theta^*}$ — несмещенные оценки θ, зависящие только от $\mathbf{T}$. По теореме 12.8

$$\widehat{\theta^*} = \widetilde{\theta^*} \quad \text{с вероятностью } 1. \tag{12.3}$$

По пункту 3) теоремы Колмогорова-Блекуэлла-Рао и из (12.3)

$$D\widehat{\theta^*} = D\widetilde{\theta^*} \leq D\tilde{\theta},$$

т.е. $\widehat{\theta^*}$ эффективна. Равенство возможно лишь тогда, когда

$$\tilde{\theta} = \widetilde{\theta^*} \quad \text{с вероятностью } 1. \tag{12.4}$$

Сравнивая (12.3) и (12.4), получаем, что равенство возможно лишь в случае, когда $\widehat{\theta^*} = \tilde{\theta}$ с вероятностью 1, т.е. эффективная оценка $\widehat{\theta}^*$ единственна с вероятностью 1. □

Пример 4. Пусть $x_1, x_2, \dots, x_n$ — выборка из распределения Бернулли с параметром p. В качестве оценки p возьмем x_1, т.е. число успехов в первом испытании. Очевидно, $Mx_1 = p$, т.е. x_1 — несмещенная оценка p. Так как она не зависит от n, то $x_1 \stackrel{P}{\to} x_1 \neq p$, т.е. эта оценка, не являясь даже состоятельной, очень плохая. Ранее мы убедились в том, что μ — полная достаточная статистика. По теореме 12.9 $\hat{p} = M(x_1/\mu)$ — единственная эффективная оценка p. По свойствам условных математических ожиданий

$$\mu = M(\mu/\mu) = M\Big(\sum_{i=1}^{n} x_i/\mu\Big) = \sum_{i=1}^{n} M(x_i/\mu) = n\cdot M(x_1/\mu),$$

откуда $\hat{p} = M(x_1/\mu) = \frac{\mu}{n}$ — относительная частота успеха. Эта оценка несмещенная, состоятельная и эффективная.

Приложение 1. Доказательство теоремы Гливенко-Кантелли.

Теорема Гливенко-Кантелли. *Эмпирическая функция распределения почти наверное равномерно сходится к теоретической функции распределения*:

$$P\Big\{\sup_{x\in\mathbb{R}}|F_n^*(x)-F(x)|\to 0 \quad \text{при} \quad n\to\infty\Big\}=1.$$

Д о к а з а т е л ь с т в о. Пусть $F(x)$ — непрерывная функция. Зафиксируем произвольное $\varepsilon > 0$ вида $\varepsilon = \frac{1}{N}$, где N — натуральное число. В силу свойств функции распределения существуют $z_0 = -\infty$, z_1, z_2, $\ldots$, z_{N-1}, $z_N = +\infty$ такие, что $F(z_0) = 0$, $F(z_1) = \frac{1}{N}$, $\ldots$, $F(z_k) = \frac{k}{N}$, $\ldots$, $F(z_N) = 1$. Поэтому для $z \in [z_k, z_{k+1})$

$$F_n^*(z)-F(z)\le F_n^*(z_{k+1})-F(z_k)=F_n^*(z_{k+1})-F(z_{k+1})+\varepsilon, \tag{П1.1}$$

$$F_n^*(z)-F(z)\ge F_n^*(z_k)-F(z_{k+1})=F_n^*(z_k)-F(z_k)-\varepsilon. \tag{П1.2}$$

Пусть $A_k = \{\omega:\ F_n^*(z_k)\to F(z_k)\}$. По усиленному закону больших чисел в форме Бореля $P(A_k) = 1$. Поэтому $\forall\ \varepsilon = \frac{1}{N} > 0$ и $\forall\ \omega \in \bigcap\limits_{k=0}^{N} A_k = A\ \ \exists\ n_0 = n_0(\omega)$ такое, что $\forall\ n \ge n_0$ выполняется неравенство

$$|F_n^*(z_k)-F(z_k)|<\varepsilon, \qquad k=\overline{0,N}. \tag{П1.3}$$

Теперь из (П1.1), (П1.2) и (П1.3) следует, что $\forall\ \varepsilon > 0$ и $\forall\ \omega \in A\ \ \exists\ n_0(\omega)$ такое, что $\forall\ n \ge n_0(\omega)$ выполняется неравенство

$$|F_n^*(z)-F(z)|<2\varepsilon.$$

Это означает, что $P(A) = 1$ $\square$.

