Оглавление

1. Задачи математической статистики

Задачи математической статистики можно рассматривать как в некотором смысле обратные к задачам теории вероятностей. В теории вероятностей считается заданным вероятностное пространство (или, например, распределения случайных величин, определённых на этом пространстве). В рамках такой модели требуется найти вероятность того или иного события. В отличие от теории вероятностей в математической статистике известно, какие события произошли, и на основании этих сведений следует высказаться о вероятностном пространстве или некоторых его свойствах.

Например, найти вероятность получить k успехов в серии из n испытаний по схеме Бернулли, если известна вероятность успеха p при единичном испытании, – это задача теории вероятностей, в которой ответ даётся формулой Бернулли. К математической статистике относится задача, в которой известно, что в результате n испытаний получено k успехов, и требуется оценить вероятность успеха в единичном испытании. Другим примером задачи математической статистики является поиск ответа на вопрос, противоречит ли полученному результату эксперимента предположение о том, что вероятность успеха в единичном испытании равна некоторой заданной величине p_0.

1.1. Основные понятия и теоремы теории случайных величин

Напомним базовые понятия и теоремы теории вероятностей, связанные со случайными величинами.

Распределения случайных величин

Случайная величина (с.в.) принимает различные числовые значения случайным образом. Поведение с.в. ξ задаётся функцией распределения

$$F(x) \stackrel{\text{def}}{=} P(\xi < x), \quad x \in \mathbb{R}.$$

Функция распределения любой с.в. определена (т. е. имеет какое-то конкретное значение) в каждой точке вещественной прямой, не убывает от $F(-\infty) = 0$ до $F(\infty) = 1$ и непрерывна слева, другими словами, $F(x-0) = F(x)$ для любого $x \in \mathbb{R}$. При этом правое предельное значение $F(x+0) = P(\xi \leqslant x)$, тем самым скачок функции распределения в точке x равен

$$F(x+0) - F(x-0) = F(x+0) - F(x) = P(\xi = x),$$

и $F(\,\cdot\,)$ непрерывна в точке x, если и только если $P(\xi = x) = 0$.

Распределение с.в. ξ называется абсолютно непрерывным, если существует функция $p(\,\cdot\,)\colon \mathbb{R} \to \mathbb{R}_+$, называемая плотностью вероятности[1)], такая что для любого $x \in \mathbb{R}$

$$dF(x) = F(x+dx) - F(x) = P(x \leqslant \xi < x + dx) = p(x)\,dx.$$

Тогда для подмножества $X \subset \mathbb{R}$

$$P(\xi \in X) = \int\limits_X p(x)\,dx.$$

Распределение с.в. ξ называется дискретным, если существует конечный $(m < \infty)$ или счётный $(m = \infty)$ набор чисел $\{b_j\}_{j=\overline{1,m}}$, такой что

$$P(\xi = b_j) = p_j > 0, \quad j = 1, \dots, m, \qquad \sum_{j=1}^{m} p_j = 1.$$

Эти равенства называют рядом или законом распределения дискретной с.в. Для подмножества $X \subset \mathbb{R}$ мы можем записать очевидное равенство

$$P(\xi \in X) = \sum_{j\colon\, b_j \in X} p_j.$$

Два указанных типа распределения не исчерпывают все возможные распределения, но в нашем курсе любая с.в. распределена либо дискретно, либо абсолютно непрерывно.

[1)] Значение плотности вероятности может не существовать в одной или нескольких точках, с другой стороны, в этих точках плотность вероятности можно доопределить, положив её равной любому конечному числу (например, по непрерывности слева или справа). Мы будем считать, что любая плотность вероятности определена на всей оси и кусочно непрерывна.

Во многих случаях мы будем иметь дело с хорошо известными и изученными распределениями. Перечислим некоторые из них и введём соответствующие обозначения.

• С.в. ξ распределена нормально (имеет нормальное распределение) с параметрами $\mu \in \mathbb{R}$ и $0 < \sigma^2 < \infty$, и тогда мы пишем $\xi \sim \mathbf{N}(\mu, \sigma^2)$, если

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < \infty. \tag{1.1}$$

• С.в. ξ распределена экспоненциально (имеет экспоненциальное распределение) с параметром $0 < a < \infty$, и мы пишем $\xi \sim \mathbf{E}(a)$, если

$$p(x) = \begin{cases} ae^{-ax}, & x \geqslant 0, \\ 0, & x < 0. \end{cases} \tag{1.2}$$

• С.в. ξ распределена равномерно (имеет равномерное распределение) на отрезке $[a, b]$, где $-\infty < a < b < \infty$, и мы пишем $\xi \sim \mathbf{U}[a, b]$, если

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases} \tag{1.3}$$

• С.в. ξ имеет распределение Пуассона с параметром $0 < \lambda < \infty$, и мы пишем $\xi \sim \mathbf{P}(\lambda)$, если

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, \dots . \tag{1.4}$$

• С.в. ξ имеет биномиальное распределение с параметрами $n = 1, 2 \dots$ и $0 < p < 1$, и мы пишем $\xi \sim \mathbf{B}(n, p)$, если

$$P(\xi = k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, \dots, n, \qquad q = 1 - p. \tag{1.5}$$

В курсе математической статистики в основном рассматривают многомерные с.в. $\xi = (\xi_1, \dots, \xi_n)$. Здесь n может принимать любое конечное натуральное значение. Далее мы называем с.в. $\xi_1, \dots, \xi_n$ составляющими многомерной с.в. ξ. В общем случае функция распределения n-мерной с.в. ξ определяется как

$$F(x) = P(\xi < x) \overset{\text{def}}{=} P(\xi_1 < x_1, \dots, \xi_n < x_n), \;\; x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Если для всех $x \in \mathbb{R}^n$

$$dF(x) = P(x_1 \leqslant \xi < x_1 + dx_1, \ldots, x_n \leqslant \xi < x_n + dx_n) = p(x)\,dx,$$

где $dx = dx_1 \ldots dx_n$, то функция $p(\,\cdot\,)\colon \mathbb{R}^n \to \mathbb{R}_+$ называется совместной плотностью вероятности n-мерной с.в. ξ.

Мы будем иметь дело только с многомерными с.в., которые имеют независимые составляющие. Это означает, что при любом $x = (x_1, \ldots, x_n) \in \mathbb{R}$ совместная функция распределения представляется как произведение одномерных функций распределения:

$$F(x) = F_{\xi_1}(x_1) \ldots F_{\xi_n}(x_n), \tag{1.6}$$

где $F_{\xi_k}(\,\cdot\,)$ – функция распределения с.в. ξ_k, $k = 1, \ldots, n$. Если существует плотность вероятности независимых с.в. $\xi_, \ldots, \xi_n$, то для неё справедливо аналогичное равенство:

$$p(x) = p_{\xi_1}(x_1) \ldots p_{\xi_n}(x_n), \tag{1.7}$$

где $p_{\xi_k}(\,\cdot\,)$ – плотность вероятности с.в. ξ_k, $k = 1, \ldots, n$. Кроме независимости, мы будем предполагать, что все $\xi_1, \ldots, \xi_n$ одинаково распределены, т. е. $P(\xi_k < x)$ не зависит от k для любого $x \in \mathbb{R}$. Тогда согласно (1.6) для задания распределения n-мерной с.в. ξ достаточно задать одномерное распределение.

Математическое ожидание с.в.

Математическое ожидание с.в. ξ – это взвешенное среднее её значений, в котором вес каждого значения тем выше, чем выше вероятность наблюдать это значение при измерении данной с.в. Если с.в. ξ имеет абсолютно непрерывное распределение с плотностью вероятности $p(\,\cdot\,)$, то математическое ожидание равно

$$M\xi = \int_{-\infty}^{+\infty} xp(x)\,dx.$$

Если с.в. ξ имеет дискретное распределение, то математическое ожидание равно

$$M\xi = \sum_{j=1}^{m} b_j P(\xi = b_j), \quad \text{где} \quad \sum_{j=1}^{m} P(\xi = b_j) = 1.$$

Математическое ожидание существует, только если интеграл/ряд сходится абсолютно.

Если с.в. η задана как функция от с.в. ξ, т. е. $\eta = g(\xi)$, то

$$M\eta = Mg(\xi) = \begin{cases} \displaystyle\int_{-\infty}^{+\infty} g(x)\,p(x)\,dx, \\ \displaystyle\sum_{j=1}^{m} g(b_j)P(\xi = b_j), \end{cases}$$

где первая строчка соответствует абсолютно непрерывному, а вторая – дискретному распределению с.в. ξ. Эти формулы справедливы и в случае n-мерной с.в. ξ: если $\eta = g(\xi)$ и $g(\,\cdot\,)\colon \mathbb{R}^n \to \mathbb{R}$ то, чтобы найти $M\eta$, нужно вычислить аналогично написанному выше n-мерный интеграл или n-мерную сумму.

Число $M\xi^k$ называется начальным моментом, $M(\xi - M\xi)^k$ – центральным моментом k-го порядка; центральный момент 2-го порядка называется дисперсией с.в. ξ,

$$D\xi = M(\xi - M\xi)^2,$$

а число

$$\operatorname{cov}(\xi_1, \xi_2) = M\big[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)\big]$$

– коэффициентом ковариации (ковариацией) случайных величин ξ_1, ξ_2. При этом $D\xi = \operatorname{cov}(\xi, \xi)$.

Математическое ожидание линейно: для любых случайных величин $\xi_1, \ldots, \xi_n$ и любых вещественных постоянных $a_1, \ldots, a_n$ и c

$$M(a_1\xi_1 + \cdots + a_n\xi_n + c) = a_1 M\xi_1 + \cdots + a_n M\xi_n + c.$$

Из линейности вытекает, что

$$D(a\xi + b) = a^2 D\xi, \qquad D\xi = M\xi^2 - (M\xi)^2,$$
$$\operatorname{cov}(\xi_1, \xi_2) = M(\xi_1\xi_2) - M\xi_1 M\xi_2.$$

Если случайные величины $\xi_1, \ldots, \xi_n$ независимы, то

$$M(\xi_1 \ldots \xi_n) = M\xi_1 \cdot \ldots \cdot M\xi_n.$$

Отсюда следует, что если с.в. ξ_1, ξ_2 независимы, то $\operatorname{cov}(\xi_1, \xi_2) = 0$; обратное утверждение, вообще говоря, неверно.

Если $P(\xi \geqslant 0) = 1$, то $M\xi \geqslant 0$. Вследствие линейности, если $P(\xi_1 \geqslant \xi_2) = 1$, то $M\xi_1 \geqslant M\xi_2$. Справедливы следующие неравенства:

$$M\xi^2 \geqslant (M\xi)^2, \tag{1.8}$$

неравенство Коши–Буняковского

$$\big|\operatorname{cov}(\xi_1, \xi_2)\big| \leqslant \sqrt{D\xi_1 D\xi_2} \tag{1.9}$$

и неравенство Чебышёва: для любого $\varepsilon > 0$

$$P\big(|\xi - M\xi| \geqslant \varepsilon\big) \leqslant \frac{D\xi}{\varepsilon^2}. \tag{1.10}$$

Из неравенства Чебышёва следует, что если $D\xi = 0$, то $\xi = M\xi$ с вероятностью единица.

Сходимости последовательностей с.в.

В нашем курсе мы встретимся с тремя типами сходимости последовательностей с.в.

Последовательность с.в. $\{\xi_n\}_{n=\overline{1,\infty}}$ сходится к с.в. ξ по вероятности, и мы пишем $\xi_n \xrightarrow[n\to\infty]{\mathrm{P}} \xi$, если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\big(|\xi_n - \xi| \geqslant \varepsilon\big) = 0 \tag{1.11}$$

или, что эквивалентно, $\lim\limits_{n\to\infty} P\big(|\xi_n - \xi| < \varepsilon\big) = 1$.

Последовательность $\{\xi_n\}_{n=\overline{1,\infty}}$ сходится к с.в. ξ по распределению, и мы пишем $\xi_n \xrightarrow[n\to\infty]{\mathrm{d}} \xi$, если имеет место поточечная сходимость функций распределения $F_{\xi_n}(\,\cdot\,)$ с.в. ξ_n, $n = 1, 2, \ldots$, к функции распределения $F_\xi(\,\cdot\,)$ с.в. ξ:

$$\lim_{n\to\infty} F_{\xi_n}(x) = F_\xi(x) \tag{1.12}$$

в каждой точке x непрерывности предельной функции $F_\xi(\,\cdot\,)$.

Последовательность $\{\xi_n\}_{n=\overline{1,\infty}}$ сходится к с.в. ξ с вероятностью единица (иначе говоря, почти наверное), и мы пишем $\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi$, если

$$P(\lim_{n\to\infty} \xi_n = \xi) = 1.$$

Закон больших чисел в форме Чебышёва. Если $\xi_1, \xi_2, \ldots$ независимы и одинаково распределены, $M\xi_k = \mu$, $D\xi_k = \sigma^2$ для $k = 1, 2, \ldots$, то

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k \underset{n\to\infty}{\longrightarrow} \mu$$

и по вероятности, и с вероятностью единица.

Закон больших чисел в форме Бернулли. Если β_n – число успехов в n испытаниях Бернулли с вероятностью успеха p, т. е. $\beta_n \sim \mathbf{B}(n, p)$, то частота успеха β_n/n сходится к вероятности успеха p при $n \to \infty$ (по вероятности или почти наверное).

Центральная предельная теорема. Если $\{\xi_n\}_{n=\overline{1,\infty}}$ – последовательность независимых одинаково распределённых случайных величин, $M\xi_k = \mu$, $D\xi_k = \sigma^2$ для $k = 1, 2, \ldots$, то

$$\eta_n = \frac{\sum_{j=1}^{n}(\xi_j - \mu)}{\sigma\sqrt{n}} \xrightarrow{\mathrm{d}} \nu_*,$$

где случайная величина ν_* имеет стандартное нормальное распределение $\mathbf{N}(0, 1)$, причём сходимость функций распределения $F_{\eta_n}(\,\cdot\,)$ с.в. η_n, $n = 1, 2, \ldots$, к функции распределения $\Phi(\,\cdot\,)$ с.в. ν_* равномерная по x на любом конечном интервале вещественной оси. Здесь и далее

$$\Phi(x) \stackrel{\mathrm{def}}{=} \frac{1}{2\pi}\int_{-\infty}^{x} e^{-t^2/2}\,dt, \qquad x \in \mathbb{R}. \tag{1.13}$$

Эта функция называется интегралом вероятности и по сути является функцией распределения с.в. $\nu_* \sim \mathbf{N}(0, 1)$.

Интегральная теорема Муавра–Лапласа. Пусть β_n – число успехов в серии n независимых испытаний Бернулли с фиксированной вероятностью успеха p. Положим $x_{k,n} = \frac{k-np}{\sqrt{npq}}$. Тогда при условии, что $|x_{k,n}| < C$ при всех значениях k для некоторого $C \in \mathbb{R}$,

$$\bigl|P(\beta_n < k) - \Phi(x_{k,n})\bigr| \to 0, \qquad n \to \infty. \tag{1.14}$$

В курсе математической статистки нам потребуются два утверждения, которые можно найти не во всех учебниках по теории вероятностей, поэтому здесь мы приведём их доказательство полностью.

Предложение 1. *Пусть функция* $f(\cdot)\colon \mathbb{R} \to \mathbb{R}$ *непрерывна в точке* $a \in \mathbb{R}$ *своей области определения. Пусть для последовательности с.в.* $\{\xi_n\}_{n=\overline{1,\infty}}$ *имеет место сходимость* $\xi_n \xrightarrow[n\to\infty]{\mathrm{P}} a$. *Тогда* $f(\xi_n) \xrightarrow[n\to\infty]{\mathrm{P}} f(a)$.

Доказательство. Из непрерывности функции $f(\cdot)$ в точке a следует, что для любого $\varepsilon > 0$ найдётся $\delta(\varepsilon) > 0$, такое что если $|x - a| < \delta(\varepsilon)$, то $|f(x) - f(a)| < \varepsilon$.

