

Введение

Математическая статистика — раздел математики, изучающий методы построения оптимальных по заданным критериям статистических выводов и анализа их свойств. Предмет математической статистики — это всевозможные статистические выводы.

Математическая статистика опирается на методы теории вероятностей, но решает иные задачи. В теории вероятностей мы, зная природу некоторого явления, выясняем, как будут вести себя (как распределены) те или иные изучаемые нами характеристики, которые можно наблюдать в экспериментах. В математической статистике наоборот — исходными являются экспериментальные данные (как правило, это наблюдения над случайными величинами), требуется вынести то или иное суждение или решение о природе рассматриваемого явления.

Как и в теории вероятностей нас будут интересовать не те эксперименты, которые позволяют делать однозначные и детерминированные выводы о рассматриваемых в природе явлениях, а эксперименты, результатами которых являются случайные события. С развитием различных отраслей науки значение такого рода задач возрастает, так как с увеличением точности исследований и экспериментов становится невозможным пренебрежение случайными факторами, связанными с различного рода помехами и ограниченностью измерительных и вычислительных возможностей. При этом возникают следующие вопросы:

1. Если мы наблюдаем одну случайную величину, то как по набору ее значений в нескольких опытах сделать как можно более точный вывод о распределении этой случайной величины?
2. Если мы наблюдаем одновременно проявление двух или более признаков, т. е. имеем набор значений нескольких случайных величин, то что можно сказать об их зависимости или о совместном распределении этих случайных величин?

Также часто бывает возможно высказать некие предположения о наблюдаемом распределении или о его свойствах. В этом случае по опытным данным требуется подтвердить или опровергнуть эти предположения (гипотезы). При этом надо помнить, что ответ "да" или "нет" может быть дан лишь с определенной степенью достоверности, и чем дольше мы можем продолжать эксперимент, тем точнее могут быть выводы. Часто оказываются заранее известными некоторые свойства наблюдаемого эксперимента и можно сформулировать какие-то априорные выводы о распределении: о наличии функциональной зависимости между наблюдаемыми величинами, о нормальности распределения, о его симметричности, о наличии у распределения плотности или о его дискретном характере и т. д. Наличие таких знаний помогает на основании результатов эксперимента делать выводы о прочих, неизвестных, свойствах распределения.

Итак, математическая статистика работает там, где есть случайный эксперимент, свойства которого частично или полностью неизвестны и который мы умеем воспроизводить в одних и тех же условиях некоторое (а лучше — неограниченное) число раз. При получении теоретических и прикладных результатов статистической обработки исследуемых данных

используется математический аппарат, который опирается на вероятностное обоснование как исходных данных, так и полученных в результате применения математических методов статистических выводов.

Глава 1

Теория статистического оценивания параметров

1.1 Основные понятия выборочного метода

Статистические данные, являющиеся исходным материалом в задачах математической статистики, представляют собой результаты наблюдений (измерений) некоторой случайной величины, характеризующей исход изучаемого явления.

Пусть проводится некоторый случайный эксперимент в котором измеряется $N \geq 1$ числовых величин $\xi_1, \dots, \xi_N \in \mathbb{R}^1$. Математической моделью такого случайного эксперимента является вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$, на котором определен случайный N -вектор $\xi = (\xi_1, \dots, \xi_N)^{\mathbf{T}} \in \mathbb{R}^N$, имеющий некоторую N -мерную функцию распределения вероятностей

$$F(z; \theta) ::= \mathbf{P}\{\xi_1 < z_1, \dots, \xi_N < z_N\}, \quad z = (z_1, \dots, z_N)^{\mathbf{T}} \in \mathbb{R}^N,$$

где $\theta = (\theta_1, \dots, \theta_m)^{\mathbf{T}} \in \Theta \subseteq \mathbb{R}^m$ — некоторый набор параметров (векторный параметр) распределения вероятностей.

Определение 1.1. Пусть осуществлено n независимых случайных экспериментов и зарегистрированы результаты: $x_1, x_2, \dots, x_n \in \mathbb{R}^N$ — первого, второго, ..., n -го эксперимента. Построенное таким образом множество результатов экспериментов

$$X = \{x_1, \dots, x_n\}$$

называется случайной выборкой объема n из N -мерного распределения вероятностей $F(\cdot; \theta)$. Число экспериментов n называется объемом выборки, результат i -го эксперимента $x_i = (x_{i1}, \dots, x_{iN})^{\mathbf{T}} \in \mathbb{R}^N$ — i -м выборочным значением.

Далее выборку X иногда будем представлять в виде составного вектора:

$$X = (x_1^{\mathbf{T}}, \dots, x_n^{\mathbf{T}})^{\mathbf{T}} \in \mathbb{R}^{nN}. \quad (1.1)$$

Определение 1.2. Множество

$$\mathcal{X} = \{X = (x_1^{\mathbf{T}}, \dots, x_n^{\mathbf{T}})^{\mathbf{T}} : X \in \mathbb{R}^{nN}\}$$

всех возможных реализаций выборки называется выборочным пространством.

Математической моделью выборки в представлении (1.1) является случайный nN -вектор $X \in \mathbb{R}^{nN}$, определенный на составном вероятностном пространстве $(\Omega^n, \mathcal{F}^n, \mathbf{P}^n)$ и имеющий nN -мерную функцию распределения

$$\Psi(X; \theta) = \prod_{i=1}^n F(x_i; \theta), \quad X \in \mathbb{R}^{nN}, \quad \theta \in \Theta.$$

Такая модель называется *параметрической моделью*, а множество Θ возможных значений параметра θ — *параметрическим множеством*.

Отметим два важных свойства случайной выборки: независимость выборочных значений и их одинаковую распределенность.

Рассмотрим подробнее составной вектор $X = (x_1^{\mathbf{T}}, \dots, x_n^{\mathbf{T}})^{\mathbf{T}} \in \mathbb{R}^{nN}$. В конкретной серии экспериментов выборка X — это некоторый набор значений векторов. Но при повторении серии экспериментов вместо набора векторов $x_1^{\mathbf{T}}, \dots, x_n^{\mathbf{T}}$ мы получим новый набор векторов, в котором, например, вместо вектора $x_1^{\mathbf{T}}$ появится другой вектор — одно из возможных значений случайного вектора $\xi \in \mathbb{R}^N$. То есть $x_1^{\mathbf{T}}$, а также и $x_2^{\mathbf{T}}$, и $x_3^{\mathbf{T}}$, и \dots , и $x_n^{\mathbf{T}}$ — не какой-то конкретный раз и навсегда заданный вектор, а переменная векторная величина, которая может принимать те же значения, что и случайный вектор ξ , и с теми же вероятностями. Итак, $x_1^{\mathbf{T}}$ — случайный вектор, одинаково распределенный с ξ , а вектор, который мы наблюдаем в данном эксперименте — одно из возможных значений случайного вектора $x_1^{\mathbf{T}}$.

Таким образом, случайная выборка $X = \{x_1, \dots, x_n\}$ объема n это:

1. В конкретной серии экспериментов — набор из n векторов, являющихся значениями (реализациями) случайного вектора ξ в n независимых экспериментах;
2. В математической модели — набор из независимых и одинаково распределенных случайных векторов (копий ξ), имеющих как и ξ , распределение $\mathcal{L}\{\xi; \theta\}$, с функцией распределения вероятностей

$$F(z; \theta) ::= \mathbf{P}\{\xi_1 < z_1, \dots, \xi_N < z_N\}, \quad z = (z_1, \dots, z_N)^{\mathbf{T}} \in \mathbb{R}^N,$$

где $\theta = (\theta_1, \dots, \theta_m)^{\mathbf{T}} \in \Theta \subseteq \mathbb{R}^m$ — некоторый набор параметров (векторный параметр) распределения вероятностей..

Определение 1.3. Любое измеримое по Борелю функциональное преобразование выборки X :

$$t = t(X) = (t_1(X), \dots, t_K(X))^{\mathbf{T}} \in \mathbb{R}^K, \quad X \in \mathbb{R}^{nN},$$

не зависящее от функции распределения $F(\cdot; \theta)$ или от ее параметра θ , называется *статистикой*.

В частности, при тождественном преобразовании $t(X) = X$ — выборка X также является статистикой.

Определение 1.4. Выводы (решения, оценки, суждения), основанные на статистиках $t = t(X)$, называются *статистическими выводами*.

1.2 Основные понятия теории статистического оценивания параметров

Пусть наблюдается случайная выборка $X = (x_1^{\mathbf{T}}, \dots, x_n^{\mathbf{T}})^{\mathbf{T}} \in \mathbb{R}^{nN}$ объема n из некоторого N -мерного распределения вероятностей с функцией распределения

$$F(z; \theta^0), \quad z \in \mathbb{R}^N, \quad \theta^0 = (\theta_1^0, \dots, \theta_m^0)^{\mathbf{T}} \in \Theta \subseteq \mathbb{R}^m.$$

Пространство \mathbb{R}^{nN} часто называют *выборочным пространством*: каждая точка в нем — выборка; Θ — *параметрическое пространство*. Предполагается, что функция распределения $F(\cdot; \theta^0)$ задана с точностью до вектора параметров $\theta^0 \in \Theta$. Задача статистического оценивания параметров заключается в том, чтобы по выборке X найти такое значение $\hat{\theta} = T(X) \in \Theta$, которое в определенном вероятностном смысле близко к θ^0 . Таким образом, возникает задача построения функционального преобразования

$$\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \Theta,$$

которое будем предполагать борелевским. Оценка $\hat{\theta} = T(X)$ задается статистикой $T(X)$ и поэтому называется *статистической оценкой*:

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_m \end{pmatrix} = T(X) = \begin{pmatrix} T_1(X) \\ \vdots \\ T_m(X) \end{pmatrix} \in \Theta \subseteq \mathbb{R}^m, \quad X \in \mathbb{R}^{nN},$$

где $\hat{\theta}_j = T_j(X)$ — статистическая оценка j -го параметра, $j \in \{1, \dots, m\}$.

Пример 1.1. Пусть $X = \{x_1, \dots, x_n\}$ — случайная выборка из одномерного ($N = 1$) нормального распределения $\mathcal{N}_1(\theta_1^0, \theta_2^0)$:

$$F(z; \theta^0) = \Phi \left(\frac{z - \theta_1^0}{\sqrt{\theta_2^0}} \right), \quad z \in \mathbb{R}^1, \quad \theta^0 = (\theta_1^0, \theta_2^0)^T.$$

Требуется оценить два параметра: математическое ожидание θ_1^0 и дисперсию θ_2^0 .

Замечание 1.1. В теории вероятностей исходят из того, что модель точно задана, т. е. $F(\cdot; \theta^0)$ известна. В математической статистике вероятностная модель известна с точностью до значения вектора параметров θ^0 или задано семейство допустимых вероятностных моделей.

Замечание 1.2. В теории вероятностей принято случайные величины и их значения обозначать разными символами. Например, ξ — случайная величина, x — ее значение. В математической статистике и случайные величины, и их значения обозначаются одними и теми же символами.

Вероятностный смысл близости оценки $\hat{\theta} = T(X)$ к истинному значению θ^0 определяется следующими тремя требованиями, предъявляемыми к статистике $T(\cdot)$.

I. Состоятельность статистической оценки

Определение 1.5. Статистическая оценка

$$\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \Theta$$

называется *состоятельной*, если, каково бы ни было истинное значение вектора параметров $\theta^0 \in \Theta$, при $n \rightarrow \infty$ имеет место сходимость оценки к истинному значению по вероятности:

$$\forall \theta^0 \in \Theta \quad \hat{\theta} \xrightarrow{\mathbf{P}} \theta^0, \quad n \rightarrow +\infty, \quad (1.2)$$

что эквивалентно соотношениям

$$\forall \theta^0 \in \Theta \quad \hat{\theta}_i \xrightarrow{\mathbf{P}} \theta_i^0, \quad i = 1, \dots, m, \quad n \rightarrow +\infty, \quad (1.3)$$

или

$$\forall \theta^0 \in \Theta, \quad \forall \varepsilon > 0 \quad \mathbf{P}_{\theta^0}\{|\hat{\theta} - \theta^0| > \varepsilon\} \rightarrow 0, \quad n \rightarrow +\infty,$$

где

$$\mathbf{P}_{\theta^0}(A) = \int_A d\Psi(X; \theta^0), \quad \Psi(X; \theta^0) = \prod_{i=1}^n F(x_i; \theta^0), \quad A \subseteq \mathcal{B}^{nN};$$

$P_{\theta^0}(\cdot)$ — вероятностная мера в выборочном пространстве, вычисленная при фиксированном истинном значении вектора параметров θ^0 ; \mathcal{B}^{nN} — борелевская σ -алгебра подмножеств из R^{nN} .

Определение 1.6. Если в соотношениях (1.2) и (1.3) имеет место сходимость почти на верное, то говорят о сильной (строгой) состоятельности оценки:

$$\forall \theta^0 \in \Theta \quad \hat{\theta} \xrightarrow{n.н.} \theta^0, \quad n \rightarrow +\infty.$$

Разумеется, из сильной состоятельности следует состоятельность по вероятности.

II. Несмещенность статистической оценки

Определение 1.7. Смещением статистической оценки $\hat{\theta} = T(X)$ называется уклонение математического ожидания этой оценки от истинного значения вектора параметров θ^0 :

$$b(n; \theta^0) = (b_1, \dots, b_m)^T ::= \mathbf{E}_{\theta^0}\{\hat{\theta}\} - \theta^0 = \int_{\mathbb{R}^{nN}} (T(X) - \theta^0) d\Psi(X; \theta^0).$$

Определение 1.8. Если, каковы бы ни были истинное значение θ^0 и объем выборки n , смещение равно нулю:

$$\forall n \geq 1, \quad \forall \theta^0 \in \Theta \quad b(n; \theta^0) = \mathbf{0}_m,$$

то оценка $\hat{\theta} = T(X)$ называется несмещенной. В противном случае — смещенной.

Определение 1.9. Если

$$\forall \theta^0 \in \Theta \quad \lim_{n \rightarrow \infty} b(n; \theta^0) = \mathbf{0}_m,$$

то оценка $\hat{\theta} = T(X)$ называется асимптотически несмещенной.

Несмещенность означает отсутствие систематической погрешности при оценивании параметров.

III. Точность статистической оценки

Определение 1.10. Матрицей вариаций статистической оценки $\hat{\theta} = T(X)$ называется $(m \times m)$ -матрица, определяемая следующим матричным соотношением:

$$V = V\{\hat{\theta}\} = (v_{jk})_{j,k=1}^m = \mathbf{E}_{\theta^0}\{(\hat{\theta} - \theta^0)(\hat{\theta} - \theta^0)^T\},$$

или покомпонентно ($j, k = 1, \dots, m$):

$$v_{jk} = \mathbf{E}_{\theta^0}\{(\hat{\theta}_j - \theta_j^0)(\hat{\theta}_k - \theta_k^0)\} = \int_{\mathbb{R}^{nN}} (T_j(X) - \theta_j^0)(T_k(X) - \theta_k^0) d\Psi(X; \theta^0).$$

Свойства матрицы вариаций

Свойство 1. Матрица вариаций симметрична: $V = V^T$.

Свойство 2. Диагональный элемент

$$v_{jj} = \mathbf{E}_{\theta^0} \{(\hat{\theta}_j - \theta_j^0)^2\} \geq 0, \quad j \in \{1, \dots, m\}, \quad -$$

среднеквадратическая ошибка оценивания j -го параметра.

Свойство 3. Пусть

$$\Sigma = \mathbf{Cov}\{\hat{\theta}, \hat{\theta}\} = \mathbf{E}_{\theta^0} \{(\hat{\theta} - \mathbf{E}_{\theta^0}\{\hat{\theta}\})(\hat{\theta} - \mathbf{E}_{\theta^0}\{\hat{\theta}\})^T\} -$$

ковариационная матрица для $\hat{\theta}$. Тогда справедливо следующее соотношение:

$$V = \Sigma + b(n; \theta^0)b^T(n; \theta^0),$$

где $b(n; \theta^0)$ — смещение оценки $\hat{\theta}$.

Свойство 4. Если $\hat{\theta}$ — несмещенная оценка, то $V = \Sigma$.

Свойство 5. Если $\text{tr}(V) \rightarrow 0$, то $\hat{\theta}$ — состоятельная оценка.

В качестве меры точности оценки $\hat{\theta}$ используются характеристики $\{v_{jj}, j = 1, \dots, m\}$, $\text{tr}(V)$, $|V|$. Чем меньше их значения, тем точнее оценка $\hat{\theta}$.

1.3 Выборочная функция распределения, выборочная характеристическая функция, выборочные моменты, гистограмма и их свойства

Рассмотрим некоторые важнейшие элементарные статистики $T = T(X)$, используемые в дальнейшем для оценивания. Пусть наблюдается случайная выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ объема n из некоторого, для простоты, одномерного распределения вероятностей с функцией распределения

$$F(x) = \mathbf{P}\{x_j < x\}, \quad x \in \mathbb{R} \quad (j \in \{1, \dots, n\}).$$

Пусть определена характеристическая функция

$$g(t) = \mathbf{E}\{e^{itx_j}\} = \int_{-\infty}^{+\infty} e^{itx} dF(x), \quad t \in \mathbb{R}, \quad (1.4)$$

где i — мнимая единица, и для некоторого $k \geq 1$ существует начальный момент k -го порядка:

$$\alpha_k = \mathbf{E}\{x_j^k\} = \int_{-\infty}^{+\infty} x^k dF(x). \quad (1.5)$$

Примем обозначения:

$$\mathbb{I}(z) = \begin{cases} 1, & z > 0; \\ 0, & z \leq 0; \end{cases} \quad - \text{единичная функция Хевисайда};$$

$$\mathbb{I}_A(z) = \begin{cases} 1, & z \in A; \\ 0, & z \notin A; \end{cases} \quad - \text{индикатор множества } A.$$

Если $F(x)$ — абсолютно непрерывная функция распределения, то соответствующую ей плотность распределения обозначаем

$$f(x) = \frac{d}{dx} F(x).$$

Рассмотрим задачу построения по выборке X статистических оценок для $F(\cdot)$, $g(\cdot)$, $\{\alpha_k\}$ и $f(\cdot)$.

Определение 1.11. Выборочной функцией распределения (эмпирической функцией распределения), построенной по выборке X объема n , называется статистика

$$\hat{F}(x) ::= F_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}(x - x_j) = \frac{L_n(x)}{n}, \quad x \in \mathbb{R}, \quad (1.6)$$

где

$$L_n(x) = \sum_{j=1}^n \mathbb{I}(x - x_j) -$$

число выборочных значений из $\{x_j\}_{j=1}^n$, для которых $x_j < x$.

Для построения графика выборочной функции распределения воспользуемся вариационным рядом выборки X : $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, где $x_{(j)}$ — j -я порядковая статистика, $j \in \{1, \dots, n\}$ (см. § ??).

Определение 1.12. Выборочной характеристической функцией для выборки X называется комплекснозначная функция действительной переменной t :

$$\hat{g}(t) ::= g_n(t) = \frac{1}{n} \sum_{j=1}^n e^{itx_j}, \quad t \in \mathbb{R}. \quad (1.7)$$

Определение 1.13. Выборочным моментом k -го порядка для выборки X называется статистика

$$\hat{\alpha}_k ::= a_k = \frac{1}{n} \sum_{j=1}^n x_j^k. \quad (1.8)$$

Формально статистики (1.7) и (1.8) получаются подстановкой (1.6) в (1.4) и (1.5) с учетом того, что

$$\frac{d}{dz} \mathbb{I}(z) ::= \delta(z) -$$

обобщенная δ -функция Дирака, для которой:

$$\int_{\mathbb{R}} G(z - b) \delta(z) dz = G(b), \quad b \in \mathbb{R},$$

где $G(\cdot)$ — произвольная функция, определенная на \mathbb{R} . Это проявление так называемого *подстановочного принципа* (plug-in) в статистике.

Теорема 1.1. Выборочная функция распределения, выборочная характеристическая функция и выборочный момент k -го порядка являются сильно состоятельными, несмещенными оценками соответственно для $F(\cdot)$, $g(\cdot)$ и α_k . Причем для вариаций оценок справедливы следующие соотношения:

$$V\{\hat{F}(x)\} = \mathbf{E}\{(\hat{F}(x) - F(x))^2\} = \frac{F(x)(1 - F(x))}{n}, \quad V\{\hat{\alpha}_k\} = \frac{\alpha_{2k} - \alpha_k^2}{n}.$$

Доказательство. для каждой из трех статистик проводится однотипно. Рассмотрим первую из статистик и по (1.6) представим ее в эквивалентном виде:

$$\hat{F}(x) = \frac{1}{n} \sum_{j=1}^n \xi_j, \quad \xi_j = \mathbb{I}(x - x_j),$$

где $\{\xi_j\}$ — случайные величины Бернулли, независимые в совокупности и одинаково распределенные, так как x — фиксированная величина, а $\{x_j\}$ — выборочные значения, независимые в совокупности и одинаково распределенные с функцией распределения $F(\cdot)$. Здесь учтено, что независимость сохраняется при функциональном преобразовании. Вероятность успеха:

$$p = \mathbf{P}\{\xi_j = 1\} = \mathbf{E}\{\xi_j\} = \mathbf{P}\{x_j < x\} = F(x).$$

Выполняются условия усиленного закона больших чисел [?], следовательно:

$$\frac{1}{n} \sum_{j=1}^n \xi_j \xrightarrow{\text{п.н.}} p = F(x), \quad n \rightarrow +\infty.$$

Далее

$$\mathbf{E}\{\hat{F}(x)\} = p = F(x),$$

т. е. оценка является несмещенной.

Вариация оценки

$$V\{\hat{F}(x)\} = \mathbf{D}\{\hat{F}(x)\} = \frac{1}{n^2} \cdot n \cdot \mathbf{D}\{\xi_1\} = \frac{p(1-p)}{n} = \frac{F(x)(1-F(x))}{n}.$$

□

Пример 1.2. Найти вариацию $V\{\hat{g}(t)\}$.

Теорема 1.2. Выборочная функция распределения $\hat{F}(x)$ и выборочный момент a_k при $n \rightarrow +\infty$ являются асимптотически нормально распределенными статистиками, причем справедливы соотношения:

1. если $0 < F(x) < 1$, то

$$\mathbf{P} \left\{ \sqrt{n} \frac{\hat{F}(x) - F(x)}{\sqrt{F(x)(1-F(x))}} < z \right\} \rightarrow \Phi(z), \quad n \rightarrow +\infty, \quad z \in \mathbb{R},$$

где $\Phi(\cdot)$ — функция распределения вероятностей стандартного нормального закона $\mathcal{N}_1(0, 1)$;

2. если момент порядка $2k$ ограничен: $\alpha_{2k} < +\infty$, и $\alpha_{2k} - \alpha_k^2 > 0$, то

$$\mathbf{P} \left\{ \sqrt{n} \frac{a_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}} < z \right\} \rightarrow \Phi(z), \quad n \rightarrow +\infty, \quad z \in \mathbb{R}.$$

Доказательство. первого утверждения основано на теореме Муавра — Лапласа, второго — на центральной предельной теореме (ЦПТ) Леви — Линдберга [?]. □

Пример 1.3. Исследовать статистику

$$g_n(t) = \text{Re}(g_n(t)) + i \cdot \text{Im}(g_n(t))$$

на асимптотическую нормальность, применив многомерный аналог ЦПТ [?] к $(\text{Re}(g_n(t)), \text{Im}(g_n(t)))$ в \mathbb{R}^2 .

Рассмотрим задачу оценивания плотности распределения

$$f(x) = \frac{d}{dx}F(x), \quad x \in \mathbb{R},$$

по выборке $X = (x_1, \dots, x_n)^T$. Используя принцип plug-in, построим подстановочную оценку

$$\tilde{f}(x) = \frac{d}{dx}F_n(x).$$

Легко видеть, что эта оценка неконструктивна:

$$\tilde{f}(x) = \frac{1}{n} \sum_{j=1}^n \delta(x - x_j) = \begin{cases} 0, & x \notin \{x_1, \dots, x_n\}; \\ \text{не существует,} & x \in \{x_1, \dots, x_n\}. \end{cases}$$

Укажем другой способ статистического оценивания плотности $f(x)$.

Полагаем, что плотность распределения $f(x)$ сосредоточена на отрезке $[x_-, x_+]$, вне его она равна нулю. Зададим некоторое натуральное число $K \geq 2$ и осуществим разбиение $[x_-, x_+)$ на K частей точками деления: $x_- = b_0 < b_1 < \dots < b_K = x_+$; k -й ячейкой гистограммы назовем числовой промежуток $\Gamma_k = [b_{k-1}, b_k)$, $k = 1, \dots, K$. Введем следующие обозначения:

$$\Delta_k = \mu(\Gamma_k) = b_k - b_{k-1} -$$

размер (мера Лебега) k -й ячейки,

$$\nu_k = \sum_{j=1}^n \mathbb{I}_{\Gamma_k}(x_j) -$$

число выборочных значений, попавших в k -ю ячейку.

Определение 1.14. *Гистограммой или гистограммной оценкой плотности $f(x)$ называется статистика*

$$\hat{f}(x) = \sum_{k=1}^K \frac{\nu_k}{n\Delta_k} \mathbb{I}_{\Gamma_k}(x), \quad x \in \mathbb{R}.$$

Заметим, что гистограмма, как и плотность распределения, удовлетворяет условию нормировки:

$$\int_{-\infty}^{\infty} \hat{f}(x) dx = \int_{-\infty}^{\infty} f(x) dx = 1.$$

Отметим также, что гистограмма — смещенная и несостоятельная оценка плотности распределения. Для ее состоятельности необходимо, чтобы при $n \rightarrow +\infty$ ячейки измельчались специальным образом:

$$K = K_n \rightarrow +\infty, \quad \max_{1 \leq k \leq K_n} \Delta_k \rightarrow 0.$$

1.4 Порядковые статистики, их свойства и применение в теории статистического оценивания

Пусть $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ — наблюдаемая случайная выборка объема n из некоторого абсолютно непрерывного распределения вероятностей с функцией распределения $F(x)$ и плотностью распределения

$$f(x) = \frac{d}{dx}F(x), \quad x \in \mathbb{R}.$$

Упорядочивая выборочные значения по возрастанию, построим *вариационный ряд*:

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

где $x_{(j)}$ — j -е по величине выборочное значение, которое называется j -й *порядковой статистикой*. В частности,

$$x_{(1)} = \min_{1 \leq k \leq n} x_k, \quad x_{(n)} = \max_{1 \leq k \leq n} x_k —$$

минимальное и максимальное выборочные значения.

