Лекция 2. Статистики первого типа. Точеченые оценки и их свойства

Буре В.М., Грауэр Л.В.

ШАД

Санкт-Петербург, 2013

Содержание

Статистики первого типа

Следуя [3] будем рассматривать функционал $G(F)$, заданный на множестве функций распределения:

$$G(F) = h\left(\int_{-\infty}^{\infty} g(x)dF(x)\right),$$

где $g : \mathbb{R} \to \mathbb{R}^m$ — заданная борелевская функция,
$h : \mathbb{R}^m \to \mathbb{R}^l$ некоторая борелевская функция, непрерывная в точке $a = \int_{-\infty}^{\infty} g(x)dF_\xi(x) \in \mathbb{R}^m$.
Тогда назовем статистику $S(X_{[n]}) = G(F_n^*)$ *статистикой первого типа*.
Таким образом,

$$S(X_{[n]}) = G(F_n^*) = h\left(\int_{-\infty}^{\infty} g(x)dF_n^*(x)\right) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right).$$

Замечание 1

Положим $h(t) \equiv t$, $g(x) = x^k$, *то* $G(F_n^*) = a_k^*$, *значит, начальный эмпирический момент порядка* k *является статистикой первого типа.*

Нетрудно заметить, что центральный эмпирический момент также является статистикой первого типа.

Покажем это для дисперсии (центрального момента второго порядка):

$$s^2 = a_2^{0*} = \frac{1}{n}\sum_{i=1}^{n}(X_i - a_1^*)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2,$$

$$a_2^0 = \int_{-\infty}^{+\infty}(x - a_1)^2 dF_\xi(x) = \int_{-\infty}^{+\infty} x^2 dF_\xi(x) - \left(\int_{-\infty}^{+\infty} x dF_\xi(x)\right)^2.$$

Выберем функции h и g следующего вида:

$$h(t_1, t_2) = t_2 - t_1^2,$$

$$g(x) = (g_1(x), g_2(x)) = (x, x^2).$$

Тогда получаем равенства:

$$a_2^0 = h\left(\int\limits_{-\infty}^{+\infty} g(x)dF_\xi(x)\right),$$

$$a_2^{0*} = h\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right).$$

Теорема 1

Пусть $S(X_{[n]}) = G(F_n^*)$ — статистика первого типа, тогда имеет место сходимость:

$$S(X_{[n]}) \xrightarrow[n\to\infty]{\text{п.н.}} h(a) = G(F_\xi).$$

Доказательство

Из усиленного закона больших чисел Колмогорова следует сходимость:

$$\frac{1}{n}\sum_{i=1}^{n} g(x_i) \xrightarrow[n\to\infty]{\text{п.н.}} Eg(\xi) = \int\limits_{-\infty}^{+\infty} g(x)dF_\xi(x),$$

так как все слагаемые взаимно независимые, одинаково распределенные случайные величины.

Функция $h(\cdot)$ непрерывна в точке a, следовательно,

$$G(F_n^*) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right) \xrightarrow[n\to\infty]{\text{п.н.}} h\left(\int_{-\infty}^{+\infty} g(x)dF_\xi(x)\right).$$

Следствие 1

Пусть у генеральной совокупности ξ существует теоретический начальный момент порядка k, $a_k = E\xi^k \in R$, тогда

$$a_k^* \xrightarrow[n\to\infty]{п.н.} a_k.$$

Следствие 2

Пусть у генеральной совокупности ξ существует теоретический центральный момент порядка k, $a_k^0 = E(\xi - E\xi)^k \in R$, тогда

$$a_k^{0*} \xrightarrow[n\to\infty]{п.н.} a_k^0.$$

Рассмотрим случайный вектор $\xi = (\xi_1, \ldots, \xi_m)^T$ с m компонентами. Пусть имеется выборка:

$$(X_1, X_2, \ldots, X_n) = \begin{pmatrix} X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & X_{22} & \ldots & X_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ X_{m1} & X_{m2} & \ldots & X_{nm} \end{pmatrix}.$$

Рассмотрим компоненты ξ_k и ξ_l, им соответствуют элементы выборки: $X_{k1}, \ldots, X_{kn}$ и $X_{l1}, \ldots, X_{ln}$.

Вычислим выборочные моменты:

$$a_{1k}^* = \bar{X}_k = \frac{1}{n}\sum_{i=1}^{n} X_{ki} \xrightarrow[n\to\infty]{\text{п.н.}} E\xi_k,$$

$$a_{1l}^* = \bar{X}_l = \frac{1}{n}\sum_{i=1}^{n} X_{li} \xrightarrow[n\to\infty]{\text{п.н.}} E\xi_l.$$

Кроме того,

$$a_{2k}^* = s_k^2 = \frac{1}{n}\sum_{i=1}^{n} (X_{ki} - \bar{X}_k)^2 \xrightarrow[n\to\infty]{\text{п.н.}} D\xi_k,$$

$$a_{2l}^* = s_l^2 = \frac{1}{n}\sum_{i=1}^{n} (X_{li} - \bar{X}_l)^2 \xrightarrow[n\to\infty]{\text{п.н.}} D\xi_l.$$

Коэффициент корреляции равен

$$\varrho(\xi_k, \xi_l) = \frac{cov(\xi_k, \xi_l)}{\sqrt{D\xi_k}\sqrt{D\xi_l}},$$

где

$$cov(\xi_k, \xi_l) = E\{(\xi_k - E\xi_k)(\xi_l - E\xi_l)\} = E(\xi_k \xi_l) - E(\xi_k)E(\xi_l).$$

Подберем эмпирический аналог ковариации. Рассмотрим элементы выборки:

$$\begin{pmatrix} X_{k1} \\ X_{l1} \end{pmatrix}, \begin{pmatrix} X_{k2} \\ X_{l2} \end{pmatrix}, \ldots, \begin{pmatrix} X_{kn} \\ X_{ln} \end{pmatrix}.$$

Надо центрировать величины:

$$\begin{pmatrix} X_{k1} - \bar{X}_k \\ X_{l1} - \bar{X}_l \end{pmatrix}, \begin{pmatrix} X_{k2} - \bar{X}_k \\ X_{l2} - \bar{X}_l \end{pmatrix}, \dots, \begin{pmatrix} X_{kn} - \bar{X}_k \\ X_{ln} - \bar{X}_l \end{pmatrix}.$$

Перемножив их, получаем выборочную ковариацию:

$$\widehat{cov}(\xi_k, \xi_l) = \frac{1}{n} \sum_{i=1}^{n} (X_{ki} - \bar{X}_k)(X_{li} - \bar{X}_l) =$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_{ki} X_{li} - \bar{X}_k \bar{X}_l \xrightarrow[n \to \infty]{\text{п.н.}} cov(\xi_k, \xi_l).$$

Так как имеют место сходимости:

$$\frac{1}{n}\sum_{i=1}^{n} X_{ki}X_{li} \xrightarrow[n\to\infty]{\text{п.н.}} E(\xi_k\xi_l),$$

$$\bar{X}_k\bar{X}_l \xrightarrow[n\to\infty]{\text{п.н.}} E\xi_k E\xi_l,$$

тогда имеет место сходимость почти наверное для выборочного коэффициента корреляции:

$$\hat{\varrho}(\xi_k\xi_l) = \frac{\widehat{cov}(\xi_k, \xi_l)}{\sqrt{s_k^2 s_l^2}} \xrightarrow[n\to\infty]{\text{п.н.}} \varrho(\xi_k\xi_l).$$

Напомним, что под одномерной выборкой понимаем взаимно независимые случайные величины, распределенные так же, как генеральная совокупность.