Выберем и зафиксируем произвольное $\varepsilon > 0$, найдём для него $\delta(\varepsilon) > 0$ и рассмотрим последовательность вероятностей

$$P_n(\varepsilon) = P(|f(\xi_n) - f(a)| < \varepsilon), \qquad n = 1, 2, \ldots .$$

Имеем

$$\begin{aligned} P_n(\varepsilon) = P(|f(\xi_n) - f(a)| < \varepsilon, |\xi_n - a| < \delta(\varepsilon)) &+ \\ + P(|f(\xi_n) - f(a)| < \varepsilon, |\xi_n - a| \geqslant \delta(\varepsilon)) &= \\ = P_{1,n}(\varepsilon) + P_{2,n}(\varepsilon). & \end{aligned} \tag{1.15}$$

Событие $|\xi_n - a| < \delta(\varepsilon)$ влечет событие $|f(\xi_n) - f(a)| < \varepsilon$, поэтому

$$P_{1,n}(\varepsilon) = P(|f(\xi_n) - f(a)| < \varepsilon, |\xi_n - a| < \delta(\varepsilon)) = P(|\xi_n - a| < \delta(\varepsilon)).$$

В силу того, что $\xi_n \xrightarrow{\mathrm{P}} a$ при $n \to \infty$, получаем

$$\lim_{n\to\infty} P_{1,n}(\varepsilon) = \lim_{n\to\infty} P(|\xi_n - a| < \delta(\varepsilon)) = 1.$$

Для второго слагаемого в (1.15) имеем

$$P_{2,n}(\varepsilon) = P(|f(\xi_n) - f(a)| < \varepsilon, |\xi_n - a| \geqslant \delta(\varepsilon)) \leqslant P(|\xi_n - a| \geqslant \delta(\varepsilon)).$$

Опять же как следствие $\xi_n \xrightarrow{\mathrm{P}} a$ получаем $P(|\xi_n - a| \geqslant \delta(\varepsilon)) \to 0$ при $n \to \infty$, поэтому

$$\lim_{n\to\infty} P_{1,n}(\varepsilon) = 0.$$

Отсюда $P_n(\varepsilon) = P_{1,n}(\varepsilon) + P_{2,n}(\varepsilon) \to 1$ при $n \to \infty$, что и доказывает сходимость $f(\xi_n) \xrightarrow{\mathrm{P}} f(a)$.

Предложение 2. *Пусть для последовательностей с.в.* $\{\xi_n\}_{n=\overline{1,\infty}}$ *и* $\{\eta_n\}_{n=\overline{1,\infty}}$ *имеют место сходимости* $\xi_n \xrightarrow[n\to\infty]{\mathrm{P}} 1$ *и* $\eta_n \xrightarrow[n\to\infty]{\mathrm{d}} \eta$. *Тогда* $\xi_n\eta_n \xrightarrow[n\to\infty]{\mathrm{d}} \eta$.

Доказательство. Обозначим как $F_{\xi_n\eta_n}(\,\cdot\,)$, $F_\eta(\,\cdot\,)$ функции распределения соответствующих с.в. Покажем, что $F_{\xi_n\eta_n}(x) \to F_\eta(x)$ при $n\to\infty$ в любой точке $x\in\mathbb{R}$ непрерывности функции $F_\eta(\,\cdot\,)$.

Выберем и зафиксируем число $\varepsilon\in(0,1)$. Тогда мы можем, как и в предыдущем утверждении, разбить искомую вероятность на два слагаемых:

$$\begin{aligned} F_{\xi_n\eta_n}(x) &= P(\xi_n\eta_n < x) = \\ &= P(\xi_n\eta_n < x, |\xi_n - 1| \geqslant \varepsilon) + P(\xi_n\eta_n < x, |\xi_n - 1| < \varepsilon) = \\ &= P_{1,n}(x,\varepsilon) + P_{2,n}(x,\varepsilon). \end{aligned} \tag{1.16}$$

Оценим каждое из слагаемых в правой части (1.16). Для первого слагаемого имеем

$$P_{1,n}(x,\varepsilon) = P(\xi_n\eta_n < x, |\xi_n - 1| \geqslant \varepsilon) \leqslant P(|\xi_n - 1| \geqslant \varepsilon) \to 0 \tag{1.17}$$

при $n\to\infty$ в силу сходимости $\xi_n \xrightarrow{\mathrm{P}} 1$.

Теперь оценим второе слагаемое в правой части (1.16). Сначала будем считать, что $x>0$. Рассмотрим систему неравенств $\xi_n\eta_n < x$, $|\xi_n - 1| < \varepsilon$ как условия, задающие область на плоскости (ξ_n,η_n). Легко видеть, что для $0<\varepsilon<1$ мы имеем следствия

$$\begin{aligned} &\begin{cases} \xi_n\eta_n < x, \\ 1-\varepsilon < \xi_n < 1+\varepsilon \end{cases} \quad \Longrightarrow \quad \eta_n < \frac{x}{1-\varepsilon}, \\ &\begin{cases} \eta_n < \dfrac{x}{1+\varepsilon}, \\ 1-\varepsilon < \xi_n < 1+\varepsilon \end{cases} \quad \Longrightarrow \quad \begin{cases} \xi_n\eta_n < x, \\ 1-\varepsilon < \xi_n < 1+\varepsilon. \end{cases} \end{aligned} \tag{1.18}$$

Другими словами, события, стоящие слева от знака следствия, влекут события, стоящие справа. Отсюда следуют оценки для вероятности $P_{2,n}(x,\varepsilon) = P(\xi_n\eta_n < x, |\xi_n - 1| < \varepsilon)$:

$$P_{2,n}(x,\varepsilon) \leqslant P\left(\eta_n < \frac{x}{1-\varepsilon}\right) = F_{\eta_n}\left(\frac{x}{1-\varepsilon}\right), \tag{1.19}$$

$$P_{2,n}(x,\varepsilon) \geqslant P\left(\eta_n < \frac{x}{1+\varepsilon}, |\xi_n - 1| < \varepsilon\right). \tag{1.20}$$

Продолжим неравенство (1.20), воспользовавшись следующим простым неравенством: для любых событий A и B

$$P(A \cap B) \geqslant P(A) - P(\overline{B}).$$

Оно следует из того, что $P(A) = P(A \cap B) + P(A \cap \overline{B})$, поэтому

$$P(A \cap B) = P(A) - P(A \cap \overline{B}) \geqslant P(A) - P(\overline{B}).$$

Положив $A = \{(1+\varepsilon)\eta_n < x\}$, $B = \{|\xi_n - 1| < \varepsilon\}$, получим из (1.20)

$$P_{2,n}(x,\varepsilon) \geqslant P\left(\eta_n < \frac{x}{1+\varepsilon}\right) - P(|\xi_n - 1| \geqslant \varepsilon) = \\ = F_{\eta_n}\left(\frac{x}{1+\varepsilon}\right) - P(|\xi_n - 1| \geqslant \varepsilon).$$

Итак, для любого $\varepsilon \in (0,1)$ и для любой точки $x > 0$ непрерывности функции $F_\eta(\,\cdot\,)$

$$F_{\eta_n}\left(\frac{x}{1+\varepsilon}\right) - P(|\xi_n - 1| \geqslant \varepsilon) \leqslant P_{2,n}(x,\varepsilon) \leqslant F_{\eta_n}\left(\frac{x}{1-\varepsilon}\right).$$

Для $x < 0$, как нетрудно показать, справедливы аналогичные оценки, только в обоих неравенствах для η_n в (1.18), а именно в

$$\eta_n < \frac{x}{1-\varepsilon}, \qquad \eta_n < \frac{x}{1+\varepsilon},$$

нужно поменять знак при ε на противоположный.

Если, наконец, $x = 0$ является точкой непрерывности функции $F_\eta(\,\cdot\,)$, то для $\varepsilon \in (0,1)$

$$P_{2,n}(0,\varepsilon) = P(\xi_n\eta_n < 0,\ 1-\varepsilon < \xi_n < 1+\varepsilon) = P(\eta_n < 0).$$

Таким образом, получаем общее неравенство

$$F_{\eta_n}\left(\frac{x}{1\pm\varepsilon}\right) - P(|\xi_n - 1| \geqslant \varepsilon) \leqslant P_{2,n}(x,\varepsilon) \leqslant F_{\eta_n}\left(\frac{x}{1\mp\varepsilon}\right), \quad (1.21)$$

где верхний знак в $\pm$ и $\mp$ отвечает $x > 0$, а нижний отвечает $x < 0$. Значение $x = 0$ можно подставить в любой вариант двойного неравенства, оно останется верным.

Итак, для любого фиксированного $\varepsilon \in (0, 1)$ и для любой точки $x \in \mathbb{R}$ непрерывности функции $F_\eta(\,\cdot\,)$

$$F_{\xi_n \eta_n}(x) = P_{1,n}(x, \varepsilon) + P_{2,n}(x, \varepsilon),$$

где $P_{1,n}(x, \varepsilon) \to 0$ при $n \to \infty$, а $P_{2,n}(x, \varepsilon)$ подчиняется неравенству (1.21). Перейдём к пределу $n \to \infty$, учитывая, что $\eta_n \xrightarrow{\mathrm{d}} \eta$ и $P(|\xi_n - 1| \geqslant \varepsilon) \to 0$. Последовательность $\{F_{\xi_n \eta_n}(x)\}_{n=\overline{1,\infty}}$ может не иметь предела, поэтому мы получаем

$$F_\eta\left(\frac{x}{1 \pm \varepsilon}\right) \leqslant \liminf_{n \to \infty} F_{\xi_n \eta_n}(x) \leqslant \limsup_{n \to \infty} F_{\xi_n \eta_n}(x) \leqslant F_\eta\left(\frac{x}{1 \mp \varepsilon}\right).$$

Теперь будем считать, что ε настолько мало, что $F_\eta(\,\cdot\,)$ непрерывна на отрезке $[x - \varepsilon, x + \varepsilon]$, и устремим ε к нулю. Получим

$$F_\eta(x) \leqslant \liminf_{n \to \infty} F_{\xi_n \eta_n}(x) \leqslant \limsup_{n \to \infty} F_{\xi_n \eta_n}(x) \leqslant F_\eta(x),$$

следовательно, $F_\eta(x) = \lim\limits_{n \to \infty} F_{\xi_n \eta_n}(x)$. Предложение доказано.

1.2. Выборочные характеристики случайных величин

Исходными данными задачи математической статистики являются результаты измерений, представленные в виде упорядоченного конечного набора числовых значений $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Эти данные называются *выборкой объема* n. Основное предположение математической статистики состоит в том, что для каждого $k = 1, \ldots, n$ число x_k есть одно из возможных значений с.в. ξ_k. С.в. $\xi_1, \ldots, \xi_n$ считаются независимыми и одинаково распределёнными. Другими словами, мы рассматриваем n-мерную с.в. $\xi = (\xi_1, \ldots, \xi_n)$ с независимыми и одинаково распределёнными составляющими, которая также называется выборкой (из этого распределения).

Сначала рассмотрим выборку ξ объёма $n = 1$ (одномерную с.в.). Назовём *реализацией* этой с.в. любое число $x \in \mathbb{R}$, такое что $p(x) \neq 0$ для абсолютно непрерывного распределения (здесь $p(\,\cdot\,)$ – это плотность вероятности с.в. ξ) или $P(\xi = x) \neq 0$ для дискретного распределения. Множество всех возможных для заданного распределения реализаций мы будем обозначать как R. Вероятность того, что результат наблюдения попадёт в множество R, равна единице:

для абсолютно непрерывного и дискретного распределений соответственно

$$\int_{\mathsf{R}} p(x)\,dx = 1 \quad \text{и} \quad \sum_{x\in\mathsf{R}} P(\xi = x) = 1. \tag{1.22}$$

Интеграл и сумма по $x \notin \mathsf{R}$, разумеется, равны нулю.

Для n-мерной выборки $\xi = (\xi_1, \dots, \xi_n)$ множество реализаций лежит в $\mathbb{R}^n$. Если каждая из с.в. ξ_k, $k = 1, \dots, n$, имеет плотность вероятности $p(\,\cdot\,)$, то

$$\mathsf{R} = \{x = (x_1, \dots, x_n)\colon p(x_k) > 0 \text{ для всех } k = 1, \dots, n\};$$

если распределение каждой из с.в. ξ_k, $k = 1, \dots, n$, дискретно, то

$$\mathsf{R} = \{x = (x_1, \dots, x_n)\colon P(\xi_k = x_k) > 0 \text{ для всех } k = 1, \dots, n\}.$$

В статистике множество R также называют генеральной совокупностью, и можно сказать, что именно из генеральной совокупности выбираются (отсюда термин *выборка*) любые результаты наблюдений.

Определение 1. Любая функция $t(\xi)$ от выборки, являющаяся с.в. и не содержащая неизвестных параметров, называется *статистикой*. Как правило, рассматривают одномерные статистики, т. е. $t(\,\cdot\,)\colon \mathbb{R}^n \to \mathbb{R}$.

По выборке можно строить самые разнообразные статистики. Рассмотрим наиболее часто встречающиеся. Всюду далее в этом разделе $\xi = (\xi_1, \dots, \xi_n)$ – выборка объема $n \geqslant 1$.

Выборочные моменты

Выборочным средним называют статистику

$$\bar{\xi} = \frac{1}{n}\sum_{k=1}^{n} \xi_k,$$

выборочным моментом k-го порядка – статистику

$$\overline{\xi^k} = \frac{1}{n}\sum_{i=1}^{n} \xi_i^k, \qquad k = 1, 2, \dots .$$

Далее будем аналогично обозначать чертой сверху любую статистику вида

$$\overline{t(\xi)} = \frac{1}{n}\sum_{k=1}^{n} t(\xi_k).$$

Выборочная дисперсия может быть определена по-разному, например как

$$S^2 = \overline{\xi^2} - (\bar{\xi})^2 = \frac{1}{n}\sum_{k=1}^{n}\xi_k^2 - \left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)^2,$$

но чаще всего рассматривают статистику (при $n > 1$)

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(\xi_k - \bar{\xi})^2.$$

Значение выборочных моментов в математической статистике состоит в том, что при определенных условиях они сходятся к соответствующим моментам с.в. ξ в силу законов больших чисел.

Заметим, что выборочное среднее имеет свойства, идентичные свойствам математического ожидания: как нетрудно показать,

$$\overline{at(\xi)+b} = a\overline{t(\xi)} + b, \qquad \overline{b\cdot t(\xi)} = b\cdot\overline{t(\xi)}$$

для любых постоянных $a, b \in \mathbb{R}$. Также можно показать (вычисления см. в примере 4 в разделе 2.2 следующей главы), что

$$\sum_{k=1}^{n}(\xi_k - \bar{\xi})^2 = \sum_{k=1}^{n}\xi_k^2 - \frac{1}{n}\sum_{k,j=1}^{n}\xi_k\xi_j,$$

откуда

$$\overline{(\xi_k - \bar{\xi})^2} = \overline{\xi^2} - (\bar{\xi})^2;$$

сравните с $M(\xi - M\xi)^2 = M\xi^2 - (M\xi)^2$.

Выборочное распределение

Для любого множества $B \subset \mathbb{R}$ зададим функцию $I_B(\,\cdot\,)$ как

$$I_B(\,\cdot\,)\colon \mathbb{R} \to \{0,1\}, \quad I_B(x) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B. \end{cases}$$

Она называется индикатором или индикаторной функцией множества B. Число элементов выборки $\xi_1, \ldots, \xi_n$, принадлежащих множеству B, и частота попадания реализаций выборки в множество B (которые, очевидно, являются с.в.) соответственно равны

$$\sum_{k=1}^{n} I_B(\xi_k), \qquad \frac{1}{n}\sum_{k=1}^{n} I_B(\xi_k) \overset{\text{def}}{=} P_n^*(B). \tag{1.23}$$

(Случайную) величину $P_n^*(B)$ естественно назвать эмпирической вероятностью попасть в множество B. Положим $p_B = P(\xi_k \in B)$ для теоретической вероятности попасть в множество B; она, разумеется, не зависит от $k = 1, \ldots, n$. Связь между эмпирической и теоретической вероятностями даётся следующим простым утверждением.

Предложение 3. *С ростом объема выборки $n \to \infty$ эмпирическая вероятность $P_n^*(B)$ сходится по вероятности к p_B.*

Доказательство . Для всех $k = 1, \ldots, n$ введём с.в. $\chi_k = I_B(\xi_k)$. По условию с.в. $\xi_1, \ldots, \xi_n$ независимы и одинаково распределены, следовательно, с.в. $\chi_1, \ldots, \chi_n$ тоже независимы и одинаково распределены,

$$\begin{aligned} &P(\chi_k = 1) = P(\xi_k \in B) = p_B, \\ &P(\chi_k = 0) = P(\xi_k \notin B) = q_B = 1 - p_B. \end{aligned}$$

При этом $M\chi_k = p_B$, $D\chi_k = p_B q_B$. Тогда согласно закону больших чисел (или усиленному закону больших чисел)

$$\lim_{n\to\infty} P_n^*(B) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1} \chi_k = p_B,$$

где сходимость понимается как сходимость по вероятности (или соответственно с вероятностью единица, т. е. почти наверное).