Найдем распределение вероятностей j -й порядковой статистики $\mathcal{L}\{x_{(j)}\}$. Введем в рассмотрение неполную бета-функцию:

$$B(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^y t^{a-1} (1-t)^{b-1} dt, \\ 0 \leq y \leq 1, \quad a \geq 1, \quad b \geq 1 \quad (B(1; a, b) \equiv 1). \quad (1.9)$$

Неполная бета-функция (1.9) связана с биномиальным распределением вероятностей:

$$B(p; a, n-a+1) = \sum_{k=a}^n C_n^k p^k (1-p)^{n-k}. \quad (1.10)$$

Теорема 1.3. Пусть $x_{(j)}$ — j -я порядковая статистика ($j \in \{1, \dots, n\}$), построенная по случайной выборке X объема n из некоторого распределения вероятностей с абсолютно непрерывной функцией распределения $F(x)$ и плотностью распределения

$$f(x) = \frac{d}{dx} F(x),$$

тогда функция распределения и плотность распределения этой порядковой статистики задаются следующими соотношениями:

$$F_{x_{(j)}}(x) = B(F(x); j, n-j+1), \quad x \in \mathbb{R}; \quad (1.11)$$

$$p_{x_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} (F(x))^{j-1} (1-F(x))^{n-j} f(x). \quad (1.12)$$

Доказательство. Действуем по определению:

$$F_{x_{(j)}}(x) = \mathbf{P}\{x_{(j)} < x\} = \mathbf{P}\{L_n(x) \geq j\}, \quad (1.13)$$

где

$$L_n(x) = \sum_{i=1}^n \mathbb{I}(x - x_i) = \sum_{i=1}^n \xi_i. \quad (1.14)$$

В (1.14) $\xi_i = \mathbb{I}(x - x_i)$, $i = 1, \dots, n$, — независимые в совокупности, одинаково распределенные случайные величины Бернулли с вероятностью успеха

$$p = \mathbf{P}\{\xi_i = 1\} = \mathbf{P}\{x_i < x\} = F(x).$$

Следовательно, $L_n(x)$ — сумма n случайных величин Бернулли — распределена по биномиальному закону $\mathcal{L}\{L_n(x)\} = Bi(n, F(x))$:

$$\mathbf{P}\{L_n(x) = k\} = C_n^k (F(x))^k (1-F(x))^{n-k}, \quad k = 0, 1, \dots, n.$$

Далее по (1.13) и (1.10) имеем

$$\begin{aligned} F_{x_{(j)}}(x) &= \sum_{k=j}^n \mathbf{P}\{L_n(x) = k\} = \sum_{k=j}^n C_n^k (F(x))^k (1 - F(x))^{n-k} = \\ &= B(F(x); j, n - j + 1), \end{aligned}$$

что совпадает с (1.11).

Соотношение (1.12) получаем дифференцированием (1.11) с учетом (1.9). \square

Следствие 1.3.1. *Распределение вероятностей первой порядковой статистики $\mathcal{L}\{x_{(1)}\}$ задается соотношениями:*

$$F_{x_{(1)}}(x) = 1 - (1 - F(x))^n, \quad p_{x_{(1)}}(x) = n(1 - F(x))^{n-1} f(x).$$

Доказательство. Полагая в (1.11) и (1.12) $j = 1$ и используя условие нормировки для биномиального распределения, получаем требуемые результаты. \square

Следствие 1.3.2. *Распределение n -й порядковой статистики $\mathcal{L}\{x_{(n)}\}$ задается соотношениями:*

$$F_{x_{(n)}}(x) = (F(x))^n, \quad p_{x_{(n)}}(x) = n(F(x))^{n-1} f(x).$$

Теорема 1.4 (о распределении r порядковых статистик). *Совместная плотность распределения вероятностей $r \geq 2$ порядковых статистик*

$$x_{(j_1)}, x_{(j_2)}, \dots, x_{(j_r)}, \quad 1 \leq j_1 < j_2 < \dots < j_r \leq n,$$

имеет вид

$$\begin{aligned} &p_{x_{(j_1)}, \dots, x_{(j_r)}}(z_1, \dots, z_r) = \\ &= \frac{n!}{(j_1 - 1)!(j_2 - j_1 - 1)! \dots (j_r - j_{r-1} - 1)!(n - j_r)!} \times \\ &\times F^{j_1-1}(z_1)(F(z_2) - F(z_1))^{j_2-j_1-1} \dots (1 - F(z_r))^{n-j_r} f(z_1) \dots f(z_r), \\ &-\infty < z_1 < z_2 < \dots < z_r < \infty. \end{aligned}$$

Следствие 1.4.1. *Совместная плотность распределения вероятностей всех n порядковых статистик имеет вид*

$$p_{x_{(1)}, \dots, x_{(n)}}(z_1, \dots, z_n) = n! f(z_1) \cdot \dots \cdot f(z_n), \quad z_1 < \dots < z_n.$$

Пример 1.4. *Доказать следствие 1.3.2, теорему 1.4 и следствие 1.4.1.*

Напомним понятие p -квантили, введенное в курсе теории вероятностей [?].

Определение 1.15. *Для некоторого p ($0 < p < 1$) p -квантилью функции распределения $F(\cdot)$ называется наименьший корень уравнения $F(\zeta) = p$, обозначаемый через ζ_p ($\zeta_p \in \mathbb{R}$).*

Определение 1.16. *Выборочной p -квантилью, вычисленной по выборке X , называется статистика*

$$Z_p = Z_p(X) = \begin{cases} x_{(k)}, & np = k - \text{целое;} \\ x_{(k+1)}, & k < np < k + 1. \end{cases} \quad (1.15)$$

Отметим, что из (1.15), (1.6) следует:

$$F_n(Z_p) = p + O\left(\frac{1}{n}\right).$$

Из данных определений и свойства состоятельности выборочной функции распределения имеем также состоятельность выборочной p -квантили: $Z_p \mathbf{P} t \zeta_p$, $n \rightarrow +\infty$.

Из теоремы 1.3 можно найти точный вид распределения $\mathcal{L}\{Z_p\}$. Для исследования асимптотических свойств выборочной p -квантили исследуем свойства $\mathcal{L}\{Z_p\}$ при $n \rightarrow \infty$.

Теорема 1.5 (о распределении выборочной квантили). Пусть выполнены условия теоремы 1.3, ζ_p — p -квантиль ($0 < p < 1$), плотность $f(x)$ — непрерывно дифференцируема в окрестности точки $x = \zeta_p$, причем $f(\zeta_p) > 0$, тогда выборочная p -квантиль Z_p при $n \rightarrow \infty$ распределена асимптотически нормально:

$$\mathcal{L}\left\{\sqrt{n}\frac{Z_p - \zeta_p}{\sqrt{p(1-p)}}f(\zeta_p)\right\} \rightarrow \mathcal{N}_1(0, 1). \quad (1.16)$$

Доказательство. Введем в рассмотрение нормированную p -квантиль:

$$\eta_n = \sqrt{n}\frac{Z_p - \zeta_p}{\sqrt{p(1-p)}}f(\zeta_p), \quad (1.17)$$

тогда утверждение (1.16) можно переписать в эквивалентном виде:

$$F_{\eta_n}(y) = \mathbf{P}\{\eta_n < y\} \rightarrow \Phi(y), \quad y \in \mathbb{R}.$$

Рассмотрим в (1.15) случай:

$$np = k, \quad \frac{k}{n} = p. \quad (1.18)$$

Согласно (1.15), в этом случае

$$Z_p = x_{(k)}. \quad (1.19)$$

Заметим, что случай, когда $k < np < k + 1$, и, следовательно, $Z_p = x_{(k+1)}$, рассматривается аналогично.

Предполагая (1.18), используя (1.19) и (1.17), получаем

$$F_{\eta_n}(y) = \mathbf{P}\left\{\sqrt{n}\frac{x_{(k)} - \zeta_p}{\sqrt{p(1-p)}}f(\zeta_p) < y\right\} = \mathbf{P}\{x_{(k)} < y_n\}, \quad y \in \mathbb{R},$$

где

$$y_n = \zeta_p + y \frac{\sqrt{p(1-p)}}{\sqrt{n}f(\zeta_p)}. \quad (1.20)$$

Воспользуемся очевидным свойством выборочной функции распределения $F_n(\cdot)$:

$$\begin{aligned} F_{\eta_n}(y) &= \mathbf{P}\{x_{(k)} < y_n\} = \mathbf{P}\left\{F_n(y_n) \geq \frac{k}{n}\right\} = \mathbf{P}\{F_n(y_n) \geq p\} = \\ &= \mathbf{P}\left\{\sqrt{n}\frac{F_n(y_n) - F(y_n)}{\sqrt{F(y_n)(1-F(y_n))}} \geq \sqrt{n}\frac{p - F(y_n)}{\sqrt{F(y_n)(1-F(y_n))}}\right\}. \end{aligned} \quad (1.21)$$

Исследуем асимптотическое поведение ($n \rightarrow +\infty$) знаменателя в (1.21). Из (1.20) имеем

$$y_n = \zeta_p + O\left(\frac{1}{\sqrt{n}}\right), \quad y_n \rightarrow \zeta_p, \quad F(y_n) \rightarrow F(\zeta_p) = F(F^{-1}(p)) = p,$$

поэтому

$$\sqrt{F(y_n)(1 - F(y_n))} \rightarrow \sqrt{p(1 - p)}.$$

Проанализируем числитель в (1.21), пользуясь линейной формулой Тейлора:

$$\begin{aligned} \sqrt{n}(p - F(y_n)) &= \sqrt{n}\left(p - \left(p + f(\zeta_p)(y_n - \zeta_p) + O\left(\frac{1}{n}\right)\right)\right) = \\ &= -\sqrt{n}f(\zeta_p)\frac{y\sqrt{p(1-p)}}{\sqrt{n}f(\zeta_p)} + O\left(\frac{1}{\sqrt{n}}\right) = -y\sqrt{p(1-p)} + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Таким образом, переходя к пределу, из (1.21) в силу асимптотической нормальности $F_n(\cdot)$ получаем

$$F_{\eta_n}(y) \rightarrow 1 - \Phi\left(-\frac{y\sqrt{p(1-p)}}{\sqrt{p(1-p)}}\right) = \Phi(y), \quad n \rightarrow +\infty.$$

□

Определение 1.17. Если при $n \rightarrow \infty$ для некоторой статистической оценки $T_n = T_n(X)$ некоторого параметра θ выполняется предельное соотношение:

$$\mathbf{P}\left\{\frac{T_n - \theta}{\sqrt{v_n}} < y\right\} \rightarrow \Phi(y), \quad y \in \mathbb{R},$$

где $v_n \rightarrow 0$, то принято говорить, что T_n является асимптотически нормальной оценкой параметра θ , а v_n называется асимптотической дисперсией этой оценки.

Следствие 1.5.1. Выборочная p -квантиль $Z_p = Z_p(X)$, определенная в (1.15), является состоятельной и асимптотически нормальной оценкой квантили $\zeta_p = F^{-1}(p)$, причем ее асимптотическая дисперсия

$$v_n = \frac{p(1-p)}{nf^2(\zeta_p)} \rightarrow 0, \quad n \rightarrow +\infty.$$

При $p = \frac{1}{2}$ квантиль $m ::= \zeta_{\frac{1}{2}} = F^{-1}\left(\frac{1}{2}\right)$ называется медианой распределения вероятностей, а $Z_{\frac{1}{2}}$ — соответственно выборочной медианой.

Пользуясь теоремой 1.5, можно найти асимптотическое распределение для выборочной медианы, учитывая, что

$$\hat{m} ::= Z_{\frac{1}{2}} = \begin{cases} x_{(k)}, & n = 2k; \\ x_{(k+1)}, & n = 2k + 1. \end{cases} \quad (1.22)$$

Если $n = 2k$ четно, то существует еще одно определение выборочной медианы:

$$\hat{m} ::= \tilde{Z}_{\frac{1}{2}} = \frac{x_{(k)} + x_{(k+1)}}{2}. \quad (1.23)$$

При $n \rightarrow \infty$ свойства выборочных медиан (1.22) и (1.23) эквивалентны.

Следствие 1.5.2. Если $f(t) > 0$, то $Z_{\frac{1}{2}}$, $\tilde{Z}_{\frac{1}{2}}$ являются состоятельными, асимптотически несмещенными и асимптотически нормальными оценками истинного значения медианы $t = \zeta_{\frac{1}{2}}$ с асимптотической дисперсией

$$v_n = \frac{1}{4nf^2(\zeta_{\frac{1}{2}})}. \quad (1.24)$$

Следствие 1.5.3. Если $X = \{x_1, \dots, x_n\}$ — случайная выборка объема n из гауссовского распределения вероятностей $\mathcal{N}_1(\theta, \sigma^2)$, то медиана совпадает с математическим ожиданием: $t = \theta$, а выборочная медиана $\hat{t} = Z_{\frac{1}{2}}$ является состоятельной и асимптотически нормальной оценкой истинного значения $t = \theta$ с асимптотической дисперсией

$$v_n = \frac{\pi \sigma^2}{2n}.$$

Доказательство. непосредственно следует из (1.24) при подстановке $f(\theta) = \frac{1}{\sqrt{2\pi}\sigma}$. \square

Пример 1.5. Сравнить по асимптотической дисперсии две статистические оценки параметра θ по выборке X объема n из $\mathcal{N}_1(\theta, \sigma^2)$: выборочное среднее $\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$ и выборочную медиану $\hat{t} = Z_{\frac{1}{2}}$.

Замечание 1.3. Выборочная медиана $\hat{t} = Z_{\frac{1}{2}}$ считается более предпочтительной, чем выборочное среднее \bar{x} , оценкой математического ожидания $\theta = \mathbf{E}\{x_k\}$ по отношению к распределениям вероятностей с хвостами, более тяжёлыми, чем у нормального распределения. Так, для распределения Лапласа с плотностью

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

убывающей на хвостах (при $|x| \rightarrow +\infty$) значительно медленнее, чем плотность

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

нормального распределения $\mathcal{N}_1(\theta, \sigma^2)$ с теми же значениями математического ожидания θ и дисперсии σ^2 , выборочная медиана оказывается оценкой максимального правдоподобия для θ (см. § ??), а выборочное среднее является таковой в случае распределения $\mathcal{N}_1(\theta, \sigma^2)$. Выборочная медиана обладает свойством робастности (см § ??).

Пример 1.6. После ознакомления с § ?? показать, что выборочная медиана $\hat{t} = Z_{\frac{1}{2}}$ является оценкой максимального правдоподобия для математического ожидания распределения Лапласа.

1.5 Неравенство информации. Неравенство Крамера — Рао. Эффективные оценки

Как отмечалось ранее, оптимальную статистическую оценку $\hat{\theta} = T(X)$ для параметра $\theta \in \mathbb{R}$ целесообразно строить исходя из критерия минимума вариации оценки:

$$V\{T(\cdot)\} = \mathbf{E}\{(T(X) - \theta)^2\} \rightarrow \min_{T(\cdot)}.$$

В этом параграфе выясним, чему равен минимум вариации: $V^* = \min_{T(\cdot)} V\{T(\cdot)\}$.

Если бы $V^* = 0$, то среднеквадратическая ошибка была бы нулевой, и оценка являлась бы точной: $\hat{\theta} \xrightarrow{\text{п.н.}} \theta$. Но этот результат, оказывается, удастся достичь лишь в вырожденных случаях. Далее будет показано, что вариация оценки ограничена снизу положительным числом:

$$V^* \geq \varepsilon, \quad \varepsilon = \varepsilon(\mathcal{I}_n) > 0,$$

где \mathcal{I}_n — величина, определяющая количество информации, содержащейся в выборке X объема n о параметре θ .

1.5.1 Информационная матрица Фишера и ее свойства

Сформулируем задачу статистического оценивания параметров в общем виде: пусть наблюдается выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из некоторого N -мерного распределения вероятностей с плотностью распределения $p(x; \theta)$, $x \in \mathbb{R}^N$, где $\theta = (\theta_1, \dots, \theta_m)^T \in \Theta \subseteq \mathbb{R}^m$ — неизвестное истинное значение векторного параметра, подлежащее оцениванию по X .

Обозначим:

$$p(X; \theta) = \prod_{l=1}^n p(x_l; \theta), \quad X \in \mathbb{R}^{nN}, \quad (1.25)$$

совместная плотность распределения выборки X , если параметр равен θ ;

$$\mathcal{P} = \{p(x; \theta), x \in \mathbb{R}^N : \theta \in \Theta \subseteq \mathbb{R}^m\} —$$

параметрическое семейство плотностей;

$$\mathbf{E}_\theta\{\psi(x_l)\} = \int_{\mathbb{R}^N} \psi(x) p(x; \theta) dx —$$

математическое ожидание, вычисленное при истинном значении параметра θ .

Аналогично:

$$\mathbf{E}_\theta\{\Psi(X)\} = \int_{\mathbb{R}^{nN}} \Psi(X) p(X; \theta) dX —$$

усреднение по распределению выборки при значении параметра θ ;

$$\nabla_\theta p(x; \theta) = \begin{pmatrix} \frac{\partial p(x; \theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial p(x; \theta)}{\partial \theta_m} \end{pmatrix} \in \mathbb{R}^m — \text{градиент функции по } \theta \text{ в точке } x;$$

$$\nabla_\theta^2 p(x; \theta) = \left(\frac{\partial^2 p(x; \theta)}{\partial \theta_j \partial \theta_k} \right)_{j,k=1}^m — (m \times m)\text{-матрица вторых производных.}$$

Определение 1.18. Параметрическое семейство плотностей распределения вероятностей \mathcal{P} называется регулярным, если выполнены следующие условия регулярности.

R1. $p(x; \theta)$ — дважды дифференцируема по θ , $\forall \theta \in \Theta$, $\forall x \in \mathbb{R}^N$.

R2. Существуют конечные математические ожидания:

$$\mathbf{E}_\theta \left\{ \left(\frac{\partial \ln p(x_l; \theta)}{\partial \theta_j} \right)^2 \right\} < +\infty, \quad \mathbf{E}_\theta \left\{ \left| \frac{\partial^2 \ln p(x_l; \theta)}{\partial \theta_j \partial \theta_k} \right| \right\} < +\infty,$$

$$j, k = 1, \dots, m, \quad \theta \in \Theta, \quad l = 1, \dots, n.$$

Р3. Допускается перестановка действий ∇_θ^j — дифференцирования j -го порядка по θ и интегрирования по $X \in \mathbb{R}^{nN}$:

$$\nabla_\theta^j \int_{\mathbb{R}^{nN}} \Psi(X) p(X; \theta) dX = \int_{\mathbb{R}^{nN}} \Psi(X) \nabla_\theta^j p(X; \theta) dX, \quad j = 1, 2.$$

Для формулировки четвертого условия регулярности введем некоторые дополнительные понятия.

Определение 1.19. Информационной матрицей Фишера для l -го выборочного значения x_l называется $(m \times m)$ -матрица $\mathcal{I} = (i_{jk})_{j,k=1}^m$, элементы которой вычисляются по формуле

$$i_{jk} = \mathbf{E}_\theta \left\{ \frac{\partial \ln p(x_l; \theta)}{\partial \theta_j} \cdot \frac{\partial \ln p(x_l; \theta)}{\partial \theta_k} \right\}, \quad j, k = 1, \dots, m,$$

или в матричном виде:

$$\mathcal{I} = \mathbf{E}_\theta \{ \nabla_\theta \ln p(x_l; \theta) (\nabla_\theta \ln p(x_l; \theta))^T \}. \quad (1.26)$$

Определение 1.20. Информационной матрицей Фишера для случайной выборки X объема n называется $(m \times m)$ -матрица

$$\mathcal{I}_n = \mathbf{E}_\theta \{ \nabla_\theta \ln p(X; \theta) (\nabla_\theta \ln p(X; \theta))^T \}. \quad (1.27)$$

Заметим, что математические ожидания в (1.26) и (1.27) существуют в силу условия **Р2**.

Р4. Информационная матрица Фишера невырождена:

$$|\mathcal{I}| \neq 0 \quad (|\mathcal{I}_n| \neq 0).$$

Следствие 1.5.4. Если выполнены условия регулярности **Р1** — **Р4**, то справедливо тождество

$$\mathbf{E}_\theta \{ \nabla_\theta \ln p(X; \theta) \} \equiv \mathbf{0}_m.$$

Доказательство. В силу (1.27), условия **Р3** при $j = 1$ и условия нормировки имеем

$$\begin{aligned} \mathbf{E}_\theta \{ \nabla_\theta \ln p(X; \theta) \} &= \mathbf{E}_\theta \left\{ \frac{1}{p(X; \theta)} \nabla_\theta p(X; \theta) \right\} = \\ &= \int_{\mathbb{R}^{nN}} \frac{1}{p(X; \theta)} \nabla_\theta p(X; \theta) p(X; \theta) dX = \int_{\mathbb{R}^{nN}} \nabla_\theta p(X; \theta) dX = \\ &= \nabla_\theta \int_{\mathbb{R}^{nN}} p(X; \theta) dX = \nabla_\theta 1 = \mathbf{0}_m. \end{aligned}$$

□

Свойства информационной матрицы Фишера

Свойство 1. Информационная матрица Фишера — симметричная матрица:

$$\mathcal{I} = \mathcal{I}^T \quad (\mathcal{I}_n = \mathcal{I}_n^T).$$

Доказательство. следует из определений для информационных матриц (1.26), (1.27).

□

Свойство 2. Для вычисления информационной матрицы Фишера могут быть использованы следующие формулы:

$$\mathcal{I} = \mathbf{E}_\theta \{-\nabla_\theta^2 \ln p(x_l; \theta)\}; \quad (1.28)$$

$$\mathcal{I}_n = \mathbf{E}_\theta \{-\nabla_\theta^2 \ln p(X; \theta)\}. \quad (1.29)$$

Доказательство. Докажем (1.28) (формула (1.29) доказывается аналогично). Пусть $l \in \{1, \dots, n\}$ — произвольный номер случайного выборочного наблюдения $x_l \in \mathbb{R}^N$. Вычислим и преобразуем правую часть (1.28) с использованием условий регулярности **R1–R4** и свойств математического ожидания $\mathbf{E}_\theta\{\cdot\}$:

$$\begin{aligned} \mathbf{E}_\theta \{-\nabla_\theta^2 \ln p(x_l; \theta)\} &= \mathbf{E}_\theta \left\{ -\nabla_\theta \left(\frac{1}{p(x_l; \theta)} \nabla_\theta p(x_l; \theta) \right) \right\} = \\ &= \mathbf{E}_\theta \left\{ \frac{1}{(p(x_l; \theta))^2} \nabla_\theta p(x_l; \theta) (\nabla_\theta p(x_l; \theta))^T \right\} - \mathbf{E}_\theta \left\{ \frac{1}{p(x_l; \theta)} \nabla_\theta^2 p(x_l; \theta) \right\} = \\ &= \mathbf{E}_\theta \left\{ \nabla_\theta \ln p(x_l; \theta) (\nabla_\theta \ln p(x_l; \theta))^T \right\} - \int_{\mathbb{R}^N} \frac{1}{p(x; \theta)} \nabla_\theta^2 p(x; \theta) p(x; \theta) dx = \\ &= \mathcal{I} - \int_{\mathbb{R}^N} \nabla_\theta^2 p(x; \theta) dx = \mathcal{I} - \nabla_\theta^2 \int_{\mathbb{R}^N} p(x; \theta) dx = \mathcal{I} - \nabla_\theta^2 1 = \mathcal{I}. \end{aligned}$$

□

Свойство 3. Информационная матрица Фишера \mathcal{I}_n для всей выборки X объема n линейно зависит от n : $\mathcal{I}_n = n\mathcal{I}$.

Доказательство. Воспользуемся соотношениями (1.29) и (1.25):

$$\mathcal{I}_n = \mathbf{E}_\theta \left\{ -\nabla_\theta^2 \ln \prod_{l=1}^n p(x_l; \theta) \right\} = \sum_{l=1}^n \mathbf{E}_\theta \{-\nabla_\theta^2 \ln p(x_l; \theta)\} = n\mathcal{I}.$$

□

Свойство 4. Информационная матрица Фишера \mathcal{I} есть ковариационная матрица для случайного вектора $\nabla_\theta \ln p(x_l; \theta) \in \mathbb{R}^m$.

Доказательство. По определению ковариационной матрицы имеем

$$\begin{aligned} \mathbf{Cov}\{\nabla_\theta \ln p(x_l; \theta), \nabla_\theta \ln p(x_l; \theta)\} &= \\ &= \mathbf{E}_\theta \{ (\nabla_\theta \ln p(x_l; \theta) - \mathbf{E}_\theta \{\nabla_\theta \ln p(x_l; \theta)\}) \times \\ &\times (\nabla_\theta \ln p(x_l; \theta) - \mathbf{E}_\theta \{\nabla_\theta \ln p(x_l; \theta)\})^T \} = \\ &= \mathbf{E}_\theta \{ \nabla_\theta \ln p(x_l; \theta) (\nabla_\theta \ln p(x_l; \theta))^T \} = \mathcal{I}, \end{aligned}$$

где учтено, что в силу следствия 1.5.4

$$\mathbf{E}_\theta \{\nabla_\theta \ln p(x_l; \theta)\} \equiv \mathbf{0}_m.$$

□

Свойство 5. Информационная матрица Фишера \mathcal{I} в условиях регулярности **R1–R4** есть матрица положительно определенная: $\mathcal{I} \succ 0$.

Доказательство. В силу **Свойства 4**, \mathcal{I} — ковариационная матрица, а любая ковариационная матрица является неотрицательно определенной: $\mathcal{I} \succeq 0$. Согласно условию регулярности **R4**, $|\mathcal{I}| \neq 0$, поэтому $\mathcal{I} \succ 0$. □

1.5.2 Неравенство информации и его следствия. Эффективные оценки

Прежде чем определить понятие эффективности статистической оценки, получим так называемое неравенство информации.