Теоремы непрерывности

Теорема 2

Пусть η — случайная величина, заданная на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, последовательность случайных величин $\{\eta_n\}$ также задана на $(\Omega, \mathcal{F}, P)$. Пусть борелевская функция $H : \mathbb{R} \longrightarrow \mathbb{R}$ непрерывна на борелевском множестве $B \in \mathcal{B}(\mathbb{R})$, $P\{\eta \in B\} = 1$, тогда справедливы утверждения:

1. *Если* $\{\eta_n\} \xrightarrow[n\to\infty]{\text{п.н.}} \eta$, *тогда* $H(\eta_n) \xrightarrow[n\to\infty]{\text{п.н.}} H(\eta)$.
2. *Если* $\{\eta_n\} \xrightarrow[n\longrightarrow\infty]{p} \eta$, *тогда* $H(\eta_n) \xrightarrow[n\longrightarrow\infty]{p} H(\eta)$.

Доказательство

Докажем первое утверждение. Определим множество A следующим образом:

$$A = \{\omega : \eta_n(\omega) \xrightarrow[n \longrightarrow \infty]{} \eta(\omega)\}.$$

По условию первого утверждения $P(A) = 1$. Рассмотрим множество $A \cap \eta^{-1}(B)$. Определим вероятность:

$$P(\overline{A \cap \eta^{-1}(B)}) = P(\overline{A} \cup \overline{\eta^{-1}(B)} \leq P(\overline{A}) + P(\overline{\eta^{-1}(B)}) = 0,$$

тогда $P(A \cap \eta^{-1}(B)) = 1$.
Теперь рассмотрим событие $\omega \in A \cap \eta^{-1}(B)$. Тогда

$$\eta_n(\omega) \xrightarrow[n \longrightarrow \infty]{} \eta(\omega) \in B.$$

Следовательно, имеет место сходимость:

$$H(\eta_n(\omega)) \xrightarrow[n \longrightarrow \infty]{} H(\eta(\omega)).$$

Докажем второе утверждение теоремы от противного.
Пусть утверждение 2 неверно. Это означает, что существует $\varepsilon > 0$, существует $\delta > 0$, и существует последовательность $\{n_k^{(1)}\}$, что справедливо неравенство

$$P\{|H(\eta_{n_k^{(1)}}) - H(\eta)| > \varepsilon\} > \delta.$$

Последовательность $\eta_{n_k^{(1)}}$ сходится по вероятности к η, $\eta_{n_k^{(1)}} \xrightarrow[k \longrightarrow \infty]{P} \eta$, следовательно, существует подпоследовательность $\{n_k^{(2)}\} \subset \{n_k^{(1)}\}$, для которой имеет место сходимость почти наверно:

$$\eta_{n_k^{(2)}} \xrightarrow[k \longrightarrow \infty]{\text{п.н}} \eta.$$

Возникло противоречие, так как $\eta_{n_k^{(2)}} \xrightarrow[k \longrightarrow \infty]{\text{п.н}} \eta$, но с другой стороны

$$P\{|H(\eta_{n_k^{(2)}}) - H(\eta)| > \varepsilon\} > \delta,$$

отсюда следует, что утверждение 2 верно.

Теорема 3

Пусть для последовательности случайных векторов имеет место сходимость по распределению:
$\eta_n = (\eta_n^{(1)}, \ldots, \eta_n^{(m)}) \xrightarrow[n \longrightarrow \infty]{d} \eta = (\eta^{(1)}, \ldots, \eta^{(m)})$. *Пусть задана функция* $H : \mathbb{R}^m \longrightarrow \mathbb{R}^l$ *непрерывная на* $\mathbb{R}^m$, *тогда* $H(\eta_n) \xrightarrow[n \longrightarrow \infty]{d} H(\eta)$.

Доказательство Для любой непрерывной и ограниченной функции $g(x)$ должно выполняться следующее предельное соотношение:

$$Eg(H(\eta_n)) \xrightarrow[n \longrightarrow \infty]{} Eg(H(\eta)).$$

Рассмотрим $g(H(t)) \equiv h(t)$, то есть h — суперпозиция двух непрерывных функций, следовательно, она непрерывна. Так как g ограничена, то отсюда следует, что h ограничена, то есть выполнено следующее соотношение:

$$Eh(\eta_n) \xrightarrow[n \longrightarrow \infty]{} Eh(\eta).$$

Теорема 4

Пусть борелевская функция $H : \mathbb{R}^m \to \mathbb{R}^k$ *непрерывна на* $B \in B(\mathbb{R}^m)$ *и* $P\{\eta \in B\} = 1$. *Пусть* $\eta_n^T = (\eta_{1n}, \ldots, \eta_{mn}) \xrightarrow[n\to\infty]{d} \eta$, *тогда* $H(\eta_n) \xrightarrow[n\to\infty]{d} H(\eta)$.

Доказательство теоремы можно найти в книге [1].

Теорема 5

Пусть последовательность случайных величин η_n *сходится по распределению к случайной величине* η, $\eta_n \xrightarrow[n\longrightarrow\infty]{d} \eta$. *Пусть функция* $H : \mathbb{R} \to \mathbb{R}$ — *борелевская функция. Числовая последовательность* $b_n \xrightarrow[n\to\infty]{} 0$, *причем* $b_n \neq 0$ *для любого* n. *Тогда справедливы утверждения:*

1. *Если функция* H *дифференцируема в точке* $a \in \mathbb{R}$, *то*

$$\frac{H(a + b_n\eta_n) - H(a)}{b_n} \xrightarrow[n\longrightarrow\infty]{d} H'(a)\eta.$$

2. *Если функция* H *дифференцируема в некоторой окрестности точки* a, $H'(a) = 0$, *и существует* $H''(a)$, *то*

$$\frac{H(a + b_n\eta_n) - H(a)}{b_n^2} \xrightarrow[n\longrightarrow\infty]{d} \frac{1}{2}H''(a)\eta^2.$$

Доказательство

Докажем утверждение 1. Введем $H_1(x)$ и $\tilde{H}_1(x,y) = H_1(x)y$, причем

$$H_1(x) = \begin{cases} \frac{H(a+x)-H(a)}{x}, & x \neq 0, \\ H'(a), & x = 0. \end{cases}$$

Из условия теоремы ясно, что $H_1(x)$ непрерывна в нуле, тогда $\tilde{H}_1(x,y)$ непрерывна на множестве $\{(0,y), y \in \mathbb{R}\}$. Покажем, что $(b_n\eta_n, \eta_n) \xrightarrow[n\to\infty]{\mathrm{d}} (0,\eta)$. Воспользуемся методом характеристических функций.