Более того, согласно центральной предельной теореме имеется сходимость по распределению:

$$\sum_{k=1}^{n} \frac{\chi_k - p_B}{\sqrt{n p_B q_B}} \xrightarrow[n\to\infty]{\text{d}} \nu_* \sim \mathbf{N}(0,1).$$

Если вернуться к исходному обозначению из (1.23),

$$P_n^*(B) = \frac{1}{n}\sum_{k=1}^{n} I_B(\xi_k) = \frac{1}{n}\sum_{k=1}^{n}\chi_k,$$

то

$$\sum_{k=1}^{n}\frac{\chi_k - p_B}{\sqrt{np_Bq_B}} = \frac{n}{\sqrt{np_Bq_B}}(P_n^*(B) - p_B) = \sqrt{\frac{n}{p_Bq_B}}\cdot(P_n^*(B) - p_B).$$

Центральная предельная теорема позволяет утверждать, что при больших объёмах выборки n

$$1 = D\nu_* \approx D\left(\sqrt{\frac{n}{p_Bq_B}}\cdot(P_n^*(B) - p_B)\right) = \frac{n}{p_Bq_B}DP_n^*(B),$$

поэтому дисперсия выборочной вероятности

$$DP_n^*(B) \approx \frac{p_Bq_B}{n}.$$

Если в качестве множества B выбирать полупрямые $(-\infty, x)$ для различных $x \in \mathbb{R}$, то получим выборочную функцию распределения с.в. ξ:

$$F_\xi^*(x) = \frac{1}{n}\sum_{k:\ \xi_k < x} I_{(-\infty,x)}(\xi_k).$$

Вариационный ряд и порядковые статистики

Пусть дана выборка[2)] $x = (x_1, \ldots, x_n)$. Чтобы построить выборочную функцию распределения $F^*(\,\cdot\,)$, удобно расположить выборочные значения в порядке неубывания. Обозначив члены переупорядоченной последовательности символами $x_{(k)}$, в которых $k = 1, \ldots, n$, получим так называемый *вариационный ряд* выборки

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)};$$

каждый член этого ряда есть функция от выборочных значений $x_1, \ldots, x_n$. Аналогичное упорядочение (независимых и одинаково

[2)] Как мы отмечали в предыдущем разделе, выборка – это не только n с.в., но и n чисел.

распределённых) с.в. $\xi_1, \dots, \xi_n$ для каждой реализации $x_1, \dots, x_n$ даёт набор из n с.в.

$$\xi_{(1)} \leqslant \xi_{(2)} \leqslant \dots \leqslant \xi_{(n)},$$

который тоже называют вариационным рядом. Члены этого ряда называются *порядковыми статистиками*: на k-м месте стоит k-я порядковая статистика $\xi_{(k)}$, случайная величина $\xi_{(1)}$ называется минимальной, а $\xi_{(n)}$ – максимальной порядковой статистикой. Понятно, что каждая порядковая статистика является некоторой функцией от всех с.в. $\xi_1, \dots, \xi_n$.

Порядковая статистика, стоящая посередине вариационного ряда, называется *выборочной медианой* распределения: если n – нечётное число, то выборочная медиана – это $\xi_{(\bar{k})}$, где $\bar{k} = \frac{n+1}{2}$. Таким образом, половина порядковых статистик лежит левее медианы, половина – правее. Можно показать, что медиана набора чисел – это число, сумма расстояний от которого до всех чисел из набора минимальна. Для чётного n медиану чаще всего задают как полусумму $(\xi_{(n/2)} + \xi_{(n/2+1)})/2$ двух стоящих в середине ряда порядковых статистик.

Пример 1. Пусть в результате измерений получена выборка $1, 6, 7, 4, 5, 3, 4, 2, 5, 4, 5, 1, 5, 6, 6$. Упорядочив этот набор, получим вариационный ряд $1, 1, 2, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 7$. Выборочная функция распределения тогда принимает вид

$$F^*_\xi(x) = \frac{1}{n} \max\{j\colon \xi_{(j)} < x\}.$$

Это кусочно-постоянная функция, непрерывная слева,

$$F^*(x) = \begin{cases} 0, & x \leqslant 1, \\ 2/15, & 1 < x \leqslant 2, \\ 1/15, & 2 < x \leqslant 3, \\ & \vdots \\ 1, & x > 7. \end{cases}$$

Предложение 4. *Для каждого $k = 1, \dots, n$ функция распределения k-й порядковой статистики имеет вид*

$$P(\xi_{(k)} < x) = \sum_{j=k}^{n} C_n^j F^j(x)(1 - F(x))^{n-j}, \tag{1.24}$$

где $F(\cdot)$ – функция распределения с.в. ξ.

Доказательство. Рассмотрим сначала $k = n$. Понятно, что

$$\max_{j=1,\ldots,n} \xi_j < x \quad \Longleftrightarrow \quad \xi_j < x \quad \text{для всех} \quad j = 1, \ldots, n.$$

Отсюда получаем

$$P(\xi_{(n)} < x) = P\left(\bigcap_{j=1}^{n} \{\xi_j < x\}\right) = \prod_{j=1}^{n} P(\xi_j < x) = \prod_{j=1}^{n} F(x),$$

где мы воспользовались тем, что все выборочные элементы ξ_k независимы и имеют одну и ту же функцию распределения $F(\cdot)$. Таким образом,

$$P(\xi_{(n)} < x) = F^n(x),$$

что соответствует (1.24) при $k = n$. Функция распределения максимальной порядковой статистики равна n-й степени функции распределения с.в. ξ.

Аналогично для $k = 1$ в силу того, что

$$\min_{k=1,\ldots,n} \xi_k \geqslant x \quad \Longleftrightarrow \quad \xi_k \geqslant x \quad \text{для всех} \quad j = 1, \ldots, n,$$

имеем

$$P(\xi_{(1)} \geqslant x) = P\left(\bigcap_{k=1}^{n} \{\xi_k \geqslant x\}\right) = \prod_{k=1}^{n} P(\xi_k \geqslant x) =$$

$$= \prod_{k=1}^{n} (1 - F(x)) = (1 - F(x))^n.$$

Запишем бином Ньютона:

$$1 = \big(F(x) + 1 - F(x)\big)^n = \sum_{j=0}^{n} C_n^j F^j(x)(1 - F(x))^{n-j} =$$

$$= (1 - F(x))^n + \sum_{j=1}^{n} C_n^j F^j(x)(1 - F(x))^{n-j}.$$

С учётом этих равенств

$$P(\xi_{(1)} < x) = 1 - P(\xi_{(1)} \geqslant x) = 1 - (1 - F(x))^n = \\ = \sum_{j=1}^{n} C_n^j F^j(x)(1 - F(x))^{n-j},$$

что соответствует (1.24) при $k = 1$.

Для вычисления функции распределения k-й порядковой статистики заметим, что событие $\{\xi_{(k)} < x\}$ эквивалентно тому, что k или более выборочных значений из n были строго меньше x. Вероятность этого события вычисляется с использованием формулы Бернулли: нужно, чтобы в серии из n независимых испытаний было получено не менее k успехов. Успехом мы считаем попадание выборочного значения в интервал $(-\infty, x)$, вероятность успеха, очевидно, равна $F(x)$. В результате имеем искомую формулу

$$P(\xi_{(k)} < x) = \sum_{j=k}^{n} C_n^j F^j(x)(1 - F(x))^{n-j}.$$

2. Точечные оценки параметров распределений

Предположим, что распределение выборки известно с точностью до одного или нескольких неизвестных исследователю параметров. Наша цель – по выборке высказаться (в той или иной форме) о значениях этих параметров. Такие задачи называются задачами параметрической статистики.

Обозначим неизвестные параметры как $(\theta_1, \ldots, \theta_s) = \theta \in \mathbb{R}^s$, $s \geqslant 1$. Предполагается, что задано множество Θ всех возможных значений параметра θ, и в нём существует единственное значение θ, которое определяет распределение выборки (мы называем это значение истинным); исследователю оно неизвестно.

При этих условиях требуется построить решающее правило, которое по выборке $\xi = (\xi_1, \ldots, \xi_n)$ делает вывод о значении $\theta \in \Theta$. Этот вывод может быть сформулирован как:

1) некоторое значение $\hat{\theta} = \hat{\theta}(\xi_1, \ldots, \xi_n)$, в определённом смысле близкое к истинному θ; в этом случае речь идет о точечном оценивании параметра θ, которому посвящена настоящая глава;

2) множество $\widehat{\Theta}(\xi) \subset \mathbb{R}^s$ с границами, определяющимися выборкой, которое с заданной вероятностью накрывает (содержит) истинное значение θ; в этом случае речь идет о задачах интервального оценивания, которым посвящена глава 3;

3) ответ на вопрос о том, верно или неверно наше предположение о том или ином значении параметра θ; в этом случае речь идет о задачах проверки статистических гипотез, которые будут рассмотрены в следующей части нашего пособия.

Окончательный ответ на поставленную задачу получается подстановкой вместо $\xi_1, \ldots, \xi_n$ набора реализаций $x_1, \ldots, x_n$, полученных в эксперименте.

2.1. Функция правдоподобия

Введём следующее важное понятие. Пусть распределение одномерной с.в. ξ зависит от параметра $\theta \in \Theta \subset \mathbb{R}^s$ и либо абсолютно непрерывно (с плотностью вероятности $p(x;\theta)$, $x \in \mathbb{R}$), либо дискретно.

Определение 1. *Функцией правдоподобия* одномерной с.в. ξ называется функция $L(\,\cdot\,) : \mathbb{R} \times \Theta \to \mathbb{R}_+$, заданная как

$$L(x,\theta) = \begin{cases} p(x;\theta) & \text{для абсолютно непрерывного распределения,} \\ P_\theta(\xi = x) & \text{для дискретного распределения,} \end{cases}$$

где $x \in \mathbb{R}$, $\theta \in \Theta$.

В этом определении и далее нижний индекс в P_θ указывает на зависимость вероятностей тех или событий от θ.

Введение функции правдоподобия связано с тем, что в задачах математической статистики более важна зависимость распределения от параметра, чем от своей «естественной» переменной, т. е. поведение $L(\,\cdot\,)$ как функции от θ при фиксированном x. Теперь мы можем записать вероятность того, что с.в. ξ попадёт в множество $X \subset \mathbb{R}^n$, как

$$P_\theta(\xi \in X) = \int\limits_X L(x,\theta)\,dx \quad \text{или} \quad P_\theta(\xi \in X) = \sum_{x \in X} L(x,\theta) \tag{2.1}$$

для соответственно абсолютно непрерывного или дискретного распределения.

Далее примем следующее соглашение: в наших теоретических выкладках для краткости мы будем использовать только первое равенство в (2.1) для абсолютно непрерывного распределения, помня, что в конкретных вычислениях для дискретного распределения интеграл необходимо заменить суммой, а плотность – вероятностью. С учётом этого соглашения для абсолютно непрерывного распределения

$$L(x,\theta)\,dx = P_\theta(x \leqslant \xi < x + dx), \tag{2.2}$$

и значение $L(x,\theta)$ тем больше, чем больше вероятность того, что с.в. ξ, имеющая распределение с параметром θ, примет значение x (точнее, окажется в бесконечно малой окрестности точки $x \in \mathbb{R}$).

Введённое выше множество R реализаций с.в. ξ может зависеть от $\theta \in \Theta$,

$$\mathsf{R}_\theta = \{x \in \mathbb{R}\colon\ L(x,\theta) \neq 0\},$$

и для любого $\theta \in \Theta$

$$P_\theta(\xi \in \mathsf{R}_\theta) = \int\limits_{\mathsf{R}_\theta} L(x,\theta)\,dx = 1, \quad P_\theta(\xi \notin \mathsf{R}_\theta) = \int\limits_{x\colon L(x,\theta)=0} L(x,\theta)\,dx = 0.$$

Посмотрим, как выглядят функция правдоподобия и множество реализаций для известных одномерных распределений. Всюду далее указаны только ненулевые значения функции правдоподобия (т.е. значения при $x \in \mathsf{R}_\theta$, $\theta \in \Theta$). При прочих значениях обоих своих аргументов (вне множества $\mathsf{R}_\theta \times \Theta$) функция правдоподобия равна нулю. В качестве множества Θ выбрано самое широкое для данного распределения, т.е. ограниченное только условием существования данного распределения.

- Нормальное распределение $\mathbf{N}(\mu,\sigma^2)$:

$$L(x,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad \theta = (\mu,\sigma^2),$$
$$\mathsf{R}_\theta = \mathbb{R}, \qquad \Theta = \big\{-\infty < \mu < \infty,\ 0 < \sigma^2 < \infty\big\}.$$

- Экспоненциальное (показательное) распределение $\mathbf{E}(a)$:

$$L(x,\theta) = ae^{-ax}, \qquad \theta = a,$$
$$\mathsf{R}_\theta = \mathbb{R}_+, \qquad \Theta = \big\{0 < a < \infty\big\}.$$

- Равномерное распределение $\mathbf{U}[a,b]$ (на отрезке $[a,b]$):

$$L(x,\theta) = \frac{1}{b-a}, \qquad \theta = (a,b),$$
$$\mathsf{R}_\theta = [a,b], \qquad \Theta = \big\{-\infty < a < b < \infty\big\}.$$

- Распределение Пуассона $\mathbf{P}(\lambda)$:

$$L(k,\theta) = \frac{\lambda^k}{k!}e^{-\lambda}, \qquad \theta = \lambda,$$
$$\mathsf{R}_\theta = \big\{0,1,2,\ldots\big\}, \qquad \Theta = \big\{0 < \lambda < \infty\big\}.$$

- Биномиальное распределение $\mathbf{B}(m,p)$:

$$L(k,\theta) = C_m^k p^k q^{m-k}, \qquad \theta = (m,p),$$
$$\mathsf{R}_\theta = \bigl\{0,1,2,\dots m\bigr\}, \quad \Theta = \bigl\{m = 1,2\dots,\, p \in (0,1)\bigr\}.$$

Заметим, что, кроме равномерного и биномиального распределений, для всех остальных распределений множество реализаций не зависит от $\theta \in \Theta$, в таких случаях мы пишем $\mathsf{R}_\theta = \mathsf{R}$.

Перейдём к функции правдоподобия выборки объёма $n > 1$. Она выводится из условия, что составляющие выборки независимы и одинаково распределены, и является функцией из $\mathbb{R}^n \times \Theta$ в $\mathbb{R}_+$. Как получить конкретные выражения, легко понять из следующих примеров. Во всех примерах множество Θ выбрано так же, как в соответствующих формулах выше.

Пример 1. Пусть ξ – n-мерная выборка из распределения Пуассона, $\xi \sim \mathbf{P}(\lambda)$. Тогда функция правдоподобия выборки равна

$$L(m,\lambda) \stackrel{\text{def}}{=} P_\lambda(\xi_1 = m_1, \dots, \xi_n = m_n) =$$
$$= \prod_{k=1}^{n} \frac{\lambda^{m_k}}{m_k!}\, e^{-\lambda} = \frac{\lambda^{\sum_{k=1}^n m_k}}{\prod\limits_{k=1}^{n} m_k!}\, e^{-n\lambda}, \qquad 0 < \lambda < \infty.$$

Здесь $m = (m_1,\dots,m_n)$ и $m_k \in \{0,1,2,\dots\}$ для всех $k = 1,\dots,n$, иначе $L(m,\lambda) = 0$.

Пример 2. Пусть ξ – выборка объема n из нормального распределения $\mathbf{N}(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2)$. Тогда аналогично предыдущему примеру функция правдоподобия выборки равна

$$L(x,\theta) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x_k-\mu)^2/2\sigma^2} =$$
$$= \frac{1}{(2\pi\sigma^2)^{n/2}}\, e^{-\sum\limits_{k=1}^{n}(x_k-\mu)^2/2\sigma^2}, \qquad \mu \in \mathbb{R}, \quad \sigma^2 > 0.$$

Здесь $x = (x_1,\dots,x_n) \in \mathbb{R}^n$.