Теорема 1.6 (неравенство информации). Пусть $\hat{\theta} = (\hat{\theta}_k)_{k=1}^m = T(X) = (T_k(X))_{k=1}^m \in \mathbb{R}^m$ — произвольная несмещенная статистическая оценка векторного параметра $\theta = (\theta_k)_{k=1}^m \in \mathbb{R}^m$ по случайной выборке X объема n из некоторого N -мерного распределения с плотностью $p(\cdot; \theta) \in \mathcal{P}$, $V = (v_{jk})_{j,k=1}^m = \mathbf{E}\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\}$ — матрица вариаций оценки $\hat{\theta}$, а \mathcal{I}_n — информационная матрица Фишера для всей выборки X . Тогда если выполнены условия регулярности **R1** — **R4** для семейства \mathcal{P} , то $(m \times m)$ -матрица $\Delta = (\Delta_{jk})_{j,k=1}^m ::= V - \mathcal{I}_n^{-1}$ неотрицательно определена:

$$\Delta = V - \mathcal{I}_n^{-1} \succeq 0. \quad (1.30)$$

Доказательство. Введем в рассмотрение два вспомогательных случайных m -вектора:

$$U_1 = T(X) - \theta = \hat{\theta} - \theta \quad (1.31)$$

случайное уклонение оценки $\hat{\theta}$;

$$U_2 = \nabla_{\theta} \ln p(X; \theta). \quad (1.32)$$

Вычислим моменты первого и второго порядка для этих случайных векторов. В силу несмещенности оценки $\hat{\theta}$

$$\mathbf{E}_{\theta}\{U_1\} = \mathbf{0}_m, \quad (1.33)$$

а из следствия 1.5.4 имеем

$$\mathbf{E}_{\theta}\{U_2\} = \mathbf{0}_m. \quad (1.34)$$

Вычислим ковариации для векторов (1.31) и (1.32):

$$A_{ij} = \mathbf{Cov}\{U_i, U_j\} = \mathbf{E}_{\theta}\{U_i U_j^T\}, \quad i, j = 1, 2.$$

Согласно (1.31), (1.32), имеем

$$\begin{aligned} A_{11} &= V, \quad A_{22} = \mathcal{I}_n, \quad A_{21} = A_{12}^T, \\ A_{12} &= \mathbf{E}_{\theta}\{(T(X) - \theta)U_2^T\} = \mathbf{E}_{\theta}\{T(X)U_2^T\} - \theta \mathbf{E}_{\theta}\{U_2^T\} = \\ &= \mathbf{E}_{\theta}\{T(X)(\nabla_{\theta} \ln p(X; \theta))^T\} = \mathbf{E}_{\theta}\left\{T(X) \frac{1}{p(X; \theta)} (\nabla_{\theta} p(X; \theta))^T\right\} = \\ &= \int_{\mathbb{R}^{nN}} p(X; \theta) T(X) \frac{1}{p(X; \theta)} (\nabla_{\theta} p(X; \theta))^T dx = \\ &= \nabla_{\theta} \int_{\mathbb{R}^{nN}} T(X) p(X; \theta) dX = \nabla_{\theta} \theta = \mathbb{I}_m, \end{aligned}$$

где \mathbb{I}_m — единичная $(m \times m)$ -матрица.

При помощи линейного преобразования над случайными векторами U_1, U_2 построим случайный вектор

$$\Xi = (\xi_k)_{k=1}^m = U_1 - \mathcal{I}_n^{-1} U_2 \in \mathbb{R}^m. \quad (1.35)$$

Из (1.33), (1.34) и (1.35) очевидно, что $\mathbf{E}_{\theta}\{\Xi\} = \mathbf{0}_m$.

Вычислим ковариационную матрицу для случайного вектора Ξ :

$$\mathbf{Cov}\{\Xi, \Xi\} = \mathbf{E}_{\theta}\{\Xi \Xi^T\} = \mathbf{E}_{\theta}\{(U_1 - \mathcal{I}_n^{-1} U_2)(U_1 - \mathcal{I}_n^{-1} U_2)^T\} =$$

$$\begin{aligned}
&= A_{11} - A_{12}\mathcal{I}_n^{-1} - \mathcal{I}_n^{-1}A_{12}^T + \mathcal{I}_n^{-1}A_{22}\mathcal{I}_n^{-1} = \\
&= V - \mathcal{I}_n^{-1} - \mathcal{I}_n^{-1} + \mathcal{I}_n^{-1}\mathcal{I}_n\mathcal{I}_n^{-1} = V - \mathcal{I}_n^{-1} = \Delta.
\end{aligned}$$

Таким образом, $\Delta = \mathbf{Cov}\{\Xi, \Xi\}$, а любая ковариационная матрица неотрицательно определена, следовательно, $\Delta \succeq 0$. \square

Следствие 1.6.1. Вариация $v_{kk} = \mathbf{E}\{(\hat{\theta}_k - \theta_k)^2\}$ оценки $\hat{\theta}_k = T_k(X)$ для параметра θ_k ограничена снизу положительным числом ($k \in \{1, \dots, m\}$):

$$v_{kk} \geq (\mathcal{I}_n^{-1})_{kk} > 0, \quad (1.36)$$

причем нижняя граница для вариации достигается тогда и только тогда, когда $\exists d_{k1}, \dots, d_{km} \in \mathbb{R}$ такие, что справедливо равенство

$$T_k(X) - \theta_k \xrightarrow{n.н.} \sum_{j=1}^m d_{kj} \frac{\partial \ln p(X; \theta)}{\partial \theta_j}. \quad (1.37)$$

При этом если (1.37) выполняется, то

$$d_{kj} = (\mathcal{I}_n^{-1})_{kj}, \quad j = 1, \dots, m.$$

Доказательство. Из теоремы 1.6 следует:

$$\Delta_{kk} = (V - \mathcal{I}_n^{-1})_{kk} = v_{kk} - (\mathcal{I}_n^{-1})_{kk} \geq 0,$$

что приводит к (1.36). Неравенство (1.36) обращается в равенство тогда и только тогда, когда $\Delta_{kk} = \mathbf{E}\{\xi_k^2\} = 0$, а это, в свою очередь, равносильно тому, что $\xi_k = (T(X) - \theta - \mathcal{I}_n^{-1}\nabla_{\theta} \ln p(X; \theta))_k \xrightarrow{П.Н.} 0$, что и означает (1.37). \square

Следствие 1.6.2. Определитель матрицы вариаций ограничен снизу положительным числом:

$$|V| \geq \frac{1}{|\mathcal{I}_n|} > 0. \quad (1.38)$$

Доказательство. Воспользуемся (1.30): $\Delta = V - \mathcal{I}_n^{-1} \succeq 0$. Умножим Δ слева на матрицу $(\mathcal{I}_n^{\frac{1}{2}})^T$, а справа — на матрицу $\mathcal{I}_n^{\frac{1}{2}}$, где $\mathcal{I}_n^{\frac{1}{2}}$ — решение матричного уравнения $Y^T \mathcal{I}_n^{-1} Y = \mathbb{I}_m$ относительно Y ($\mathcal{I}_n = \mathcal{I}_n^{\frac{1}{2}}(\mathcal{I}_n^{\frac{1}{2}})^T$) [?, ?]. Для получившейся неотрицательно определенной матрицы определим спектр собственных значений $\{\lambda_1, \dots, \lambda_m\}$. Составим характеристическое уравнение:

$$|(\mathcal{I}_n^{\frac{1}{2}})^T V \mathcal{I}_n^{\frac{1}{2}} - \mathbb{I}_m| = 0, \quad \lambda_j \geq 0, \quad j = 1, \dots, m,$$

или, что эквивалентно:

$$|(\mathcal{I}_n^{\frac{1}{2}})^T V \mathcal{I}_n^{\frac{1}{2}} - \mu_j \mathbb{I}_m| = 0, \quad \mu_j = 1 + \lambda_j \geq 1, \quad j = 1, \dots, m.$$

При этом $\{\mu_j\}_{j=1}^m$ — характеристические числа матрицы $(\mathcal{I}_n^{\frac{1}{2}})^T V \mathcal{I}_n^{\frac{1}{2}}$.

Воспользуемся свойством характеристических чисел матрицы и свойством определителя матрицы:

$$|(\mathcal{I}_n^{\frac{1}{2}})^T V \mathcal{I}_n^{\frac{1}{2}}| = |(\mathcal{I}_n^{\frac{1}{2}})^T| |V| |\mathcal{I}_n^{\frac{1}{2}}| = |\mathcal{I}_n^{\frac{1}{2}}| |(\mathcal{I}_n^{\frac{1}{2}})^T| |V| = |\mathcal{I}_n| |V| = \prod_{j=1}^m \mu_j \geq 1,$$

откуда и следует (1.38). \square

Определение 1.21. Эффективностью несмещенной статистической оценки $\hat{\theta} = T(X)$ векторного параметра $\theta \in \mathbb{R}^m$ называется величина

$$\mathcal{E}_n = \frac{\inf_{T(\cdot)} |V|}{|V|} = \frac{1}{|V||\mathcal{I}_n|}.$$

Из (1.38) следует, что $0 \leq \mathcal{E}_n \leq 1$.

Определение 1.22. Несмещенная оценка $\hat{\theta} = T(X)$ называется эффективной, если $\forall \theta \in \mathbb{R}^m, \forall n \geq 1$ эффективность $\mathcal{E}_n = 1$, иначе говоря, если неравенство (1.38) обращается в равенство.

Следствие 1.6.3. Если выполнены условия регулярности **R1–R4** и $\hat{\theta} = T(X)$ — эффективная оценка, то ее матрица вариаций имеет вид

$$V = \frac{1}{n} \mathcal{I}^{-1}.$$

Доказательство. Из хода доказательства следствия 1.6.2 заключаем:

$$\mu_j = 1 \Rightarrow \lambda_j = 0, \quad j = 1, \dots, m.$$

Следовательно, $\Delta = V - \mathcal{I}_n^{-1} = \mathbf{0}_m \Rightarrow V = \mathcal{I}_n^{-1}$. Отсюда, воспользовавшись свойством информационной матрицы Фишера: $\mathcal{I}_n = n\mathcal{I}$, получаем доказываемое. \square

Следствие 1.6.4 (критерий эффективности оценки). Для того чтобы в условиях регулярности **R1–R4** несмещенная оценка $\hat{\theta} = T(X)$ была эффективной, необходимо и достаточно, чтобы существовала такая положительно определенная матрица $\bar{D} \succ 0$, для которой выполняется соотношение

$$\nabla_{\theta} \ln p(X; \theta) \xrightarrow{n.н.} \bar{D}(T(X) - \theta). \quad (1.39)$$

При этом если (1.39) выполняется, то $\bar{D} = \mathcal{I}_n$ — информационная матрица Фишера для всей выборки.

Доказательство. Из определения эффективной оценки и следствия 1.6.1, а именно из системы соотношений (1.37), имеем

$$T(X) - \theta \xrightarrow{n.н.} \mathcal{I}_n^{-1} \nabla_{\theta} \ln p(X; \theta).$$

Умножая левую и правую части на \mathcal{I}_n , получаем (1.39). \square

Следствие 1.6.5. В условиях регулярности **R1–R4** эффективная оценка состоятельна.

Доказательство. В силу следствия 1.6.3:

$$V = \frac{1}{n} \mathcal{I}_n^{-1} \rightarrow 0, \quad n \rightarrow +\infty \Rightarrow \hat{\theta} \xrightarrow{P} \theta.$$

\square

1.5.3 Неравенство Крамера — Рао

Неравенство Крамера — Рао — частный случай неравенства информации, когда оценивается одномерный (скалярный) параметр $\theta \in \Theta \subseteq \mathbb{R}^1$, т. е. $m = 1$.

Полагая $m = 1$ в первых двух пунктах данного параграфа, упростим ранее введенные обозначения:

$$\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^1 —$$

некоторая несмещенная статистическая оценка параметра θ ;

$$V\{\hat{\theta}\} = \mathbf{E}_{\theta}\{(\hat{\theta} - \theta)^2\} \geq 0 —$$

вариация оценки $\hat{\theta}$;

$$\mathcal{I} = - \int_{\mathbb{R}^N} \frac{\partial^2 \ln p(x; \theta)}{\partial \theta^2} p(x; \theta) dx = \int_{\mathbb{R}^N} \left(\frac{\partial \ln p(x; \theta)}{\partial \theta} \right)^2 p(x; \theta) dx > 0 —$$

количество информации по Фишеру о параметре θ , содержащейся в отдельном выборочном значении $x_l \in \mathbb{R}^N$ ($l = 1, \dots, n$);

$$\mathcal{I}_n = n\mathcal{I} —$$

количество информации по Фишеру о параметре θ , содержащейся во всей выборке X объема n .

Теорема 1.7 (неравенство Крамера — Рао). *Если плотность распределения $p(x; \theta)$ удовлетворяет ранее приведенным условиям регулярности **R1–R4**, то вариация любой несмещенной статистической оценки $\hat{\theta} = T(X)$ ограничена снизу положительным числом:*

$$V\{\hat{\theta}\} \geq \frac{1}{\mathcal{I}_n} = \frac{1}{n\mathcal{I}} > 0. \quad (1.40)$$

Доказательство. Положим $m = 1$ в теореме 1.6 либо в следствии 1.6.2. □

Замечание 1.4. *В условиях регулярности при конечном объеме выборки $n < +\infty$ невозможно оценить параметр θ безошибочно, т. е.*

$$V\{\hat{\theta}\} \geq \varepsilon > 0, \quad \varepsilon = \frac{1}{n\mathcal{I}}.$$

Замечание 1.5. *Увеличивая n , можно достичь наперед заданной точности оценивания, поскольку $\varepsilon \rightarrow 0$ при $n \rightarrow \infty$.*

Следствие 1.7.1. *Для того, чтобы несмещенная оценка $\hat{\theta} = T(X) \in \mathbb{R}^1$ была эффективной, т. е. чтобы неравенство Крамера — Рао (1.40) обращалось в равенство, необходимо и достаточно, чтобы существовало такое число $\overline{D} > 0$, что справедливо соотношение*

$$\frac{\partial \ln p(X; \theta)}{\partial \theta} \xrightarrow{n.н.} \overline{D}(T(X) - \theta). \quad (1.41)$$

При этом если (1.41) выполняется, то $\overline{D} = \mathcal{I}_n = n\mathcal{I}$.

Доказательство. Положим $m = 1$ в следствии 1.6.4. □

1.6 Метод моментов. Оценки по методу моментов и их свойства

Метод моментов — простейший из методов построения статистических оценок, предложенный в 1900 г. английским статистиком К. Пирсоном.

Пусть наблюдается случайная выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ объема n из некоторого, для простоты, одномерного распределения вероятностей с функцией распределения $F(x; \theta^0)$, $x \in \mathbb{R}^1$, $\theta^0 \in \mathbb{R}^m$, где θ^0 — истинное значение вектора параметров, которое неизвестно. Задача заключается в том, чтобы по выборке X построить статистическую оценку $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)^T = T(X) = (T_1(X), \dots, T_m(X))^T \in \mathbb{R}^m$.

Будем предполагать, что функция распределения $F(\cdot; \theta)$ такова, что $\forall \theta = (\theta_1, \dots, \theta_m)^T \in \mathbb{R}^m$ существуют начальные моменты этого распределения до m -го порядка включительно. Пусть α_k — начальный момент k -го порядка:

$$\begin{aligned} \alpha_k &= \alpha_k(\theta_1, \dots, \theta_m) ::= \mathbf{E}_\theta \{x_l^k\} = \\ &= \int_{-\infty}^{+\infty} x^k dF(x; \theta), \quad k = 1, \dots, m. \end{aligned} \quad (1.42)$$

Будем дополнительно предполагать, что функции (1.42) задают взаимно-однозначное и непрерывное функциональное соответствие $\{\theta_1, \dots, \theta_m\} \leftrightarrow \{\alpha_1, \dots, \alpha_m\}$. Если в (1.42) истинные значения моментов известны: $\alpha_k^0 = \alpha_k(\theta_1^0, \dots, \theta_m^0)$, $k = 1, \dots, m$, то решая систему уравнений

$$\alpha_k(\theta_1, \dots, \theta_m) = \alpha_k^0, \quad k = 1, \dots, m,$$

относительно θ , однозначно определяем $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T$.

В качестве оценок начальных моментов будем использовать выборочные моменты (1.8):

$$\hat{\alpha}_k = a_k = \frac{1}{n} \sum_{l=1}^n x_l^k, \quad k = 1, \dots, m,$$

для которых ранее в теореме 1.1 была установлена сильная состоятельность:

$$a_k \xrightarrow{\text{П.Н.}} \alpha_k, \quad n \rightarrow +\infty, \quad k = 1, \dots, m. \quad (1.43)$$

Кроме того, эти оценки — несмещенные и асимптотически нормальные.

Определение 1.23 (принцип метода моментов). В качестве оценки $\hat{\theta} = T(X)$ надлежит принимать такое значение $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)^T$, которое является решением системы из m уравнений

$$\alpha_k(\theta_1, \dots, \theta_m) = a_k, \quad k = 1, \dots, m. \quad (1.44)$$

Построенная таким образом оценка $\hat{\theta} = T(X)$ называется оценкой по методу моментов (ОММ).

Теорема 1.8. Пусть функция распределения $F(x; \theta)$, $x \in \mathbb{R}^1$, $\theta \in \mathbb{R}^m$, такова, что:

1. существуют начальные моменты $\alpha_k = \alpha_k(\theta)$, $k = 1, \dots, m$;
2. функциональное отображение $\alpha = \alpha(\theta) : \mathbb{R}^m \rightarrow \mathbb{R}^m$, задаваемое формулой (1.42), является взаимно-однозначным и непрерывным.

Тогда ОММ $\hat{\theta} = T(X)$, определяемая (1.44), является состоятельной оценкой, т. е. имеет место сходимость по вероятности:

$$\hat{\theta} \xrightarrow{\mathbf{P}} \theta^0, \quad n \rightarrow +\infty.$$

Доказательство. Результат теоремы вытекает из (1.43) и известного свойства сходимости по вероятности: если $\xi_n \xrightarrow{\mathbf{P}} \xi$, $n \rightarrow +\infty$, и $f(\cdot)$ — непрерывная функция, то $f(\xi_n) \xrightarrow{\mathbf{P}} f(\xi)$, $n \rightarrow +\infty$. \square

Замечание 1.6. Если $\nabla_{\theta}\alpha(\theta)|_{\theta=\theta^0} \neq \mathbf{0}_m$, то ОММ $\hat{\theta}$ — асимптотически нормальная оценка.

Замечание 1.7. Несмещенность и эффективность оценок по методу моментов в общем случае показать не удастся. Они не обладают высокой точностью и на практике используются как начальные приближения.

Пример 1.7. Наблюдается выборка $X = (x_1, \dots, x_n)^{\mathbf{T}} \in \mathbb{R}^n$ из нормального распределения $\mathcal{N}_1(\theta^0, \sigma^2)$, где математическое ожидание $\theta^0 \in \mathbb{R}^1$ неизвестно, а дисперсия $\sigma^2 > 0$ известна. Показать, что ОММ для θ^0 имеет вид

$$\hat{\theta} = T(X) = a_1 = \bar{x} = \frac{1}{n} \sum_{l=1}^n x_l.$$

1.7 Метод максимального правдоподобия и его свойства

1.7.1 Принцип максимального правдоподобия. Пример

Метод максимального правдоподобия был предложен в 1930 г. английским статистиком Р. Фишером. Этот метод является универсальным для построения статистических оценок.

Рассмотрим задачу статистического оценивания параметров в общем виде. Пусть наблюдается случайная выборка $X = (x_1^{\mathbf{T}}, \dots, x_n^{\mathbf{T}})^{\mathbf{T}} \in \mathbb{R}^{nN}$ объема n из некоторого N -мерного распределения вероятностей с плотностью распределения $p(x; \theta^0)$, $x \in \mathbb{R}^N$, где $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^{\mathbf{T}} \in \Theta \subseteq \mathbb{R}^m$ — неизвестное истинное значение вектора параметров. Обозначим: Θ^* — замыкание области Θ .

Определение 1.24. Функцией правдоподобия Фишера называется плотность распределения выборки X , причем X считается фиксированной, а вектор параметров $\theta \in \Theta^*$ — переменной величиной:

$$L = L(\theta) ::= p(X; \theta) = \prod_{k=1}^n p(x_k; \theta) \geq 0, \quad \theta \in \Theta^*. \quad (1.45)$$

Логарифмической функцией правдоподобия Фишера называется функция

$$l = l(\theta) ::= \ln L(\theta) = \sum_{k=1}^n \ln p(x_k; \theta), \quad \theta \in \Theta^*. \quad (1.46)$$

В случае дискретного распределения вероятностей

$$P(z; \theta) = P_{\theta}\{x_k = z\}, \quad z \in Z = \{z_1, \dots, z_M\},$$

функция правдоподобия строится аналогично (1.45):

$$L(\theta) = \prod_{k=1}^n P(x_k; \theta), \quad \theta \in \Theta^*. \quad (1.47)$$

Проясним содержательный смысл функции правдоподобия. Из (1.47) видно, что $L(\theta)$ — это вероятность получить выборку X в предположении, что параметр равен θ . Эта величина рассматривается как степень правдоподобия выборки X при данном θ или степень правдоподобия θ при данной X .

Если $L(\theta^{**}) > L(\theta^*)$, то для данной выборки X значение параметра $\theta = \theta^{**}$ более правдоподобно, чем θ^* .

Пример 1.8. Менеджеру компьютерной фирмы предложили партию чипов. Поставщик утверждает, что доля годных чипов в этой партии составляет 95 %, т. е. $\theta^0 = 0,95$. У менеджера есть предположение, что $\theta^0 = 0,7$. Чтобы разрешить этот спор, открывают контейнер, наудачу выбирают $n = 3$ коробки и проверяют чипы до обнаружения первого брака. Допустим, $x_1 = 4$, $x_2 = 6$, $x_3 = 3$ (x_k — число извлеченных небракованных чипов до первого обнаружения брака из k -й коробки). Что же более правдоподобно: предположение поставщика или менеджера?

Извлечение чипов — схема независимых испытаний Бернулли. Случайная величина x_k — случайное число попыток до первой неудачи (обнаружения первого брака в данном случае) — имеет следующее распределение вероятностей (известное как геометрическое распределение):

$$P_\theta\{x_k = z\} = \theta^z(1 - \theta), \quad z = 0, 1, 2, \dots; \quad \theta \in \Theta^* = \{0,95; 0,7\}.$$

Запишем функцию правдоподобия (1.47):

$$L(\theta) = \prod_{k=1}^3 (\theta^{x_k}(1 - \theta)) = (1 - \theta)^3 \theta^{13};$$

$$L(0,95) = 6,4 \cdot 10^{-5}, \quad L(0,7) = 2,6 \cdot 10^{-4}.$$

Таким образом, предположение менеджера более правдоподобно, чем утверждение поставщика.

Определение 1.25 (принцип максимума правдоподобия). В качестве статистической оценки параметра θ^0 надлежит использовать такое значение $\hat{\theta} = T(X)$, при котором функция правдоподобия (1.45) (или (1.47)) достигает максимума:

$$L(\hat{\theta}) = \max_{\theta \in \Theta^*} L(\theta), \quad \hat{\theta} = \arg \max_{\theta \in \Theta^*} L(\theta), \quad (1.48)$$

или, что эквивалентно, достигается максимальное значение логарифмической функции правдоподобия (1.46):

$$l(\hat{\theta}) = \max_{\theta \in \Theta^*} l(\theta), \quad \hat{\theta} = \arg \max_{\theta \in \Theta^*} l(\theta). \quad (1.49)$$

Построенная таким образом статистическая оценка $\hat{\theta} = T(X)$ называется оценкой максимального правдоподобия (ОМП, МП-оценкой).

1.7.2 Вычисление МП-оценки. Примеры

В условиях регулярности **R1–R4** функция $l(\theta)$ дважды непрерывно дифференцируема по $\theta = (\theta_1, \dots, \theta_m)^T$, поэтому для нахождения максимума в (1.49) можно использовать необходимые и достаточные условия максимума. Необходимое условие имеет вид

$$\nabla_\theta l(\theta) = \mathbf{0}_m, \quad (1.50)$$

или покомпонентно:

$$\begin{cases} \frac{\partial l(\theta)}{\partial \theta_1} = 0, \\ \vdots \\ \frac{\partial l(\theta)}{\partial \theta_m} = 0. \end{cases} \quad (1.51)$$

Определение 1.26. Система m уравнений (1.51) относительно m неизвестных $\theta_1, \dots, \theta_m$ называется системой уравнений правдоподобия.

Решая систему (1.51) (векторное уравнение (1.50)), находим стационарные точки, и чтобы среди них найти локальный максимум, необходимо проверить достаточное условие:

$$\nabla_{\theta}^2 l(\theta)|_{\theta=\hat{\theta}} \prec 0. \quad (1.52)$$

Если локальных максимумов несколько, то глобальный находится перебором значений функции $l(\cdot)$ в этих точках.

Пример 1.9 (продолжение примера 1.8). *Построить оценку доли годных чипов в партии по методу максимального правдоподобия, предполагая, что $\theta \in \Theta = (0, 1)$, $\theta^* = [0, 1]$.*

Логарифмическая функция правдоподобия имеет вид

$$l(\theta) = \ln \prod_{k=1}^n (\theta^{x_k} (1 - \theta)) = n \ln(1 - \theta) + n\bar{x} \ln \theta, \quad \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

Воспользуемся (1.49), (1.50), (1.52), составим и решим уравнение правдоподобия ($m = 1$):

$$\begin{aligned} \nabla_{\theta} l(\theta) &= -\frac{n}{1 - \theta} + \frac{n\bar{x}}{\theta} = 0; \\ \hat{\theta} &= \frac{\bar{x}}{\bar{x} + 1} = \frac{\frac{13}{3}}{\frac{13}{3} + 1} = \frac{13}{16} \approx 0,8. \end{aligned}$$

Так как $\nabla_{\theta}^2 l(\theta)|_{\theta=\hat{\theta}} < 0$, то $\hat{\theta}$ — оценка максимального правдоподобия.