Надо показать, что

$$\varphi_{(b_n\eta_n,\eta_n)}(t_1,t_2) \xrightarrow[n\longrightarrow\infty]{} \varphi_{(0,\eta)}(t_1,t_2). \qquad (1)$$

Рассмотрим левую часть (1):

$$\varphi_{(b_n\eta_n,\eta_n)}(t_1,t_2) = Ee^{i(t_1 b_n\eta_n + t_2\eta_n)} = Ee^{i(t_1 b_n + t_2)\eta_n} = \varphi_{\eta_n}(b_n t_1 + t_2)\pm$$
$$\pm\varphi_\eta(b_n t_1 + t_2) = (\varphi_{\eta_n}(b_n t_1 + t_2) - \varphi_\eta(b_n t_1 + t_2)) + \varphi_\eta(b_n t_1 + t_2) \xrightarrow[n\to\infty]{}$$
$$\xrightarrow[n\to\infty]{} \varphi_\eta(t_2) = Ee^{it_2\eta + it_1 0} = \varphi_{(0,\eta)}(t_1,t_2),$$

так как t_1, t_2 фиксированы, то $(b_n t_1 + t_2) \in [a,b]$, то есть, сходимость на данном промежутке равномерная, тогда
$(\varphi_{\eta_n}(b_n t_1 + t_2) - \varphi_\eta(b_n t_1 + t_2)) \xrightarrow[n\to\infty]{} 0$.
Таким образом, показали, что

$$\varphi_{(b_n\eta_n,\eta_n)}(t_1,t_2) \xrightarrow[n\longrightarrow\infty]{} \varphi_{(0,\eta)}(t_1,t_2).$$

Заметим, что $P\{(0,\eta) \in \{(0,y) : y \in R\}\} = 1$.

Подставим в $\tilde{H}_1$ случайный вектор $(b_n\eta_n, \eta_n)$, получим:

$$\tilde{H}_1(b_n\eta_n, \eta_n) = \frac{H(a + b_n\eta_n) - H(a)}{b_n},$$

$$\tilde{H}_1(b_n\eta_n, \eta_n) \xrightarrow[n\to\infty]{d} \tilde{H}_1(0, \eta) = H'(a)\eta$$

как следует из теоремы 4.

Докажем второе утверждение теоремы. Запишем формулу Тейлора второго порядка:

$$H(a + x) = H(a) + \frac{1}{2}H''(a)x^2 + o(x^2).$$

Введем функцию:

$$H_2(x) = \begin{cases} \frac{H(a+x)-H(a)}{x^2}, & x \neq 0, \\ \frac{1}{2}H''(a), & x = 0. \end{cases}$$

Введем функцию $\tilde{H}_2(x,y) = H_2(x)y^2$, функция непрерывна на множестве $\{(0,y) : y \in \mathbb{R}\}$. Подставим в $\tilde{H}_2$ случайный вектор $(b_n\eta_n, \eta_n)$:

$$\tilde{H}_2(b_n\eta_n, \eta_n) \xrightarrow[n\longrightarrow\infty]{d} \tilde{H}_2(0,\eta) = \frac{1}{2}H''(a)\eta^2,$$

$$\tilde{H}_2(b_n\eta_n, \eta_n) = \frac{H(a+b_n\eta_n) - H(a)}{b_n^2}.$$

Утверждение 2 доказано.

Замечание 2

Теорема 5 допускает обобщение на многомерный случай.

Теорема 6

Пусть задана функция $H : \mathbb{R}^m \to \mathbb{R}$ *и последовательность случайных векторов* $\eta_n = (\eta_n^{(1)}, \ldots, \eta_n^{(m)}) \xrightarrow[n \longrightarrow \infty]{d} \eta = (\eta^{(1)}, \ldots, \eta^{(m)})$. *Пусть числовая последовательность* $b_n \xrightarrow[n \to \infty]{} 0$ *такая, что* $b_n \neq 0$ *для любого* n. *Тогда справедливы утверждения:*

1. *Если существует* $\frac{\partial H}{\partial t}\Big|_{t=a} = \left(\frac{\partial H}{\partial t_1}, \ldots, \frac{\partial H}{\partial t_m}\right)\Big|_{t=a}$, $a \in \mathbb{R}^m$, *то*
$$\frac{H(a + b_n \eta_n) - H(a)}{b_n} \xrightarrow[n \to \infty]{d} H'(a)\eta^T.$$
2. *Если* $H'(a) = 0$, *и существует* $H''(a) = \left(\frac{\partial^2 H}{\partial t_i \partial t_j}\right)\Big|_{t=a}$, *то*
$$\frac{H(a + b_n \eta_n) - H(a)}{b_n^2} \xrightarrow[n \longrightarrow \infty]{d} \frac{1}{2}\eta H''(a)\eta^T.$$

Можно продолжить обобщение теоремы 6 на случай $H : \mathbb{R}^m \to \mathbb{R}^l$.

Предельное распределение статистик первого типа

Пусть задана статистика первого типа:

$$S(X_{[n]}) = G(F_n^*) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right),$$

где $G(F) = h\left(\int_{-\infty}^{\infty} g(x)dF(x)\right)$, $\int_{-\infty}^{\infty} g(x)dF_\xi(x) = a \in \mathbb{R}^m$, т. е. $G(F_\xi) = h(a)$.

Теорема 7

Пусть имеется выборка $X_{[n]}$ *из генеральной совокупности* ξ *с функцией распределения* F_ξ *и* $S(X_{[n]}) = h\left(\frac{1}{n}\sum\limits_{i=1}^{n} g(X_i)\right)$ *— статистика I типа, борелевские функции* $h: \mathbb{R} \longrightarrow \mathbb{R}$, $g: \mathbb{R} \longrightarrow \mathbb{R}$, *тогда справедливы утверждения:*

1. *Если существует* $h'(a)$, *то*

$$\sqrt{n}\left(S(X_{[n]}) - h(a)\right) \xrightarrow[n\longrightarrow\infty]{d} h'(a)\zeta,$$

где ζ *— случайная величина, распределенная нормально с параметрами* $(0, Dg(\xi))$, $\zeta \sim N(0, Dg(\xi))$.

2. *Если* $h'(a) = 0$ *и существует* $h''(a)$, *то*

$$n\left(S(X_{[n]}) - h(a)\right) \xrightarrow[n\longrightarrow\infty]{d} \frac{1}{2}h''(a)\zeta^2.$$

Доказательство

Применима теорема 5. Преобразуем функцию $G(F^*)$:

$$G(F^*) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right) = h\left(a + \frac{1}{n}\sum_{i=1}^{n}(g(X_i) - a)\right) =$$

$$= h\left(a + \frac{1}{\sqrt{n}\sqrt{n}}\sum_{i=1}^{n}(g(X_i) - a)\right),$$

где $a = Eg(X_1)$.

По центральной предельной теореме для одинаково распределенных слагаемых, справедливо:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(X_i) - a) = \xi_n \xrightarrow[n\longrightarrow\infty]{\mathrm{d}} \zeta \sim N(0, \sigma^2),$$

где $\sigma^2 = Dg(\xi)$. Заметим, что $\frac{1}{\sqrt{n}} = b_n$.