Пример 3. Пусть ξ – выборка объема n. Каждая случайная величина ξ_k для $k = 1,\dots,n$ принимает значение 1 с вероятностью θ

и значение 0 с вероятностью $1-\theta$, $0<\theta<1$. Запишем функцию правдоподобия выборки там, где она отлична от нуля:

$$L(m,\theta) = P_\theta(\xi_1 = m_1)\dots P_\theta(\xi_n = m_n),$$

где $m_k \in \{0,1\}$ для всех $k = 1,\dots,n$. В этом произведении каждый сомножитель $P_\theta(\xi_k = m_k)$ в функции правдоподобия равен либо θ (если $m_k = 1$), либо $1-\theta$ (если $m_k = 0$). Так как количество сомножителей, равных θ, в точности равно числу составляющих выборки $m_1,\dots,m_n$, равных единице, а остальные составляющие равны нулю, мы можем найти это количество как

$$m_1 + \dots + m_n = n\overline{m}, \quad \text{где} \quad \overline{m} = \frac{1}{n}\sum_{k=1}^{n} m_k,$$

Отсюда

$$L(m,\theta) = \theta^{n\overline{m}}(1-\theta)^{n-n\overline{m}}, \qquad 0<\theta<1.$$

Перейдём к решению задач параметрической статистики и начнём с задач точечного оценивания.

Точечная оценка – это такая статистика $t(\xi) = t(\xi_1,\dots,\xi_n)$, которая в определённом смысле близка к истинному значению параметра θ. Напомним, что статистика не должна зависеть от неизвестных параметров, и тогда мы можем найти её значение по заданной выборке, не зная параметр. Если нам будет важно, что объём выборки равен n, будем использовать для статистики обозначение $t_n(\xi)$.

Рассмотрим более общую задачу точечного оценивания: в которой интерес представляет не само значение θ, а значение, которое принимает заданная функция $\tau(\,\cdot\,)$ от аргумента θ, иначе говоря, нас интересует значение $\tau(\theta)$. Например, если задана выборка из нормального распределения $\mathbf{N}(\mu,\sigma^2)$, то θ можно рассматривать как двумерный параметр (μ,σ^2) и интересоваться значениями $\tau_1(\theta) = \mu^2+\sigma^2$, $\tau_2(\theta) = \mu$ или $\tau_3(\theta) = (\mu,\sigma^2)$ (в последнем случае функция $\tau_3(\theta) = \theta$ – тождественная функция, равная своему аргументу).

Далее мы рассматриваем функции $\tau(\,\cdot\,)\colon \Theta \to \mathbb{R}$, т. е. функции с числовыми значениями, соответственно статистика $t(\xi)$ – это одномерная с.в. В сущности, точечной оценкой значения $\tau(\theta)$ можно назвать любую статистику $t(\xi)$, но понятно, что нас будут интересовать те, которые обладают какими-то полезными и желательными свойствами. Ниже мы дадим формальные определения таких «хороших» свойств оценок.

2.2. Несмещённость и состоятельность точечных оценок

Определение 2. Статистика $t(\xi)$ называется *несмещённой оценкой* значения $\tau(\theta)$ функции $\tau(\,\cdot\,)$ параметра θ, если для любого $\theta \in \Theta$ выполнено равенство $M_\theta\, t(\xi) = \tau(\theta)$.

Здесь и далее индекс θ в обозначении M_θ математического ожидания означает, что математическое ожидание вычисляется по распределению, определяемому значением параметра $\theta \in \Theta$:

$$M_\theta\, t(\xi) = \int_{\mathbb{R}^n} t(x) L(x, \theta)\, dx = \int_{\mathbb{R}^n} t(x) \prod_{k=1}^{n} p(x_k; \theta)\, dx$$

(здесь, как обычно, $dx = dx_1 \dots dx_n$), где $p(\,\cdot\,; \theta)$ – плотность вероятности любой из с.в. ξ_k с параметром θ. Напомним, что мы условились писать формулы, отвечающие абсолютно непрерывному распределению с.в. ξ, тогда функция правдоподобия – это совместная плотность вероятности с.в. $\xi_1, \dots, \xi_n$. В случае дискретного распределения интегралы, конечно, нужно заменять суммами.

Примером несмещённой оценки математического ожидания любой из (одинаково распределённых) с.в. ξ_k является выборочное среднее $\bar{\xi} = \frac{1}{n}\sum_{k=1}^{n} \xi_k$. Действительно, если это математическое ожидание существует и равно μ, то в силу линейности математического ожидания для любого μ

$$M_\mu\, \bar{\xi} = M_\mu \left(\frac{1}{n} \sum_{k=1}^{n} \xi_k \right) = \frac{1}{n} \sum_{k=1}^{n} M_\mu\, \xi_k = \mu.$$

Оценка может не являться несмещённой, но её математическое ожидание может стремиться к оцениваемой величине при стремлении объёма выборки к бесконечности.

Определение 3. Статистика $t_n(\xi)$ называется *асимптотически несмещённой оценкой* для $\tau(\theta)$, если для любого $\theta \in \Theta$ имеет место сходимость $M_\theta\, t_n(\xi) \to \tau(\theta)$ при $n \to \infty$, где значение n равно объёму выборки.

Пример 4. Пусть $\mu = M\xi_k$, $\sigma^2 = D\xi_k$. Найдём множитель C, при котором статистика

$$C \cdot \sum_{k=1}^{n} (\xi_k - \bar{\xi})^2$$

является несмещённой оценкой параметра σ^2.

Простые алгебраические преобразования с учётом равенства $\sum_{k=1}^{n} (\xi_k - \mu) = n(\bar{\xi} - \mu)$ дают

$$\begin{aligned}
\sum_{k=1}^{n} (\xi_k - \bar{\xi})^2 &= \sum_{k=1}^{n} [(\xi_k - \mu) - (\bar{\xi} - \mu)]^2 = \\
&= \sum_{k=1}^{n} (\xi_k - \mu)^2 - 2n(\bar{\xi} - \mu)(\bar{\xi} - \mu) + n(\bar{\xi} - \mu)^2 = \\
&= \sum_{k=1}^{n} (\xi_k - \mu)^2 - n(\bar{\xi} - \mu)^2.
\end{aligned}$$

Вычисляем математические ожидания двух выражений в правой части:

$$\begin{aligned}
&M \sum_{k=1}^{n} (\xi_k - \mu)^2 = \sum_{k=1}^{n} M(\xi_k - \mu)^2 = n\sigma^2, \\
&M(\bar{\xi} - \mu)^2 = M\left(\frac{1}{n} \sum_{k=1}^{n} (\xi_k - \mu)\right)^2 = \frac{1}{n^2} M \sum_{k=1}^{n} (\xi_k - \mu) \sum_{j=1}^{n} (\xi_j - \mu) = \\
&\qquad = \frac{1}{n^2} \left(\sum_{k=1}^{n} M(\xi_k - \mu)^2 + \sum_{k \neq j} M(\xi_k - \mu)(\xi_j - \mu) \right) = \\
&\qquad = \frac{1}{n^2} n\sigma^2 = \frac{1}{n} \sigma^2,
\end{aligned}$$

где на последнем этапе мы учли, что в силу независимости элементов выборки $M(\xi_k - \mu)M(\xi_j - \mu) = 0$ при $k \neq j$. Отсюда

$$M\left(\sum_{k=1}^{n} (\xi_k - \bar{\xi})^2 \right) = M \sum_{k=1}^{n} (\xi_k - \mu)^2 - nM(\bar{\xi} - \mu)^2 = (n\sigma^2 - \sigma^2),$$

следовательно, $C = \frac{1}{n-1}$, т. е. несмещённой оценкой дисперсии является

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(\xi_k - \bar{\xi})^2,$$

введённая нами в разделе 1.2.

Понятно, что статистика

$$S^2 = \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \bar{\xi})^2 = \overline{\xi^2} - (\bar{\xi})^2$$

не является несмещённой оценкой дисперсии, её математическое ожидание равно $\frac{n-1}{n}\sigma^2$, но эта оценка является асимптотически несмещённой.

Заметим также, что и $\hat{\sigma}^2$, и S^2 не зависят от математического ожидания μ элементов выборки. Следовательно, их численное значение можно найти, не зная μ, в отличие от, например, функции $\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \mu)^2$, которая также является несмещённой оценкой дисперсии, но совершенно бесполезна для решения нашей задачи, если мы не знаем μ.

Замечание 1. Требование несмещённости может оказаться невыполнимым. Приведём пример. Пусть ξ – число успехов в серии n независимых испытаний, θ – вероятность успеха при единичном испытании. Требуется несмещённо оценить $\tau(\theta) = 1/\theta$ по выборке $\xi = k$ объема единица, т. е. в единственной серии из n испытаний получено k успехов. Условие несмещённости оценки $t(\xi)$ принимает вид

$$M_\theta\, t(\xi) = \sum_{j=0}^{n} t(k) C_n^k \theta^k (1-\theta)^{n-k} = \theta^{-1}, \quad \theta \in [0,1].$$

Это равенство для всех $\theta \in [0,1]$ невыполнимо при любой функции $t(k)$, $k = 0, 1, \ldots, n$, ибо левая часть равенства – полином от аргумента θ, его значение ограничено в точке $\theta = 0$, а правая часть при $\theta \to 0$ стремится к бесконечности.

Несмещённость оценки по сути означает отсутствие систематической погрешности оценивания, но при этом мы ничего не можем сказать о близости оценки к истинному значению параметра. С этой точки зрения важно следующее свойство.

Определение 4. Статистика $t(\xi_1, \ldots, \xi_n) = t_n(\xi)$ называется *состоятельной оценкой* для $\tau(\theta)$, если при любом $\theta \in \Theta$ имеет место сходимость по вероятности $t_n(\xi) \to \tau(\theta)$ при $n \to \infty$.

Заметим, что если в определении асимптотически несмещённой оценки речь идёт о сходимости числовой последовательности математических ожиданий $Mt_n(\xi)$ статистики $t_n(\xi)$, то в определении состоятельности оценки фигурирует сходимость по вероятности последовательности самих оценок $t_n(\xi)$ при $n \to \infty$. Например, состоятельность оценки $t_n(\xi) = \bar{\xi}$ математического ожидания с.в. ξ в случае существования $D\xi_j$ является следствием закона больших чисел.

2.3. Несмещённые оценки минимальной дисперсии

Свойства несмещённости и состоятельности оценок не определяют близость оценки $t_n(\xi)$ к оцениваемой величине $\tau(\theta)$ при фиксированном объёме выборки. Чтобы численно охарактеризовать погрешность оценивания, будем использовать отличие оценки от оцениваемой величины в среднем квадратичном (с.к.).

Будем говорить, что оценка $t(\xi)$ лучше оценки $\tilde{t}(\xi)$ в с.к., если для любого $\theta \in \Theta$

$$M_\theta(\tau(\theta) - t(\xi))^2 \leqslant M_\theta(\tau(\theta) - \tilde{t}(\xi))^2$$

и хотя бы при одном $\theta \in \Theta$ это неравенство строгое.

Существует ли оценка, наилучшая в с.к. для всех $\theta \in \Theta$? Если не ограничивать класс возможных оценок, то ответ на этот вопрос отрицательный. Действительно, пусть $\tau(\theta) = \theta$ и $t(\xi)$ – наилучшая в с.к. оценка для θ, т.е. для любой другой оценки $\tilde{t}(\xi)$ при любом $\theta \in \Theta$ выполнено неравенство

$$M_\theta(t(\xi) - \theta)^2 \leqslant M_\theta(\tilde{t}(\xi) - \theta)^2.$$

Пусть θ_1 – произвольный элемент множества Θ. Рассмотрим статистику $t_1(\xi) \equiv \theta_1$, не зависящую от выборки ξ. Тогда в силу предположения

$$M_\theta(t(\xi) - \theta)^2 \leqslant M_\theta(t_1(\xi) - \theta)^2 = M_\theta(\theta_1 - \theta)^2, \qquad \theta \in \Theta.$$

В частности, при $\theta = \theta_1$ получим $M_\theta(t(\xi) - \theta)^2 \leqslant 0$, что, очевидно, равносильно $M_\theta(t(\xi) - \theta)^2 = 0$. Равенство нулю математического ожидания неотрицательной с.в. возможно только в случае, когда она равна нулю с вероятностью единица, т. е. когда $t(\xi) = \theta$ для любого $\theta \in \Theta$ (оценка в точности отгадывает неизвестный параметр). Это в свою очередь возможно только для неинтересной с точки зрения математической статистики задачи, в которой множество Θ состоит из одного априори известного значения параметра.

Однако если сузить класс точечных оценок, рассматривая только несмещённые оценки для $\tau(\theta)$, то в ряде случаев можно указать наилучшую в с.к. несмещённую оценку. Поскольку для несмещённых оценок $M_\theta\, t(\xi) = \tau(\theta)$, с.к. погрешность такой оценки равна её дисперсии: $M_\theta(t(\xi) - \tau(\theta))^2 = D_\theta(t(\xi)$.

Всюду далее мы рассматриваем оценки с конечным вторым моментом $M_\theta\, t^2(\xi)$ или, что то же самое, с конечной дисперсией, поскольку бесконечная дисперсия означает бесконечную с.к. погрешность оценивания, и полезность такой оценки невелика.

Определение 5. Оценка, наилучшая в с.к. в классе несмещённых оценок, называется *несмещённой оценкой минимальной дисперсии* (НОМД).

Формально НОМД $t^*(\xi)$ является общим для всех $\theta \in \Theta$ решением задачи на условный экстремум

$$M_\theta(t^*(\xi) - \tau(\theta))^2 = \min_{t:\ M_\theta t(\theta) = \tau(\theta)} M_\theta(t(\xi) - \tau(\theta))^2. \tag{2.3}$$

Такие оценки весьма привлекательны, поскольку они не содержат систематической погрешности и имеют минимальный в с.к. разброс значений вокруг оцениваемой величины. Однако в общем случае найти такую оценку нелегко.

Замечание 2. Требование несмещённости не может уменьшить с.к. ошибку оценивания. Действительно, если K – некоторый класс точечных оценок, то

$$\min_{t \in K} M_\theta(t(\xi) - \tau(\theta))^2 \leqslant \min_{\substack{t(\xi) \in K, \\ M_\theta t(\xi) = \tau(\theta)}} M_\theta(t(\xi) - \tau(\theta))^2,$$

так как в правой части этого неравенства минимум ищется по более узкому классу $K \cap \{t(\xi) \colon\ M_\theta\, t(\xi) = \tau(\theta)\}$.

Теорема 1. *Если оценки $t_1(\xi)$ и $t_2(\xi)$ являются НОМД для значения $\tau(\theta)$, то $t_1(\xi) = t_2(\xi)$ с вероятностью единица.*

Доказательство. Для краткости формул не будем писать нижний индекс θ у математических ожиданий и дисперсий, поскольку все наши расуждения верны для любого θ.

Образуем новую оценку

$$t(\xi) = \frac{t_1(\xi) + t_2(\xi)}{2},$$
$$Mt(\xi) = Mt_1(\xi) = Mt_2(\xi) = \tau(\theta),$$

т. е. $t(\xi)$, как и $t_1(\xi)$, $t_2(\xi)$, – несмещённая оценка для $\tau(\theta)$. Вычислим её дисперсию:

$$Dt(\xi) = \frac{Dt_1(\xi) + Dt_2(\xi) + 2\operatorname{cov}(t_1(\xi), t_2(\xi))}{4}.$$

Обозначим дисперсию НОМД как $Dt_1(\xi) = Dt_2(\xi) = d^2 = d^2(\theta)$, очевидно, что дисперсии обеих оценок равны, так обе они совпадают с минимальной дисперсией. Далее, пользуясь неравенством Коши–Буняковского для ковариаций $|\operatorname{cov}(t_1(\xi), t_2(\xi))| \leqslant \sqrt{d^2 \cdot d^2}$, получим

$$Dt(\xi) \leqslant \frac{d^2 + d^2 + 2d^2}{4} = d^2.$$

Однако $Dt(\xi)$ не может быть меньше d^2, так как это минимальное возможное значение дисперсии несмещённой оценки. Следовательно, выполняется равенство

$$d^2 = Dt(\xi) = \frac{d^2 + d^2 + 2\operatorname{cov}(t_1(\xi), t_2(\xi))}{4},$$

а это равносильно тому, что $\operatorname{cov}(t_1(\xi), t_2(\xi)) = \sqrt{d^2 \cdot d^2}$. По свойствам неравенства Коши–Буняковского тогда с вероятностью единица выполняется равенство

$$t(\xi) - \tau(\theta) = a(t_2(\xi) - \tau(\theta)),$$

где коэффициент a неслучаен, но может зависеть от θ. Но тогда $d^2 = \operatorname{cov}(t_1(\xi), t_2(\xi)) = ad^2$, откуда $a = 1$, и поэтому $t_1(\xi) = t_2(\xi)$ с вероятностью единица.