Пример 1.10. *Наблюдается случайная выборка $X = (x_1, \dots, x_n)^T$ объема n из нормального распределения $\mathcal{N}_1(\theta^0, \sigma^2)$, где математическое ожидание θ^0 неизвестно, а дисперсия σ^2 известна. Построить МП-оценку для θ^0 и исследовать ее свойства.*

Логарифмическая функция правдоподобия имеет вид

$$l(\theta) = \ln \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_k - \theta)^2}{2\sigma^2}} = n \ln \frac{1}{\sqrt{2\pi\sigma}} - \frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \theta)^2.$$

Составим и решим уравнение правдоподобия:

$$\nabla_{\theta} l(\theta) = \frac{2}{2\sigma^2} \sum_{k=1}^n (x_k - \theta) = 0;$$

$$\frac{1}{\sigma^2} n(\bar{x} - \theta) = 0;$$

$$\hat{\theta} = \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

Так как $\nabla_{\theta}^2 l(\theta)|_{\theta=\hat{\theta}} = -\frac{n}{\sigma^2} < 0$, то $\hat{\theta}$ — оценка максимального правдоподобия.

Таким образом, искомая МП-оценка построена. Проведем исследование ее свойств. Так как $\hat{\theta} = \bar{x}$ — выборочное среднее, т. е. выборочный момент первого порядка, то $\hat{\theta} \xrightarrow{\text{п.н.}} \theta^0$ при $n \rightarrow \infty$, и имеем сильную состоятельность для найденной оценки. Далее $\mathbf{E}_{\theta^0}(\hat{\theta}) = \theta^0$, что означает ее несмещенность. Исследуем полученную оценку на эффективность, для этого воспользуемся критерием эффективности:

$$\frac{\partial \ln p(X; \theta)}{\partial \theta} = \nabla_{\theta} l(\theta) = \frac{n}{\sigma^2} (\hat{\theta} - \theta) = \mathcal{I}_n(\hat{\theta} - \theta),$$

следовательно, $\hat{\theta}$ — эффективная оценка. Кроме того, $\mathcal{I}_n = \frac{n}{\sigma^2}$ — количество информации по Фишеру во всей выборке о параметре θ . Так как $\hat{\theta}$ линейно зависит от гауссовских случайных величин $\{x_k\}_{k=1}^n$, то эта оценка имеет гауссовское распределение:

$$\mathcal{L}_{\theta^0}\{\hat{\theta}\} = \mathcal{N}_1\left(\theta^0, \frac{\sigma^2}{n}\right).$$

1.7.3 Свойства МП-оценки

Свойство 1. Инвариантность МП-оценки.

Пусть $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)^T$ — единственная МП-оценка для параметров $(\theta_1, \dots, \theta_m)^T$, а $(\tau_1, \dots, \tau_m)^T$ — некоторый набор m "новых" параметров, которые связаны со "старыми" параметрами некоторым взаимно-однозначным функциональным преобразованием:

$$\begin{aligned} \tau_1 &= f_1(\theta_1, \dots, \theta_m), \\ &\dots \\ \tau_m &= f_m(\theta_1, \dots, \theta_m), \end{aligned} \tag{1.53}$$

тогда оценка максимального правдоподобия для $(\tau_1, \dots, \tau_m)^T$ также единственна и находится функциональным преобразованием оценки максимального правдоподобия "старых" параметров:

$$\begin{aligned} \hat{\tau}_1 &= f_1(\hat{\theta}_1, \dots, \hat{\theta}_m), \\ &\dots \\ \hat{\tau}_m &= f_m(\hat{\theta}_1, \dots, \hat{\theta}_m). \end{aligned}$$

Доказательство. Это свойство доказывается с помощью замены переменных (1.53) в экстремальной задаче (1.48), (1.49). \square

Свойство 2. Связь МП-оценки с эффективной оценкой.

Пусть выполнены условия регулярности **R1–R4** и существует эффективная оценка $\hat{\theta}^* = T^*(X)$, тогда оценка максимального правдоподобия $\hat{\theta}$ совпадает с $\hat{\theta}^*$:

$$\hat{\theta} = \hat{\theta}^* = T^*(X).$$

Доказательство. Воспользуемся критерием эффективности оценки:

$$\nabla_{\theta} l(\theta) = \mathcal{I}_n(\hat{\theta}^* - \theta).$$

С другой стороны,

$$\nabla_{\theta} l(\theta) = \mathbf{0}_m -$$

система уравнений правдоподобия. Так как $\mathcal{I}_n \succ 0$ и $\nabla_{\theta}^2 l(\theta)|_{\theta=\hat{\theta}^*} \prec 0$, то единственная оценка максимального правдоподобия — это $\hat{\theta} = \hat{\theta}^*$. \square

Следствие 1.8.1. В условиях регулярности метод максимального правдоподобия является способом построения эффективных оценок.

Свойство 3. Сильная состоятельность МП-оценки.

Будем предполагать, что выполнены два дополнительных условия:

1. $\theta = \theta^*$, т. е. параметрическое пространство — компакт;
2. $\theta^0 \in \theta$ — внутренняя точка параметрического пространства.

Определение 1.27. Информационной функцией Кульбака для случайного наблюдения $x_l \in \mathbb{R}^N$ с плотностью распределения $p(x; \theta)$, $x \in \mathbb{R}^N$, где $\theta \in \Theta \subset \mathbb{R}^m$, называется функция 2-х переменных:

$$\begin{aligned} \mathcal{J}(\theta^0, \theta) &::= \mathbf{E}_{\theta^0} \left\{ \ln \frac{p(x_l; \theta^0)}{p(x_l; \theta)} \right\} = \\ &= \int_{\mathbb{R}^N} p(x; \theta^0) \ln \frac{p(x; \theta^0)}{p(x; \theta)} dx, \quad \theta^0, \theta \in \Theta. \end{aligned} \quad (1.54)$$

Лемма 1.1. Справедливо неравенство

$$\mathcal{J}(\theta^0, \theta) \geq 0, \quad \theta^0, \theta \in \Theta, \quad (1.55)$$

причем

$$\mathcal{J}(\theta^0, \theta) = 0 \Leftrightarrow \int_{\mathbb{R}^N} |p(x; \theta) - p(x; \theta^0)| dx = 0. \quad (1.56)$$

Доказательство. Воспользуемся (1.54), неравенством Иенсена для выпуклой функции $y = g(x) = -\ln x$ и условием нормировки:

$$\begin{aligned} \mathcal{J}(\theta^0, \theta) &= \mathbf{E}_{\theta^0} \left\{ -\ln \frac{p(x_l; \theta)}{p(x_l; \theta^0)} \right\} \geq -\ln \mathbf{E}_{\theta^0} \left\{ \frac{p(x_l; \theta)}{p(x_l; \theta^0)} \right\} = \\ &= -\ln \int_{\mathbb{R}^N} p(x; \theta^0) \frac{p(x; \theta)}{p(x; \theta^0)} dx = -\ln 1 = 0, \end{aligned}$$

т. е. выполняется неравенство (1.55). В силу неравенства Иенсена:

$$\mathcal{J}(\theta^0, \theta) = 0 \Leftrightarrow \frac{p(x_l; \theta)}{p(x_l; \theta^0)} \stackrel{\text{п.н.}}{=} 1,$$

что эквивалентно (1.56). □

Определение 1.28. Если

$$\int_{\mathbb{R}^N} |p(x; \theta) - p(x; \theta^0)| dx = 0 \Leftrightarrow \theta = \theta^0,$$

то значение параметра θ^0 называется идентифицируемым, в противном случае — неидентифицируемым.

Следствие 1.8.2. Если θ^0 — идентифицируемое значение, то

$$\theta^0 = \arg \min_{\theta \in \Theta} \mathcal{J}(\theta^0, \theta). \quad (1.57)$$

Доказательство. Соотношение (1.57) следует из (1.55) и (1.56). □

Теорема 1.9 (о сильной состоятельности ОМП). Пусть θ^0 — идентифицируемое значение параметра — является внутренней точкой Θ , а функция плотности распределения вероятностей $p(x; \theta)$, $x \in \mathbb{R}^N$, непрерывна по θ и такова, что $\forall \theta \neq \theta^0$ существует $U_\theta \subset \Theta$ — окрестность точки θ , для которой

$$\mathbf{E}_{\theta^0} \left\{ \inf_{\theta^* \in U_\theta} \ln \frac{p(x_l; \theta^0)}{p(x_l; \theta^*)} \right\} > -\infty. \quad (1.58)$$

Тогда МП-оценка $\hat{\theta}$ сильно состоятельна:

$$\hat{\theta} \xrightarrow{\text{п.н.}} \theta^0, \quad n \rightarrow \infty. \quad (1.59)$$

Доказательство. Представим оценку максимального правдоподобия в виде, эквивалентном (1.49):

$$\begin{aligned}\hat{\theta} &= \arg \max_{\theta \in \Theta} \left(\frac{1}{n} (l(\theta) - l(\theta^0)) \right) = \arg \min_{\theta \in \Theta} \left(\frac{1}{n} (l(\theta^0) - l(\theta)) \right) = \\ &= \arg \min_{\theta \in \Theta} \mathcal{J}_n(\theta^0, \theta),\end{aligned}$$

где

$$\mathcal{J}_n(\theta^0, \theta) = \frac{1}{n} \sum_{k=1}^n \ln \frac{p(x_k; \theta^0)}{p(x_k; \theta)}.$$

Сравним этот результат с (1.57). По усиленному закону больших чисел имеем:

$$\mathcal{J}_n(\theta^0, \theta) \xrightarrow{\text{П.Н.}} \mathcal{J}(\theta^0, \theta), \quad n \rightarrow \infty.$$

Из сходимости целевых функций, вообще говоря, не следует сходимость точек минимума (1.59). Для того чтобы это выполнялось, достаточно, чтобы имело место (1.58). \square

Следствие 1.9.1. Оценка максимального правдоподобия состоятельна, т. е. $\hat{\theta} \xrightarrow{\text{П.Н.}} \theta^0$, $n \rightarrow \infty$.

Свойство 4. Асимптотическая несмещенность и нормальность оценки максимального правдоподобия.

Теорема 1.10. Пусть выполнены условия теоремы 1.9, а также условия регулярности **R1–R4**, причем условие **R1** выполняется в усиленном смысле: $p(x; \theta)$, $x \in \mathbb{R}^N$, — трижды дифференцируема по θ и существует окрестность $U_{\theta^0} \subset \Theta$ точки θ^0 такая, что

$$\sup_{\theta \in U_{\theta^0}} \mathbf{E}_{\theta^0} \left\{ \left| \frac{\partial^3 \ln p(x_i; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \right\} < +\infty.$$

Тогда при $n \rightarrow \infty$ оценка максимального правдоподобия $\hat{\theta}$ асимптотически нормально распределена с вектором математического ожидания θ^0 и ковариационной матрицей $\mathcal{I}_n^{-1} = \frac{1}{n} \mathcal{I}^{-1}$, где $\mathcal{I} \succ 0$ — информационная матрица Фишера для отдельного наблюдения. Другими словами:

$$\mathcal{L}\{(\mathcal{I}_n^{-1})^{\frac{1}{2}}(\hat{\theta} - \theta^0)\} \rightarrow \mathcal{N}_m(\mathbf{0}_m, \mathcal{I}_m), \quad n \rightarrow +\infty,$$

где \mathcal{I}_n^{-1} — решение матричного уравнения $Y^{\text{T}} \mathcal{I}_n^{-1} Y = \mathcal{I}_m$ по Y ($\mathcal{I}_n = \mathcal{I}_n^{-1} (\mathcal{I}_n^{-1})^{\text{T}}$).

Доказательство. Для простоты проведем доказательство в одномерном случае: $m = 1$, $\theta^0 \in \mathbb{R}^1$. Так как выполняются условия теоремы 1.9, то $\hat{\theta} \xrightarrow{\text{П.Н.}} \theta^0$, $n \rightarrow \infty$, и, следовательно, $\exists \bar{n}$, что $\forall n \geq \bar{n}$, $\hat{\theta} \in U_{\theta^0}$ с вероятностью 1. Дальнейший анализ проводим именно для таких значений $n \geq \bar{n}$.

Запишем уравнение правдоподобия ($m = 1$):

$$\nabla_{\theta} l(\theta) = \frac{d}{d\theta} l(\theta) = l^{(1)}(\theta) = 0,$$

причем в силу **R1** функция $l^{(1)}(\theta)$ дважды непрерывно дифференцируема. Применим к левой части этого уравнения квадратичную формулу Тейлора в окрестности U_{θ^0} :

$$l^{(1)}(\theta^0) + (\hat{\theta} - \theta^0) l^{(2)}(\theta^0) + \frac{1}{2} (\hat{\theta} - \theta^0)^2 l^{(3)}(\bar{\theta}) = 0,$$

где $l^{(k)}(\theta)$ — k -я производная по θ , а $\bar{\theta} \in \theta$ — "промежуточная" точка: $|\bar{\theta} - \theta^0| < |\hat{\theta} - \theta^0|$, поэтому $\bar{\theta} \xrightarrow{\text{п.н.}} \theta^0$ при $n \rightarrow \infty$.

Выразим случайное уклонение $\hat{\theta} - \theta^0$ из полученного выше уравнения следующим образом:

$$\hat{\theta} - \theta^0 = \frac{l^{(1)}(\theta^0)}{-l^{(2)}(\theta^0) - \frac{1}{2}(\hat{\theta} - \theta^0)l^{(3)}(\bar{\theta})}.$$

Умножим левую и правую части полученного соотношения на $\sqrt{n\mathcal{I}}$, а числитель и знаменатель правой части — на $\frac{1}{n\mathcal{I}}$:

$$\sqrt{n\mathcal{I}}(\hat{\theta} - \theta^0) = \frac{\frac{1}{\sqrt{n\mathcal{I}}}l^{(1)}(\theta^0)}{\frac{1}{n\mathcal{I}}(-l^{(2)}(\theta^0) - \frac{1}{2}(\hat{\theta} - \theta^0)l^{(3)}(\bar{\theta}))}. \quad (1.60)$$

Согласно усиленному закону больших чисел:

$$-\frac{1}{n}l^{(2)}(\theta^0) = \frac{1}{n} \sum_{k=1}^n \left(-\frac{\partial^2 \ln p(x_k; \theta^0)}{(\partial \theta^0)^2} \right) \xrightarrow{\text{п.н.}} \mathcal{I};$$

$$-\frac{1}{n}l^{(3)}(\bar{\theta}) \xrightarrow{\text{п.н.}} C,$$

где $|C| < +\infty$ — некоторая константа. Поскольку $\hat{\theta} - \theta^0 \xrightarrow{\text{п.н.}} 0$, то знаменатель сходится к 1 почти наверное.

Числитель в (1.60) при $n \rightarrow \infty$ распределен асимптотически нормально:

$$\mathcal{L}_{\theta^0} \left\{ \frac{1}{\sqrt{n\mathcal{I}}}l^{(1)}(\theta^0) \right\} \rightarrow \mathcal{N}_1(0, 1),$$

в силу ЦПТ Леви — Линдеберга. □

1.8 Достаточные статистики и их свойства

Рассмотрим задачу статистического оценивания параметров в общем виде, акцентируя внимание на вычислительной сложности. Пусть имеется параметрическое семейство плотностей

$$\mathcal{P} = \{p(x; \theta), x \in \mathbb{R}^N : \theta \in \Theta \subseteq \mathbb{R}^m\}$$

и наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из некоторого распределения вероятностей с плотностью $p(x; \theta^0)$, $x \in \mathbb{R}^N$, где $\theta^0 \in \Theta$ — неизвестное истинное значение вектора параметров. Необходимо построить статистическую оценку

$$\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^m.$$

Проанализируем эту задачу с точки зрения ее вычислительной сложности. Необходимо построить m функций, каждая из которых зависит от nN переменных. Упростим задачу следующим образом. Будем строить оценку в два этапа.

Этап 1. Сжатие данных — вырожденное функциональное преобразование выборки X :

$$t = t(X) = \begin{pmatrix} t_1(X) \\ \vdots \\ t_K(X) \end{pmatrix} : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K, \quad K < nN. \quad (1.61)$$

Этап 2. Построение оценки:

$$\hat{\theta} = \psi(t) = \begin{pmatrix} \psi_1(t) \\ \vdots \\ \psi_m(t) \end{pmatrix} : \mathbb{R}^K \rightarrow \mathbb{R}^m. \quad (1.62)$$

Статистика, которая позволяет осуществить сжатие данных $t = t(X)$ так, чтобы построенная на ее основе оценка $\tilde{\theta}$ была бы так же точна, как и $\hat{\theta}$, называется *достаточной статистикой*.

Теперь перейдем к строгому математическому определению достаточной статистики. Для обобщения поставленной задачи статистического оценивания предположим, что θ^0 — случайный вектор с некоторой плотностью распределения вероятностей $\Pi(\theta)$, $\theta \in \Theta$. Определим условную плотность распределения вероятностей выборки:

$$p(X|\theta) = p(X; \theta) = \prod_{l=1}^n p(x_l; \theta).$$

Вычислим аналогично по формуле Байеса апостериорную плотность распределения вероятностей случайного вектора θ^0 при условии, что наблюдается выборка X :

$$p(\theta|X) = \frac{p(X|\theta)\Pi(\theta)}{\int_{\Theta} p(X|\theta^*)\Pi(\theta^*)d\theta^*}. \quad (1.63)$$

По аналогии с (1.63) найдем условную плотность распределения θ^0 при условии наблюдения статистики $t = t(X)$:

$$q(\theta|t) = \frac{q(t|\theta)\Pi(\theta)}{\int_{\Theta} q(t|\theta^*)\Pi(\theta^*)d\theta^*}, \quad (1.64)$$

где $q(t|\theta)$ — условная плотность распределения случайной величины $t = t(X)$ при условии $\theta^0 = \theta$.

Определение 1.29. K -мерная статистика (1.61) называется *достаточной статистикой* для параметрического семейства \mathcal{P} , если условные плотности (1.63) и (1.64) совпадают:

$$q(\theta|t(X)) = p(\theta|X), \quad X \in \mathbb{R}^{nN}, \quad \theta \in \Theta. \quad (1.65)$$

При этом достаточная статистика с наименьшей размерностью K^* называется *минимальной достаточной статистикой*.

Замечание 1.8. Минимальная достаточная статистика «наибольшим образом» сжимает данные.

Замечание 1.9. Соотношение (1.65) означает, что статистика $t = t(X)$ несет столько же информации о параметре θ , сколько и вся выборка X .

Замечание 1.10. Если выборка X получена из дискретного распределения вероятностей, то это определение остается в силе, только вместо плотности распределения необходимо везде использовать соответствующие вероятности.

Теорема 1.11 (критерий факторизации Неймана - Фишера). Для того чтобы статистика $t = t(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K$ была достаточной статистикой для семейства плотностей распределения вероятностей \mathcal{P} , необходимо и достаточно, чтобы совместная

плотность распределения выборки допускала следующую факторизацию, то есть следующее разделение переменных ($X \in \mathbb{R}^{nN}$, $\theta \in \Theta$):

$$p(X; \theta) = \prod_{k=1}^n p(x_k; \theta) = g(X)h(t(X); \theta), \quad (1.66)$$

где $g(X) \geq 0$ зависит только от выборки X , но не зависит от параметра θ , а $h(t(X); \theta) \geq 0$ зависит от θ , но зависимость от выборки X имеет место только через статистику $t(X)$.

Доказательство. Необходимость. Пусть $t = t(X)$ — достаточная статистика, тогда выполняется (1.65). Выразим условную плотность распределения вероятностей выборки с учетом (1.63)—(1.65):

$$p(X|\theta) = p(X; \theta) = \int_{\Theta} p(X; \theta^*) \Pi(\theta^*) d\theta^* \frac{q(t|\theta)}{\int_{\Theta} q(t|\theta^*) \Pi(\theta^*) d\theta^*}, \quad t = t(X).$$

Легко видеть, что первый сомножитель не зависит от θ , а второй может выступать в качестве $h(t(X); \theta)$, так что выполняется факторизация (1.66).

Достаточность. Пусть имеет место факторизация (1.66). Проверим выполнение (1.65). По (1.63) из (1.66) имеем

$$p(\theta|X) = \frac{g(X)h(t; \theta)\Pi(\theta)}{\int_{\Theta} g(X)h(t; \theta^*)\Pi(\theta^*)d\theta^*} = \frac{h(t; \theta)\Pi(\theta)}{\int_{\Theta} h(t; \theta^*)\Pi(\theta^*)d\theta^*}. \quad (1.67)$$

Вычислим апостериорную плотность $q(\theta|t)$, используя (1.64). Для этого сначала найдем $q(t|\theta)$. Воспользуемся методом функциональных преобразований случайных величин. Так как $K < nN$, то функциональное преобразование (1.61) вырожденное. Введем вспомогательное преобразование: $\tau = \tau(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^{nN-K}$, причем функцию $\tau(\cdot)$ подберем таким образом, чтобы составное функциональное преобразование $X \rightarrow (t^T, \tau^T)^T$ было взаимно-однозначным, то есть чтобы существовало единственное обратное преобразование $X = f(t, \tau)$ с якобианом преобразования $\mathcal{J}(t, \tau) \neq 0$. Тогда по правилам функционального преобразования случайных величин имеем

$$\begin{aligned} q(t|\theta) &= \int_{\mathbb{R}^{nN-K}} p(f(t, \tau); \theta) |\mathcal{J}(t, \tau)| d\tau = \\ &= h(t; \theta) \int_{\mathbb{R}^{nN-K}} g(f(t, \tau)) |\mathcal{J}(t, \tau)| d\tau = h(t; \theta) B(t), \end{aligned}$$

где

$$B(t) = \int_{\mathbb{R}^{nN-K}} g(f(t, \tau)) |\mathcal{J}(t, \tau)| d\tau.$$

Подставляя это выражение в (1.64), получаем искомую плотность:

$$q(\theta|t) = \frac{B(t)h(t; \theta)\Pi(\theta)}{\int_{\Theta} B(t)h(t; \theta^*)\Pi(\theta^*)d\theta^*} = \frac{h(t; \theta)\Pi(\theta)}{\int_{\Theta} h(t; \theta^*)\Pi(\theta^*)d\theta^*}. \quad (1.68)$$

Сравнивая (1.67) и (1.68), видим, что они совпадают, то есть выполняется (1.65), следовательно, $t = t(X)$ — достаточная статистика. \square

Пример 1.11. Пусть $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ — случайная выборка объема n из гауссовского распределения $\mathcal{N}_1(\theta_1, \theta_2)$, $\theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2$. Построить достаточную статистику для θ .

Воспользуемся критерием факторизации Неймана — Фишера и выполним факторизацию (1.66).

1. Построим совместную плотность распределения выборки:

$$p(X; \theta) = \prod_{k=1}^n \left(\frac{1}{\sqrt{2\pi}\theta_2} e^{-\frac{1}{2\theta_2}(x_k - \theta_1)^2} \right).$$

2. Найдем достаточную статистику $t(X)$, осуществив разделение переменных X и θ с помощью эквивалентных преобразований:

$$p(X; \theta) = \left(\frac{1}{\sqrt{2\pi}} \right)^n \left(\frac{1}{\theta_2} \right)^{\frac{n}{2}} e^{-\frac{1}{2\theta_2} \sum_{k=1}^n x_k^2 + \frac{\theta_1}{\theta_2} \sum_{k=1}^n x_k - \frac{n\theta_1^2}{2\theta_2}} = g(X)h(t(X); \theta);$$

$$g(X) = \left(\frac{1}{\sqrt{2\pi}} \right)^n, \quad t(X) = (t_1(X), t_2(X))^T, \quad t_l(X) = \sum_{k=1}^n x_k^l, \quad l = 1, 2;$$

$$h(t; \theta) = \left(\frac{1}{\theta_2} \right)^{\frac{n}{2}} e^{-\frac{t_2(x)}{2\theta_2} + \frac{\theta_1}{\theta_2} t_1(x) - \frac{n\theta_1^2}{2\theta_2}} \geq 0.$$

Таким образом, найденная достаточная статистика является двухмерной: $t(X) = (t_1(X), t_2(X))$ ($K = 2$).

Пример 1.12. Наблюдается выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ из $R[\theta_1, \theta_2]$. Показать, что достаточной статистикой является следующая двумерная статистика:

$$t = t(X) = (t_1(X), t_2(X))^T, \quad t_1(X) = \min_{1 \leq k \leq n} x_k, \quad t_2(X) = \max_{1 \leq k \leq n} x_k.$$

Свойства достаточной статистики

Свойство 1. Эквивалентная формулировка критерия факторизации.

Статистика $t = t(X)$ является достаточной для семейства \mathcal{P} тогда и только тогда, когда для $\forall \theta, \theta^* \in \Theta, \theta \neq \theta^*$, отношение правдоподобия зависит от выборки X только через статистику $t(X)$:

$$\frac{p(X; \theta)}{p(X; \theta^*)} = H(t(X); \theta, \theta^*), \quad X \in \mathbb{R}^{nN}. \quad (1.69)$$

Доказательство. Покажем, что (1.69) эквивалентно критерию факторизации (1.66). Выберем произвольное $\theta^* \in \Theta$ и зафиксируем его. Тогда из (1.69) следует:

$$p(X; \theta) = p(X; \theta^*)H(t(X); \theta, \theta^*) = g(X)h(t(X); \theta),$$

что эквивалентно (1.66). □

Свойство 2. Свойство статистики $t = t(X)$ быть достаточной не зависит от вида априорной плотности $\Pi(\cdot)$.

Доказательство. Действительно, $\Pi(\cdot)$ не входит в критерий (1.66). □

Свойство 3. Инвариантность достаточной статистики.

Достаточная статистика $t = t(X)$ единственна с точностью до произвольной биекции. Другими словами, если $\phi = \phi(t) : \mathbb{R}^K \rightarrow \mathbb{R}^K$ — некоторая биекция, то $\phi(t(X)) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K$ также является достаточной статистикой.

Доказательство. Этот результат можно получить с помощью замены переменных в (1.66). \square

Свойство 4. Если семейство \mathcal{P} удовлетворяет условиям регулярности R1 — R4 и существует эффективная оценка $\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^m$, то статистика $T = T(X)$ является достаточной статистикой.