Теорема 8

Пусть задана статистика I типа $S(X_{[n]}) = h\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right)$ *и борелевские функции* $h : \mathbb{R}^m \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}^m$, *тогда справедливы утверждения:*

1. *Если существует* $h'(a) = \left(\frac{\partial h}{\partial t_1}, \ldots, \frac{\partial h}{\partial t_m}\right)\Big|_{t=a}$, *где* $a = Eg(\xi) = (Eg_1(\xi), \ldots, Eg_m(\xi))$, *то*

$$\sqrt{n}\left(S(X_{[n]}) - h(a)\right) \xrightarrow[n\to\infty]{d} h'(a)\zeta^T,$$

где случайный вектор $\zeta = (\zeta_1, \ldots, \zeta_m)$ *подчиняется многомерному нормальному распределению с параметрами* $(0, Dg(\xi))$, $\zeta \sim N(0, Dg(\xi))$.

2. *Если* $h'(a) = 0$ *и существует* $h''(a)$, *то*

$$n\left(S(X_{[n]}) - h(a)\right) \xrightarrow[n\longrightarrow\infty]{d} \frac{1}{2}\zeta h''(a)\zeta^T.$$

Доказательство проводится аналогично доказательству теоремы 7. Применяется центральная предельная теорема для одинаково распределенных случайных векторов.

Пример

Рассмотрим генеральную совокупность ξ, для которой $E\xi = \alpha > 0$, $D\xi = \sigma^2$. Получена выборка $X_{[n]} = (X_1, \ldots, X_n)$ из ξ. Найдем асимптотическое распределение статистики $1/\bar{X}$, где $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = a_1^*$. Покажем, что $1/\bar{X}$ — статистика первого типа. Для доказательства достаточно взять $h(t) = 1/t$, $g(x) = x$ и заметить, что

$$1/\bar{X} = h\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right).$$

Очевидно, что

$$a = \int\limits_{-\infty}^{+\infty} g(x) dF_{\xi}(x) = Eg(\xi) = E\xi = \alpha,$$

при этом, $h(\alpha) = 1/\alpha$, $1/\bar{X} = h(\frac{1}{n}\sum_{i=1}^{n} g(x_i))$ является статистикой первого типа.

Из теоремы 7 следует, что

$$\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\alpha}\right) \xrightarrow[n\to\infty]{\mathrm{d}} \zeta\left(\frac{-1}{\alpha^2}\right) = -\zeta\frac{1}{\alpha^2},$$

где $\zeta \sim N(0, \sigma^2)$.

Точеченые оценки

Рассмотрим выборку $X_{[n]} = (X_1, \dots, X_n)$, генеральную совокупность ξ и ее функцию распределения $F_\xi(x, \theta)$, где $\theta = (\theta_1, \dots, \theta_m)$ — неизвестные параметры в распределении случайной величины ξ. По имеющейся выборке можно построить оценку для этих параметров.

Определение 1

Пусть $\theta \in \Theta \subset \mathbb{R}$. Под оценкой понимается статистика $\hat{\theta}(X_{[n]})$ такая, что получившееся значение можно рассматривать как точечную оценку параметра θ ($\hat{\theta}(X_{[n]}) \sim \theta$).

Невозможно найти численное значение вероятности $P\left\{\left|\hat{\theta}(X_{[n]}) - \theta\right| > \varepsilon\right\}$ для произвольного ε, так как вероятность содержит неизвестный параметр θ. Тогда какую оценку считать «хорошей»?

Свойства точечных оценок

1. **Несмещенность.**

Определение 2

Пусть параметр $\theta \in \Theta \subset \mathbb{R}$. *Говорят, что оценка* $\hat{\theta}(X_{[n]})$ *является несмещенной оценкой параметра* θ, *если*

$$E\hat{\theta}(X_{[n]}) = \theta \qquad (2)$$

для любого $\theta \in \Theta$.

Определение 3

Говорят, что оценка $\hat{\theta}(X_{[n]})$ *является асимптотически несмещенной оценкой параметра* θ, *если*

$$E\hat{\theta}(X_{[n]}) \xrightarrow[n\to\infty]{} \theta \qquad (3)$$

для любого $\theta \in \Theta$.

Замечание 3

Свойство несмещенности позволяет агрегировать информацию, накопленную в различных научных центрах. Рассмотрим следующий пример. Пусть $\hat{\theta}_1$ — несмещенная оценка параметра θ, полученная в некотором научном центре, $\hat{\theta}_2$ — несмещенная оценка того же параметра, полученная в другом научном центре. Предполагая, что техническая оснащенность научных центров одинаковая, будем считать, что дисперсии оценок одинаковы:

$$D(\hat{\theta}_i) = E(\hat{\theta}_i - \theta)^2 = \sigma^2(\theta),$$

$$E(\hat{\theta}_i) = \theta, \quad i = 1, 2.$$

Рассмотрим новую оценку:

$$\hat{\theta} = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}, \quad E\hat{\theta} = \frac{E\hat{\theta}_1 + E\hat{\theta}_2}{2} = \theta,$$

тогда имеют место равенства:

$$D\hat{\theta} = E(\hat{\theta} - \theta)^2 = \frac{1}{4} E\{(\hat{\theta}_1 - \theta) + (\hat{\theta}_2 - \theta)\}^2 = \frac{\sigma^2(\theta)}{2}.$$

Как видим, агрегированная оценка оказывается более точной, дисперсия уменьшилась в два раза.

Пример 1

Выборочное среднее является несмещенной оценкой для математического ожидания:

$$E\bar{X} = E\left\{\frac{1}{n}\sum_{k=1}^{n} X_k\right\} = \frac{1}{n}\sum_{k=1}^{n} EX_k = E\xi = a_1.$$

Рассмотрим выборочную дисперсию, проверим, выполнено ли свойство несмещенности:

$$Es^2 = E\left\{\frac{1}{n}\sum_{k=1}^{n}\left(X_k - a_1 - \frac{1}{n}\sum_{i=1}^{n}(X_i - a_1)\right)\right\}^2 =$$

$$= E\left\{\frac{1}{n}\sum_{k=1}^{n}(X_k - a_1)^2 - \left(\frac{1}{n}\sum_{k=1}^{n}(X_k - a_1)\right)^2\right\} =$$

$$= \frac{1}{n}\sum_{k=1}^{n}E(X_k - a_1)^2 - \frac{1}{n^2}\sum_{k=1}^{n}\sigma^2 = \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2,$$

при выводе формулы учитывалось следующее соотношение:

$$E\{(X_k - a_1)(X_l - a_1)\} = \sigma^2\delta_{kl} = \begin{cases} \sigma^2, & k = l; \\ 0, & k \neq l. \end{cases}$$

Таким образом, получаем равенство:

$$Es^2 = \frac{n-1}{n}\sigma^2.$$

Следовательно, s^2 — смещенная оценка, однако она является асимптотически несмещенной оценкой: $Es^2 \xrightarrow[n\to\infty]{} \sigma^2$.
Рассмотрим исправленную оценку дисперсии:

$$\tilde{s}^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2,$$

что доказывает, что $\tilde{s}^2$ — несмещенная оценка дисперсии.

2. **Состоятельность.**

Определение 4

Пусть параметр $\theta \in \Theta \subset \mathbb{R}$. Говорят, что оценка $\hat{\theta}(X_{[n]})$ состоятельна, если

$$\hat{\theta}(X_{[n]}) \xrightarrow[n\to\infty]{p} \theta \tag{4}$$

для любого $\theta \in \Theta$.