Докажем одно вспомогательное утверждение из теории вероятностей.

Лемма 1. *Рассмотрим класс с.в.*

$$\mathcal{H} = \{\eta\colon M\eta = 0,\ M\eta^2 < \infty\}.$$

С.в. $\alpha_* \in \mathcal{H}$ *удовлетворяет условию* $M\alpha_*^2 \leqslant M\alpha^2$ *для любой* $\alpha \in \mathcal{H}$ *тогда и только тогда, когда* $M\alpha_*\eta = 0$ *для всех* $\eta \in \mathcal{H}$.

Доказательство. Возьмём любую $\alpha \in \mathcal{H}$ и положим $\eta = \alpha - \alpha_*$. Тогда $M\eta = M\alpha - M\alpha_* = 0$, и по неравенству Коши–Буняковского

$$M\eta^2 = M\alpha^2 - 2M\alpha\alpha_* + M\alpha_*^2 \leqslant M\alpha^2 + 2\sqrt{M\alpha^2 \cdot M\alpha_*^2} + M\alpha_*^2 = \\ = \left[\sqrt{M\alpha^2} + \sqrt{M\alpha_*^2}\,\right]^2 < \infty,$$

т. е. $\eta \in \mathcal{H}$. Рассмотрим разность

$$\Delta \stackrel{\text{def}}{=} M\alpha^2 - M\alpha_*^2 = M(\alpha_* + \eta)^2 - M\alpha_*^2 = 2M\alpha_*\eta + M\eta_*^2.$$

Если $M\alpha_*\eta = 0$ для всех $\eta \in \mathcal{H}$, то $\Delta = M\eta_*^2 \geqslant 0$ и $M\alpha^2 \geqslant M\alpha_*^2$ для любой $\alpha \in \mathcal{H}$.

Пусть теперь, наоборот, $\Delta \geqslant 0$ для любой $\alpha \in \mathcal{H}$. Предположим, что найдётся $\eta \in \mathcal{H}$, для которой $M\alpha_*\eta \neq 0$. Тогда возьмём произвольное неслучайное число $a \in \mathbb{R}$ и рассмотрим $\alpha = \alpha_* + a\eta$. Очевидно, что $\alpha \in \mathcal{H}$ и для такой α

$$\Delta = M\alpha^2 - M\alpha_*^2 = 2a \cdot M\alpha_*\eta + a^2 M\eta_*^2.$$

График правой части этого равенства как функция от a представляет собой параболу, ветви которой направлены вверх и которая (при условии $M\alpha_*\eta \neq 0$) имеет ровно два различных корня: $a = 0$ и $a = -M\alpha_*\eta/M\eta_*^2$. При значениях a, лежащих между корнями, правая часть неравенства строго отрицательна, и при этом для $\alpha = \alpha_* + a\eta$ мы имеем $\Delta < 0$, что противоречит изначальному предположению. Таким образом, не могут существовать $\eta \in \mathcal{H}$, для которых $M\alpha_*\eta \neq 0$. Лемма доказана.

Из этой леммы, положив $\alpha_* = t(\xi) - \tau(\theta)$ для несмещённой оценки $t(\xi)$ (тогда $M\alpha_* = 0$, $M\alpha_*^2 = D_\theta\, t(\xi)$), напрямую получаем следующее утверждение.

Теорема 2. *Пусть $t(\xi)$ – несмещённая оценка для $\tau(\theta)$ для любого θ. Оценка $t(\xi)$ является НОМД тогда и только тогда, когда для любой статистики $\eta(\xi)$, такой что $M_\theta \eta(\xi) = 0$, $M_\theta\, \eta^2(\xi) < \infty$, при каждом θ выполнено равенство $M_\theta\, t(\xi)\eta(\xi) = 0$.*

2.4. Эффективные оценки

В этом разделе мы покажем, что дисперсию несмещённых оценок можно ограничить снизу. Это означает, что при определённых условиях нельзя получить с.к. точность несмещённого оценивания выше некоторого предела. Затем представим метод построения НОМД определённого типа, называемых эффективными оценками.

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – выборка из (абсолютно непрерывного) распределения с.в. с функцией правдоподобия

$$L(x,\theta) = p(x_1;\theta)\ldots p(x_n;\theta), \qquad x = (x_1,\ldots,x_n) \in \mathbb{R}^n,$$

и $\mathsf{R} = \{x \in \mathbb{R}^n \colon L(x,\theta) > 0\}$ – множество реализаций.

Теорема 3 (Крамера–Рао)**.** *Пусть $\theta \in \mathbb{R}$, $t(\xi)$ – несмещённая оценка значения $\tau(\theta)$, множество реализаций R не зависит от $\theta \in \mathbb{R}$ и существуют производные $\frac{\partial \ln L(x,\theta)}{\partial\theta}$, $\tau'(\theta) = \frac{d\tau(\theta)}{d\theta}$ при всех $x \in \mathsf{R}$, $\theta \in \mathbb{R}$. Предположим, что для данных $L(\,\cdot\,)$ и $t(\,\cdot\,)$ допустимо вносить производные под знак интегралов:*

$$\frac{d}{d\theta}\int_{\mathsf{R}} L(x,\theta)\,dx = \int_{\mathsf{R}} \frac{\partial L(x,\theta)}{\partial\theta}\,dx, \tag{2.4}$$

$$\frac{d}{d\theta}\int_{\mathsf{R}} t(x)L(x,\theta)\,dx = \int_{\mathsf{R}} t(x)\frac{\partial L(\xi,\theta)}{\partial\theta}\,dx. \tag{2.5}$$

Тогда

$$D_\theta\, t(\xi) \geqslant \frac{[\tau'(\theta)]^2}{M_\theta\big(\frac{\partial \ln L(\xi,\theta)}{\partial\theta}\big)^2}\,, \tag{2.6}$$

причём равенство здесь выполняется тогда и только тогда, когда с вероятностью единица

$$\frac{\partial \ln L(\xi,\theta)}{\partial\theta} = a(\theta)(t(\xi) - \tau(\theta)) \tag{2.7}$$

для некоторой функции $a(\theta)$, $\theta \in \mathbb{R}$.

Доказательство. При условиях теоремы из равенств (2.4) и (2.5) следует, что

$$\int_{\mathsf{R}} \frac{\partial \ln L(x,\theta)}{\partial\theta} L(x,\theta)\,dx = \frac{d}{d\theta}\int_{\mathsf{R}} L(x,\theta)\,dx = \frac{d}{d\theta}1 = 0,$$

$$\int_{\mathsf{R}} t(x)\frac{\partial \ln L(\xi,\theta)}{\partial\theta} L(x,\theta)\,dx = \frac{d}{d\theta}\int_{\mathsf{R}} t(x) L(x,\theta)\,dx = \\ = \frac{d\,M_\theta\,t(\xi)}{d\theta} = \tau'(\theta).$$

Умножая первое равенство на $\tau(\theta)$ и вычитая из второго, получаем

$$\int_{\mathsf{R}} (t(x) - \tau(\theta))\frac{\partial \ln L(x,\theta)}{\partial\theta} L(x,\theta)dx = \tau'(\theta)$$

или

$$\tau'(\theta) = M_\theta\left[(t(\xi) - \tau(\theta)) \cdot \frac{\partial \ln L(\xi,\theta)}{\partial\theta}\right]. \tag{2.8}$$

Следовательно, возведя обе части неравенства в квадрат и применив неравенство Коши–Буняковского вида $(M\alpha\beta)^2 \leqslant M\alpha^2 M\beta^2$ для двух сомножителей в правой части (2.8), имеем

$$[\tau'(\theta)]^2 \leqslant M_\theta[(t(\xi) - \tau(\theta))]^2 \cdot M_\theta\left[\frac{\partial \ln L(\xi,\theta)}{\partial\theta}\right]^2 = \\ = D_\theta\,t(\xi) \cdot M_\theta\left[\frac{\partial \ln L(\xi,\theta)}{\partial\theta}\right]^2, \tag{2.9}$$

отсюда после простейших преобразований получаем неравенство Крамера–Рао (2.6).

Равенство в неравенстве Коши–Буняковского, т. е. в (2.9) или, эквивалентно, в (2.6) имеет место, если и только если в использованном неравенстве Коши–Буняковского достигается равенство. Это эквивалентно тому, что с.в. $\frac{\partial \ln L(\xi,\theta)}{\partial\theta}$ и $(t(\xi) - \tau(\theta))$ линейно зависимы, т. е. с вероятностью единица выполнено равенство (2.7) для некоторого коэффициента $a(\theta)$, который не зависит от ξ, но может зависеть от θ.

Определение 6. Несмещённая оценка $t(\xi)$ называется *эффективной*, если для неё выполнены условия теоремы Крамера–Рао и

$$D_\theta\,t(\xi) = \frac{[\tau'(\theta)]^2}{M_\theta\left(\frac{\partial \ln L(x,\theta)}{\partial\theta}\right)^2}. \tag{2.10}$$

Иными словами, эффективная оценка – это такая несмещённая оценка, для которой неравенство Крамера–Рао превращается в равенство. Для эффективной оценки, как мы знаем, с вероятностью единица справедливо равенство (2.7), откуда

$$a^2(\theta) \cdot M_\theta(t(\xi) - \tau(\theta))^2 = M_\theta\left(\frac{\partial \ln L(\xi, \theta)}{\partial \theta}\right)^2.$$

Подставляя выражение для математического ожидания в правой части этого равенства в знаменатель формулы (2.10), получаем, что для эффективной оценки

$$D_\theta\, t(\xi) = \frac{[\tau'(\theta)]^2}{a^2(\theta)\, D_\theta\, t(\xi)}. \tag{2.11}$$

Разрешая это уравнение относительно $D_\theta\, t(\xi) \geqslant 0$, получаем ещё одно выражение для дисперсии эффективной оценки через коэффициент $a(\theta)$:

$$D_\theta\, t(\xi) = \left|\frac{\tau'(\theta)}{a(\theta)}\right|. \tag{2.12}$$

Если выполнены условия теоремы Крамера–Рао и $t(\xi)$ – эффективная оценка для $\tau(\theta)$, то $t(\xi)$ является НОМД. Обратное утверждение неверно, так как эффективная оценка существует только для таких $\tau(\theta)$, для которых выполнены условия теоремы Крамера–Рао и равенство (2.7). Это налагает существенные ограничения на вид функции правдоподобия и функции $\tau(\theta)$, $\theta \in \Theta$.

Теорема Крамера–Рао верна и для дискретных распределений. В этом случае плотность вероятности $p(\,\cdot\,)$ с.в. ξ следует заменить на вероятность $P_\theta(\xi = x)$, а интегрирование – на суммирование.

Пример 5. Пользуясь дискретным вариантом теоремы, построим эффективную оценку неизвестного параметра $\lambda \in (0, \infty)$ распределения Пуассона по выборке $\xi_1, \ldots, \xi_n$ объема n. Найдём логарифм функции правдоподобия: для $\lambda \in (0, \infty)$, $m = (m_1, \ldots, m_n)$

$$\ln L(m, \lambda) = \ln \prod_{k=1}^{n} \frac{\lambda^{m_k}}{m_k!}\, e^{-\lambda} = \sum_{k=1}^{n} (-\lambda + m_k \ln \lambda - \ln(m_k!)),$$

где $m_k \in \{0, 1, \ldots\}$ для $k = 1, \ldots, n$. Тогда

$$\frac{\partial \ln L(m, \lambda)}{\partial \lambda} = \sum_{k=1}^{n}\left(-1 + \frac{m_k}{\lambda}\right) = \frac{n}{\lambda}\left(\frac{1}{n}\sum_{k=1}^{n} m_k - \lambda\right).$$

Подставляя вместо m аргумент ξ, получаем

$$\frac{\partial \ln L(\xi,\lambda)}{\partial\lambda} = \frac{n}{\lambda}\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k - \lambda\right) = \frac{n}{\lambda}(\bar{\xi}-\lambda).$$

Видно, что для $t(\xi)=\bar{\xi}$ равенство (2.7) выполнено при $a(\lambda)=n/\lambda$, поэтому $\bar{\xi}$ – эффективная оценка параметра λ, а её дисперсия равна λ/n.

Пример 6. Пусть $\xi=(\xi_1,\ldots,\xi_n)$ – выборка объема n из нормального распределения $\mathbf{N}(\theta,\sigma^2)$ с известной дисперсией; неизвестным параметром распределения является математическое ожидание θ. Запишем логарифм функции правдоподобия и его производную по θ:

$$\ln L(x,\theta) = \ln\left[\frac{1}{(\sqrt{2\pi\sigma^2}\,)^n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2\right)\right] =$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2,$$

$$\frac{\partial \ln L(x,\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta) = \frac{n}{\sigma^2}\left(\frac{1}{n}\sum_{i=1}^{n}x_i-\theta\right).$$

Последняя формула при $\tau(\theta)=\theta$ и $x=\xi$ имеет вид (2.7), когда

$$t(\xi)=\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i, \qquad a(\theta)=\frac{n}{\sigma^2}.$$

Следовательно, $\bar{\xi}$ – эффективная оценка параметра μ при известной σ^2, дисперсия этой оценки, определяющая ее с.к. погрешность, согласно (2.12) равна

$$D_\theta\, t(\xi) = \left|\frac{\tau'}{a}\right| = \frac{\sigma^2}{n},$$

что, впрочем, и так очевидно.

Пример 7. Пусть теперь в условиях предыдущего примера математическое ожидание нормального распределения известно, а неизвестным параметром является корень из дисперсии, $\theta=\sigma$. Пусть

$\tau(\sigma) = \sigma^2$. Тогда, найдя производную по σ логарифма функции правдоподобия,

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{n}(x_j-\mu)^2 = \frac{n}{\sigma^3}\left(-\sigma^2 + \frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)^2\right),$$

получим соотношение (2.12) sпри

$$t(\xi) = \frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)^2, \qquad a(\theta) = \frac{n}{\sigma^3}.$$

Следовательно, $t(\xi)$ – эффективная оценка для σ^2 при известном μ, и её дисперсия равна

$$D_\theta\, t(\xi) = \left|\frac{\tau'}{a}\right| = \frac{2\sigma}{n}\sigma^3 = \frac{2\sigma^4}{n}.$$

Экспоненциальное семейство распределений

Предыдущие примеры можно обобщить, рассмотрев так называемое экспоненциальное семейство, состоящее из распределений с плотностями вероятности вида

$$p(x,\theta) = e^{a(\theta)b(x)+c(\theta)+d(x)}, \qquad \theta\in\Theta\subseteq\mathbb{R}, \quad x\in\mathbb{R},$$

где $a(\,\cdot\,)$, $c(\,\cdot\,)$ – заданные функции параметра θ, дифференцируемые при всех $\theta\in\Theta$, а $b(\,\cdot\,)$ и $d(\,\cdot\,)$ – заданные функции из $\mathbb{R}$ в $\mathbb{R}$. Для выборок размера n из таких распределений функция правдоподобия имеет вид

$$L(x,\theta) = \exp\left\{a(\theta)\sum_{k=1}^{n}b(x_k) + nc(\theta) + \sum_{k=1}^{n}d(x_k)\right\},$$

отсюда

$$\frac{\partial \ln L(x,\theta)}{\partial\theta} = a'(\theta)\sum_{k=1}^{n}b(x_k) + nc'(\theta) = na'(\theta)\left(\frac{1}{n}\sum_{k=1}^{n}b(x_k) + \frac{c'(\theta)}{a'(\theta)}\right).$$

Следовательно, согласно (2.12) статистика $t(\xi) = \frac{1}{n}\sum_{j=1}^{n}b(\xi_j)$ является эффективной оценкой для $-\frac{c'(\theta)}{a'(\theta)}$, и ее дисперсия равна

$$D_\theta\, t = \left|\left(\frac{c'(\theta)}{a'(\theta)}\right)'\frac{1}{na'(\theta)}\right|,$$

если функция $L(\,\cdot\,)$, а также функции

$$\tau(\theta) = -\frac{c'(\theta)}{a'(\theta)}, \quad \theta \in \Theta, \qquad t(x) = \frac{1}{n}\sum_{k=1}^{n} b(x_k), \quad x_k \in \mathbb{R},$$

удовлетворяют условиям теоремы Крамера–Рао.

К экспоненциальному семейству относятся нормальное распределение, распределения Бернулли, Пуассона, хи-квадрат и др.