Доказательство. Воспользуемся критерием эффективности оценки:

$$\nabla_{\theta} \ln p(X; \theta) = \mathcal{I}_n(\theta)(T(X) - \theta).$$

Проинтегрируем это векторное соотношение по θ . В результате получим следующее выражение для плотности:

$$p(X; \theta) = e^{A^T(\theta)T(X) + B(\theta) + C(X)}, \quad (1.70)$$

где $C(X)$ — функция выборки, появившаяся в результате интегрирования, а $A(\theta)$ и $B(\theta)$ должны удовлетворять следующим условиям:

$$\nabla_{\theta} A(\theta) = \mathcal{I}_n(\theta), \quad \nabla_{\theta} B(\theta) = -\mathcal{I}_n(\theta)\theta.$$

Сравнивая (1.66) и (1.70), заключаем, что в качестве $g(X)$ и $h(T(X); \theta)$ можно выбрать:

$$g(X) = e^{C(X)} \geq 0, \quad h(T(X); \theta) = e^{A^T(\theta)T(X) + B(\theta)} \geq 0.$$

Следовательно, в силу выполнения критерия факторизации $T = T(X)$ — достаточная статистика. \square

Свойство 5. *Эквивалентное определение достаточной статистики.*

Статистика $t = t(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K$ является достаточной статистикой для семейства \mathcal{P} тогда и только тогда, когда условная плотность распределения выборки X при условии $t(X) = t^*$ не зависит от параметра θ , то есть

$$p(X|t(X) = t^*; \theta) = \tilde{g}(X; t^*) = \text{invar}_{\theta}.$$

Свойство 6. Это свойство достаточной статистики сформулируем в виде теоремы.

Теорема 1.12 (Дармуа). *Плотность распределения вероятностей $p(x; \theta)$, $x \in \mathbb{R}^N$, с m -мерным параметром $\theta \in \mathbb{R}^m$, допускающая m -мерную достаточную статистику $t = t(X) \in \mathbb{R}^m$, имеет вид*

$$p(x; \theta) = \exp \left(\sum_{j=1}^m a_j(\theta) w_j(x) + b(\theta) + c(x) \right), \quad (1.71)$$

где $\{a_j(\cdot)\}, b(\cdot), c(\cdot), \{w_j(\cdot)\}$ — функции указанных выше переменных. При этом достаточная статистика имеет следующий аддитивный вид:

$$t = t(X) = (t_1(X), \dots, t_m(X))^T,$$

$$t_j(X) = \sum_{k=1}^n w_j(x_k), \quad j = 1, \dots, m. \quad (1.72)$$

Определение 1.30. Семейство плотностей вида (1.71) называется экспоненциальным семейством.

Следствие 1.12.1. Для экспоненциального семейства минимальная достаточная статистика является m -мерной и имеет вид (1.72).

Свойство 7. Сформулируем теорему, которая позволяет улучшать оценки с помощью достаточных статистик.

Теорема 1.13 (Колмогорова – Рао – Блекуэлла). Пусть $\hat{\theta} = T(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^1$ — некоторая несмещенная статистическая оценка параметра $\theta^0 \in \mathbb{R}^1$ по случайной выборке X , а $t = t(X) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^K$, $K < nN$, — достаточная статистика для семейства плотностей \mathcal{P} . Пусть далее с помощью условного математического ожидания определена борелевская функция (1.62):

$$\psi = \psi(t^*) ::= \mathbf{E}\{T(X)|t(X) = t^*\}, \quad t^* \in \mathbb{R}^K, \quad \psi(\cdot) \in \mathbb{R}^1. \quad (1.73)$$

Тогда композиция функций

$$\tilde{\theta} = \psi(t(X)) : \mathbb{R}^{nN} \rightarrow \mathbb{R}^1 \quad (1.74)$$

определяет несмещенную оценку параметра θ^0 , причем вариация этой оценки не превосходит вариации исходной оценки:

$$V\{\tilde{\theta}\} \leq V\{\hat{\theta}\}.$$

Доказательство. С учетом (1.73), (1.74) и определения условного математического ожидания представим оценку (1.74) в эквивалентном виде:

$$\tilde{\theta} = \mathbf{E}\{T(X)|t(X)\}.$$

Проверим условие несмещенности с использованием формулы полного математического ожидания:

$$\mathbf{E}\{\tilde{\theta}\} = \mathbf{E}\{\mathbf{E}\{T(X)|t(X)\}\} = \mathbf{E}\{T(X)\} = \mathbf{E}\{\hat{\theta}\} = \theta^0.$$

Сравним вариации оценок:

$$\begin{aligned} V\{\tilde{\theta}\} &= \mathbf{E}\{(\tilde{\theta} - \theta^0)^2\} = \mathbf{E}\{(\mathbf{E}\{T(X) - \theta^0|t(X)\})^2\} \leq \\ &\leq \mathbf{E}\{\mathbf{E}\{(T(X) - \theta^0)^2|t(X)\}\} = \mathbf{E}\{(T(X) - \theta^0)^2\} = V\{\hat{\theta}\}. \end{aligned}$$

Здесь использованы известное свойство [?]:

$$(\mathbf{E}\{\xi\})^2 \leq \mathbf{E}\{\xi^2\},$$

и формула полного математического ожидания. □

1.9 Интервальное оценивание параметров

1.9.1 Понятие об интервальном оценивании

Рассмотрим следующую задачу статистического оценивания параметров. Пусть в \mathbb{R}^N наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ из некоторого абсолютно непрерывного распределения вероятностей с некоторой плотностью $p(x; \theta)$, $x \in \mathbb{R}^N$, где $\theta \in \mathbb{R}^1$ — неизвестное значение параметра. Примем следующие обозначения:

$$p(X; \theta) = \prod_{k=1}^n p(x_k; \theta) —$$

совместная плотность распределения вероятностей всей выборки;

$$\mathbf{P}_\theta(A) = \int_A p(X; \theta) dX, \quad A \subseteq \mathcal{B}^{nN}, —$$

вероятностная мера при условии, что параметр равен θ .

Рассмотрим задачу оценивания θ по X . Геометрический смысл статистического оценивания состоит в том, что по выборке X пытаются построить точку $\hat{\theta}$, где $\hat{\theta} = T(X) \in \mathbb{R}^1$ — некоторая статистика. В силу того что оценка интерпретируется точкой, такое оценивание называется точечным. В условиях регулярности известно следующее свойство точечных оценок:

$$V\{\hat{\theta}\} = \mathbf{E}_\theta\{|\hat{\theta} - \theta|^2\} \geq \frac{1}{n\mathcal{I}} > 0,$$

где $\mathcal{I} > 0$ — количество информации по Фишеру.

Отсюда следует, что $\mathbf{P}_\theta\{\hat{\theta} = \theta\} = 0$ — точечная оценка с вероятностью единица ошибочна! Однако если возьмем какой-либо интервал $(\hat{\theta} - \delta, \hat{\theta} + \delta)$ длины $2\delta > 0$, то

$$\mathbf{P}_\theta\{\hat{\theta} - \delta < \theta < \hat{\theta} + \delta\} > 0, \quad \forall \delta > 0.$$

На этом факте и основана идея интервального оценивания.

Определение 1.31. Пусть $0 < \varepsilon < \frac{1}{2}$. Случайный интервал $(\underline{\theta}, \bar{\theta}) \subset \mathbb{R}^1$, границы которого $\underline{\theta} = \underline{\theta}(X)$, $\bar{\theta} = \bar{\theta}(X)$ (причем $\underline{\theta}(X) < \bar{\theta}(X)$) являются статистиками, называется доверительным интервалом для параметра θ , если этот интервал «накрывает» истинное значение θ с наперед заданной достаточно большой вероятностью $1 - \varepsilon$:

$$\forall \theta \in \mathbb{R} \quad \mathbf{P}_\theta\{\underline{\theta} < \theta < \bar{\theta}\} = 1 - \varepsilon. \quad (1.75)$$

При этом величина $1 - \varepsilon$ называется доверительной вероятностью (коэффициентом доверия, доверительным уровнем), а статистики $\underline{\theta}(X)$ и $\bar{\theta}(X)$ — соответственно нижней и верхней доверительными границами.

На практике доверительные вероятности обычно выбираются стандартным образом:

$$1 - \varepsilon \in \{0,8; 0,85; 0,9; 0,95; 0,99; 0,995; 0,999\}.$$

Определение 1.32. Интервальной оценкой параметра θ по выборке X называется основанное на (1.75) вероятностное утверждение следующего вида:

$$\text{с вероятностью } 1 - \varepsilon \quad \theta \in (\underline{\theta}, \bar{\theta}).$$

Замечание 1.11. Иногда доверительным уровнем называется следующая величина: $(1 - \varepsilon)100\%$, то есть доверительный уровень иногда измеряется в процентах.

Запишем условие (1.75) в эквивалентном виде:

$$\forall \theta \in \mathbb{R} \quad \mathbf{P}_\theta\{\theta \leq \underline{\theta}\} = \varepsilon_1, \quad \mathbf{P}_\theta\{\theta \geq \bar{\theta}\} = \varepsilon_2, \quad \varepsilon_1 + \varepsilon_2 = \varepsilon.$$

В зависимости от соотношения между ε_1 и ε_2 известна следующая классификация типов доверительных интервалов.

Определение 1.33. *Существуют три типа доверительных интервалов:*

1. если $\varepsilon_1 = \varepsilon$, $\varepsilon_2 = 0$, то интервальная оценка имеет вид: с вероятностью $1 - \varepsilon$ $\theta > \underline{\theta}$, и доверительный интервал $(\underline{\theta}, \infty)$ называется левосторонним доверительным интервалом;
2. если $\varepsilon_1 = 0$, $\varepsilon_2 = \varepsilon$, то интервальная оценка имеет вид: с вероятностью $1 - \varepsilon$ $\theta < \bar{\theta}$, и доверительный интервал $(-\infty, \bar{\theta})$ называется правосторонним доверительным интервалом;
3. если $\varepsilon_1 = \varepsilon_2 = \frac{\varepsilon}{2}$, то интервальная оценка принимает вид:

$$\forall \theta \in \mathbb{R} \quad \mathbf{P}_\theta\{\theta \leq \underline{\theta}\} = \frac{\varepsilon}{2}, \quad \mathbf{P}_\theta\{\theta \geq \bar{\theta}\} = \frac{\varepsilon}{2}; \quad (1.76)$$

при этом доверительный интервал $(\underline{\theta}, \bar{\theta})$ называется центральным доверительным интервалом (ЦДИ).

1.9.2 Методы построения доверительных интервалов

Рассмотрим три основных метода построения доверительных интервалов. Основное внимание будет уделено центральным доверительным интервалам, определяемым (1.76).

I. Метод обратной функции

Пусть некоторым образом построена состоятельная точечная оценка параметра θ : $\hat{\theta} = T(X)$. Предположим, что для нее вычислена функция распределения $F(z; \theta) = \mathbf{P}_\theta\{\hat{\theta} < z\}$, $z \in \mathbb{R}$, причем предполагаем, что $F(\cdot; \theta)$ монотонна по θ . Будем искать доверительные границы в виде

$$\underline{\theta} = \underline{\theta}(X) = f_1(\hat{\theta}), \quad \bar{\theta} = \bar{\theta}(X) = f_2(\hat{\theta}), \quad (1.77)$$

где $f_1(\cdot)$, $f_2(\cdot)$ — некоторые, пока не известные, монотонно возрастающие функции.

Подберем $f_1(\cdot)$, $f_2(\cdot)$ в (1.77) таким образом, чтобы выполнялись соотношения (1.76). Для первого из них имеем

$$\forall \theta \in \mathbb{R} \quad \mathbf{P}_\theta\{\theta \leq \underline{\theta}\} \equiv \mathbf{P}_\theta\{\theta \leq f_1(\hat{\theta})\} = \mathbf{P}_\theta\{f_1^{-1}(\theta) \leq \hat{\theta}\} = 1 - F(\theta_1; \theta) = \frac{\varepsilon}{2},$$

где $\theta_1 = f_1^{-1}(\theta)$, что эквивалентно $\theta = f_1(\theta_1)$.

Аналогично действуя со вторым соотношением, получаем

$$F(\theta_2 + 0; \theta) = \frac{\varepsilon}{2}, \quad \theta = f_2(\theta_2).$$

Таким образом, функции $f_1(\cdot)$, $f_2(\cdot)$ однозначно определяются следующими уравнениями:

$$F(\theta_1; \theta) = 1 - \frac{\varepsilon}{2}, \quad \text{где } \theta = f_1(\theta_1); \quad (1.78)$$

$$F(\theta_2 + 0; \theta) = \frac{\varepsilon}{2}, \quad \text{где } \theta = f_2(\theta_2). \quad (1.79)$$

Метод обратной функции порождает следующий алгоритм построения центрального доверительного интервала.

1. Строим состоятельную оценку $\hat{\theta} = T(X)$.
2. Находим функцию распределения этой оценки $F(\cdot; \theta)$.
3. Составляем и решаем по θ следующие два уравнения:

$$F(\hat{\theta}; \theta) = 1 - \frac{\varepsilon}{2};$$

$$F(\hat{\theta} + 0; \theta) = \frac{\varepsilon}{2}.$$

4. Определяем доверительные границы:

$$\underline{\theta} = \underline{\theta}(X) \text{ — наименьший из корней,}$$

$$\bar{\theta} = \bar{\theta}(X) \text{ — наибольший из корней.}$$

Теорема 1.14. Пусть $\hat{\theta} = T(X)$ — некоторая состоятельная асимптотически несмещенная и асимптотически нормальная оценка параметра θ с асимптотической дисперсией $\sigma_n^2(\theta) \rightarrow 0$, $n \rightarrow \infty$, то есть случайная величина $\tau = \tau(X; \theta) = (\hat{\theta} - \theta)/\sigma_n(\theta)$ при $n \rightarrow \infty$ распределена асимптотически нормально по закону $\mathcal{N}_1(0, 1)$:

$$F_\tau(z) = \mathbf{P}_\theta\{\tau < z\} \rightarrow \Phi(z), \quad z \in \mathbb{R}. \quad (1.80)$$

Тогда при $n \rightarrow \infty$ границы $(1 - \varepsilon)100\%$ ЦДИ определяются из уравнений

$$\frac{\hat{\theta} - \theta}{\sigma_n(\theta)} = \pm g_{1-\frac{\varepsilon}{2}}. \quad (1.81)$$

Доказательство. Применим метод обратной функции и воспользуемся асимптотикой (1.80). Составим уравнения (1.78) и (1.79), найдем

$$F(z; \theta) = \mathbf{P}_\theta\{\hat{\theta} < z\} = \mathbf{P}_\theta\left\{\frac{\hat{\theta} - \theta}{\sigma_n(\theta)} < \frac{z - \theta}{\sigma_n(\theta)}\right\} = F_\tau\left(\frac{z - \theta}{\sigma_n(\theta)}\right).$$

Учитывая (1.78), (1.79) и используя (1.80), получаем асимптотические соотношения:

$$\Phi\left(\frac{\hat{\theta} - \theta}{\sigma_n(\theta)}\right) = 1 - \frac{\varepsilon}{2}; \quad \Phi\left(\frac{\hat{\theta} - \theta}{\sigma_n(\theta)}\right) = \frac{\varepsilon}{2}.$$

Действуя дальше аналогично примеру ??, получаем (1.81). □

Следствие 1.14.1. Если оценка $\hat{\theta}$ имеет нормальное распределение: $\mathcal{L}\{\hat{\theta}\} = \mathcal{N}_1(\theta, \sigma_n^2(\theta))$, то границы, определяемые (1.81), являются точными при любом n .

Следствие 1.14.2. Если дисперсия $\sigma_n^2(\theta) = \sigma_n^2$ и не зависит от θ , то доверительные границы имеют явный вид:

$$\underline{\theta} = \hat{\theta} - g_{1-\frac{\varepsilon}{2}} \sigma_n, \quad \bar{\theta} = \hat{\theta} + g_{1-\frac{\varepsilon}{2}} \sigma_n.$$

Теорема 1.15 (интервальная оценка вероятности успеха). Пусть рассматривается схема n независимых испытаний Бернулли, и $X = (x_1, \dots, x_n)^T \in \{0, 1\}^n$ — случайная выборка объема n из распределения $Bi(1, \theta)$, где $\theta \in (0, 1)$ — неизвестная вероятность успеха. Если $m = \sum_{i=1}^n x_i$ — зарегистрированное число успехов в n испытаниях, причем $m \in \{1, \dots, n-1\}$, то $(1 - \varepsilon)100\%$ ЦДИ имеет границы $\underline{\theta}$, $\bar{\theta}$, являющиеся корнями следующих уравнений:

$$\sum_{i=m}^n C_n^i \theta^i (1 - \theta)^{n-i} = \frac{\varepsilon}{2}; \quad (1.82)$$

$$\sum_{i=0}^m C_n^i \theta^i (1 - \theta)^{n-i} = \frac{\varepsilon}{2}. \quad (1.83)$$

Доказательство. Применим метод обратной функции. Известна состоятельная оценка для θ : $\hat{\theta} = \frac{m}{n}$ — относительная частота, для которой

$$\mathcal{L}\{m\} = Bi(n, \theta), \quad F\left(\frac{k}{n}; \theta\right) = \mathbf{P}_\theta\{m < k\} = \sum_{i=0}^{k-1} C_n^i \theta^i (1 - \theta)^{n-i}.$$

Составляем уравнение (1.78):

$$F\left(\frac{m}{n}; \theta\right) = \sum_{i=0}^{m-1} C_n^i \theta^i (1 - \theta)^{n-i} = 1 - \sum_{i=m}^n C_n^i \theta^i (1 - \theta)^{n-i} = 1 - \frac{\varepsilon}{2},$$

откуда получаем (1.82).

Запишем уравнение (1.79):

$$F\left(\frac{m}{n} + 0; \theta\right) = \sum_{i=0}^m C_n^i \theta^i (1 - \theta)^{n-i} = \frac{\varepsilon}{2},$$

что совпадает с (1.83). □

Следствие 1.15.1. Если число успехов $m = 0$, то имеем правосторонний доверительный интервал:

с вероятностью $1 - \varepsilon$ $0 < \theta < \bar{\theta}$, где $\bar{\theta} = 1 - \sqrt[n]{\varepsilon}$.

Следствие 1.15.2. Если число успехов $m = n$, то имеем левосторонний доверительный интервал:

с вероятностью $1 - \varepsilon$ $\underline{\theta} < \theta < 1$, где $\underline{\theta} = \sqrt[n]{\varepsilon}$.

II. Метод стьюдентизации

Метод стьюдентизации был предложен английским статистиком В. Госсетом (псевдоним Student). Он обобщает метод обратной функции и состоит из следующих трех шагов.

1. Строится функция $\tau = \tau(X; \theta)$ такая, что выполняются два условия:

- а) $\mathcal{L}\{\tau\} = \text{invar}_\theta$ — не зависит от параметра;
- б) $\tau(\cdot; \theta)$ — непрерывна и монотонна по θ .

Например, в следствии 1.14.1 теоремы 1.14 была построена функция

$$\tau(X; \theta) = \frac{\hat{\theta} - \theta}{\sqrt{\mathbf{D}_\theta\{\hat{\theta}\}}}.$$

Если $\hat{\theta}$ нормально распределена, то $\mathcal{L}\{\tau\} = \mathcal{N}_1(0, 1)$, так что условия а) и б) выполняются.

2. По заданному доверительному уровню $1 - \varepsilon$ выбираются два числа $g_-(\varepsilon)$, $g_+(\varepsilon)$ такие, что

$$\mathbf{P}_\theta\{g_-(\varepsilon) < \tau < g_+(\varepsilon)\} = 1 - \varepsilon, \quad \forall \theta \in \mathbb{R}.$$

Когда числа $g_-(\varepsilon)$, $g_+(\varepsilon)$ находятся неоднозначно, используют еще одно дополнительное условие: $g_+(\varepsilon) - g_-(\varepsilon) \rightarrow \min$.

Если τ имеет симметричное распределение вероятностей относительно $\tau = 0$, то $g_\pm(\varepsilon) = \pm g(\varepsilon)$. Например, если $\mathcal{L}\{\tau\} = \mathcal{N}_1(0, 1)$, то $g(\varepsilon) = \Phi^{-1}(1 - \frac{\varepsilon}{2})$.

3. Решаются два уравнения относительно θ :

$$\tau(X; \theta) = g_\pm(\varepsilon).$$

Нижняя доверительная граница $\underline{\theta}_\tau$ равна наименьшему из корней, а верхняя доверительная граница $\bar{\theta}_\tau$ равна наибольшему из корней.

III. Метод построения асимптотически наикратчайших доверительных интервалов

Определение 1.34. Пусть с помощью некоторой функции $\tau = \tau(X; \theta)$ построен ЦДИ $(\underline{\theta}_\tau, \bar{\theta}_\tau)$ с доверительным уровнем $1 - \varepsilon$. Средней длиной этого доверительного интервала называется число

$$\Delta = \Delta(\tau(\cdot)) ::= \mathbf{E}\{\bar{\theta}_\tau - \underline{\theta}_\tau\} \geq 0.$$

Возникает задача построения наикратчайшего доверительного интервала, имеющего минимальную длину:

$$\Delta(\tau(\cdot)) \rightarrow \min_{\tau(\cdot)}.$$

Приведем асимптотический результат.

Теорема 1.16. Пусть в условиях регулярности рассматривается семейство центральных доверительных интервалов доверительного уровня $1 - \varepsilon$ $(\underline{\theta}_\tau, \bar{\theta}_\tau)$, основанных на функции $\tau = \tau(X; \theta)$ следующего вида:

$$\tau(X; \theta) = \frac{1}{\sqrt{n\mathbf{D}_\theta\{h(x_1; \theta)\}}} \sum_{k=1}^n h(x_k; \theta),$$

где $h(\cdot; \theta)$ — произвольная борелевская функция такая, что математическое ожидание $\mathbf{E}_\theta\{h(x_1; \theta)\} = 0$, дисперсия $\mathbf{D}_\theta\{h(x_1; \theta)\} < +\infty$; при этом доверительные границы определяются из уравнений

$$\tau(X; \theta) = \pm g_{1-\frac{\varepsilon}{2}}, \quad g_\alpha = \Phi^{-1}(\alpha).$$

Тогда при $n \rightarrow \infty$ асимптотически наикратчайший центральный доверительный интервал определяется функцией

$$h(x; \theta) = h^*(x; \theta) = \ln p(x; \theta), \quad x \in \mathbb{R}^N,$$

где $p(\cdot; \theta)$ — плотность распределения вероятностей выборочных значений, а доверительные границы этого асимптотически наикратчайшего ЦДИ являются корнями уравнений

$$\frac{1}{\sqrt{n\mathcal{I}}} \frac{dl(\theta)}{d\theta} = \pm g_{1-\frac{\varepsilon}{2}},$$

где

$$l(\theta) = \sum_{k=1}^n \ln p(x_k; \theta) -$$

логарифмическая функция правдоподобия;

$$\mathcal{I} = \mathbf{E}_{\theta} \left\{ -\frac{d^2 \ln p(x_k; \theta)}{d\theta^2} \right\} > 0 -$$

количество информации по Фишеру.

Отметим, что одномерные доверительные интервалы могут быть обобщены на многомерный случай: $\theta \in \mathbb{R}^m$, $m \geq 1$.

Определение 1.35. Пусть $0 < \varepsilon < 1$, тогда случайная область $V_{\varepsilon} = V_{\varepsilon}(X) \subset \mathbb{R}^m$, граница которой зависит от выборки X , называется доверительной областью для векторного параметра $\theta \in \mathbb{R}^m$, если эта область содержит θ с наперед заданной достаточно большой вероятностью $1 - \varepsilon$, то есть

$$\forall \theta \mathbf{P}_{\theta} \{ \theta \in V_{\varepsilon}(X) \} = 1 - \varepsilon.$$

1.10 Метод наименьших квадратов и его свойства

1.10.1 Регрессионная модель. Принцип метода наименьших квадратов

Метод наименьших квадратов (МНК) применяется для оценивания параметров стохастических зависимостей. Идея этого метода предложена немецким математиком К. Гауссом, а математические основы заложены русским математиком А. Марковым.

Рассмотрим следующую математическую модель. Пусть наблюдается случайная величина η , которая зависит от некоторой другой величины x . Зависимость задается в виде следующего стохастического уравнения:

$$\eta = f(x; \theta^0) + \varepsilon, \quad (1.84)$$

где $f(x; \theta^0)$ – некоторая функция, вид которой известен; $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^{\mathbf{T}} \in \Theta \subset \mathbb{R}^m$ – неизвестное истинное значение вектора параметров; $\eta = (\eta_1, \dots, \eta_N)^{\mathbf{T}} \in \mathbb{R}^N$ – наблюдаемый случайный вектор, который часто называют вектором зависимых переменных, где η_i – i -я зависимая переменная ($i = 1, \dots, N$); $x = (x^1, \dots, x^M)^{\mathbf{T}} \in \mathbb{R}^M$ – вектор-столбец M независимых переменных (предикторов, регрессоров). Относительно ненаблюдаемого случайного вектора ошибок наблюдений $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^{\mathbf{T}} \in \mathbb{R}^N$, без потери общности, предполагаем:

$$\mathbf{E}\{\varepsilon\} = \mathbf{0}_N, \quad \Sigma = \mathbf{Cov}\{\varepsilon, \varepsilon\} = \mathbf{E}\{\varepsilon\varepsilon^{\mathbf{T}}\} \succ 0.$$

Определение 1.36. Функция

$$\bar{\eta} ::= \mathbf{E}\{\eta\} = f(x; \theta^0), \quad x \in \mathbb{R}^M, \quad \theta^0 \in \Theta, \quad \bar{\eta} \in \mathbb{R}^N,$$

задающая зависимость математического ожидания $\bar{\eta} = \mathbf{E}\{\eta\}$ от регрессоров x , называется функцией регрессии. При этом математическая модель наблюдений (1.84) называется регрессионной моделью наблюдений.

Определение 1.37. В зависимости от значений размерностей M и N существует следующая классификация регрессионных моделей.

1. Если $N > 1$, регрессионная модель (1.84) называется многомерной.
2. Если $N = 1$, $M > 1$, регрессионная модель (1.84) называется множественной.
3. Если $N = M = 1$, регрессионная модель (1.84) называется простой.