Определение 5

Оценка $\hat{\theta}(X_{[n]})$ называется сильно состоятельной оценкой параметра θ, если

$$\hat{\theta}(X_{[n]}) \xrightarrow[n\to\infty]{\text{п.н.}} \theta \tag{5}$$

для любого $\theta \in \Theta$.

Замечание 4

В случае, когда $\hat{\theta}(X_{[n]})$ — векторная оценка, свойство состоятельности и сильной состоятельности рассматриваются покомпонентно.

Замечание 5

В определении оценки предполагалось, что n фиксировано, но обычно под оценкой понимают некоторое правило, по которому можно построить оценку для любого n.

Замечание 6

Пусть существует $E\xi^k$, тогда a_k^* — статистика первого рода, где a_k^* — эмпирический момент порядка k. Тогда $a_k^* \to a_k = E\xi^k$, то есть, a_k^* является сильно состоятельной оценкой.

Пусть существует $E(\xi - E\xi)^k$, *тогда*

$$a_k^{0*} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

— сильно состоятельная оценка для теоретического момента, то есть:

$$a_k^{0*} \xrightarrow[n\to\infty]{\text{п.н.}} a_k^0 = E(\xi - E\xi)^k.$$

Пример 2

Выборочное среднее и выборочная дисперсия представляют собой сильно состоятельные оценки соответствующих числовых характеристик случайной величины при условии, что они существуют и конечны.

$$a_1^* = \bar{x} \xrightarrow[n\to\infty]{п.н} a_1 = E\xi,$$

$$s^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{X})^2 = a_2^{0*} \xrightarrow[n\to\infty]{п.н} a_2^0 = D\xi.$$

3. **Эффективность.**

Пусть в распределении генеральной совокупности имеется неизвестный параметр $\theta \in \Theta \subset \mathbb{R}$. Рассмотрим некоторый класс оценок $K = \{\hat{\theta}(X_{[n]})\}$ параметра θ.

Определение 6

Говорят, что оценка $\theta^*(X_{[n]}) \in K$ является эффективной оценкой параметра θ в классе K, если для любой другой оценки $\hat{\theta} \in K$ имеет место неравенство:

$$E(\theta^* - \theta)^2 \leqslant E(\hat{\theta} - \theta)^2 \qquad (6)$$

для любого $\theta \in \Theta$.

Класс несмещенных оценок обозначим через

$$K_0 = \left\{ \hat{\theta}(X_{[n]}) : E\hat{\theta} = \theta, \forall \theta \in \Theta \right\}.$$

Рассмотрим случай, когда $m > 1$, то есть, $\theta = (\theta_1, \ldots, \theta_m)$. Для любого $y \in \mathbb{R}^m$ определим $\alpha_y = (\theta, y) = \theta_1 y_1 + \ldots + \theta_m y_m$. Тогда $\alpha_y^* = (\theta^*, y)$ — оценка параметра α_y.

Определение 7

Будем говорить, что оценка $\theta^* \in K$ является эффективной оценкой параметра $\theta = (\theta_1, \ldots, \theta_m)$ в классе K, если для любой другой оценки $\hat{\theta} \in K$ и любого $y \in \mathbb{R}^m$ при любом допустимом значении $\theta \in \Theta$ имеет место неравенство:

$$E(\alpha_y^* - \alpha_y)^2 \leqslant E(\hat{\alpha}_y - \alpha_y)^2, \tag{7}$$

где $\hat{\alpha}_y = (\hat{\theta}, y)$.

Теорема 9

Пусть несмещенные оценки $\hat{\theta}_1$ и $\hat{\theta}_2$ параметра $\theta \in \Theta \subset \mathbb{R}$ являются эффективными, тогда оценки $\hat{\theta}_1$ и $\hat{\theta}_2$ почти наверное совпадают.

Доказательство

Рассмотрим $\tilde{\theta} = \frac{\hat{\theta}_1+\hat{\theta}_2}{2}$. Нетрудно показать несмещенность данной оценки. Справедливо тождество:

$$\frac{(\hat{\theta}_1-\theta)^2}{2}+\frac{(\hat{\theta}_2-\theta)^2}{2}=\left(\frac{(\hat{\theta}_1-\theta)}{2}-\frac{(\hat{\theta}_2-\theta)}{2}\right)^2+$$
$$+\left(\frac{(\hat{\theta}_1-\theta)}{2}+\frac{(\hat{\theta}_2-\theta)}{2}\right)^2.$$

Из условия теоремы следует, что для любого $\theta \in \Theta$ имеет место равенство:

$$D\hat{\theta}_1 = D\hat{\theta}_2 = d^2(\theta).$$

Кроме того, для любого $\tilde{\theta} \in K_0$ должно выполняться неравенство: $D\tilde{\theta} \geq d^2(\theta)$. Найдем для тождества математическое ожидание:

$$d^2(\theta) = \frac{1}{4}E(\hat{\theta}_1 - \hat{\theta}_2)^2 + D\tilde{\theta}.$$

Следовательно, $E(\hat{\theta}_1 - \hat{\theta}_2)^2 = 0$ (так как если это неверно, то $D\tilde{\theta} < d^2(\theta)$, чего быть не может), но тогда $\hat{\theta}_1 \overset{\text{п.н}}{=} \hat{\theta}_2$.

Замечание 7

Утверждение теоремы переносится на многомерный случай.

Рассмотрим случай, когда θ вектор. Пусть $\theta = (\theta_1, \ldots, \theta_k)^T \in \Theta \subset \mathbb{R}^k$, $\hat{\theta} \in K_0$, то есть $E\hat{\theta} = \theta$ для любого $\theta \in \Theta$. Возьмем любой вектор $y \in \mathbb{R}^k$, рассмотрим скалярное произведение:

$$(\hat{\theta}, y) = \sum_{i=1}^{k} \hat{\theta}_i(X_{[n]})y_i,$$

как оценку для скалярного произведения (θ, y). Оценка $(\hat{\theta}, y)$ будет несмещенной оценкой:

$$E(\hat{\theta}, y) = (\theta, y)$$

для любого $\theta \in \Theta$ и $y \in \mathbb{R}^k$.
Оценка $(\hat{\theta}, y)$ эффективна, если $D(\hat{\theta}, y) \leq D(\tilde{\theta}, y)$ для любого $\tilde{\theta} \in K_0$ и любого $\theta \in \Theta$, и $y \in \mathbb{R}^k$.
Вычислим левую и правую части неравенства, получим:

$$D(\hat{\theta}, y) = y^T D\hat{\theta} y,$$

где $D\hat{\theta}$ — ковариацинонная матрица вектора $\hat{\theta}$,

$$D(\tilde{\theta}, y) = y^T D\tilde{\theta} y,$$

где $D\tilde{\theta}$ — ковариацинонная матрица вектора $\tilde{\theta}$. Следовательно, справедливо неравенство:

$$y^T (D\tilde{\theta} - D\hat{\theta}) y \geq 0.$$

Тогда, так как y — любое, получаем, что матрица коэффициентов квадратичной формы является неотрицательно определенной матрицей, т. е. $D\tilde{\theta} - D\hat{\theta} \succeq 0$. В результате приходим к определению 8, которое эквивалентно определению 7 для класса несмещенных оценок.