Отметим, что для доказательства теоремы в случае разрывной функции надо использовать то, что множества точек непрерывности функции F всюду плотно в расширенной числовой прямой.

Приложение 2. Доказательство теоремы о свойствах решения уравнения правдоподобия.

Предположим выполненными следующие условия.

1. Параметр θ и его истинное значение θ_0 лежат внутри интервала (θ_1, θ_2), в котором существуют производные $\frac{\partial p}{\partial \theta}$, $\frac{\partial^2 p}{\partial \theta^2}$, $\frac{\partial^3 p}{\partial \theta^3}$.

2. Допустимо двукратное дифференцирование под знаком интеграла $\int\limits_{-\infty}^{+\infty} p(x/\theta)dx = 1$.

3. $i(\theta_0) = M\left(\frac{\partial \ln p}{\partial \theta}\right)^2\Big|_{\theta=\theta_0} > 0$, $\left|\frac{\partial^3 p}{\partial \theta^3}\right| \leq H(x)$, $MH(\xi) \leq M$, где M не зависит от θ.

Теорема П2.1. *Если выполнены условия 1,2,3, то уравнение правдоподобия*

$$\frac{\partial L(x/\theta)}{\partial \theta} = 0$$

имеет решение $\hat{\theta}$, являющееся состоятельной оценкой θ_0. Эта оценка максимального правдоподобия асимптотически нормальна с параметрами $(\theta_0, \frac{1}{n \cdot i(\theta_0)})$ и асимптотически эффективна.

Доказательство. Уравнение правдоподобия равносильно следующему уравнению для логарифмической функции правдоподобия:

$$\frac{1}{L}\frac{\partial L}{\partial \theta} = \frac{\partial \ln L}{\partial \theta} = \sum_{k=1}^{n} \frac{\partial \ln p(x_k/\theta)}{\partial \theta} = 0. \qquad (\text{П}2.1)$$

Разложим $\frac{\partial \ln p(x/\theta)}{\partial \theta}$ по формуле Тейлора в окрестности точки θ_0:

$$\frac{\partial \ln p}{\partial \theta} = \frac{\partial \ln p}{\partial \theta}\Big|_{\theta=\theta_0} + (\theta-\theta_0)\frac{\partial^2 \ln p}{\partial \theta^2}\Big|_{\theta=\theta_0} + \frac{(\theta-\theta_0)^2}{2}\delta\cdot H(x), \tag{П2.2}$$

где $|\delta| \le 1$. Разделив (П2.1) на n и используя разложение (П2.2), получим

$$\frac{1}{n}\frac{\partial \ln L}{\partial \theta} = B_0 + B_1(\theta-\theta_0) + \frac{1}{2}\delta\cdot B_2(\theta-\theta_0)^2, \tag{П2.3}$$

где

$$B_0 = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial \ln p(x_k/\theta)}{\partial \theta}\Big|_{\theta=\theta_0},$$

$$B_1 = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial^2 \ln p(x_k/\theta)}{\partial \theta^2}\Big|_{\theta=\theta_0},$$

$$B_2 = \frac{1}{n}\sum_{k=1}^{n} H_k(x).$$

Поскольку

$$\frac{\partial \ln p}{\partial \theta} = \frac{1}{p}\frac{\partial p}{\partial \theta},$$

$$\frac{\partial^2 \ln p}{\partial \theta^2} = \frac{\partial}{\partial \theta}\Big(\frac{1}{p}\frac{\partial p}{\partial \theta}\Big) = \frac{1}{p}\frac{\partial^2 p}{\partial \theta^2} - \frac{1}{p^2}\Big(\frac{\partial p}{\partial \theta}\Big)^2 = \frac{1}{p}\frac{\partial^2 p}{\partial \theta^2} - \Big(\frac{\partial \ln p}{\partial \theta}\Big)^2,$$