Определение 1.38. В зависимости от свойств функции $f(\cdot; \theta^0)$ существует следующая классификация регрессионных моделей: если функция $f(\cdot; \theta^0)$ линейная по параметру θ^0 , то регрессионная модель линейная, в противном случае – нелинейная.

Пример 1.13. Рассмотрим коммерческую фирму, которая производит и продает некоторый товар. Доход фирмы за один месяц $\eta \in \mathbb{R}^1$ считаем случайной величиной. Через $x \in \mathbb{R}^1$ обозначим расходы фирмы на рекламу товара за прошлый месяц. Предположим, что средний доход за месяц линейно зависит от x : $\bar{\eta} = \mathbf{E}\{\eta\} = f(x; \theta^0) = \theta_1^0 + \theta_2^0 x$, $\theta^0 = (\theta_1^0, \theta_2^0)^T$. По приведенной классификации имеем простую линейную модель, а в результате экспериментов получаем множество точек, лежащих вблизи прямой $\bar{\eta} = \theta_1^0 + \theta_2^0 x$.

Сформулируем статистическую задачу. Пусть проведена серия n независимых случайных экспериментов вида (1.84). Для простоты предполагаем, что $N = 1$, то есть рассматриваем множественную регрессионную модель; i -й случайный эксперимент заключается в следующем: фиксируется некоторое значение независимых переменных $x = x_i \in \mathbb{R}^M$, проводится эксперимент и регистрируется его результат $\eta = y_i \in \mathbb{R}^1$, причем между этими величинами имеет место соотношение вида (1.84), то есть

$$y_i = f(x_i; \theta^0) + \varepsilon_i, \quad i = 1, \dots, n. \quad (1.85)$$

Предполагается, что ошибки измерения $\{\varepsilon_i\}$ независимы в совокупности и $\mathbf{E}\{\varepsilon_i\} = 0$, $\mathbf{D}\{\varepsilon_i\} = \sigma^2 < +\infty$. Задача заключается в том, чтобы по имеющимся наблюдениям $\{y_i\}$ и значениям факторов $\{x_i\}$ построить статистические оценки параметра θ^0 и дисперсии ошибок σ^2 .

Определение 1.39. Случайные эксперименты вида (1.85) называются активными, если экспериментатор имеет возможность выбирать значения факторов $x_1, \dots, x_n \in \mathbb{R}^M$ по своему усмотрению, в противном случае – пассивными.

Попытаемся решить поставленную задачу, пользуясь методом максимального правдоподобия. Для этого дополнительно предположим, что случайные ошибки наблюдения подчиняются нормальному закону распределения вероятностей:

$$\mathcal{L}\{\varepsilon_i\} = \mathcal{N}_1(0, \sigma^2), \quad i = 1, \dots, n. \quad (1.86)$$

Построим логарифмическую функцию правдоподобия. В силу формулы (1.85) и свойств нормального распределения из (1.86) имеем

$$\mathcal{L}\{y_i\} = \mathcal{N}_1(f(x_i; \theta^0), \sigma^2), \quad i = 1, \dots, n.$$

Тогда логарифмическая функция правдоподобия

$$\begin{aligned} l(\theta) &= \ln \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - f(x_i; \theta))^2}{2\sigma^2}} \right) = \\ &= -\frac{1}{2}n \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} R^2(\theta), \end{aligned} \quad (1.87)$$

где

$$R^2(\theta) = \sum_{i=1}^n (\delta y_i)^2 = \sum_{i=1}^n (y_i - f(x_i; \theta))^2 \geq 0. \quad (1.88)$$

Определение 1.40. Величина $\delta y_i = y_i - f(x_i; \theta)$ называется невязкой (остатком) в i -м эксперименте ($i = 1, \dots, n$) при значении параметра θ , а функция $R^2(\theta)$, определяемая по (1.88), называется суммой квадратов невязок (остатков) при значении параметра θ .

Оценка максимального правдоподобия

$$\hat{\theta} = \arg \max_{\theta} l(\theta)$$

с учетом (1.87) имеет вид

$$\hat{\theta} = \arg \min_{\theta} R^2(\theta). \quad (1.89)$$

Определение 1.41 (принцип метода наименьших квадратов). Принцип метода наименьших квадратов состоит в том, что в качестве оценки $\hat{\theta}$ вектора параметров θ^0 принимается значение (1.89), которое минимизирует сумму квадратов остатков (1.88). Такая статистическая оценка $\hat{\theta}$ называется МНК-оценкой.

Укажем геометрический смысл МНК-оценки. В семействе поверхностей $F = \{y = f(x; \theta), x \in \mathbb{R}^M : \theta \in \mathbb{R}^m\}$ с помощью (1.89) необходимо найти такую поверхность $y = f(x; \hat{\theta}), x \in \mathbb{R}^M$, для которой сумма квадратов отклонений (вдоль оси Y) экспериментальных точек $\{y_i\}$ от этой поверхности минимальна.

Замечание 1.12. Соотношение (1.89) определяет МНК-оценку без ограничений, то есть для $\theta^0 \in \mathbb{R}^m$. Если же $\theta^0 \in \Theta \subset \mathbb{R}^m$, то решается задача на условный минимум:

$$R^2(\theta) \rightarrow \min_{\theta \in \Theta}.$$

1.10.2 Оптимальные свойства МНК-оценки для линейной множественной регрессионной модели

Рассмотрим часто используемую на практике линейную множественную регрессионную модель.

Определение 1.42. Линейная множественная регрессионная модель является частным случаем модели (1.84):

$$\eta = \sum_{j=1}^m x_j \theta_j^0 + \varepsilon, \quad (1.90)$$

где η – зависимая переменная, $x = (x^1, \dots, x^m)^T \in \mathbb{R}^m$ – m -вектор независимых переменных (факторов, предикторов, регрессоров), $m = M$.

Замечание 1.13. Модель (1.90) линейна не только по θ , но и без потери общности линейна по $\{x_j\}$, так как любая модель вида

$$\eta = \sum_{j=1}^m g_j(x) \theta_j^0 + \varepsilon$$

приводится к (1.90) заменой:

$$x_j ::= g_j(x), \quad j = 1, \dots, m.$$

Модель наблюдений, согласно (1.85), (1.90), примет вид

$$y_i = \sum_{j=1}^m x_{ij}\theta_j^0 + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.91)$$

где y_i – значение зависимой переменной в i -м эксперименте; x_{ij} – значение j -го фактора в i -м эксперименте ($x_i = (x_{i1}, \dots, x_{im})^T$); ε_i – случайная ошибка в i -м эксперименте.

Замечание 1.14. В п. 1.10.1 метод наименьших квадратов получен как частный случай метода максимального правдоподобия (в предположении гауссовости случайных ошибок $\{\varepsilon_i\}$), но на практике метод наименьших квадратов используют и тогда, когда $\{\varepsilon_i\}$ не являются гауссовскими случайными величинами.

С учетом замечания 1.14 в дальнейшем относительно $\{\varepsilon_i\}$ предполагаем только следующее:

- 1) $\{\varepsilon_i\}$ некоррелированы, то есть $\mathbf{E}\{\varepsilon_i\varepsilon_j\} = 0, i \neq j$;
- 2) $\mathbf{E}\{\varepsilon_i\} = 0$;
- 3) $\mathbf{D}\{\varepsilon_i\} = \sigma^2 < \infty$.

Удобно перейти к следующим матричным обозначениям:

$Y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$ – вектор-столбец результатов наблюдений;

$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T \in \mathbb{R}^n$ – вектор-столбец ошибок наблюдений;

$X = (x_1 : \dots : x_n)^T$ – план эксперимента ($(n \times m)$ -матрица);

$\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \mathbb{R}^m$ – вектор-столбец оцениваемых параметров.

Тогда (1.91) можно переписать в матричном виде:

$$Y = X\theta^0 + \varepsilon - \quad (1.92)$$

множественная линейная регрессионная модель, где относительно случайного вектора ε предполагается

$$\mathbf{E}\{\varepsilon\} = \mathbf{0}_n, \quad \text{Cov}\{\varepsilon, \varepsilon\} = \sigma^2 \mathbf{1}_n. \quad (1.93)$$

Теорема 1.17. Пусть имеет место линейная множественная регрессионная модель (1.92), выполняются соотношения (1.93) и план эксперимента X – невырожденный:

$$|X^T X| \neq 0. \quad (1.94)$$

Тогда МНК-оценка $\hat{\theta}$ вектора параметров θ^0 единственна:

$$\hat{\theta} = C_0 Y, \quad C_0 = (X^T X)^{-1} X^T, - \quad (1.95)$$

линейна по наблюдениям Y , несмещенная и имеет матрицу вариаций:

$$V = V\{\hat{\theta}\} = \mathbf{E}_{\theta^0}\{(\hat{\theta} - \theta^0)(\hat{\theta} - \theta^0)^T\} = \sigma^2 (X^T X)^{-1}.$$

Доказательство. Получим МНК-оценку из (1.89), (1.88) с учетом модели (1.92):

$$R^2(\theta) = (Y - X\theta)^T(Y - X\theta) = Y^T Y - \theta^T X^T Y - Y^T X\theta + \theta^T X^T X\theta \rightarrow \min_{\theta}.$$

Необходимым условием минимума является

$$\nabla_{\theta} R^2(\theta) = \mathbf{0}_m. \quad (1.96)$$

Воспользуемся свойствами оператора ∇ . Для произвольной константы C , произвольного вектора b и матрицы A :

$$\nabla_{\theta}(C) = \mathbf{0}_m, \quad \nabla_{\theta}(b^T \theta) = \nabla_{\theta}(\theta^T b) = b, \quad \nabla_{\theta}(\theta^T A \theta) = 2A\theta.$$

Из (1.96) получим

$$\nabla_{\theta} R^2(\theta) = \mathbf{0}_m - 2X^T Y + 2X^T X \theta = \mathbf{0}_m.$$

Умножая на $(X^T X)^{-1}$ ($X^T X$ невырождена согласно (1.94)), имеем (1.95). Заметим, что

$$\nabla_{\theta}^2 R^2(\theta) = 2X^T X \succ 0.$$

Следовательно, найден единственный минимум.

Подставим (1.92) в (1.95):

$$\hat{\theta} = (X^T X)^{-1} X^T (X \theta^0 + \varepsilon) = \theta^0 + C_0 \varepsilon. \quad (1.97)$$

Отсюда найдем смещение оценки:

$$\mathbf{E}_{\theta^0} \{\hat{\theta} - \theta^0\} = \mathbf{E}_{\theta^0} \{C_0 \varepsilon\} = \mathbf{0}_m,$$

следовательно, оценка несмещенная. Аналогично вычислим матрицу вариаций:

$$\begin{aligned} V = V\{\hat{\theta}\} &= \mathbf{E}_{\theta^0} \{C_0 \varepsilon (C_0 \varepsilon)^T\} = C_0 \mathbf{E}_{\theta^0} \{\varepsilon \varepsilon^T\} C_0^T = C_0 \sigma^2 \mathbf{1}_n C_0^T = \\ &= \sigma^2 (X^T X)^{-1}. \end{aligned}$$

□

Обозначим: $\lambda_{\min}(A)$, $\lambda_{\max}(A)$ – соответственно наименьшее и наибольшее характеристические числа неотрицательно определенной матрицы A .

Следствие 1.17.1. Если выполнено условие Эйкера:

$$\lambda_{\min}(X^T X) \rightarrow +\infty, \quad n \rightarrow +\infty, \quad (1.98)$$

то МНК-оценка $\hat{\theta}$, определяемая (1.95), является состоятельной оценкой:

$$\hat{\theta} \xrightarrow{\mathbf{P}} \theta^0, \quad n \rightarrow +\infty. \quad (1.99)$$

Доказательство. Оценим следующую величину:

$$\begin{aligned} \mathbf{E}_{\theta^0} \{|\hat{\theta} - \theta^0|^2\} &= \mathbf{E}_{\theta^0} \{(\hat{\theta} - \theta^0)^T (\hat{\theta} - \theta^0)\} = \mathbf{E}_{\theta^0} \{\text{tr}((\hat{\theta} - \theta^0)(\hat{\theta} - \theta^0)^T)\} = \\ &= \text{tr}(V\{\hat{\theta}\}) = \sigma^2 \text{tr}((X^T X)^{-1}) \leq \sigma^2 m \lambda_{\max}((X^T X)^{-1}). \end{aligned}$$

Известно свойство характеристических чисел матрицы: если выполняется (1.98), то $\lambda_{\max}((X^T X)^{-1}) \rightarrow 0$, следовательно:

$$\mathbf{E}_{\theta^0} \{|\hat{\theta} - \theta^0|^2\} \rightarrow 0, \quad n \rightarrow +\infty,$$

а значит, выполняется (1.99). □

Замечание 1.15. Не при любом плане эксперимента X МНК-оценка является состоятельной.

Рассмотрим теперь задачу оценивания дисперсии σ^2 .

Определение 1.43. Остаточной суммой квадратов называется статистика

$$R_{\min}^2 = \sum_{i=1}^n (y_i - f(x_i; \hat{\theta}))^2 = R^2(\hat{\theta}) \geq 0,$$

где $\hat{\theta}$ – МНК-оценка.

Учитывая линейную модель (1.92), легко получить

$$R_{\min}^2 = (Y - X\hat{\theta})^T(Y - X\hat{\theta}) \geq 0. \quad (1.100)$$

Теорема 1.18. Если выполнены условия теоремы 1.17 и число наблюдений больше числа оцениваемых параметров: $n > m$, то несмещенной оценкой дисперсии σ^2 является статистика

$$\hat{\sigma}^2 = \frac{1}{n-m} R_{\min}^2. \quad (1.101)$$

Доказательство. Используя теорему 1.17, имеем

$$Y - X\hat{\theta} = Y - X\theta^0 - XC_0\varepsilon = \varepsilon - XC_0\varepsilon = (\mathbf{1}_n - X(X^T X)^{-1}X^T)\varepsilon.$$

Подставляя в (1.100), получаем

$$R_{\min}^2 = \varepsilon^T(\mathbf{1}_n - X(X^T X)^{-1}X^T)^2\varepsilon.$$

Установим несмещенность оценки (1.101), учитывая (1.93) и легко проверяемое свойство идемпотентности матрицы $\mathbf{1}_n - X(X^T X)^{-1}X^T$:

$$\begin{aligned} \mathbf{E}\{\hat{\sigma}^2\} &= \frac{1}{n-m} \mathbf{E}\{\text{tr}((\mathbf{1}_n - X(X^T X)^{-1}X^T)^2\varepsilon\varepsilon^T)\} = \\ &= \frac{1}{n-m} \sigma^2 \text{tr}(\mathbf{1}_n - X(X^T X)^{-1}X^T) = \frac{\sigma^2}{n-m} (n-m) = \sigma^2. \end{aligned}$$

□

Следствие 1.18.1. Несмещенной оценкой для матрицы вариаций $V = V\{\hat{\theta}\}$ является следующая статистика:

$$\hat{V} = \frac{R_{\min}^2}{n-m} (X^T X)^{-1}.$$

Теорема 1.19 (Гаусса – Маркова). Если выполнены условия теоремы 1.17, то в семействе \mathcal{S} всевозможных линейно несмещенных статистических оценок вектора параметров $\theta^0 = (\theta_1^0, \dots, \theta_m^0)^T \in \mathbb{R}^m$ МНК-оценка, определяемая теоремой 1.17: $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_m)^T = C_0 Y$, характеризуется наименьшими вариациями.

Доказательство. Определим семейство \mathcal{S} , о котором говорится в теореме:

$$\mathcal{S} = \{\tilde{\theta} = CY : C - \text{произвольная матрица}, \mathbf{E}_{\theta^0}\{\tilde{\theta}\} = \theta^0, \forall \theta^0\}.$$

Представим матрицу C в виде: $C = C_0 + \Delta$, где $\Delta = (\Delta_{ij})$ – произвольная $(m \times n)$ -матрица. По теореме 1.17 для МНК-оценки имеем

$$\hat{\theta} = C_0 Y, \quad \mathbf{E}_{\theta^0}\{\hat{\theta}\} = \theta^0, \quad \forall \theta^0;$$

$$V\{\hat{\theta}\} = (v_{ij}) = \mathbf{E}_{\theta^0}\{(\hat{\theta} - \theta^0)(\hat{\theta} - \theta^0)^T\} = \sigma^2 (X^T X)^{-1}.$$

Отсюда следует, что $\hat{\theta} \in \mathcal{S}$.

Получим ограничение на матрицу Δ , обусловленное несмещенностью МНК-оценки (1.95). Для любого $\theta^0 \in \mathbb{R}^m$ должно выполняться

$$\begin{aligned} \mathbf{E}_{\theta^0}\{\tilde{\theta}\} &= \mathbf{E}_{\theta^0}\{(C_0 + \Delta)Y\} = \mathbf{E}_{\theta^0}\{\hat{\theta} + \Delta Y\} = \theta^0 + \Delta \mathbf{E}_{\theta^0}\{Y\} = \\ &= \theta^0 + \Delta X \theta^0 = \theta^0. \end{aligned}$$

Следовательно,

$$\Delta X = \mathbf{0},$$

и имеем эквивалентное представление для оценки $\tilde{\theta} \in \mathcal{S}$:

$$\tilde{\theta} = (C_0 + \Delta)Y = \hat{\theta} + \Delta(X\theta^0 + \varepsilon) = \hat{\theta} + \Delta\varepsilon.$$

С учетом этого вычислим матрицу вариаций:

$$\begin{aligned}\tilde{V} = (\tilde{v}_{ij}) &= V\{\tilde{\theta}\} = \mathbf{E}_{\theta^0}\{(\tilde{\theta} - \theta^0)(\tilde{\theta} - \theta^0)^{\mathbf{T}}\} = \\ &= \mathbf{E}_{\theta^0}\{((\hat{\theta} - \theta^0) + \Delta\varepsilon)((\hat{\theta} - \theta^0)^{\mathbf{T}} + \varepsilon^{\mathbf{T}}\Delta^{\mathbf{T}})\} = \\ &= V\{\hat{\theta}\} + \Delta\mathbf{E}\{\varepsilon(\hat{\theta} - \theta^0)^{\mathbf{T}}\} + \mathbf{E}\{(\Delta\varepsilon(\hat{\theta} - \theta^0)^{\mathbf{T}})^{\mathbf{T}}\} + \Delta\mathbf{E}_{\theta^0}\{\varepsilon\varepsilon^{\mathbf{T}}\}\Delta^{\mathbf{T}}.\end{aligned}$$

Учитывая (1.93) и представление (1.97) для МНК-оценки (1.95), полученное при доказательстве теоремы 1.17, имеем

$$\Delta\mathbf{E}\{\varepsilon(\hat{\theta} - \theta^0)^{\mathbf{T}}\} = \Delta\mathbf{E}_{\theta^0}\{\varepsilon\varepsilon^{\mathbf{T}}C_0^{\mathbf{T}}\} = \sigma^2\Delta C_0^{\mathbf{T}} = \sigma^2\Delta X(X^{\mathbf{T}}X)^{-1}X^{\mathbf{T}} = \mathbf{0}.$$

В результате получим

$$\tilde{V} = (\tilde{v}_{ij}) = V + \sigma^2\Delta\Delta^{\mathbf{T}},$$

откуда имеем выражения для диагональных элементов:

$$\tilde{v}_{ii} = v_{ii} + \sigma^2 \sum_{j=1}^n \Delta_{ij}^2 \geq v_{ii}, \quad i = 1, \dots, m,$$

причем равенство $\tilde{V} = V$ выполняется тогда и только тогда, когда $\Delta = \mathbf{0} \Leftrightarrow \tilde{\theta} = \hat{\theta}$. \square

1.10.3 Выравнивание экспериментальных данных прямой линией. Применение метода наименьших квадратов в эконометрике

Необходимо построить прямую $y = \hat{\theta}_1 + \hat{\theta}_2 x$ по методу наименьших квадратов для следующей линейной регрессионной модели:

$$y_i = \theta_1 + \theta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n.$$

Сведем эту модель к модели из п. 1.10.2:

$$x_{i1} := 1, \quad x_{i2} := x_i,$$

получим модель из п. 1.10.2, в которой $\theta = (\theta_1, \theta_2)^{\mathbf{T}}$ – подлежащий оцениванию вектор параметров ($m = 2$).

Для вычисления оценки метода наименьших квадратов определим:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}; \quad (X^{\mathbf{T}}X)^{-1} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}^{-1};$$

$$X^{\mathbf{T}}Y = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n x_i y_i \right)^{\mathbf{T}}; \quad \hat{\theta} = (X^{\mathbf{T}}X)^{-1}X^{\mathbf{T}}Y; \quad V\{\hat{\theta}\} = \sigma^2(X^{\mathbf{T}}X)^{-1}.$$

Пример 1.14 (оценивание производственной функции). Рассмотрим производственную функцию Кобба – Дугласа:

$$Y_i = \beta_1 K_i^{\beta_2} L_i^{\beta_3},$$

где i – номер года ($i = 1, 2, \dots, n, \dots$); Y_i – суммарная стоимость ВВП за i -й год; K_i – суммарная стоимость капиталовложений в i -м году; L_i – суммарная стоимость трудовых затрат в i -м году; $\beta_1, \beta_2, \beta_3 \geq 0$ – неизвестные параметры.

По данным $\{Y_i, K_i, L_i\}$ за n лет необходимо оценить параметры $\beta_1, \beta_2, \beta_3$. Приведем эту модель к линейной регрессионной модели с помощью логарифмирования:

$$y_i = \theta_1 x_{i1} + \theta_2 x_{i2} + \theta_3 x_{i3} + \varepsilon_i, \quad m = 3, \quad \theta = (\theta_1, \theta_2, \theta_3)^T,$$

где $y_i = \ln Y_i$; $x_{i1} \equiv 1$; $x_{i2} = \ln K_i$; $x_{i3} = \ln L_i$; $\theta_1 = \ln \beta_1$; $\theta_2 = \beta_2$; $\theta_3 = \beta_3$; $\{\varepsilon_i\}$ – случайные ошибки наблюдения.

Вектор оценок $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)^T$ можно получить по (1.95) из теоремы 1.17.

После того как модель построена, можно построить прогноз ВВП в $(n+1)$ -м году:

$$Y_{n+1} = e^{\hat{\theta}_1} K_{n+1}^{\hat{\theta}_2} L_{n+1}^{\hat{\theta}_3},$$

где K_{n+1}, L_{n+1} – плановые значения капитальных вложений и трудовых затрат в $(n+1)$ -м году.

Заметим, что иногда применяется нелинейный метод наименьших квадратов:

$$\sum_{i=1}^n (Y_i - \beta_1 K_i^{\beta_2} L_i^{\beta_3})^2 \rightarrow \min_{\beta_1, \beta_2, \beta_3}.$$

Глава 2

Теория статистической проверки гипотез

2.1 Основные понятия теории статистической проверки гипотез

Рассмотрим следующую статистическую задачу. Пусть имеется случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathcal{X} \subseteq \mathbb{R}^{nN}$ из некоторого N -мерного распределения вероятностей $P_\theta(\cdot)$, заданного на измеримом пространстве (Ω, \mathcal{F}) , где $\theta \in \Theta \subseteq \mathbb{R}^m$ — неизвестное истинное значение векторного параметра, Θ — параметрическое пространство, $\mathcal{X} \subseteq \mathbb{R}^{nN}$ — выборочное пространство. Задано некоторое натуральное число $K \geq 2$ и определено некоторое разбиение параметрического пространства Θ на K областей:

$$\Theta = \bigcup_{k=0}^{K-1} \Theta_k, \quad \Theta_k \cap \Theta_l = \emptyset, \quad k \neq l.$$

Определение 2.1. Утверждение о неизвестном значении параметра θ , состоящее в том, что $\theta \in \Theta_k$, называется k -й гипотезой:

$$H_k : \theta \in \Theta_k, \quad k = 0, \dots, K-1.$$

Определение 2.2. Гипотеза H_k называется простой, если мощность $|\Theta_k| = 1$, в противном случае — сложной (составной).

Определение 2.3. Пусть на самом деле значение параметра $\theta \in \Theta_\nu$, где $\nu \in \{0, \dots, K-1\}$, тогда говорят, что гипотеза H_ν истинна.

Задача статистической проверки гипотез H_0, \dots, H_{K-1} состоит в том, чтобы по наблюдаемой выборке X неким оптимальным образом оценить номер ν истинной гипотезы: $d = d(X) = k$ — выносим решение в пользу гипотезы H_k ($d = d(X)$ — статистическая оценка для ν). Возможно K решений ($k \in \{0, 1, \dots, K-1\}$). Множество возможных решений обозначим

$$D = \{0, 1, \dots, K-1\}, \quad |D| = K,$$

и назовем *пространством решений*.

Определение 2.4. Решающим правилом (решающей функцией, критерием, тестом) в выше сформулированной задаче статистической проверки гипотез называется функциональное отображение выборочного пространства \mathcal{X} в пространство решений D :

$$\mathcal{X} \xrightarrow{d(\cdot)} D. \quad (2.1)$$

Существуют два основных типа решающих правил (РП) (2.1).

Определение 2.5. *Нерандомизированным РП называется отображение (2.1) следующего вида:*

$$d = d(X) = \begin{cases} 0, & X \in \mathcal{X}_0; \\ \vdots & \\ K-1, & X \in \mathcal{X}_{K-1}, \end{cases}$$

где $\{\mathcal{X}_0, \dots, \mathcal{X}_{K-1}\}$ — некоторое детерминированное борелевское разбиение выборочного пространства:

$$\mathcal{X} = \bigcup_{k=0}^{K-1} \mathcal{X}_k, \quad \mathcal{X}_k \cap \mathcal{X}_l = \emptyset, \quad k \neq l.$$

При этом, если выборка X фиксирована, то решение $d = d(X)$ не случайно.

Определение 2.6. *Рандомизированным РП называется случайное отображение (2.1) следующего вида:*

$$d = d(X, \omega), \quad \omega \in \Omega, \quad X \in \mathcal{X}, \quad d \in D,$$

причем если выборка X фиксирована, то решение $d = d(X, \omega)$ является дискретной случайной величиной с множеством значений D и некоторым дискретным распределением вероятностей:

$$\varphi_i = \varphi_i(X) = \mathbf{P}\{d = i | X\}, \quad i \in D.$$

При этом борелевские функции $\varphi_i = \varphi_i(X)$, $i \in D$, удовлетворяют следующим ограничениям:

$$0 \leq \varphi_i(X) \leq 1, \quad i \in D; \quad \sum_{i \in D} \varphi_i(X) = 1, \quad X \in \mathcal{X},$$

и называются критическими функциями.