Определение 8

Оценка $\hat{\theta}$ эффективна в классе K_0, или просто эффективна, если $D\tilde{\theta} - D\hat{\theta} \succeq 0$ (неотрицательно определенная матрица), где $\tilde{\theta} \in K_0$ для любого $\theta \in \Theta \subset \mathbb{R}^k$.

4. **Асимптотическая нормальность.**

Определение 9

Пусть оценивается параметр $\theta \in \Theta \subset \mathbb{R}$. Оценка $\hat{\theta}$ называется асимптотически нормальной оценкой параметра θ с коэффициентом рассеивания $\sigma^2(\theta)$, если

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{d} \zeta \sim N(0, \sigma^2(\theta)). \tag{8}$$

Из этого определения следует, что для любого $x \in \mathbb{R}$ имеет место сходимость:

$$P\left\{\sqrt{n}(\hat{\theta} - \theta) \leqslant x\right\} \xrightarrow[n\longrightarrow\infty]{} \frac{1}{\sqrt{2\pi}\sigma(\theta)} \int\limits_{-\infty}^{x} e^{-\frac{y^2}{2\sigma^2(\theta)}} dy.$$

Определение 10

Пусть оценивается параметр $\theta \in \Theta \subset \mathbb{R}^m$. *Оценка* $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_m)$ *называется асимптотически нормальной с матрицей рассеивания* $\Sigma(\theta)$, *если имеет место сходимость по распределению:*

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{d} \eta \sim N(0, \Sigma(\theta)).$$

5. **Асимптотическая эффективность.**

Определение 11

Оценка $\hat{\theta}$ *называется асимптотически эффективной в класс* K *оценок параметра* $\theta \in \Theta \subset \mathbb{R}$, *если*

$$\overline{\lim_{n\to\infty}} \frac{E(\hat{\theta}-\theta)^2}{E(\tilde{\theta}-\theta)^2} \leqslant 1$$

для любого параметра $\theta \in \Theta \subset \mathbb{R}$ *и любой оценки* $\tilde{\theta} \in K$.

Статистическая оценка считается «хорошей», если она обладает хотя бы некоторыми из свойств 1-5.

Методы построения точечных оценок

Рассмотрим сначала **метод моментов**. Пусть требуется оценить параметр $\theta \in \Theta \subset \mathbb{R}$ по имеющейся выборке $X_{[n]} = (X_1, \ldots, X_n)$. Рассмотрим борелевскую функцию $g(x) : \mathbb{R} \to \mathbb{R}$ и определим функцию $m(\theta) = \int\limits_{-\infty}^{\infty} g(x) dF_\xi(x; \theta)$.

Далее положим, что

$$\int\limits_{-\infty}^{\infty} g(x) dF_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} g(X_i) = \bar{g}. \tag{9}$$

Получим уравнение

$$m(\theta) = \bar{g} = \frac{1}{n} \sum_{i=1}^{n} g(X_i). \tag{10}$$

Предположим, что уравнение (10) имеет единственное решение $\hat{\theta}(X_{[n]})$, тогда будем это решение называть оценкой $\hat{\theta}$ неизвестного параметра θ, полученной по методу моментов:

$$\hat{\theta}(X_{[n]}) = m^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right).$$

Введем следующее обозначение: $h(\cdot) = m^{-1}(\cdot)$, оценка по методу моментов при некоторых условиях является статистикой первого типа. По теореме о предельном поведении статистики первого типа получаем сходимость:

$$\hat{\theta}(X_{[n]}) \xrightarrow[n\to\infty]{\text{п.н.}} \theta.$$

Свойства оценок, построенных по методу моментов:

1. Если функция $m^{-1}(y)$ непрерывна на всей области определения, то оценка по методу моментов сильно состоятельна.
2. Если $m'(\theta) \neq 0$ для всех $\theta \in \Theta$, тогда оценка по методу моментов асимптотически нормальна с коэффициентом рассеяния $\frac{Dg(\xi)}{(m'(\theta))^2}$, где θ — истинное значение параметра.

Метод моментов легко обобщить на многомерный случай, при этом, $g(x) = (g_1(x), \ldots, g_k(x))$, где k — число неизвестных параметров, то есть, $\theta = (\theta_1, \ldots, \theta_k)^T \in \Theta \subset \mathbb{R}^k$.

Пример 3

Пусть $\xi \sim N(a, \sigma^2)$, *тогда* $\theta = (a, \sigma^2)^T \in \Theta = \mathbb{R} \times \mathbb{R}_+$. *Выберем* $g(x) = (x, x^2)$, *тогда*

$$Eg(\xi) = \begin{pmatrix} E\xi \\ E\xi^2 \end{pmatrix} = \begin{pmatrix} a \\ \sigma^2 + a^2 \end{pmatrix},$$

так как $\sigma^2 = D\xi = E\xi^2 - (E\xi)^2 = E\xi^2 - a^2$. *Нетрудно показать, что*

$$\bar{g} = \begin{pmatrix} \frac{1}{n}\sum\limits_{k=1}^{n} X_k \\ \frac{1}{n}\sum\limits_{k=1}^{n} X_k^2 \end{pmatrix} = \begin{pmatrix} \bar{X} \\ s^2 + \bar{X}^2 \end{pmatrix},$$

так как $s^2 = \frac{1}{n}\sum\limits_{k=1}^{n} X_k^2 - (\bar{X})^2$.

Таким образом, получили систему:

$$\begin{cases} a = \bar{X} \\ \sigma^2 + a^2 = s^2 + \bar{X}^2. \end{cases}$$

Следовательно, оценки по методу моментов имеют следующий вид:

$$\begin{cases} \hat{a} = \bar{X}, \\ \hat{\sigma}^2 = s^2. \end{cases}$$

Пример 4

Рассмотрим равномерно распределенную случайную величину ξ *с плотностью распределения:*

$$f_\xi(x) = \begin{cases} \frac{1}{\theta}, & x \in [0, \theta]; \\ 0, & x \notin [0, \theta]. \end{cases}$$

Так как неизвестный параметр один, то $g(x) = x$. *Вычислим математическое ожидание:*

$$Eg(\xi) = E\xi = \int_0^\theta x\frac{1}{\theta}dx = \frac{1}{2\theta}x^2 = \frac{\theta}{2}.$$

Уравнение имеет вид:

$$\frac{\theta}{2} = \bar{X},$$

откуда получаем оценку:

$$\hat{\theta} = 2\bar{X} = 2\frac{1}{n}\sum_{k=1}^{n} X_i.$$

Может оказаться, что $\frac{1}{n}\sum_{k=1}^{n} X_i > \theta/2$, *тогда* $\hat{\theta} > \theta$. *Данный метод может дать сильно завышенную оценку.*

Рассмотрим теперь **метод максимального правдоподобия** построения точечных оценок. Пусть задана генеральная совокупность ξ с функцией распределения F_ξ и плотностью распределения f_ξ (будем предполагать, что плотность распределения существует). Задана выборка $X_{[n]} = (X_1, \ldots, X_n)$. Совместная плотность распределения выборки имеет вид:

$$f_{X_{[n]}}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\xi(x_i). \tag{11}$$

В плотности распределения выборки существует неизвестный параметр θ, поэтому ниже будем рассматривать совместную плотность распределения в виде:

$$f_{X_{[n]}}(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f_\xi(x_i, \theta) \quad \text{или} \quad f_{X_{[n]}}(X_{[n]}|\theta) = \prod_{i=1}^{n} f_\xi(X_i, \theta),$$

где $\theta = (\theta_1, \ldots, \theta_m) \in \Theta$.