то умножая обе части полученного равенства на p и интегрируя по $x \in \mathbb{R}$, получим:

$$\int_{-\infty}^{+\infty}\frac{\partial^2 \ln p(x/\theta)}{\partial \theta^2}p(x/\theta)dx =$$

$$= \int_{-\infty}^{+\infty}\frac{\partial^2 p(x/\theta)}{\partial \theta^2}dx - \int_{-\infty}^{+\infty}\Big(\frac{\partial \ln p(x/\theta)}{\partial \theta}\Big)^2 p(x/\theta)dx. \tag{П2.4}$$

В силу условия 2) теоремы

$$\int_{-\infty}^{+\infty} \frac{\partial^2 p(x/\theta)}{\partial \theta^2} dx = \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{+\infty} p(x/\theta) dx = \frac{\partial^2 1}{\partial \theta^2} = 0.$$

Поэтому при выполнении условия 2) теоремы из (П2.4) получаем еще одно выражение для информации по Фишеру о параметре θ_0 в наблюдении:

$$i(\theta_0) = M\Big(\frac{\partial \ln p(x/\theta)}{\partial \theta}\Big)^2\Big|_{\theta=\theta_0} =$$

$$= \int_{-\infty}^{+\infty} \Big(\frac{\partial \ln p(x/\theta)}{\partial \theta}\Big)^2 p(x/\theta) dx\Big|_{\theta=\theta_0} =$$

$$= -\int_{-\infty}^{+\infty} \frac{\partial^2 \ln p(x/\theta)}{\partial \theta^2} p(x/\theta) dx\Big|_{\theta=\theta_0} = -M\frac{\partial^2 \ln p(x/\theta)}{\partial \theta^2}\Big|_{\theta=\theta_0}.$$

По закону больших чисел в форме Хинчина

$$B_0 \to MB_0 = 0, \quad B_1 \to MB_1 = -i(\theta_0),$$

$$B_2 \to MB_2, \quad |MB_2| \le M.$$

Пусть теперь $h > 0$ и $\varepsilon > 0$ фиксированы. Выбираем n_0 таким, чтобы при всех $n \ge n_0$

$$P\{|B_0| \ge h^2\} < \frac{\varepsilon}{3}, \quad P\Big\{B_1 > -\frac{i(\theta_0)}{2}\Big\} < \frac{\varepsilon}{3},$$

$$P\{|B_2| > 2M\} < \frac{\varepsilon}{3}. \qquad (\text{П}2.5)$$

Обозначим через S событие, состоящее в том, что одновременно выполняются неравенства

$$|B_0| \le h^2, \quad B_1 \le -\frac{i(\theta_0)}{2}, \quad |B_2| \le 2M.$$

В силу (П2.5) $P(\overline{S}) < \varepsilon$ и $P(S) > 1 - \varepsilon$. При $\theta = \theta_0 \pm h$ уравнение (П2.3) приобретает вид

$$B_0 \mp B_1 h + \frac{1}{2}\delta B_2 h^2 = 0. \qquad (\text{П}2.6)$$

В множестве S

$$\left| B_0 + \frac{1}{2}\delta B_2 h^2 \right| \le (M+1)h^2,$$

и при $h < \frac{i(\theta_0)}{2(M+1)}$ знак левой части (П2.6) определяется знаком члена $\mp B_1 h$. Так как $\frac{1}{n}\frac{\partial \ln L}{\partial \theta}$ непрерывно зависит от θ, то в интервале $(\theta_0 - h, \theta_0 + h)$ при $n \ge n_0$ с вероятностью большей или равной $1 - \varepsilon$ существует корень $\hat{\theta}$. Таким образом, мы доказали первую часть теоремы. Перепишем равенство

$$B_0 + B_1(\hat{\theta} - \theta_0) + \frac{B_2 \delta}{2}(\hat{\theta} - \theta_0)^2 = 0$$