Укажем алгоритм принятия решения с помощью рандомизированного решающего правила.

1. По выборке X вычисляем значения критических функций: $\varphi_i = \varphi_i(X)$, $i \in D$, и определяем дискретное распределение вероятностей $\{\varphi_0(X), \varphi_1(X), \dots, \varphi_{K-1}(X)\}$.
2. Проводим случайный эксперимент (жребий) со множеством исходов D и дискретным распределением вероятностей, найденным на шаге 1.
3. Регистрируем исход k этого жребия и принимаем решение $d = k$.

Нерандомизированное решающее правило есть частный случай рандомизированного решающего правила, если критические функции принимают одно из двух возможных значений:

$$\varphi_i(X) \in \{0, 1\}, \quad X \in \mathcal{X}; \quad \mathcal{X}_i = \{X : \varphi_i(X) = 1\}, \quad i \in D.$$

2.2 Решающее правило Неймана – Пирсона

Рассмотрим задачу проверки двух гипотез ($K = 2$). Пусть имеется случайная выборка $X = (x_1^{\mathbf{T}}, \dots, x_n^{\mathbf{T}})^{\mathbf{T}} \in \mathbb{R}^{nN}$, образованная наблюдениями с плотностью распределения вероятностей $p(x; \theta)$, $x \in \mathbb{R}^N$, $\theta \in \Theta$, где θ – неизвестное истинное значение параметра. Обозначим:

$$p(X; \theta) = \prod_{l=1}^n p(x_l; \theta) -$$

совместная плотность распределения всей выборки X ;

$$\mathbf{P}_\theta(A) = \int_A p(X; \theta) dX, \quad A \in \mathcal{B}^{nN}, \quad -$$

вероятностная мера, определенная при значении параметра θ .

Пространство параметров Θ разбито на две непересекающиеся области:

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset.$$

Определены две гипотезы:

$$\begin{aligned} H_0 &: \theta \in \Theta_0, \\ H_1 &: \theta \in \Theta_1. \end{aligned}$$

Так как $H_1 = \overline{H_0}$, то H_1 называется *альтернативой*, а H_0 — *нулевой гипотезой*.

Задача заключается в построении решающего правила (теста) для проверки гипотез H_0, H_1 по выборке X . Построим РП общего вида — рандомизированное решающее правило:

$$d = d(X, \omega), \quad X \in \mathbb{R}^{nN}, \quad \omega \in \Omega, \quad d \in D = \{0, 1\}; \quad (2.2)$$

$$\mathbf{P}\{d(X, \omega) = 1 | X\} = \varphi(X), \quad \mathbf{P}\{d(X, \omega) = 0 | X\} = 1 - \varphi(X), \quad (2.3)$$

где $0 \leq \varphi(X) \leq 1$ — некоторая произвольная критическая функция.

Существует бесконечно много РП $d(\cdot)$, различающихся критическими функциями $\varphi(\cdot)$. Как найти оптимальную критическую функцию $\varphi_*(\cdot)$? Сформулируем критерий оптимальности.

Определение 2.7. Принято говорить, что при принятии решений имеет место ошибка I рода, если на самом деле верна гипотеза H_0 ($\nu = 0$), а принято решение $d = 1$ в пользу H_1 . При этом вероятностью ошибки I рода называется число

$$\begin{aligned} \alpha = \alpha(\theta) &::= \mathbf{P}_\theta\{d(X, \omega) = 1\} = \mathbf{E}_\theta\{\varphi(X)\} = \\ &= \int_{\mathbb{R}^{nN}} \varphi(X) p(X; \theta) dX, \quad \theta \in \Theta_0. \end{aligned}$$

Определение 2.8. Принято говорить, что имеет место ошибка II рода, если на самом деле верна гипотеза H_1 ($\nu = 1$), а принято решение $d = 0$ в пользу H_0 . При этом вероятностью ошибки II рода называется число

$$\begin{aligned} \beta = \beta(\theta) &::= \mathbf{P}_\theta\{d(X, \omega) = 0\} = \mathbf{E}_\theta\{1 - \varphi(X)\} = \\ &= 1 - \int_{\mathbb{R}^{nN}} \varphi(X) p(X; \theta) dX, \quad \theta \in \Theta_1. \end{aligned}$$

Определение 2.9. Мощностью решающего правила $d = d(X, \omega)$ называется вероятность правильного принятия альтернативы H_1 ($\nu = d = 1$):

$$w = w(\theta) = \int_{\mathbb{R}^{nN}} \varphi(X) p(X; \theta) dX, \quad \theta \in \Theta_1.$$

Отметим, что $w(\theta) = 1 - \beta(\theta)$.

Было бы целесообразно найти функцию $\varphi_*(\cdot)$ так, что $\alpha = \beta = 0$. Очевидно, что в условиях регулярности это невозможно (как и в § ?? невозможно было построить статистическую оценку с нулевой вариацией $V = 0$). Как поступать в этой ситуации?

Определение 2.10 (принцип оптимальности). Критическую функцию $\varphi(X)$ в рандомизированном решающем правиле (2.2), (2.3) надлежит выбирать таким образом, чтобы вероятность ошибки I рода не превосходила некоторого наперед заданного числа $\varepsilon \in (0, 1)$, а вероятность ошибки II рода была минимальной, то есть как решение следующей экстремальной задачи:

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \varepsilon, \quad \sup_{\theta \in \Theta_1} \beta(\theta) \rightarrow \min_{0 \leq \varphi(\cdot) \leq 1},$$

что эквивалентно:

$$\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \varepsilon, \quad \inf_{\theta \in \Theta_1} w(\theta) \rightarrow \max_{0 \leq \varphi(\cdot) \leq 1}. \quad (2.4)$$

При этом оптимальная критическая функция $\varphi_*(X)$ и соответствующее ей решающее правило $d_*(X, \omega)$ называются соответственно критической функцией Неймана — Пирсона и решающим правилом Неймана — Пирсона, величина $\sup_{\theta \in \Theta_0} \alpha(\theta)$ — размером теста, а ε — уровнем значимости теста.

Отметим, что экстремальная задача (2.4) — задача вариационного исчисления. Нейман и Пирсон решили эту задачу лишь для случая, когда H_0, H_1 — простые гипотезы: $\Theta = \{\theta_0, \theta_1\}$, а $\Theta_0 = \{\theta_0\}$ и $\Theta_1 = \{\theta_1\}$ — одноточечные множества.

Упростим обозначения:

$$H_i : \theta = \theta_i, \quad i = 0, 1; \quad \theta_0, \theta_1 \in \Theta, \quad \theta_0 \neq \theta_1;$$

$$p_i(X) = p(X; \theta_i), \quad X \in \mathbb{R}^{nN};$$

$$\mathbf{P}_i(B) = \mathbf{P}_{\theta_i}(B), \quad Q(B) = \frac{1}{2}(\mathbf{P}_0(B) + \mathbf{P}_1(B)), \quad B \in \mathcal{B}^{nN};$$

$$\mathbf{E}\{\cdot\} \equiv \mathbf{E}_{\theta_i}\{\cdot\}.$$

Определение 2.11. Статистикой отношения правдоподобия для проверки простых гипотез H_0, H_1 называется статистика

$$L = L(X) = \frac{p_1(X)}{p_0(X)} \geq 0.$$

Теорема 2.1 (Неймана — Пирсона). Для выше сформулированной задачи проверки простых гипотез H_0, H_1 справедливы следующие три утверждения:

1. Для любого наперед заданного ε , $0 < \varepsilon < 1$, найдутся такие постоянные $c_* > 0$, $\varkappa_* \in [0, 1]$, что решающее правило (тест) Неймана — Пирсона $d = d_*(X, \omega)$ с критической функцией

$$\varphi_*(X) = \begin{cases} 0, & L(X) < c_*; \\ \varkappa_*, & L(X) = c_*; \\ 1, & L(X) > c_*, \end{cases} \quad (2.5)$$

имеет размер, в точности равный ε ;

2. Тест Неймана — Пирсона $d = d_*(X, \omega)$, определяемый (2.5), имеет наибольшую мощность w_* среди всех тестов, размер которых не превосходит ε ;
3. Тест $d = d_*(X, \omega)$ — единственный с точностью до множества B нулевой меры: $Q(B) = 0$.

Доказательство. Запишем экстремальную задачу (2.4) с учетом определений 2.7, 2.9 в явном виде:

$$\begin{aligned}\alpha &= \int_{\mathbb{R}^{nN}} \varphi(X) p_0(X) dX \leq \varepsilon, \\ w &= \int_{\mathbb{R}^{nN}} \varphi(X) p_1(X) dX \rightarrow \max_{\varphi(\cdot)}.\end{aligned}\quad (2.6)$$

Доказательство разобьем на три части.

1. Введем в рассмотрение функцию распределения статистики $L = L(X)$ при верной гипотезе H_0 :

$$F_0(y) = \mathbf{P}_0\{L(X) < y\}, \quad y \geq 0.$$

Считая произвольными c_* и \varkappa_* , вычислим для теста (2.5) вероятность ошибки I рода:

$$\begin{aligned}\alpha &= \mathbf{E}_0\{\varphi(X)\} = \varkappa_* \mathbf{P}_0\{L = c_*\} + \mathbf{P}_0\{L > c_*\} = \\ &= \varkappa_*(F_0(c_* + 0) - F_0(c_*)) + (1 - F_0(c_* + 0)).\end{aligned}\quad (2.7)$$

В этом соотношении подберем произвольные константы c_* , \varkappa_* таким образом, чтобы $\alpha = \alpha_* = \varepsilon$. Для этого выберем c_* так, чтобы

$$F_0(c_*) \leq 1 - \varepsilon \leq F_0(c_* + 0).$$

Возможны два случая:

- c_* — точка непрерывности функции $F_0(\cdot)$; тогда

$$c_* = F_0^{-1}(1 - \varepsilon),$$

и из (2.7) имеем

$$\alpha_* = \varepsilon \quad \forall \varkappa_*;$$

- c_* — точка разрыва; в этом случае выберем

$$\varkappa_* = \frac{F_0(c_* + 0) - (1 - \varepsilon)}{F_0(c_* + 0) - F_0(c_*)}.$$

Подставляя это значение в (2.7), получаем

$$\alpha_* = \varepsilon.$$

2. Для критической функции $\varphi_*(\cdot)$ имеем

$$\alpha_* = \varepsilon, \quad w_* = \int_{\mathbb{R}^{nN}} \varphi_*(X) p_1(X) dX. \quad (2.8)$$

Выберем произвольную критическую функцию $\tilde{\varphi}(X)$, для которой обозначим

$$\tilde{\alpha} = \int_{\mathbb{R}^{nN}} \tilde{\varphi}(X) p_0(X) dX \leq \varepsilon, \quad \tilde{w} = \int_{\mathbb{R}^{nN}} \tilde{\varphi}(X) p_1(X) dX.$$

Покажем: $\tilde{w} \leq w_*$. Легко проверить поточечно, используя (2.5), что

$$\Delta = \Delta(X) ::= (p_1(X) - c_* p_0(X))(\varphi_*(X) - \tilde{\varphi}(X)) \geq 0, \quad \forall X \in \mathbb{R}^{nN}.$$

Отсюда с учетом (2.8) получаем

$$\int_{\mathbb{R}^{nN}} \Delta(X) dX = w_* - c_* \alpha_* - \tilde{w} + c_* \tilde{\alpha} = w_* - \tilde{w} - c_*(\varepsilon - \tilde{\alpha}) \geq 0.$$

Следовательно,

$$w_* \geq \tilde{w} + c_*(\varepsilon - \tilde{\alpha}) \geq \tilde{w}.$$

3. В экстремальной задаче (2.6) значения интегралов не изменятся, если значение $\varphi_*(X)$ изменить на множестве B : $Q(B) = 0$.

□

Следствие 2.1.1. Если функция распределения $F_0(\cdot)$ статистики отношения правдоподобия $L = L(X)$ при гипотезе H_0 — непрерывная функция, то РП Неймана — Пирсона является нерандомизированным и имеет следующий вид:

$$d = d_*(X) = \begin{cases} 0, & L(X) < c_*; \\ 1, & L(X) \geq c_*, \end{cases}$$

где

$$c_* = F_0^{-1}(1 - \varepsilon).$$

Доказательство. Обратимся к первому пункту доказанной выше теоремы 2.1. Если $F_0(\cdot)$ непрерывна, то \varkappa_* можно выбирать произвольно на отрезке $[0, 1]$, поэтому выберем $\varkappa_* = 1$, тогда, согласно (2.5), $\varphi_*(X) \in \{0, 1\}$, то есть жребий исчезает и имеет место детерминированное (нерандомизированное) решающее правило. □

2.3 Проверка гипотез о распределении вероятностей. Критерии согласия

Пусть наблюдается случайная выборка $X = (x_1, \dots, x_n)^T \in \mathbb{R}^n$ из одномерного распределения вероятностей с неизвестной функцией распределения $F(x)$, $x \in \mathbb{R}^1$. Пусть $F_0(x)$, $x \in \mathbb{R}^1$, — некоторая гипотетическая (предполагаемая) функция распределения вероятностей. В связи с этим определены две гипотезы:

$$\begin{aligned} H_0 &: F(\cdot) \equiv F_0(\cdot), \\ H_1 &: F(\cdot) \not\equiv F_0(\cdot). \end{aligned}$$

Нулевая гипотеза H_0 — простая, альтернатива H_1 — сложная гипотеза.

Определение 2.12. Гипотеза H_0 означает, что выборка X согласуется с распределением, задаваемым гипотетической функцией распределения вероятностей $F_0(\cdot)$, и поэтому называется гипотезой согласия; решающее правило $d = d(X)$ для проверки гипотез H_0, H_1 — критерием согласия.

2.3.1 χ^2 -критерий согласия Пирсона

Воспользуемся той же схемой группировки выборки X , что и при построении гистограммы.

Разобьем числовую прямую на $K \geq 2$ ячеек точками деления $-\infty = b_0 < b_1 < b_2 < \dots < b_K = +\infty$:

$$\Gamma_k = [b_{k-1}, b_k), \quad k = 1, 2, \dots, K.$$

Вычислим два набора величин:

1. Теоретические вероятности попадания случайного наблюдения x_i в ячейки $\{\Gamma_k\}_{k=1}^K$ при условии, что верна H_0 ($k = 1, \dots, K$):

$$p_k^0 = \mathbf{P}_{H_0}\{x_i \in \Gamma_k\} = \mathbf{P}_{H_0}\{b_{k-1} \leq x_i < b_k\} = F_0(b_k) - F_0(b_{k-1}) > 0;$$

$$\sum_{k=1}^K p_k^0 = 1;$$

2. Относительные и абсолютные частоты попадания в ячейки, вычисленные по выборке X :

$$\hat{p}_k = \frac{\nu_k}{n}, \quad \nu_k = \sum_{i=1}^n \mathbf{1}_{\Gamma_k}(x_i), \quad k = 1, \dots, K; \quad \sum_{k=1}^K \nu_k = n, \quad \sum_{k=1}^K \hat{p}_k = 1.$$

Определение 2.13. χ^2 -статистикой Пирсона называется статистика

$$\chi^2 = \sum_{k=1}^K \frac{(\nu_k - np_k^0)^2}{np_k^0} = n \sum_{k=1}^K \frac{(\hat{p}_k - p_k^0)^2}{p_k^0} \geq 0. \quad (2.9)$$

Из (2.9) видно, что χ^2 — это с точностью до множителя n взвешенная с весами $\{\frac{1}{p_k^0}\}$ сумма квадратов отклонений частот \hat{p}_k от теоретических значений p_k^0 .

Лемма 2.1. Если верна гипотеза H_0 , то при $n \rightarrow \infty$ имеет место сходимость к χ^2 -распределению с $K - 1$ степенью свободы:

$$\mathcal{L}_{H_0}\{\chi^2\} \rightarrow \chi_{K-1}^2,$$

или

$$\mathbf{P}_{H_0}\{\chi^2 < y\} \rightarrow F_{\chi_{K-1}^2}(y), \quad n \rightarrow +\infty, \quad y \geq 0, \quad (2.10)$$

где $F_{\chi_{K-1}^2}(\cdot)$ — функция χ^2 -распределения с $K - 1$ степенью свободы.

Доказательство. Обозначим нормированные частоты, входящие в (2.9):

$$\nu_k^*(n) = \frac{\nu_k - np_k^0}{\sqrt{n}}, \quad k = 1, \dots, K; \quad \sum_{k=1}^K \nu_k^*(n) \equiv 0. \quad (2.11)$$

Определим $(K - 1)$ -вектор нормированных частот

$$\nu^*(n) = (\nu_1^*(n), \dots, \nu_{K-1}^*(n))^{\mathbf{T}} \in \mathbb{R}^{K-1}.$$

Заметим, что в силу условия нормировки в (2.11) $\nu_K^*(n)$ линейно зависит от $\nu^*(n)$:

$$\nu_K^*(n) = - \sum_{l=1}^{K-1} \nu_l^*(n). \quad (2.12)$$

В асимптотике $n \rightarrow +\infty$ воспользуемся центральной предельной теоремой для $\nu^*(n)$ (многомерный аналог теоремы Муавра — Лапласа):

$$\mathcal{L}_{H_0}\{\nu^*(n)\} \rightarrow \mathcal{N}_{K-1}(\mathbf{0}_{K-1}, \Sigma), \quad (2.13)$$

где $\Sigma = (\sigma_{ij})$ — асимптотическая ковариационная матрица, для элементов которой справедлива формула ($i, j = 1, \dots, K - 1$)

$$\sigma_{ij} = p_i^0(\delta_{ij} - p_j^0).$$

Непосредственной проверкой можно убедиться, что в этом случае для обратной матрицы справедливо выражение

$$\Sigma^{-1} = (\bar{\sigma}_{ij}), \quad \bar{\sigma}_{ij} = \frac{1}{p_K^0} + \frac{\delta_{ij}}{p_i^0}. \quad (2.14)$$

Из (2.9), (2.11)–(2.14) имеем представление для χ^2 -статистики в виде квадратичной формы:

$$\begin{aligned} \chi^2 &= \sum_{k=1}^{K-1} \frac{(\nu_k^*(n))^2}{p_k^0} + \frac{(\sum_{l=1}^{K-1} \nu_l^*(n))^2}{p_K^0} = \sum_{k,l=1}^{K-1} \nu_k^*(n) \nu_l^*(n) \bar{\sigma}_{kl} = \\ &= (\nu^*(n))^T \Sigma^{-1} \nu^*(n). \end{aligned}$$

Введем в рассмотрение случайный вектор

$$\xi(n) = (\xi_k(n)) = (\Sigma^{1/2})^{-1} \nu^*(n) \in \mathbb{R}^{K-1},$$

где матрица $\Sigma^{1/2}$ — решение по Y матричного уравнения $Y^T \Sigma^{-1} Y = \mathbf{I}_{K-1}$ ($\Sigma = \Sigma^{1/2} (\Sigma^{1/2})^T$). Тогда χ^2 -статистика представима в виде суммы квадратов:

$$\chi^2 = \xi^T(n) \xi(n) = \sum_{k=1}^{K-1} \xi_k^2(n),$$

причем в силу (2.13)

$$\mathcal{L}_{H_0} \{\xi(n)\} \rightarrow \mathcal{N}_{K-1}(\mathbf{0}_{K-1}, \mathbf{I}_{K-1}).$$

Отсюда по определению χ^2 -распределения получаем (2.10). \square

Определение 2.14. χ^2 -критерий согласия Пирсона имеет вид

$$d = d(X) = \begin{cases} 0, & \chi^2 < F_{\chi_{K-1}^2}^{-1}(1 - \varepsilon); \\ 1, & \chi^2 \geq F_{\chi_{K-1}^2}^{-1}(1 - \varepsilon), \end{cases} \quad (2.15)$$

где $F_{\chi_{K-1}^2}^{-1}(1 - \varepsilon)$ — квантиль χ^2 -распределения с $K - 1$ степенью свободы уровня $1 - \varepsilon$, $0 < \varepsilon < 1$.

Теорема 2.2. При $n \rightarrow \infty$ асимптотический размер теста Пирсона (2.15) совпадает с наперед заданным уровнем значимости ε :

$$\alpha = \mathbf{P}_{H_0} \{d(X) = 1\} \rightarrow \varepsilon, \quad n \rightarrow +\infty.$$

Доказательство. основано на лемме 2.1 и состоит в вычислении α с использованием (2.10). \square

Замечание 2.1. χ^2 -критерий согласия Пирсона справедлив и для N -мерных ($N > 1$) функций распределения $F(x)$, $x \in \mathbb{R}^N$.

Замечание 2.2. χ^2 -критерий Пирсона допускает обобщение для случая сложных гипотез согласия:

$$H_0 : F(\cdot) \in F_0,$$

где

$$F_0 = \{F_0(x; \theta), x \in \mathbb{R}^N : \theta \in \Theta \subseteq \mathbb{R}^m\} -$$

некоторое параметрическое семейство функций распределения с m -мерным параметром θ . Появляется дополнительное ограничение: $K > m + 1$, и критерий согласия имеет вид

$$d = d^0(X) = \begin{cases} 0, & \inf_{\theta} \chi^2(\theta) < F_{\chi^2_{K-m-1}}^{-1}(1 - \varepsilon); \\ 1, & \inf_{\theta} \chi^2(\theta) \geq F_{\chi^2_{K-m-1}}^{-1}(1 - \varepsilon), \end{cases}$$

где

$$\chi^2(\theta) = \sum_{k=1}^K \frac{(\nu_k - np_k(\theta))^2}{np_k(\theta)} \geq 0, \quad p_k(\theta) = \mathbf{P}_{\theta}\{x_i \in \Gamma_k\}, \quad k = 1, \dots, K.$$

Заметим также, что статистика

$$\tilde{\theta} = \arg \min_{\theta} \chi^2(\theta)$$

называется χ^2 -оценкой параметра θ .

2.3.2 Критерий Колмогорова

Будем предполагать, что гипотеза согласия $H_0 : F(\cdot) \equiv F_0(\cdot)$ — простая, причем функция $F_0(\cdot)$ непрерывна. Построим по наблюдаемой выборке X выборочную функцию распределения, изученную ранее:

$$\hat{F}(x) = F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x - x_i) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(x - x_{(i)}), \quad x \in \mathbb{R}^1,$$

где $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ — вариационный ряд выборки.

Как было установлено, выборочная функция распределения — сильно состоятельная оценка истинной функции распределения:

$$F_n(x) \xrightarrow{\text{П.Н.}} F(x), \quad n \rightarrow +\infty, \quad x \in \mathbb{R}^1.$$

Определение 2.15. Расстоянием Колмогорова между выборочной функцией распределения $F_n(x)$ и гипотетической функцией распределения $F_0(x)$ называется величина

$$D_n = \sup_{-\infty < x < +\infty} |F_n(x) - F_0(x)| \quad (0 \leq D_n \leq 1). \quad (2.16)$$

Лемма 2.2. Если верна гипотеза H_0 , то случайные величины $u_1 = F_0(x_1)$, $u_2 = F_0(x_2)$, \dots , $u_n = F_0(x_n)$ независимы и одинаково распределены со стандартным равномерным распределением: $\mathcal{L}_{H_0}\{u_i\} = \mathcal{R}[0, 1]$, $i = 1, \dots, n$.

Доказательство. Независимость $\{u_i\}$ вытекает из независимости $\{x_i\}$ и соответствующей теоремы о борелевских функциональных преобразованиях. Вычислим функцию распределения для случайной величины u_i при условии, что верна гипотеза H_0 ($0 \leq y \leq 1$):

$$F_{u_i}(y) ::= \mathbf{P}_{H_0}\{u_i < y\} = \mathbf{P}_{H_0}\{x_i < F_0^{-1}(y)\} = F_0(F_0^{-1}(y)) = y.$$

Это и есть функция распределения стандартного равномерного закона $\mathcal{R}[0, 1]$. \square

Теорема 2.3. Если верна нулевая гипотеза H_0 , то распределение вероятностей статистики Колмогорова D_n , определяемой формулой (2.16) не зависит от гипотетической функции распределения $F_0(\cdot)$:

$$\mathcal{L}_{H_0}\{D_n\} = \mathcal{L}\{D_n^*\}, \quad D_n^* = \sup_{0 \leq u \leq 1} |\Psi_n(u) - u|;$$

$$\Psi_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(u - u_i),$$

где $\{u_i\}$ — независимые в совокупности случайные величины, имеющие стандартное равномерное распределение: $\mathcal{L}\{u_i\} = \mathcal{R}[0, 1]$, $i = 1, \dots, n$.

Доказательство. вытекает из леммы 2.2 и замены переменных $x = F_0^{-1}(u)$, $u = F_0(x)$ в экстремальной задаче (2.16). \square

Теорема 2.4. Если верна гипотеза H_0 , то при бесконечном увеличении объема выборки n ($n \rightarrow +\infty$) функция распределения нормированной статистики $\sqrt{n}D_n$ сходится к распределению Колмогорова:

$$\mathbf{P}_{H_0}\{\sqrt{n}D_n < z\} \rightarrow K(z) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2 z^2}, \quad z \geq 0.$$

Определение 2.16. Пусть ε — некоторое число: $0 < \varepsilon < 1$, а $K^{-1}(1 - \varepsilon)$ — квантиль распределения Колмогорова уровня $1 - \varepsilon$. Тогда тест

$$d = d(X) = \begin{cases} 0, & \sqrt{n}D_n < K^{-1}(1 - \varepsilon); \\ 1, & \sqrt{n}D_n \geq K^{-1}(1 - \varepsilon), \end{cases} \quad (2.17)$$

называется критерием согласия Колмогорова.

Замечание 2.3. Распределение статистики D_n не зависит от гипотетической функции распределения $F_0(\cdot)$, в результате чего имеет место универсальность теста Колмогорова.

Следствие 2.4.1. При $n \rightarrow \infty$ асимптотический размер теста Колмогорова (2.17) совпадает с наперед заданным уровнем значимости $\varepsilon \in (0, 1)$.