2.5. Достаточные статистики

Одной из существенных проблем статистического анализа является проблема анализа «больших выборок», т. е. таких, которые не вмещаются в память одного компьютера. Одним из подходов к её решению является такое сжатие данных, которое бы не приводило к потере информации, важной для решения поставленной задачи.

Как видно из рассмотренных примеров, часто для оценивания параметров распределения не требуется знание всей выборки, достаточно знать, чему равна некоторая функция от выборочных значений. Так, для оценки математического ожидания нормально распределённой с.в. достаточно знать сумму $\xi_1 + \cdots + \xi_n = n\bar{\xi}$ всех элементов выборки, а не значение каждой с.в. ξ_k выборки в отдельности. Для оценки дисперсии с.в. ξ при известном математическом ожидании μ достаточно знать сумму квадратов отклонений $(\xi_1-\mu)^2+\cdots+(\xi_n-\mu)^2$ выборочных значений от их математических ожиданий, а при неизвестном μ – сумму квадратов выборочных значений $\xi_1^2 + \cdots + \xi_n^2 = n\overline{\xi^2}$ и $n\bar{\xi}$.

Оказывается, что во многих случаях для оценивания параметра θ распределения выборки $\xi = (\xi_1, \ldots, \xi_n)$ можно указать некоторую статистику

$$\mathcal{T}(\xi) = \bigl(T_1(\xi), \ldots, T_m(\xi)\bigr) \in \mathbb{R}^m, \qquad m < n, \tag{2.13}$$

в известном смысле содержащую всю информацию о параметре θ.

Определение 7. m-Мерная статистика $\mathcal{T}(\xi)$ называется *достаточной* для семейства функций правдоподобия $L(\,\cdot\,,\theta)$, $\theta \in \Theta$, если условное распределение выборки $\xi = (\xi_1, \ldots, \xi_n)$ при условии $\mathcal{T}(\xi) = \mathrm{const}$ не зависит от θ.

Напомним, что такое условное распределение. Пусть две (для простоты одномерные) с.в. α и β имеют совместное распределение. Тогда условное распределение с.в. α при фиксированной $\beta = b$ задаётся следующим образом: в дискретном случае – как ряд условных вероятностей

$$P(\alpha = a_j \,|\, \beta = b) = \frac{P(\alpha = a_j, \beta = b)}{P(\beta = b)}, \quad j = 1, \ldots, M \leqslant \infty, \quad (2.14)$$

а в абсолютно непрерывном случае – как условная плотность вероятности

$$p_{\alpha|\beta}(a|b) = \frac{p_{\alpha,\beta}(a, b)}{p_\beta(b)}, \qquad a \in \mathbb{R}, \qquad (2.15)$$

или, в дифференциальной форме,

$$p_{\alpha|\beta}(a|b)\, da = P(a \leqslant \alpha < a + da \,|\, b \leqslant \beta < b + db)$$

(сравните с $p_\alpha(a)\, da = P(a \leqslant \alpha < a + da)$). Условные распределения определены для любых значений b, для которых в дискретном случае $P(\beta = b) \neq 0$, а в абсолютно непрерывном случае $p_\beta(b) \neq 0$. Как функции от значения с.в. α условные распределения при фиксированном b обладают всеми свойствами обычных распределений, в частности удовлетворяют условию нормировки

$$\sum_{j=1}^{n} P(\alpha = a_j \,|\, \beta = b) = 1, \qquad \int_{-\infty}^{\infty} p_{\alpha|\beta}(a|b)\, da = 1$$

для соответственно дискретного и абсолютно непрерывного распределений. Также справедлива формула полной вероятности

$$P(\alpha = a_j) = \sum_b P(\alpha = a_j \,|\, \beta = b) P(\beta = b),$$
$$p_\alpha(a) = \int\limits_{b:\ p_\beta(b) \neq 0} p_{\alpha|\beta}(a|b) p_\beta(b)\, db = \int\limits_{-\infty}^{\infty} p_{\alpha|\beta}(a|b) p_\beta(b)\, db.$$

Если с.в. α и β независимы, то условные распределения совпадают с безусловными.

Применим эти конструкции к понятию достаточной статистики. Мы имеем две с.в. – выборку, т. е. n-мерную с.в. ξ, и m-мерную статистику $\mathcal{T}(\xi)$. Как устроено условное распределение ξ при фиксированной $\mathcal{T}(\xi) = s$, проще всего понять в случае дискретного распределения выборки. Пусть заданы вероятности

$$P_\theta(\xi_k = x_j), \quad j = 1, 2, \ldots, M \leqslant \infty, \qquad \sum_{j=1}^{M} P_\theta(\xi_k = x_j) = 1, \quad (2.16)$$

для каждого представителя ξ_k, $k = 1, \ldots, n$, выборки ξ размера n. Функция правдоподобия и вероятность события $\mathcal{T}(\xi) = s$ в этом случае задаются как

$$L(x, \theta) = \prod_{k=1}^{n} P_\theta(\xi_k = x_j), \qquad P_\theta(\mathcal{T}(\xi) = s) = \sum_{x:\ \mathcal{T}(x) = s} L(x, \theta)$$

для $x = (x_1, \ldots, x_n)$. Если $P_\theta(\mathcal{T}(\xi) = s)$ отлична от нуля, то

$$P_\theta(\xi = x \,|\, \mathcal{T}(\xi) = s) = \frac{P_\theta(\xi = x, \mathcal{T}(\xi) = s)}{P_\theta(\mathcal{T}(\xi) = s)}. \quad (2.17)$$

Если статистика $\mathcal{T}(\xi)$ является достаточной, то дробь в правой части не зависит от θ.

Легко понять, что если x таков, что $\mathcal{T}(x) \neq s$, то в числителе (2.17) стоят несовместные события и тогда

$$P_\theta(\xi = x \,|\, \mathcal{T}(\xi) = s) = 0,$$

что, конечно, не зависит от θ. Если x таков, что $\mathcal{T}(x) = s$, то событие $\xi = x$ влечёт $\mathcal{T}(\xi) = s$, и тогда

$$P_\theta(\xi = x, \mathcal{T}(\xi) = s) = P_\theta(\xi = x).$$

Таким образом,

$$\begin{gathered} P(\xi = x \,|\, \mathcal{T}(\xi) = s) = \begin{cases} \dfrac{P_\theta(\xi = x)}{P_\theta(\mathcal{T}(\xi) = s)}, & \mathcal{T}(x) = s, \\ 0, & \mathcal{T}(x) \neq s, \end{cases} \\ P_\theta(\mathcal{T}(\xi) = s) = \sum_{x:\ \mathcal{T}(x) = s} P_\theta(\xi = x). \end{gathered} \quad (2.18)$$

Если значение достаточной статистики фиксировано, $\mathcal{T}(\xi) = s$, то изменение параметра θ не влияет на условное распределение случайной выборки ξ. Можно сказать, что вся информация о параметре θ содержится в значении достаточной статистики $\mathcal{T}(\xi)$.

В некоторых случаях доказать свойство достаточности статистики можно прямым вычислением условного распределения выборки при $\mathcal{T}(\xi) = s \in \mathbb{R}^m$. Далее для простоты мы рассматриваем одномерные $(m = 1)$ достаточные статистики и обозначаем их как $T(\xi)$.

Пример 8. Рассмотрим с.в., принимающую значение 1 с вероятностью p и значение 0 с вероятностью q, $q+p = 1$ (распределение Бернулли). Выборка объёма n из такого распределения представляет собой последовательность из n нулей и единиц. Числовая статистика $T(\xi) = \sum_{i=1}^{n} \xi_i$ даёт число успехов в серии n независимых испытаний Бернулли. Покажем, что $T(\xi)$ является достаточной статистикой для семейства распределений Бернулли с параметром $\theta = p$.

Запишем распределение выборки $\xi = (\xi_1, \ldots, \xi_n)$ при $T(\xi) = k$, пользуясь определением условной вероятности:

$$P(\xi_1 = k_1, \ldots, \xi_n = k_n \,|\, T(\xi) = k) = \\ = \frac{P(\xi_1 = k_1, \ldots, \xi_n = k_n, T(\xi) = k)}{P(T(\xi) = k)}, \tag{2.19}$$

где $k_i \in \{0, 1\}$ для $i = 1, \ldots, n$. При $k_1 + \cdots + k_n = k$ числитель дроби в (2.19) равен $p^k q^{n-k}$. Если же $k_1 + \cdots + k_n \neq k$, то события $\{\xi_1 = k_1, \ldots, \xi_n = k_n\}$ и $T(\xi) = k$ несовместны, и числитель равен нулю. Величина знаменателя даётся формулой Бернулли:

$$P(T(\xi) = k) = C_n^k p^k q^{n-k},$$

в результате имеем

$$P(\xi_1 = k_1, \ldots, \xi_n = k_n \,|\, T(\xi) = k) = \begin{cases} 1/C_n^k, & T(\xi) = k, \\ 0, & T(\xi) \neq k, \end{cases}$$

и условное распределение не зависит от p, а значит, число успехов $T(\xi)$ в серии из n независимых испытаний Бернулли является достаточной статистикой для параметра p.

Прямая проверка независимости условного распределения выборки от θ не всегда удобна, но существует необходимое и достаточное условие достаточности статистики $\mathcal{T}(\xi)$.

Теорема 4 (о факторизации)**.** *Статистика $\mathcal{T}(\xi)$ является достаточной для семейства функций правдоподобия $L(\,\cdot\,,\theta)$, $\theta \in \Theta$, тогда и только тогда, когда функция правдоподобия на множестве реализаций R_θ может быть представлена в виде*

$$L(x,\theta) = g(\mathcal{T}(x),\theta)h(x), \qquad x \in \mathsf{R}_\theta, \quad \theta \in \Theta. \tag{2.20}$$

Доказательство. Приведем доказательство для дискретного распределения, считая, что выполнено (2.16). Пусть $\theta \in \Theta$ произвольно, но фиксировано.

Пусть выполнено условие (2.20), покажем, что статистика $\mathcal{T}(\xi)$ является достаточной для θ. Для этого вычислим условное распределение $P_\theta(\xi = x \,|\, \mathcal{T}(\xi) = s)$, пользуясь формулой (2.18). Если $\mathcal{T}(x) \neq s$, то искомая условная вероятность равна нулю и не зависит от θ. Если x таков, что $\mathcal{T}(x) = s$, то в силу представления (2.20) для $L(x,\theta) = P_\theta(x)$ имеем

$$P(\xi = x|\mathcal{T}(\xi) = s) = \frac{L(x,\theta)}{\sum\limits_{x:\,\mathcal{T}(x)=s} L(x,\theta)} = \frac{g(s,\theta)h(x)}{\sum\limits_{x:\,\mathcal{T}(x)=s} g(\mathcal{T}(x),\theta)h(x)} = \\ = \frac{g(s,\theta)h(x)}{\sum\limits_{x:\,\mathcal{T}(x)=s} g(s,\theta)h(x)} = \frac{h(x)}{\sum\limits_{x:\,\mathcal{T}(x)=s} h(x)},$$

т. е. условное распределение выборки $\xi_1, \ldots, \xi_n$ при фиксированной $\mathcal{T}(x) = s$ не зависит от θ, что и требовалось доказать.

Пусть теперь $\mathcal{T}(\xi)$ – достаточная статистика для θ, покажем, что тогда выполнено (2.20). Тогда для любого $x \in \mathsf{R}$ событие $\xi = x$ влечёт событие $\mathcal{T}(\xi) = \mathcal{T}(x) \stackrel{\text{def}}{=} s$, и мы имеем

$$L(x,\theta) = P_\theta(\xi = x) = P_\theta(\xi = x, \mathcal{T}(\xi) = s) = \\ = P(\xi = x \,|\, \mathcal{T}(\xi) = s)P_\theta(\mathcal{T}(\xi) = s), \quad s = \mathcal{T}(x). \tag{2.21}$$

Условное распределение выборки ξ при фиксированном $\mathcal{T}(x) = s$ не зависит от θ, поэтому мы не пишем индекс θ в первом множителе в правой части (2.21). Второй множитель полностью определяется значением θ и значением s достаточной статистики, и его можно записать как $g(s,\theta)$ для некоторой функции $g(\,\cdot\,)$, которая в сущности задаёт зависимость вероятности $P_\theta(\mathcal{T}(\xi) = s)$ от θ и s. Аналогично первый множитель зависит от x и s, но не зависит от θ

в силу достаточности статистики $\mathcal{T}(\xi)$. Поэтому можно записать его как $H(x, s)$ для некоторой функции $H(\,\cdot\,)$. В результате (2.21) принимает вид

$$L(x, \theta) = g(s, \theta) H(x, s), \qquad x \in \mathsf{R}, \quad s = \mathcal{T}(x),$$

Введём функцию $h(x) = H(x, \mathcal{T}(x))$, $x \in \mathsf{R}$, тогда, подставляя $\mathcal{T}(x)$ вместо s получаем,

$$L(x, \theta) = g(\mathcal{T}(x), \theta) h(x), \quad x \in \mathsf{R}, \quad \theta \in \Theta.$$

Замечание 3. Если $\mathcal{T}(\xi)$ – достаточная статистика для θ, то любая взаимно однозначная функция от $\mathcal{T}$ также является достаточной статистикой для θ.

Пример 9. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – выборка объема n из нормального распределения $\mathbf{N}(\theta, 1)$. Как показывают простые алгебраические преобразования, её функцию правдоподобия можно привести к нужному виду,

$$L(x, \theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2\right\} = g(T(x), \theta) h(x),$$

если положить

$$T(x) = \sum_{i=1}^{n} x_i = n\bar{x}, \qquad \begin{aligned} g(T(x), \theta) &= \exp\left\{\theta \sum_{i=1}^{n} x_i - \frac{n}{2}\theta^2\right\}, \\ h(x) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right\}. \end{aligned}$$

Следовательно, $T(\xi) = n\bar{\xi}$ – достаточная статистика для θ.

Замечание 4. Если при выполнении условия теоремы Крамера–Рао $t(\xi)$ – эффективная оценка, то $t(\xi)$ также является достаточной статистикой. В самом деле, мы имеем для всех $x \in \mathsf{R}$ и $\theta \in \Theta$

$$\frac{\partial \ln L(x, \theta)}{\partial \theta} = a(\theta)(t(x) - \tau(\theta)) = a(\theta)t(x) - a(\theta)\tau(\theta) = a(\theta)t(x) - b(\theta),$$

где $b(\theta) = a(\theta)\tau(\theta)$. Проинтегрируем равенство для производной по θ:

$$\ln L(x, \theta) = t(x)A(\theta) - B(\theta) + c(x),$$

где $A(\,\cdot\,)$, $B(\,\cdot\,)$ – первообразные функций $a(\,\cdot\,)$, $b(\,\cdot\,)$ на Θ, а $c(x)$ – «постоянная» интегрирования. Тогда

$$L(x,\theta) = e^{t(x)A(\theta)-B(\theta)} \cdot e^{c(x)},$$

и условие факторизации (2.20) для функции правдоподобия $L(x,\theta)$ выполнено при

$$g(t(x),\theta) = e^{t(x)A(\theta)-B(\theta)}, \qquad h(x) = e^{c(x)}.$$

Если НОМД существует, то её можно найти как функцию от достаточной статистики. Этот факт является следствием теоремы Блекуэлла–Колмогорова (часто её также называют теоремой Рао–Блекуэлла–Колмогорова). Прежде чем мы перейдём к формулировке и доказательству этой теоремы, вспомним, что такое условное математическое ожидание.

Мы имеем выборку ξ как n-мерную с.в. и одномерную достаточную статистику $T(\xi)$. Рассмотрим условное математическое ожидание статистики $t(\xi)$ при фиксированном значении $T(\xi) = s$, для ясности считая, что выборка ξ распределена дискретно:

$$\begin{aligned} M(t(\xi)|T(\xi)=s) &= \sum_{x\in\mathsf{R}} t(x)P(\xi = x\,|\,T(\xi)=s) = \\ &= \sum_{x\colon T(x)=s} t(x)P(\xi = x\,|\,T(\xi)=s). \end{aligned} \tag{2.22}$$

Здесь мы учли, что $P(\xi = x\,|\,T(\xi) = s) \neq 0$, только если $s = T(x)$. Условное математическое ожидание не зависит от θ в силу определения достаточной статистики, но, конечно, зависит от s. Оно является специфической функцией на множестве S значений достаточной статистики: $M(t(\xi)|T(\xi) = s) = E(s)$, $s \in \mathsf{S}$. Тогда мы можем ввести с.в. $E(T)$ (функцию от достаточной статистики) по стандартному правилу: если $T(\xi) = s$, то эта с.в. равна $E(s)$.