в следующем виде:

$$(\hat{\theta} - \theta_0)\sqrt{ni(\theta_0)} = \frac{\frac{1}{\sqrt{ni(\theta_0)}} \sum\limits_{k=1}^{n} \frac{\partial \ln p(x_k/\theta)}{\partial \theta}\Big|_{\theta=\theta_0}}{-\frac{B_1}{i(\theta_0)} - \frac{1}{2}\delta B_2 \frac{\hat{\theta} - \theta_0}{i(\theta_0)}}. \qquad (\text{П}2.7)$$

Числитель в (П2.7) по центральной предельной теореме асимптотически нормален с параметрами (0,1), а знаменатель при $n \to \infty$ сходится по вероятности к 1. Поэтому случайная величина $(\hat{\theta} - \theta_0)\sqrt{ni(\theta_0)}$ асимптотически нормальна с параметрами (0,1), что и доказывает теорему. □

Литература

1. Барра Ж.Р. Основные понятия математической статистики. — М.: Мир, 1974. 277с.

2. Большев Л.Н., Смирнов Н.В.. Таблицы математической статистики. — М.: Наука, 1983. 416с.

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

4. Де Гроот М. Оптимальные статистические решения. — М.: Мир, 1974. 493с.

5. Закс Ш. Теория статистических выводов. — М.: Мир, 1975. 776с.

6. Ивченко Г.И., Медведев Ю.И., Чистяков А.В. Сборник задач по математической статистике. — М.: Высшая школа, 1989. 255с.

7. Кендалл М., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973. 900с.

8. Климов Г.П. Теория вероятностей и математическая статистика. — М.: МГУ, 1983. 328с.

9. Климов Г.П., Кузьмин А.Д. Вероятность, процессы, статистика. Задачи с решениями. — М.: МГУ, 1985. 232с.

10. Кокс Д., Хинкли Д. Теоретическая статистика. — М.: Мир, 1978. 560с.

11. Кокс Д., Хинкли Д. Задачи по теоретической статистике с решениями. — М.: Мир, 1981. 225с.

12. Крамер Г. Математические методы статистики. — М.: Мир, 1976. 648 с.

13. Леман Э. Проверка статистических гипотез. — М.: Наука, 1979. 408 с.

14. Рао С.Р. Линейные статистические методы и их применения. — М.: Наука, 1968. 548с.

15. Смирнов Н.В., Дунин-Барковский. Курс теории вероятностей и математической статистики для технических приложений. — М.: Наука, 1965. 512с.

16. Соле Ж.-Л. Основные структуры математической статистики. — М.: Мир, 1972. 104с.

17. Шметтерер Л. Введение в математическую статистику. — М.: Наука, 1976. 520с.

18. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. // под редакцией Свешникова А.А. — М.: Наука, 1970. 656с.

Содержание

Учебное издание

МАЛИНКОВСКИЙ Юрий Владимирович

Теория вероятностей и математическая статистика (часть 2. Математическая статистика). Учебное пособие

Подписано в печать 5.07.2004. ЛВ № 02330/0133208 от 30 апреля 2004г. Формат 60×84 1/16. Бумага офсетная. Печать офсетная. Усл. п. л. 6,5 Уч.-изд. л. 20. Тираж 250 экз. Заказ № 59.

Отпечатано на полиграфической технике УО “ГГУ им. Ф.Скорины”
Лицензия ЛВ № 02330/0056611.
246019 г.Гомель, ул.Советская 104