Доказательство. основывается на соотношениях (2.17), теореме 2.4 и заключается в проверке следующего факта: $\alpha \rightarrow \varepsilon$, где α — вероятность ошибки I рода. \square

2.4 Критерий отношения правдоподобия для проверки сложных гипотез

Ранее были рассмотрены методы проверки специальных сложных гипотез — гипотез согласия. Рассмотрим теперь универсальный критерий проверки сложных гипотез — критерий отношения правдоподобия, суть которого состоит в следующем.

Пусть наблюдается случайная выборка $X \in \mathbb{R}^{nN}$ объема n из некоторого N -мерного распределения вероятностей с плотностью $p(x; \theta)$, $x \in \mathbb{R}^N$, $\theta \in \Theta \subseteq \mathbb{R}^m$. Определены две, в общем случае сложные, гипотезы:

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \subset \Theta; \\ H_1 &: \theta \in \Theta_1 = \Theta \setminus \Theta_0. \end{aligned}$$

Аналогично критерию Неймана — Пирсона введем в рассмотрение статистику отношения правдоподобия:

$$\tilde{\Lambda}_n = \tilde{\Lambda}_n(X; \Theta_0) = \frac{\sup_{\theta \in \Theta_1} p(X; \theta)}{\sup_{\theta \in \Theta_0} p(X; \theta)}, \quad (2.18)$$

где

$$p(X; \theta) = \prod_{k=1}^n p(x_k; \theta) = L(\theta) —$$

функция правдоподобия. Часто рассматривают эквивалентные статистики:

$$\Lambda_n = \Lambda_n(X; \Theta_0) = \frac{\sup_{\theta \in \Theta} p(X; \theta)}{\sup_{\theta \in \Theta_0} p(X; \theta)} \geq 1; \quad (2.19)$$

$$\lambda_n = \lambda_n(X; \Theta_0) = \frac{1}{\Lambda_n(X; \Theta_0)} = \frac{\sup_{\theta \in \Theta_0} p(X; \theta)}{\sup_{\theta \in \Theta} p(X; \theta)} \in [0, 1].$$

Статистики (2.19) взаимно-однозначно функционально связаны между собой и статистикой отношения правдоподобия (2.18):

$$\Lambda_n = \max\{1, \tilde{\Lambda}_n\}, \quad \lambda_n = \min\left\{1, \frac{1}{\tilde{\Lambda}_n}\right\} \in [0, 1].$$

Определение 2.17. Критерием отношения правдоподобия (КОП) в задаче проверки сложных гипотез H_0, H_1 по выборке X объема n называется следующий статистический критерий ($C \in [0, 1]$):

$$d = d(X) = \begin{cases} 0, & \lambda_n(X; \Theta_0) > C; \\ 1, & \lambda_n(X; \Theta_0) \leq C, \end{cases}$$

или в эквивалентном виде

$$d = d(X) = \begin{cases} 0, & -2 \ln \lambda_n(X; \Theta_0) < \delta; \\ 1, & -2 \ln \lambda_n(X; \Theta_0) \geq \delta, \end{cases} \quad (2.20)$$

где критическое (пороговое) значение $\delta \geq 0$ выбирается так, чтобы критерий имел наперед заданный уровень значимости ε :

$$\begin{aligned} \mathbf{P}_\theta\{d(X) = 1\} &= \int_{\lambda_n(X; \Theta_0) \leq C} p(X; \theta) dX = \\ &= \mathbf{P}_\theta\{-2 \ln \lambda_n \geq \delta\} \leq \varepsilon, \quad \forall \theta \in \Theta_0. \end{aligned} \quad (2.21)$$

Как было показано ранее (теорема 2.1, известная как фундаментальная лемма Неймана — Пирсона), для простых гипотез КОП оптимален при любом n . Для сложных гипотез это, вообще говоря, не так. Тем не менее КОП широко применяется на практике. Оказывается, что он асимптотически оптимален при $n \rightarrow \infty$.

Рассмотрим вначале случай простой гипотезы H_0 и сложной альтернативы H_1 .

Теорема 2.5. Пусть определены простая гипотеза $H_0 : \theta = \theta_0$, где $\theta_0 = (\theta_{0i}) \in \Theta$ — фиксированная точка t -мерной области Θ , и сложная альтернатива $H_1 : \theta \neq \theta_0$, тогда при выполнении условий регулярности, принятых в теории оценок максимального правдоподобия, асимптотический ($n \rightarrow \infty$) размер КОП (2.20), в котором $\delta = F_{\chi_m^2}^{-1}(1 - \varepsilon)$ — квантиль уровня $1 - \varepsilon$ от χ^2 -распределения с t степенями свободы, совпадает с ε :

$$\mathbf{P}_{\theta_0}\{d(X) = 1\} = \mathbf{P}_{\theta_0}\{-2 \ln \lambda_n(X; \theta_0) \geq \delta\} \rightarrow \varepsilon. \quad (2.22)$$

Доказательство. В силу (2.21) для доказательства (2.22) достаточно показать сходимость к χ^2 -распределению:

$$\mathcal{L}_{H_0}\{-2 \ln \lambda_n(X; \theta_0)\} \rightarrow \chi_m^2, \quad n \rightarrow \infty,$$

где из (2.19) следует

$$-2 \ln \lambda_n(X; \theta_0) = 2 \left(l(\hat{\theta}) - l(\theta_0) \right), \quad (2.23)$$

$l(\theta) = \ln p(X; \theta)$ — логарифмическая функция правдоподобия, а $\hat{\theta} = \arg \max_{\theta \in \Theta} p(X; \theta)$ — ОМП, которая в условиях регулярности обладает свойством сильной состоятельности:

$$\hat{\theta} \xrightarrow{\text{П.Н.}} \theta^0, \quad n \rightarrow \infty.$$

Следовательно, $\forall \varepsilon > 0 \exists \bar{n} = \bar{n}(\varepsilon)$, что $\forall n \geq \bar{n}$ с вероятностью единица: $|\hat{\theta} - \theta_0| \leq \varepsilon$. В силу условий регулярности применим к $l(\theta_0)$ в (2.23) квадратичную формулу Тейлора с остаточным членом в форме Лагранжа (учитывая, что $\nabla_{\theta} l(\theta)|_{\theta=\hat{\theta}} = \mathbf{0}_m$):

$$-2 \ln \lambda_n(X; \theta_0) = n(\hat{\theta} - \theta_0)^T \left(-\frac{1}{n} \nabla_{\theta^*}^2 l(\theta^*) \right) (\hat{\theta} - \theta_0).$$

Здесь θ^* — «промежуточная точка»: $|\theta^* - \theta_0| < |\hat{\theta} - \theta_0|$, следовательно, $\theta^* \xrightarrow{\text{П.Н.}} \theta_0$. В силу условий регулярности выполняется усиленный закон больших чисел:

$$\begin{aligned} -\frac{1}{n} \nabla_{\theta^*}^2 l(\theta^*) &= \frac{1}{n} \sum_{k=1}^n \left(-\nabla_{\theta^*}^2 \ln p(x_k; \theta^*) \right) \xrightarrow{\text{П.Н.}} \mathbf{E}_{\theta_0} \left\{ -\nabla_{\theta_0}^2 \ln p(x_k; \theta_0) \right\} = \\ &= \mathcal{I}(\theta_0), \end{aligned}$$

где $\mathcal{I} = \mathcal{I}(\theta_0)$ — информационная матрица Фишера для выборочного значения. Из свойств ОМП

$$\mathcal{L}_{H_0}\{\sqrt{n}(\hat{\theta} - \theta_0)\} \rightarrow \mathcal{N}_m(\mathbf{0}_m, \mathbb{I}^{-1}(\theta_0)), \quad n \rightarrow \infty,$$

и по свойствам линейного преобразования многомерного нормального распределения:

$$\mathcal{L}_{H_0}\{\sqrt{n}(\mathcal{I}^{\frac{1}{2}}(\theta_0))^T(\hat{\theta} - \theta_0)\} = \mathcal{L}\{\xi_n\} \rightarrow \mathcal{N}_m(\mathbf{0}_m, \mathbf{I}_m) = \mathcal{L}\{\xi\}.$$

В результате имеем

$$\mathcal{L}_{H_0}\{-2 \ln \lambda_n(X; \theta_0)\} \rightarrow \mathcal{L}\{\xi^T \xi\} = \chi_m^2, \quad n \rightarrow \infty.$$

□

Замечание 2.4. Если H_0 является сложной гипотезой вида

$$H_0 : \theta \in \Theta_0, \quad \Theta_0 = \{\theta : h_j(\theta) = 0, \quad j = \overline{1, m-s}\},$$

где $1 \leq s \leq m-1$, $\{h_j(\cdot)\}_{j=0}^{m-s}$ — фиксированные непрерывные функции, то результат теоремы 2.5 остается в силе, только

$$\delta = F_{\chi_{m-s}^2}^{-1}(1 - \varepsilon).$$

2.5 Байесовское решающее правило

Предположим, что параметр θ – случайная величина, принимающая одно из двух возможных значений:

$$\theta \in \Theta = \{\theta_0, \theta_1\}; \mathbf{P}\{\theta = \theta_i\} = \Pi_i, \quad 0 < \Pi_i < 1, \quad i = 0, 1; \quad \Pi_0 + \Pi_1 = 1.$$

Наблюдается случайная выборка $X = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ объема n из некоторого распределения вероятностей с условной плотностью $p(x|\theta)$, $x \in \mathbb{R}^N$, $\theta \in \Theta$.

Обозначим:

$$p_i(X) = \prod_{j=1}^n p(x_j|\theta_i), \quad i = 0, 1, \quad -$$

условная плотность распределения выборки X при условии, что $\theta = \theta_i$.

Истинное значение θ неизвестно и определены две простые гипотезы:

$$H_i : \theta = \theta_i, \quad i = 0, 1.$$

Задача заключается в построении теста для проверки H_0, H_1 по выборке X .

Эту задачу можно решить с помощью теста Неймана-Пирсона, но в таком случае игнорируется информация о $\{\Pi_i\}$. Поэтому рассмотрим другой подход.

Построим рандомизированное решающее правило:

$$d = d(X, \omega) \in D = \{0, 1\}, \quad X \in \mathbb{R}^{nN}, \quad \omega \in \Omega; \quad (2.24)$$

$$\mathbf{P}\{d(X, \omega) = 1|X\} = \varphi(X), \quad \mathbf{P}\{d(X, \omega) = 0|X\} = 1 - \varphi(X),$$

где $\varphi(X)$ – произвольная критическая функция ($0 \leq \varphi(X) \leq 1$).

Обозначим: $\nu = \nu(\omega) \in \{0, 1\}$ – случайная величина Бернулли – номер истинной гипотезы H_ν . В силу случайности θ

$$\mathbf{P}(H_i) = \mathbf{P}\{\theta = \theta_i\} = \Pi_i, \quad i = 0, 1,$$

поэтому Π_i принято называть *априорной вероятностью* i -й гипотезы.

Определение 2.18. Функцией потерь в рассматриваемой задаче проверки двух гипотез H_0, H_1 называется функция двух переменных:

$$w = w(i, j) \geq 0, \quad i, j \in D = \{0, 1\},$$

где $w(i, j)$ – величина потерь, которые несет статистик в ситуации, когда на самом деле $\nu = i$ (верна H_i), а принято решение $d = j$ в пользу гипотезы H_j .

Определение 2.19. Принято говорить, что имеет место $(0-1)$ -функция потерь, если

$$w(i, j) = 1 - \delta_{ij} = \begin{cases} 0, & i = j; \\ 1, & i \neq j. \end{cases}$$

Функцию потерь удобно задавать в виде матрицы потерь: $W = (w_{ij})$, $w_{ij} = w(i, j)$. В случае $(0-1)$ -матрицы потерь имеем

$$W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Определение 2.20. *Функционалом риска называется математическое ожидание случайных потерь (средние потери)*

$$r = r(\varphi(\cdot)) = \mathbf{E}\{w(\nu, d(X, \omega))\} \geq 0. \quad (2.25)$$

Определение 2.21 (байесовский принцип оптимальности). *Критическую функцию $\varphi(\cdot)$ в рандомизированном решающем правиле (2.24) надлежит выбирать таким образом, чтобы функционал риска (2.25) достигал минимального значения:*

$$r(\varphi^*(\cdot)) = \inf_{\varphi(\cdot)} r(\varphi(\cdot)). \quad (2.26)$$

При этом критическая функция $\varphi^*(\cdot)$, определяемая (2.26), называется байесовской критической функцией, а соответствующее решающее правило $d^*(X, \omega)$, определяемое (2.24), – байесовским решающим правилом (БРП).

Теорема 2.6. *Пусть в сформулированной выше задаче проверки простых гипотез H_0, H_1 функция потерь имеет следующий вид:*

$$w(i, j) = \begin{cases} 0, & i = j; \\ w_0, & i = 0, j = 1; \\ w_1, & i = 1, j = 0, \end{cases} \quad (2.27)$$

где $w_0 > 0, w_1 > 0$ – некоторые заданные величины.

Тогда байесовская критическая функция задается соотношением ($X \in \mathbb{R}^{nN}$):

$$\varphi^*(X) = \begin{cases} 0, & L(X) < C^*; \\ \kappa^*, & L(X) = C^*; \\ 1, & L(X) > C^*, \end{cases} \quad (2.28)$$

где

$$L(X) = \frac{p_1(X)}{p_0(X)} \geq 0, \quad C^* = \frac{\Pi_0 w_0}{\Pi_1 w_1} \geq 0, \quad \kappa^* \in [0, 1]. \quad (2.29)$$

Доказательство. Прежде всего заметим, что в силу произвола в выборе κ^* байесовская критическая функция неоднозначна. Сформулируем экстремальную задачу (2.26) в явном виде и решим ее. Из (2.25) и (2.27) имеем

$$\begin{aligned} r &= r(\varphi(\cdot)) = \\ &= 0 \cdot \mathbf{P}\{d(X, \omega) = \nu\} + w_0 \cdot \mathbf{P}\{\nu = 0, d = 1\} + w_1 \cdot \mathbf{P}\{\nu = 1, d = 0\} = \\ &= w_0 \cdot \mathbf{P}\{\nu = 0\} \cdot \mathbf{P}\{d = 1|H_0\} + w_1 \cdot \mathbf{P}\{H_1\} \cdot \mathbf{P}\{d = 0|H_1\} = \\ &= w_0 \cdot \Pi_0 \cdot \alpha + w_1 \cdot \Pi_1 \cdot \beta. \end{aligned}$$

Используя интегральные представления вероятностей ошибок α, β из § 2.2, приходим к задаче:

$$\begin{aligned} r &= r(\varphi(\cdot)) = \\ &= \Pi_0 w_0 \int_{\mathbb{R}^{nN}} \varphi(X) p_0(X) dX + \Pi_1 w_1 \int_{\mathbb{R}^{nN}} (1 - \varphi(X)) p_1(X) dX = \\ &= \Pi_1 w_1 - \int_{\mathbb{R}^{nN}} \varphi(X) G(X) dX \rightarrow \min_{\varphi(\cdot)}, \end{aligned} \quad (2.30)$$

где

$$G(X) = \Pi_1 w_1 p_1(X) - \Pi_0 w_0 p_0(X).$$

Из (2.30) получаем эквивалентную экстремальную задачу:

$$\int_{\mathbb{R}^{nN}} \varphi(X) G(X) dX \rightarrow \max_{\varphi(\cdot)},$$

являющуюся линейной задачей вариационного исчисления, причем ограничения носят точечный, локальный характер: $0 \leq \varphi(X) \leq 1$, $X \in \mathbb{R}^{nN}$, поэтому решение очевидно:

$$\varphi^*(X) = \begin{cases} 0, & G(X) < 0; \\ \varkappa^*, & G(X) = 0; \\ 1, & G(X) > 0. \end{cases}$$

С учетом обозначений (2.29) последнее соотношение эквивалентно (2.28). Найдем минимум риска, который при этом достигается:

$$r^* = r(\varphi^*) = \Pi_1 w_1 - \int_{G(X) > 0} G(X) dX.$$

□

Следствие 2.6.1. Среди байесовских решающих правил (2.28) существует нерандомизированное решающее правило:

$$d = d^*(X) = \begin{cases} 0, & L(X) < C^*; \\ 1, & L(X) \geq C^*. \end{cases}$$

Доказательство. Для доказательства достаточно выбрать $\varkappa^* = 1$.

□

Следствие 2.6.2. Если имеет место $(0-1)$ -функция потерь, то есть $w_0 = w_1 = 1$, и гипотезы H_0, H_1 равновероятны ($\Pi_0 = \Pi_1 = \frac{1}{2}$), то БРП имеет вид

$$d = d^*(X) = \begin{cases} 0, & p_0(X) > p_1(X); \\ 1, & p_1(X) \geq p_0(X). \end{cases} \quad (2.31)$$

Доказательство. Из формулы (2.29) имеем: $C^* = 1, \varkappa^* = 1$.

□

Заметим в заключение, что решающее правило (2.31) часто называют *тестом максимального правдоподобия*.

2.6 Понятие о последовательном анализе Вальда

Ранее рассмотренные статистические критерии Неймана – Пирсона, Байеса, χ -квадрат, Колмогорова, отношения правдоподобия характеризуются общим свойством: объем n наблюдаемой выборки $X_n = (x_1^T, \dots, x_n^T)^T \in \mathbb{R}^{nN}$ фиксируется априори. Такая ситуация возникает при обработке ранее полученных (архивных) экспериментальных данных, в так называемых пассивных экспериментах. Однако на практике часто встречаются ситуации, когда проверку гипотез необходимо осуществлять в ходе активного эксперимента и решать вопрос о его продолжении или прекращении (в режиме «on-line»). В 1950 г. американский статистик А. Вальд выдвинул идею о том, что априорно фиксировать объем выборки n необязательно. Целесообразно определять необходимый объем выборки в ходе самого эксперимента, так что сам объем выборки становится случайной величиной, зависящей от

экспериментальных данных: $n = n(X)$. Методы проверки гипотез, основанные на этой идее, получили название *последовательного анализа*. Рассмотрим простейшую модель последовательного анализа Вальда.

Пусть определена случайная последовательность наблюдений $x_1, x_2, \dots, x_n, \dots \in \mathbb{R}^N$ – независимых в совокупности одинаково распределенных случайных векторов, имеющих плотность $p(x; \theta)$, $x \in \mathbb{R}^N$, где $\theta \in \Theta \in \{\theta_0, \theta_1\}$ – неизвестное истинное значение параметра, принимающего одно из двух различных возможных значений. Определены две простые гипотезы:

$$\begin{aligned} H_0 &: \theta = \theta_0; \\ H_1 &: \theta = \theta_1. \end{aligned}$$

Обозначим: $X_m = (x_1^T, \dots, x_m^T)^T \in \mathbb{R}^{mN}$ – составная выборка, полученная после проведения m экспериментов ($m = 1, 2, \dots$); $d_m \in D = \{0, 1, 2\}$ – возможное решение после проведения m экспериментов.

Решение $d_m = 0(1)$ означает, что по результатам m экспериментов X_m принимается гипотеза $H_0(H_1)$, а решение $d_m = 2$ – что для вынесения решения об истинности H_0 или H_1 данных, содержащихся в X_m , недостаточно, и необходимо осуществить $(m + 1)$ -й эксперимент.

Процесс увеличения числа экспериментов m продолжается до тех пор, пока для некоторого случайного n не будет принято «окончательное» решение $d_n \in \{0, 1\}$. Задать последовательное решающее правило (*последовательный критерий*) – значит задать последовательность решающих правил:

$$d_m = d_m(X_m), \quad X_m \in \mathbb{R}^{mN}, \quad d_m \in \{0, 1, 2\}, \quad m = 1, 2, \dots$$

Точность последовательного критерия характеризуют условные вероятности ошибок I и II рода:

$$\alpha = \mathbf{P}_{\theta_0}\{d_n = 1\}, \quad \beta = \mathbf{P}_{\theta_1}\{d_n = 0\},$$

а также условные математические ожидания объемов выборок (средние длительности экспериментов):

$$W_i = \mathbf{E}_{\theta_i}\{n\}, \quad i = 0, 1.$$

Определение 2.22 (принцип оптимальности Вальда). *Последовательное решающее правило $\{d_m = d_m(X_m)\}$ надлежит выбирать таким образом, чтобы вероятности ошибок I и II рода α и β не превосходили соответственно наперед заданных достаточно малых уровней α_0 и β_0 , а условные средние объемы выборок W_0, W_1 были бы минимальны.*

Этому принципу, оказывается, удовлетворяет последовательный критерий отношения правдоподобия Вальда следующего вида:

$$d_m = d_m(X_m) = \begin{cases} 0, & L_m(X_m) < B; \\ 2, & B \leq L_m(X_m) < A; \\ 1, & L_m(X_m) \geq A, \end{cases} \quad (2.32)$$

где

$$L_m(X_m) = \frac{p(X_m; \theta_1)}{p(X_m; \theta_0)} - \quad (2.33)$$

статистика отношения правдоподобия;

$$p(X_m; \theta_i) = \prod_{j=1}^m p(x_j; \theta_i), \quad i = 0, 1, - \quad (2.34)$$

плотность распределения случайной выборки X_m при $\theta = \theta_i$ ($i = 0, 1$), а $A, B \in \mathbb{R}$ ($B < A$) – некоторые, пока не определенные, числовые параметры правила, называемые соответственно *верхним порогом* и *нижним порогом* и выбираемые из принципа оптимальности Вальда.

Решающее правило Вальда (2.32)–(2.34) допускает эквивалентное представление, удобное для дальнейшего анализа:

$$d_m = d_m(Z_m) = \begin{cases} 0, & Z_m < \ln B; \\ 2, & \ln B \leq Z_m < \ln A; \\ 1, & Z_m \geq \ln A, \end{cases} \quad (2.35)$$

где

$$Z_m = \sum_{j=1}^m z_j, \quad z_j = \ln \frac{p(x_j; \theta_1)}{p(x_j; \theta_0)}.$$

Теорема 2.7. Пусть конечны условные дисперсии:

$$0 < \sigma_i^2 = \mathbf{D}_{\theta_i} \left\{ \ln \frac{p(x_j; \theta_1)}{p(x_j; \theta_0)} \right\} < \infty, \quad i = 0, 1,$$

и

$$|\ln A| < \infty, \quad |\ln B| < \infty,$$

тогда последовательный критерий Вальда (2.35) представляет собой шаговый процесс, заканчивающийся с вероятностью единица принятием одной из гипотез H_0 или H_1 .

Доказательство. Пусть на самом деле верна гипотеза H_i ($i \in \{0, 1\}$). Тогда по условию теоремы существует условное математическое ожидание, определяемое с учетом (2.35):

$$\mu_i = \mathbf{E}_{\theta_i} \{z_j\} = \int_{\mathbb{R}^N} p(x; \theta_i) \ln \frac{p(x; \theta_1)}{p(x; \theta_0)} dx,$$

причем в силу свойства функции Кульбака: $\mu_1 > 0$, $\mu_0 < 0$. К сумме Z_m при $m \rightarrow \infty$ применима ЦПТ Леви – Линдберга:

$$\mathbf{P}_{\theta_i} \left\{ \frac{Z_m - m\mu_i}{\sqrt{m}\sigma_i} < y \right\} \rightarrow \Phi(y), \quad y \in \mathbb{R}, \quad (2.36)$$

где $\Phi(\cdot)$ – функция распределения вероятностей закона $\mathcal{N}_1(0, 1)$.

Исследуем асимптотику для вероятности \bar{Q}_m «незавершения шагового процесса принятия решений за m шагов». Согласно (2.35):

$$\bar{Q}_m = \mathbf{P}_{\theta_i} \{ \ln B \leq Z_1 < \ln A, \ln B \leq Z_2 < \ln A, \dots, \ln B \leq Z_m < \ln A \}.$$

В силу свойства вероятности

$$\mathbf{P} \left\{ \bigcap_{k=1}^m C_k \right\} \leq \mathbf{P} \{C_l\}, \quad l \in \{1, \dots, m\},$$

имеем

$$\bar{Q}_m \leq \mathbf{P}_{\theta_i} \{ \ln B \leq Z_m < \ln A \},$$

и вероятность Q_m «завершения шагового процесса за m шагов» оценивается с учетом (2.36) снизу следующим образом:

$$Q_m = 1 - \bar{Q}_m \geq 1 - \mathbf{P}_{\theta_i} \{ \ln B \leq Z_m < \ln A \} =$$

$$\begin{aligned}
&= 1 - \mathbf{P}_{\theta_i} \left\{ \frac{\ln B - m\mu_i}{\sqrt{m}\sigma_i} \leq \frac{Z_m - m\mu_i}{\sqrt{m}\sigma_i} < \frac{\ln A - m\mu_i}{\sqrt{m}\sigma_i} \right\} = \\
&= 1 - \mathbf{P}_{\theta_i} \left\{ \frac{1}{\sqrt{m}} \cdot \frac{\ln B}{\sigma_i} - \sqrt{m} \cdot \frac{\mu_i}{\sigma_i} \leq Z_{mi}^0 < \frac{1}{\sqrt{m}} \cdot \frac{\ln A}{\sigma_i} - \sqrt{m} \frac{\mu_i}{\sigma_i} \right\} \rightarrow 1.
\end{aligned}$$

□

Важным вопросом является выбор пороговых значений A, B в (2.35). Можно показать, что если выбирать их по принципу Вальда, то

$$A \leq \frac{1 - \beta_0}{\alpha_0}, \quad B \geq \frac{\beta_0}{1 - \alpha_0}.$$

При этом если берется знак «=», то $\alpha + \beta \leq \alpha_0 + \beta_0$. Заметим, что точные выражения $A = A(\alpha_0, \beta_0)$, $B = B(\alpha_0, \beta_0)$ получить не удастся. Условные средние объемы выборок правила Вальда при таком упрощенном выборе A, B определяются следующими соотношениями:

$$\begin{aligned}
W_0 &= \frac{(1 - \alpha_0) \ln \frac{\beta_0}{1 - \alpha_0} + \alpha_0 \ln \frac{1 - \beta_0}{\alpha_0}}{\mu_0}; \\
W_1 &= \frac{\beta_0 \ln \frac{\beta_0}{1 - \alpha_0} + (1 - \beta_0) \ln \frac{1 - \beta_0}{\alpha_0}}{\mu_1}.
\end{aligned}$$