Справедлива формула, которую можно назвать последовательным усреднением:

$$M_\theta\, t(\xi) = \sum_{s\in\mathsf{S}} M(t(\xi)|T(\xi)=s)P_\theta(T(\xi)=s) = \sum_{s\in\mathsf{S}} E(s)P_\theta(T(\xi)=s),$$

другими словами, мы сначала усредняем $t(\xi)$, считая $T(\xi)$ фиксированной, а затем усредняем по всем значениям $T(\xi)$. Доказательство этой формулы простое: в силу леммы о суммировании по блокам

$$
\begin{aligned}
M_\theta\, t(\xi) &= \sum_{x\in\mathsf{R}} t(x)P_\theta(\xi = x) = \sum_{s\in\mathsf{S}}\ \sum_{x\colon T(x)=s} t(x)P_\theta(\xi = x) = \\
&= \sum_{s\in\mathsf{S}}\Bigg(\sum_{x\colon T(x)=s} t(x)\frac{P_\theta(\xi = x)}{P_\theta(T(\xi) = s)}\Bigg)P_\theta(T(\xi) = s) = \\
&= \sum_{s\in\mathsf{S}}\Bigg(\sum_{x\colon T(x)=s} t(x)P(\xi = x\,|\,T(\xi) = s)\Bigg)P_\theta(T(\xi) = s).
\end{aligned}
$$

С учетом (2.22) мы понимаем, что в круглых скобках стоит в точности $M(t(\xi)|T = s)$, а суммирование по всем s с весами $P_\theta(T(\xi) = s)$ даёт математическое ожидание по распределению с.в. $T(\xi)$. Формулу последовательного усреднения можно записать как

$$
M_\theta\, t(\xi) = M_\theta\, E(T), \quad E(s) = M(t(\xi)|T(\xi) = s), \quad s \in \mathsf{S}, \tag{2.23}
$$

где математическое ожидание $M_\theta\, t(\xi)$ вычисляется по распределению с.в. $t(\xi)$, а математическое ожидание $M_\theta\, E(T)$ – по распределению с.в. $T(\xi)$.

До сих пор мы рассматривали $M(t(\xi)|T(\xi) = s)$ как функцию $E(s)$ от значения достаточной статистики, но это математическое ожидание также является функцией от значения выборки: если $\xi = x$, то $M(t(\xi)|T(\xi) = s) = E(s)$ при $s = T(x)$, другим словами $M(t(\xi)|T(\xi) = s) = E(T(x))$. Таким образом, мы имеем с.в. $\hat{t}(\xi) = E(T(\xi))$, и эта с.в. зависит только от выборки и не зависит от θ, т. е. является статистикой.

Теперь мы готовы доказать теорему Блекуэлла–Колмогорова (для дискретного распределения выборки).

Теорема 5. *Пусть $t(\xi)$ — несмещённая оценка для $\tau(\theta) \in \mathbb{R}$, $T(\xi)$ – одномерная достаточная статистика для θ. Тогда условное математическое ожидание $\hat{t}(\xi) = M_\theta(t(\xi)|T(\xi))$ удовлетворяет неравенству*

$$
M_\theta[t(\xi) - \tau(\theta)]^2 \geqslant M_\theta[\hat{t}(\xi) - \tau(\theta)]^2. \tag{2.24}
$$

Доказательство. Пусть $\theta \in \Theta$ произвольно, но фиксировано.

Если $t(\xi)$ – несмещённая оценка для $\tau(\theta)$, то $\hat{t}(\xi)$ – тоже несмещённая оценка. Это напрямую вытекает из (2.23): с учётом того, что $\hat{t}(\xi) = E(T)$,

$$
\tau(\theta) = M_\theta\, t(\xi) = M_\theta\, E(T) = M_\theta\, \hat{t}(\xi).
$$

Теперь оценим дисперсию статистики $\hat{t}(\xi)$. Для простоты формул будем писать $T = s$ вместо $T(\xi) = s$. Имеем

$$D_\theta \hat{t}(\xi) = M_\theta \big[E(T) - \tau(\theta)\big]^2 = \sum_{s \in \mathsf{S}} \big[E(s) - \tau(\theta)\big]^2 P_\theta(T = s) =$$
$$= \sum_{s \in \mathsf{S}} \big[M(t(\xi) \,|\, T = s) - \tau(\theta)\big]^2 P_\theta(T = s). \tag{2.25}$$

Рассмотрим выражение в квадратных скобках. Для любой с.в. α и любой неслучайной c мы имеем $M\alpha - c = M(\alpha - c)$ независимо от того, по какому распределению (обычному или условному) вычисляется математическое ожидание. Поэтому для любого $s \in \mathsf{S}$

$$M(t(\xi) \,|\, T = s) - \tau(\theta) = M(t(\xi) - \tau(\theta) \,|\, T = s).$$

Также всегда верно неравенство $[M\alpha]^2 \leqslant M\alpha^2$, поэтому

$$\big[M(t(\xi) - \tau(\theta) \,|\, T = s)\big]^2 \leqslant M\big[(t(\xi) - \tau(\theta))^2 \,|\, T = s\big].$$

Подставим эти соотношения в (2.25), получим

$$\begin{aligned} D_\theta \hat{t}(\xi) &\leqslant \sum_{s \in \mathsf{S}} M\big[(t(\xi) - \tau(\theta))^2 \,|\, T = s\big] P_\theta(T = s) = \\ &= \sum_{s \in \mathsf{S}} \Bigg(\sum_{x:\, T(x) = s} (t(x) - \tau(\theta))^2 P(\xi = x \,|\, T = s) \Bigg) P_\theta(T = s) = \\ &= \sum_{s \in \mathsf{S}} \sum_{x:\, T(x) = s} (t(x) - \tau(\theta))^2 P_\theta(\xi = x) = \\ &= \sum_{x \in \mathsf{R}} (t(x) - \tau(\theta))^2 P_\theta(\xi = x) = D_\theta t(\xi). \end{aligned}$$

Неравенство $D_\theta \hat{t}(\xi) \leqslant D_\theta t(\xi)$ доказано.

Замечание 5. Для доказательства теоремы не требуется достаточность статистики $T(\xi)$, но это условие гарантирует, что мы можем вычислить значение статистики $\hat{t}(x)$ при реализации $\xi = x$, не зная параметр θ. Действительно, по определению достаточной статистики условное распределение $t(\xi)$ при условии $T(\xi) = s$ не зависит от θ, следовательно, условное математическое ожидание $t(\xi)$ также не зависит от θ.

Доказанная теорема позволяет утверждать, что НОМД следует искать как функцию от достаточной статистики (конечно, если она существует). В самом деле, пусть $t(\xi)$ – НОМД для $\tau(\theta)$. Поскольку условное математическое ожидание $M_\theta(t(\xi) \,|\, T(\xi))$ статистики $t(\xi)$ также является несмещённой оценкой для $\tau(\theta)$ и не увеличивает её дисперсию в силу доказанной теоремы, но и не уменьшает её (так как $t(\xi)$ – НОМД), то в силу единственности НОМД $t(\xi) = M_\theta(t(\xi) \,|\, T(\xi))$ с вероятностью единица, и поэтому $t(\xi)$ зависит от ξ только через $T(\xi)$.

2.6. Полные достаточные статистики

Ещё один способ построения НОМД может быть применён для специального класса статистик, называемых полными.

Определение 8. Статистика $T(\xi)$ называется *полной*, если всякая функция от неё, имеющая при всех $\theta \in \Theta$ нулевое математическое ожидание, равна нулю с вероятностью единица (для любого $\theta \in \Theta$).

Если $T(\xi)$ – полная статистика, то, во-первых, как любая статистика она является несмещённой оценкой своего математического ожидания $M_\theta T(\xi)$, а во-вторых, в силу полноты не существует двух функций от $T(\xi)$, несмещённо оценивающих $M_\theta T(\xi)$. В самом деле, тогда разность этих функций есть функция от $T(\xi)$, и её математическое ожидание равно нулю, но для полных статистик $T(\xi)$ разность таких оценок равна нулю с вероятностью единица.

Всякая функция $f(T(\xi))$ от полной достаточной статистики является НОМД математического ожидания $M_\theta T(\xi)$ для всех $\theta \in \Theta$. Действительно, как было показано выше, НОМД следует искать в классе функций $f(T(\xi))$ от достаточной статистики. Поскольку статистика $f(T(\xi))$ полная, не существует двух различных функций от $f(T(\xi))$, несмещённо оценивающих $M_\theta T(\xi)$.

Если существует полная достаточная статистика $T(\xi)$ для параметра $\theta \in \Theta$, то для получения НОМД значения $\tau(\theta)$ можно построить любую несмещённую оценку $t(\xi)$, затем найти её условное математическое ожидание $f(s) = M_\theta(t(\xi) \,|\, T(\xi) = s)$. Тогда функция $f(s)$, $s \in \mathbb{R}$, не зависит от θ и $f(T(\xi))$ – НОМД для $\tau(\theta)$.

Рассмотрим пример использования полноты статистики для получения НОМД для разных функций $\tau(\theta)$. В предыдущем разделе было показано, что в серии $\xi_1, \ldots, \xi_n$ из n независимых испытаний Бернулли для вероятности $\theta \in (0,1)$ успеха в единичном испытании достаточной статистикой является число успехов $\xi = \xi_1 + \cdots + \xi_n$. Покажем, что $T(\xi) = \xi$ является полной статистикой. Для этого выберем произвольную статистику $f(\xi)$, для которой $M_\theta\, f(\xi) = 0$, другими словами,

$$M_\theta\, f(\xi) = \sum_{k=0}^{n} f(k) C_n^k \theta^k (1-\theta)^{n-k} = 0.$$

Это соотношение означает, что полином степени n от аргумента θ равен нулю для всех $\theta \in (0,1)$, что возможно только тогда, когда все его коэффициенты равны нулю. Значит, ξ – полная достаточная статистика, и любая функция от нее является НОМД своего математического ожидания. В частности, поскольку

$$M_\theta\, \frac{\xi}{n} = \theta, \quad M_\theta\, \frac{\xi(\xi-1)}{n(n-1)} = \theta^2, \quad M_\theta\, \frac{\xi(n-\xi)}{n(n-1)} = \theta(1-\theta),$$

получаем, что

$$\frac{\xi}{n}, \quad \frac{\xi(\xi-1)}{n(n-1)}, \quad \frac{\xi(n-\xi)}{n(n-1)}$$

являются НОМД для θ, θ^2 и $\theta(1-\theta)$ соответственно.

2.7. Эвристические методы построения оценок

Опишем два способа построения точечных оценок, которые основаны на неформальных соображениях, но тем не менее в ряде случаев обладают полезными свойствами.

Оценки максимального правдоподобия

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – выборка объема n из абсолютно непрерывного распределения с плотностью $p(\,\cdot\,;\theta)$, $\theta \in \Theta$. Функция правдоподобия такой выборки равна

$$L(x,\theta) = \prod_{k=1}^{n} p(x_k;\theta), \qquad \theta \in \Theta, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n. \tag{2.26}$$

Тогда

$$L(x,\theta)\, dx = P_\theta(x_1 \leqslant \xi_1 < x_1 + dx_1, \ldots\, x_n \leqslant \xi_n < x_n + dx_n).$$

При фиксированном $x \in \mathbb{R}^n$ эта вероятность зависит от $\theta \in \Theta$, и величина $L(x, \theta)$ определяет, насколько вероятно получить выборку $\xi = x$ для данного значения θ. Рассмотрим теперь с.в. $L(\xi, \theta)$, которая зависит от случайной выборки. С учётом смысла реализаций $L(x, \theta)$ этой с.в. разумно выбрать такое значение параметра θ, при котором значение функции правдоподобия максимально при данном x. Другими словами, необходимо решить следующую задачу: требуется найти

$$\max_{\theta \in \Theta} L(x, \theta) = L(x, \hat{\theta}(x)). \tag{2.27}$$

Подставляя вместо x с.в. ξ, получаем оценку $\hat{\theta}(\xi)$, к которой (если она является статистикой) применимы все понятия и методы анализа из предыдущих разделов.

Определение 9. Оценка $\hat{\theta}(\xi)$, доставляющая максимум в (2.27), называется *оценкой максимального правдоподобия* (ОМП) для параметра $\theta \in \Theta$.

Таким образом, принцип максимального правдоподобия состоит в том, что в качестве значения неизвестного параметра предлагается использовать ОМП, т. е. такое значение параметра, при котором максимальна вероятность получить в результате наблюдений имеющуюся реализацию выборки.

Если требуется найти ОМП параметра дискретного распределения, то в функции правдоподобия (2.26) следует заменить значение плотности $p(x_k; \theta)$ на $P_\theta(\xi_k = x_k)$.

Если функция правдоподобия дифференцируема по параметру и максимум достигается во внутренней точке множества Θ, то для достижения максимума в (2.27) необходимо равенство нулю производной $L(x, \theta)$ по θ. В ряде случаев удобнее искать максимум логарифма функции правдоподобия; в силу монотонности логарифма мы получим ту же точку максимума $\hat{\theta}$. Необходимое условие максимума принимает вид

$$\left. \frac{\partial \ln L(x, \theta)}{\partial \theta} \right|_{\theta = \hat{\theta}(x)} = 0. \tag{2.28}$$

Это уравнение называется уравнением максимального правдоподобия. Если в точке максимума функция правдоподобия $L(x, \theta)$ не дифференцируема по θ, то ОМП $\hat{\theta}$ тем не менее может существовать, но воспользоваться уравнением максимального правдоподобия уже не удастся. Наконец, если максимум в (2.27) не достигается, то ОМП для θ не существует.

Если $\theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^m$ и $m > 1$, то уравнение максимального правдоподобия (2.28) записывается как система уравнений

$$\frac{\partial \ln L(x, \theta_j)}{\partial \theta_j} = 0, \qquad j = 1, \ldots, m.$$

Её решение $\hat{\theta}(x) = (\hat{\theta}_1(x), \ldots, \hat{\theta}_m(x))$ даёт ОМП.

Отметим, что если существует эффективная оценка $t(\xi)$ числового параметра θ, то она может быть найдена как ОМП. Действительно, условие эффективности оценки записывается как

$$\frac{\partial \ln L(x, \theta)}{\partial \theta} = a(\theta)(t(x) - \theta), \qquad x \in \mathsf{R}. \tag{2.29}$$

Пусть для каждого $x \in \mathsf{R}$ значение $\hat{\theta}(x)$ определяется условием

$$\left.\frac{\partial \ln L(x, \theta)}{\partial \theta}\right|_{\theta = \hat{\theta}(x)} = 0,$$

тогда $t(x) - \hat{\theta}(x) = 0$ для всех $x \in \mathsf{R}$ (при условии $a(\theta) \neq 0$ для всех $\theta \in \Theta$). Это означает, что эффективная оценка $t(\xi) = \hat{\theta}(\xi)$ с вероятностью единица.

Второе полезное свойство ОМП состоит в том, что если существует достаточная статистика $T(\xi)$ для θ, то ОМП является функцией от этой статистики. Действительно, согласно теореме 4 тогда $L(x, \theta) = g(T(x), \theta)h(x)$, следовательно, ОМП доставляет максимум по θ функции $g(T(x), \theta)$, зависящий от x только через $T(x)$.

Кроме того, ОМП при определенных условиях обладают рядом других полезных свойств, проявляющихся при большом объеме выборки (состоятельность, асимптотическая нормальность).

Пример 10. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – выборка объема n из распределения $\mathbf{N}(\mu, \sigma^2)$ и $\theta = (\mu, \sigma^2)$ – двумерный параметр. Построим для него ОМП. Имеем

$$\ln L(x,\mu,\sigma^2) = \sum_{i=}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x_i-\mu)^2/2\sigma^2}\right) = \\ = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Система уравнений максимального правдоподобия имеет вид

$$\frac{\partial \ln L(x,\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0,$$
$$\frac{\partial \ln L(x,\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0,$$

откуда следует ОМП для (μ,σ^2):

$$\hat\mu = \bar\xi = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar\xi\,)^2.$$

Заметим, что при неизвестном μ ОМП для σ^2 не обладает свойством несмещённости, однако и $\hat\mu$, и $\widehat{\sigma^2}$ являются состоятельными оценками.