Определение 12

Если генеральная совокупность имеет плотность распределения f_ξ, то функцией правдоподобия выборки $X_{[n]}$ будем называть функцию

$$L(X_{[n]}, \theta) = \prod_{i=1}^{n} f_\xi(X_i, \theta).$$

Определение 13

Если генеральная совокупность ξ — дискретная случайная величина с возможными значениями $\{z_i\}$ и соответствующими вероятностями $p_\xi(z_i, \theta)$, то функцией правдоподобия выборки $X_{[n]}$ будем называть функцию

$$L(X_{[n]}, \theta) = \prod_{i=1}^{n} p_\xi(X_i, \theta).$$

Будем считать функцию правдоподобия функцией неизвестного параметра θ. Для нахождения оценки параметра θ решаем задачу:

$$\max_{\theta \in \Theta} L(X_{[n]}, \theta).$$

Определение 14

Оценкой максимального правдоподобия параметра θ называется оценка

$$\hat{\theta}(X_{[n]}) = \arg\max_{\theta \in \Theta} L(X_{[n]}, \theta), \tag{12}$$

если решение задачи максимизации существует и единственно.

Свойства оценок максимального правдоподобия:

1. Предположим, что существует взаимно однозначное соответствие $\beta : \Theta \leftrightarrow B$, пусть
$$\hat{b}(X_{[n]}) = \arg\max_{b \in B} L(X_{[n]}, \beta^{-1}(b)). \qquad (13)$$
Если решение (12) существует и единственно, то существует и единственно решение (13), причем, имеет место равенство:
$$\hat{\theta} = \beta^{-1}(\hat{b}).$$
2. Если функция правдоподобия непрерывно дифференцируема, и выполнены некоторые условия гладкости, то можно доказать, что оценки метода максимального правдоподобия — сильно состоятельны, асимптотически эффективны и асимптотически нормальны [3], [?], [?].

Замечание 8

Часто вместо функции $L(X_{[n]}, \theta)$ *рассматривают функцию* $\ln L(X_{[n]}, \theta)$, *поскольку функция* $\ln(t)$ *является строго возрастающей функцией своего аргумента* t, *и данный переход правомерен.*

Пример 5

Рассмотрим случайную величину $\xi \sim N(a, \sigma^2)$ *с плотностью распределения*

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Функция правдоподобия имеет вид:

$$L(X_{[n]}, a, \sigma^2) = \prod_{i=1}^{n} f_\xi(X_i, a, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{\frac{-\sum\limits_{i=1}^{n}(X_i-a)^2}{2\sigma^2}}.$$

Тогда

$$\ln L = \ln \frac{1}{((2\pi)^{\frac{1}{2}}\sigma)^n} - \frac{\sum\limits_{i=1}^{n}(X_i - a)^2}{2\sigma^2},$$

продифференцируем по a*:* $\partial \ln L/\partial a = 0$*, или* $\sum_{i=1}^{n} X_i - an = 0$*, откуда* $\hat{a} = \bar{X}$.

Продифференцируем по σ:

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum\limits_{i=1}^{n}(X_i - a)^2}{\sigma^3} = 0,$$

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

откуда находим решение:

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

Нетрудно проверить, что $\bar{X}$ и s^2 доставляют максимум функции правдоподобия.

Пример 6

Пусть случайная величина ξ подчиняется равномерному распределению с плотностью:

$$f(x,\theta) = \begin{cases} \frac{1}{\theta}, & x \in [0,\theta]; \\ 0, & x \notin [0,\theta]. \end{cases}$$

Запишем функцию правдоподобия:

$$L(X_{[n]},\theta) = \prod_{i=1}^{n} f(X_i,\theta) = \begin{cases} \frac{1}{\theta^n}, & \text{если для } \forall i : X_i \in [0,\theta]; \\ 0, & \text{если } \exists i : X_i \notin [0,\theta]. \end{cases}$$

Построим вариационный ряд $X_{(1)} \leq \ldots \leq X_{(n)}$. Таким образом, получаем:

$$L(X_{[n]},\theta) = \begin{cases} \frac{1}{\theta^n}, & X_{(n)} \in [0,\theta]; \\ 0, & \exists k : X_{(k)} \notin [0,\theta]. \end{cases}$$

Очевидно, что оценка максимального правдоподобия $\hat{\theta}(X_{[n]}) = X_{(n)}$.

Неравенство Рао-Крамера

Пусть имеется выборка $X_{[n]}$ из генеральной совокупности ξ с функцией распределения $F_\xi(x,\theta)$ и плотностью распределения $f_\xi(x,\theta)$, где $\theta \in \Theta \subset \mathbb{R}$ — неизвестный параметр.

Замечание 9

Все результаты этого параграфа можно перенести на дискретный случай.

Функция правдоподобия имеет вид:

$$L(X_{[n]},\theta) = \prod_{i=1}^{n} f_\xi(X_i,\theta),$$

совместная плотность выборки имеет вид:

$$f_{X_{[n]}}(x,\theta) = L(x,\theta) = \prod_{i=1}^{n} f_\xi(x_i,\theta).$$

Имеет место следующее равенство:

$$\int\limits_{\mathbb{R}^n} L(x,\theta)dx = 1. \qquad (14)$$

Пусть имеется оценка $\hat{\theta}(X_{[n]})$ неизвестного параметра θ, и справедливо следующее равенство:

$$E\hat{\theta} = \int\limits_{\mathbb{R}^n} \hat{\theta}(x_1,\ldots,x_n)L(x,\theta)dx = h(\theta). \qquad (15)$$

Обозначим через $I_n(\theta)$ математическое ожидание:

$$I_n(\theta) = E\left(\frac{\partial \ln L(X_{(n)},\theta)}{\partial\theta}\right)^2 = \int\limits_{\mathbb{R}^n}\left(\frac{\partial \ln L(x,\theta)}{\partial\theta}\right)^2 L(x,\theta)dx.$$

Определение 15

Величина $I_n(\theta)$, если математическое ожидание существует и конечно, называется информационным количеством Фишера (соответствующим выборке объема n).

Будем предполагать, что выполнены *условия регулярности*:

- Для информационного количества Фишера выполнено неравенство $0 < I_n(\theta) < \infty$ для любого $\theta \in \Theta$.
- Равенства (14) и (15) можно продифференцировать и получить следующие уравнения:

$$\int\limits_{\mathbb{R}^n} \frac{\partial L(x, \theta)}{\partial \theta} dx = 0, \tag{16}$$

$$\int\limits_{\mathbb{R}^n} \hat{\theta}(x_1, \ldots, x_n) \frac{\partial L(x, \theta)}{\partial \theta} dx = h'(\theta). \tag{17}$$

- Множество $N = \{x \in R^n : L(x, \theta) = 0\}$ не зависит от θ.

Теорема 10

(Неравенство Рао-Крамера) Пусть имеется генеральная совокупность ξ *с функцией распределения* $F_\xi(y,\theta)$, *где* $\theta \in \Theta \subset \mathbb{R}$. *Задана выборка* $X_{[n]}$ *из генеральной совокупности* ξ, *и выполнены условия регулярности, тогда имеет место неравенство:*

$$D\hat{\theta} \geqslant \frac{(h'(\theta))^2}{I_n(\theta)}. \tag{18}$$

Доказательство

Перепишем 16 и 17 следующим образом:

$$\int\limits_{\mathbb{R}^n} \frac{\partial \ln L(x,\theta)}{\partial \theta} L(x,\theta)dx = E\left\{\frac{\partial \ln L(X_{[n]},\theta)}{\partial \theta}\right\} = 0,$$

$$\int\limits_{\mathbb{R}^n} \hat{\theta}(x)\frac{\partial \ln L(x,\theta)}{\partial \theta} L(x,\theta)dx = E\left\{\hat{\theta}(X_{[n]})\frac{\partial \ln L(X_{[n]},\theta)}{\partial \theta}\right\} = h'(\theta).$$

Заметим, что множество $N = \{x : L(x, \theta) = 0\}$ — множество меры нуль, так как

$$P\{X_{[n]} \in N\} = \int\limits_N L(x, \theta) dx = 0.$$

Умножим первое равенство на $E\hat{\theta}$ и вычтем из второго:

$$E\left\{(\hat{\theta}(X_{[n]}) - E\hat{\theta})\frac{\partial \ln L(X_{[n]}, \theta)}{\partial \theta}\right\} = h'(\theta).$$

Сделаем обозначения:

$$\eta_1 = (\hat{\theta}(X_{[n]}) - E\hat{\theta}),$$

$$\eta_2 = \frac{\partial \ln L(X_{[n]}, \theta)}{\partial \theta}.$$

Воспользуемся неравенством Коши-Шварца-Буняковского:

$$(h'(\theta))^2 = (E\{\eta_1\eta_2\})^2 \leq E\eta_1^2 E\eta_2^2 = D\hat{\theta} I_n(\theta).$$

Теорема доказана.

Замечание 10

Неравенство (18) выполнено как равенство тогда и только тогда, когда $\eta_1 \overset{\text{п.н.}}{=} A(\theta)\eta_2$ *(следует из неравенства Коши-Шварца-Буняковского), то есть,*

$$(\hat{\theta}(X_{[n]}) - \theta) \overset{\text{п.н.}}{=} A(\theta)\frac{\partial \ln L}{\partial \theta}.$$

Замечание 11

Если $E\hat{\theta} = \theta$ *для любого* $\theta \in \Theta$ *(то есть, оценка — несмещенная), то справедливо неравенство:*

$$D\hat{\theta} \geq \frac{1}{I_n(\theta)}.$$

Замечание 12

По определению $I_n(\theta) = E\left(\frac{\partial \ln L}{\partial \theta}\right)^2$. *Заметим, что*

$$E\left\{\frac{\partial \ln L(X_{[n]},\theta)}{\partial \theta}\right\} = 0.$$

Следовательно, имеют место равенства:

$$I_n(\theta) = E\left(\frac{\partial \ln L}{\partial \theta}\right)^2 = D\left(\frac{\partial \ln L(X_{[n]},\theta)}{\partial \theta}\right) =$$

$$= D\left(\sum_{i=1}^{n} \frac{\partial \ln f_\xi(X_i,\theta)}{\partial \theta}\right) = \sum_{i=1}^{n} D\frac{\partial \ln f_\xi(X_i,\theta)}{\partial \theta} = nI_1(\theta),$$

где $I_1(\theta)$ *— информационное количество Фишера, соответствующее одному наблюдению. Как видим, наблюдается линейный рост информации.*

- *Из неравенства Рао-Крамера можно сделать вывод, что в регулярном случае дисперсия не может убывать быстрее чем* $1/n$.
- *Для «хороших» оценок дисперсия должна убывать, разброс должен становиться меньше с ростом* n.
- *Для несмещенных оценок при выполнении условий регулярности оценка эффективна, если неравенство Рао-Крамера выполнено как равенство.*

Справедлива еще одна формула для вычисления $I_n(\theta)$ при выполнении дополнительных условий, которые необходимы для корректного проведения всех последующих преобразований. Продифференцируем равенство

$$\int\limits_{\mathbb{R}^n} \frac{\partial \ln L(x,\theta)}{\partial\theta} L(x,\theta)dx = 0$$

еще один раз, получаем следующее равенство:

$$\int\limits_{\mathbb{R}^n} \left(\frac{\partial \ln L(x,\theta)}{\partial\theta}\right)^2 L(x,\theta)dx + \int\limits_{\mathbb{R}^n} \frac{\partial^2 \ln L(x,\theta)}{\partial\theta} L(x,\theta)dx = 0,$$

или, что тоже самое:

$$E\left\{\frac{\partial \ln L(X_{[n]},\theta)}{\partial\theta}\right\}^2 + E\left\{\frac{\partial^2 \ln L(X_{[n]},\theta)}{\partial\theta^2}\right\} = 0.$$

Откуда сразу получаем равенство:

$$I_n(\theta) = -E\left\{\frac{\partial^2 \ln L(X_{[n]},\theta)}{\partial\theta^2}\right\}.$$

Замечание 13

Справедлив аналог неравенства Рао-Крамера для многомерного случая. Пусть $\theta \in \Theta \subset \mathbb{R}^k$*, пусть* $E\hat{\theta}(X_{[n]}) = \theta$*, тогда*

$$D\hat{\theta} \geq (I_n(\theta))^{-1},$$

где I_n *- информационная матрица Фишера, а* $D\hat{\theta}$ *- ковариацинонная матрица*

$$I_n(\theta) = E\left\{\frac{\partial \ln L(X_{[n]}, \theta)}{\partial \theta}\left(\frac{\partial \ln L}{\partial \theta}\right)^T\right\},$$

$$D\hat{\theta} = E\left\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\right\}$$

То есть, матрица $D\hat{\theta} - (I_n(\theta))^{-1} \succeq 0$ *— неотрицательно определенная.*

Пример 7

Пусть $\xi \sim N(a, \sigma^2)$, *методы максимального правдоподобия и моментов дали оценку* $\hat{a} = \bar{X}$, *которая является сильно состоятельной, несмещенной, асимптотически нормальной оценкой. Выясним, обращается ли неравенство Рао-Крамера в равенство. Действительно, можно заметить, что*

$$\frac{\partial \ln L}{\partial a} = \frac{\sum\limits_{i=1}^{n}(X_i - a)}{\sigma^2} = \frac{n}{\sigma^2}(\bar{X} - a),$$

где

$$L(X_{[n]}, a) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n} e^{-\frac{\sum\limits_{i=1}^{n}(X_i - a)^2}{2\sigma^2}}.$$

Как следует из замечания 10, случайные величины пропорциональны, следовательно, $\bar{X}$ *— эффективная оценка.*

Биллингсли П.Сходимость вероятностных мер. — М.: Изд. Наука, 1977.

Боровков А.А.Теория вероятностей. — М.: Изд. Наука, 1986.

Боровков А.А.Математическая статистика. — М.: Изд. Наука, 1984.