Пример 11. Построим ОМП для параметра θ равномерного на $[0,\theta]$ распределения с.в. по выборке объема $n=1$. В этом случае

$$L(\xi,\theta) = \begin{cases} 1/\theta, & \theta \geqslant \xi, \\ 0, & \theta < \xi, \end{cases}$$

максимум функции правдоподобия достигается в точке $\theta=\xi$, где функция правдоподобия терпит разрыв, но мы тем не менее имеем ОМП $\hat\theta(\xi)=\xi$ для θ.

Метод моментов

Пусть $\xi=(\xi_1,\ldots,\xi_n)$ – выборка объема n из распределения, зависящего от неизвестного параметра $\theta\in\Theta$. Рассмотрим некоторую функцию $f(\,\cdot\,):\mathbb{R}\to\mathbb{R}$, такую что $f(\xi_k)$ (здесь $k=1,\ldots,n$

любое, но фиксированное) является с.в. и имеет конечное математическое ожидание $M_\theta\, f(\xi_k)$. Все составляющие выборки имеют одно и то же распределение, поэтому $M_\theta\, f(\xi_k)$ не зависит от k, но, конечно, зависит от неизвестного параметра: $M_\theta\, f(\xi_k) = g(\theta)$. Рассмотрим это равенство как уравнение относительно θ, и пусть оно разрешимо единственным образом. Обозначим его решение как $g^{(-1)}(M_\theta\, f(\xi_k))$. Заменяя математическое ожидание с.в. $f(\xi_k)$ его выборочным значением $\frac{1}{n}\sum_{k=1}^{n} f(\xi_k)$, получим оценку параметра θ *методом моментов*:

$$\tilde{\theta} = g^{(-1)}\left(\frac{1}{n}\sum_{k=1}^{n} f(\xi_k)\right).$$

Функцию $f(\,\cdot\,)$ называют пробной. Чаще всего в качестве пробной функции выбирают степенную функцию $f(\xi_k) = \xi_k^r$ (если существует $M_\theta\, \xi_k^r$).

Положительным свойством метода моментов является его состоятельность. Если $M_\theta\, f(\xi_k) = g(\theta)$ и для последовательности с.в. $f(\xi_k)$, $k = 1, 2, \ldots$, выполнены условия закона больших чисел,

$$\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) \xrightarrow[n\to\infty]{\mathrm{P}} g(\theta),$$

то в случае непрерывности функции $g^{(-1)}(\,\cdot\,)$ по свойству 1, доказанному в разделе 1.1, оценка $\tilde{\theta}$ метода моментов является состоятельной:

$$\tilde{\theta} = g^{(-1)}\left(\frac{1}{n}\sum_{i=1}^{n} f(\xi_i)\right) \xrightarrow[n\to\infty]{\mathrm{P}} g^{(-1)}(M_\theta\, f(\xi)) = g^{(-1)}(g(\theta)) = \theta.$$

Если $\theta = (\theta_1, \ldots, \theta_m)$ с $m > 1$, то в качестве пробной берут вектор-функцию $f(x)\colon \mathbb{R} \to \mathbb{R}^m$ со значениями $(f_1(x), \ldots, f_m(x))$ для $x \in \mathbb{R}$. Тогда, найдя математические ожидания каждой $f_i(\xi_k)$, мы получаем для $\theta \in \Theta \subseteq \mathbb{R}^m$ набор равенств

$$M_\theta\, f_i(\xi_k) = g_i(\theta), \qquad i = 1, \ldots, m$$

(напомним, что математическое ожидание не зависит от k). Далее подставляем вместо $M_\theta\, f_i(\xi_k)$ их выборочные значения $\overline{f_i(\xi)}$. Получаем систему уравнений

$$\overline{f_i(\xi)} = g_i(\theta), \qquad i = 1, \ldots, m,$$

с известными функциями $g_i(\,\cdot\,)$ на Θ. Она должна иметь единственное решение $\tilde{\theta}_i(\xi)$, $i = 1, \ldots, m$. Это решение даёт многомерную оценку для θ методом моментов.

Пример 12. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – выборка объема n из равномерного на отрезке $[0, \theta]$ распределения. Найдём оценку параметра θ методом моментов, выбрав пробную функцию $f(\xi_1) = \xi_1$. Поскольку $M_\theta\, \xi_1 = \theta/2$, мы можем записать $\theta = 2\, M_\theta\, \xi_1$, и, заменяя математическое ожидание выборочным средним, получаем, что оценка методом моментов равна

$$\tilde{\theta} = \frac{2}{n} \sum_{k=1}^{n} \xi_k.$$

Если в качестве пробной функции $f(\,\cdot\,)$ взять r-й момент с.в. ξ_1, то в силу того, что

$$M_\theta\, \xi_1^r = \int_0^\theta \frac{z^r}{\theta}\, dz = \frac{\theta^r}{r+1},$$

мы имеем

$$\tilde{\theta} = \sqrt[r]{\frac{r+1}{n} \sum_{k=1}^{n} \xi_k^r}.$$

Пример 13. Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – выборка объема n из нормального распределения $\mathbf{N}(\mu, \sigma^2)$. Найдём оценку двумерного параметра $\theta = (\mu, \sigma^2)$ методом моментов. Выберем пробную функцию $f(\xi_1) = (\xi_1, \xi_1^2)$. Имеем

$$M_\theta \xi_1 = \mu, \qquad M_\theta \xi_1^2 = \mu^2 + \sigma^2,$$

откуда, заменяя $M_\theta \xi_1$, $M_\theta \xi_1^2$ на соответственно $\bar{\xi}$, $\overline{\xi^2}$, получаем оценки методом моментов

$$\tilde{\mu} = \bar{\xi}, \qquad \tilde{\sigma}^2 = \overline{\xi^2} - (\bar{\xi}\,)^2.$$

Далее мы перейдём к задачам интервального оценивания параметров распределений с.в.

3. Интервальные оценки параметров распределений

3.1. Интервальные оценки параметров распределений случайных величин

Точечные оценки параметров распределений, даже если они обладают какими-либо полезными свойствами, имеют один недостаток: при конечном размере выборки в общем случае значение точечной оценки совпадает с оцениваемой величиной с вероятностью ноль. Другими словами, если $t(\xi)$ – точечная оценка для θ, то мы не можем утверждать, что $t(x) = \theta$ практически ни для какой реализации $\xi = x$. Причина этого в том, что $t(\xi)$ – случайная точка в пространстве параметров (потому оценка называется точечной), и вероятность того, что она «упадёт» в точности в точку θ, в общем случае равна нулю. Справиться с этой проблемой можно, заменив точечную оценку $t(\xi)$ на некое множество $\widehat{\Theta}(\xi)$, границы которого определяются выборкой, такое что вероятность $P_\theta(\widehat{\Theta}(\xi) \ni \theta)$ достаточно велика для всех $\theta \in \Theta$.

Далее мы рассмотрим случай одномерного параметра θ, тогда в качестве множества $\widehat{\Theta}(\xi)$ на прямой естественно выбрать интервал $(t_1(\xi), t_2(\xi))$. Данный случайный интервал содержит истинное значение параметра θ с вероятностью

$$P_\theta(t_1(\xi) < \theta < t_2(\xi)) = P_\theta(t_1(\xi) < \theta,\, t_2(\xi) > \theta).$$

Нам требуется, чтобы эта вероятность была достаточно велика для всех $\theta \in \Theta$. Чтобы не анализировать поведение вероятности в зависимости от θ, потребуем, чтобы она вовсе не зависела от θ. Мы приходим к следующему определению.

Определение 1. *Интервальной оценкой* (или *доверительным интервалом*) для одномерного параметра θ называется случайный интервал $(t_1(\xi), t_2(\xi))$, такой что вероятность

$$\gamma = P_\theta(t_1(\xi) < \theta < t_2(\xi)) \tag{3.1}$$

не зависит от $\theta \in \Theta$. Величина γ называется *уровнем доверия* интервальной оценки.

Границы интервальной оценки представляют собой некоторые функции от выборки (статистики). Интервальная оценка тем лучше, чем выше уровень доверия. Случайная величина $t_2(\xi) - t_1(\xi)$ равна длине интервала, и понятно, что желательно, чтобы её распределение было сосредоточено в области малых значений – тогда мы можем говорить о высокой точности оценивания. Очевидно, что в общем случае чем выше уровень доверия, тем ниже точность оценивания (длинным интервалом легче накрыть точку, чем коротким).

Найти статистики $t_1(\xi)$, $t_2(\xi)$, для которых вероятность (3.1) не зависит от θ, – это сложная задача. Однако можно попробовать свести эту задачу к более простой. А именно, если нам удастся найти функцию $T(x;\theta)$, $x \in \mathbb{R}^n$, $\theta \in \Theta$, такую что

- распределение с.в. $T(\xi;\theta)$ известно и не зависит от $\theta \in \Theta$;
- неравенство $a_1 < (\xi;\theta) < a_2$, выполненное с вероятностью γ, можно решить относительно θ как $t_1(\xi) < \theta < t_2(\xi)$,

то случайный интервал $(t_1(\xi), t_2(\xi))$ удовлетворяет условию (3.1).

Этот подход используется при построении самых известных интервальных оценок – оценок параметров нормального распределения.

3.2. Интервальные оценки параметров нормального распределения

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – выборка из распределения $\mathbf{N}(\mu, \sigma^2)$. Построение интервальных оценок параметров нормального распределения основано на следующих определениях и свойствах этого распределения, известных из теории вероятностей.

Определение 2. Мы говорим, что с.в. χ_n^2 имеет *распределение хи-квадрат с n степенями свободы* (или, иначе, *распределение Пирсона*), и мы пишем $\chi_n^2 \sim \mathbf{X}_n^2$, если

$$\chi_n^2 = \sum_{k=1}^{n} \xi_k^2,$$

где ξ_k, $k = 1, \ldots, n$, имеют распределение $\mathbf{N}(0,1)$ и независимы.

Определение 3. Мы говорим, что с.в. τ_n имеет *распределение Стьюдента с n степенями свободы* (или, иначе, T-распределение), и мы пишем $\tau_n \sim \mathbf{T}_n$, если

$$\tau_n = \frac{\xi_0}{\sqrt{\frac{1}{n}\sum_{k=1}^n \xi_k^2}}, \qquad \xi_k \sim \mathbf{N}(0,1), \quad k = 0, 1, \ldots, n,$$

где ξ_k, $k = 0, 1, \ldots, n$, независимы. Альтернативное определение связано с распределением хи-квадрат:

$$\tau_n = \frac{\xi_0}{\sqrt{\frac{1}{n}\chi_n^2}} \sim \mathbf{T}_n, \quad \text{если} \quad \xi_0 \sim \mathbf{N}(0,1), \quad \chi_n^2 \sim \mathbf{X}_n^2$$

и ξ_0, χ_n^2 независимы.

Следующие свойства показывают, что определённые комбинации нормальных с.в. имеют упомянутые распределения.

Свойство N1. *Если с.в.* $\xi_k \sim \mathbf{N}(\mu_k, \sigma_k^2)$, $k = 1, \ldots, n$, *независимы, то*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{\xi_k - \mu_k}{\sqrt{\sigma_k^2}} \sim \mathbf{N}(0,1).$$

В частности, если $\mu_k = \mu$ *и* $\sigma_k^2 = \sigma^2$ *для всех* $k = 1, \ldots, n$, *то*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{\xi_k - \mu}{\sqrt{\sigma^2}} = \sqrt{\frac{n}{\sigma^2}}(\bar{\xi} - \mu) \sim \mathbf{N}(0,1).$$

Свойство N2. *Если с.в.* $\xi_k \sim \mathbf{N}(\mu_k, \sigma_k^2)$, $k = 1, \ldots, n$, *независимы, то для любых постоянных* $b, a_1, \ldots, a_n$

$$a_1\xi_1 + \cdots + a_n\xi_n + b \sim \mathbf{N}(\mu, \sigma^2), \qquad \begin{cases} \mu = a_1\mu_1 + \cdots + a_n\mu_n + b, \\ \sigma^2 = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2. \end{cases}$$

Свойство N3. *Если с.в.* $\xi_k \sim \mathbf{N}(\mu, \sigma^2)$, $k = 1, \ldots, n$, *независимы, то*

$$\frac{1}{\sigma^2} \sum_{k=1}^n (\xi_k - \mu)^2 \sim \mathbf{X}_n^2.$$

Свойство N4. *Если с.в.* $\xi_k \sim \mathbf{N}(\mu, \sigma^2)$, $k = 1, \ldots, n$, *независимы, то*

$$\frac{1}{\sigma^2} \sum_{k=1}^{n} (\xi_k - \bar{\xi})^2 \sim \mathbf{X}^2_{n-1}.$$

Свойство N5. *Если с.в.* $\xi_k \sim \mathbf{N}(\mu, \sigma^2)$, $k = 1, \ldots, n$, *независимы, то*

$$\frac{\frac{1}{\sqrt{n}}(\bar{\xi} - \mu)}{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(\xi_j - \bar{\xi})^2}} \sim \mathbf{T}_{n-1}.$$

Эти свойства лежат в основе интервальных оценок параметров нормального распределения.

Интервальная оценка математического ожидания при известной дисперсии

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – выборка из распределения $\mathbf{N}(\mu, \sigma^2)$. Построим случайный интервал, накрывающий значение математического ожидания μ. В качестве функции $T(\xi; \mu)$, распределение которой не зависит от оцениваемого параметра[1], выберем

$$T(\xi; \mu) = \sqrt{\frac{n}{\sigma^2}}(\bar{\xi} - \mu).$$

Согласно свойству $\mathbf{N}1$ имеем $T(\xi; \mu) \sim \mathbf{N}(0, 1)$. Тогда

$$P(b_1 < T(\xi; \mu) < b_2) = \Phi(b_1) - \Phi(-b_2) = \int_{b_1}^{b_2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx,$$

Таким образом, если мы выберем $b_1 < b_2$ так, чтобы выполнялось равенство $\Phi(b_1) - \Phi(-b_2) = \gamma$, то

$$\gamma = P\left(b_1 < \sqrt{\frac{n}{\sigma^2}}(\bar{\xi} - \mu) < b_2\right) = P\left(\bar{\xi} - \sqrt{\frac{\sigma^2}{n}}\, b_2 < \mu < \bar{\xi} - \sqrt{\frac{\sigma^2}{n}}\, b_1\right).$$

Уравнение $\Phi(b_1) - \Phi(-b_2) = \gamma$ имеет бесконечно много решений $b_1 < b_2$; чтобы выделить одно, наложим дополнительное условие $b_1 = -b_2$. Другими словами, будем решать относительно $b > 0$ уравнение

$$P(-b < T(\xi; \mu) < b) = \Phi(b) - \Phi(-b) = \gamma. \tag{3.2}$$

[1] См. идею построения интервальных оценок в конце предыдущего раздела.

Тогда для любого $\gamma \in (0,1)$ существует единственное $b = b_\gamma > 0$, являющееся корнем уравнения (3.2). В результате интервальную оценку для μ при известной σ^2 с уровнем доверия $\gamma \in (0,1)$ можно построить следующим образом. По заданному γ находим $b = b_\gamma$, решив уравнение

$$\Phi(b) - \Phi(-b) = \int_{-b}^{b} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \gamma, \tag{3.3}$$

тогда

$$P\left(\bar{\xi} - \sqrt{\frac{\sigma^2}{n}}\, b_\gamma < \mu < \bar{\xi} + \sqrt{\frac{\sigma^2}{n}}\, b_\gamma\right) = \gamma. \tag{3.4}$$

Численные границы интервальной оценки находятся после подстановки вместо $\bar{\xi}$ выборочного среднего $\bar{x}$ полученных в эксперименте данных. При решении уравнения (3.3) полезно иметь в виду очевидные равенства, вытекающие из чётности подынтегральной функции:

$$\Phi(-b) = 1 - \Phi(b) = \frac{1-\gamma}{2}.$$

Замечание 1. В оценке (3.4) центр интервала, накрывающего истинное значение μ, совпадает с выборочным средним, границы интервала суть

$$t_1(\xi) = \bar{\xi} - \sqrt{\frac{\sigma^2}{n}}\, b_\gamma, \qquad t_2(\xi) = \bar{\xi} + \sqrt{\frac{\sigma^2}{n}}\, b_\gamma,$$

длина интервала неслучайна и равна

$$t_2(\xi) - t_1(\xi) = 2\sqrt{\frac{\sigma^2}{n}}\, b_\gamma,$$

так как не зависит от выборки. Можно показать, что среди всех пар чисел (b_1, b_2), удовлетворяющих условию $\Phi(b_1) - \Phi(-b_2) = \gamma$, пара $b_1 = -b_\gamma$, $b_2 = b_\gamma$ даёт минимальную длину доверительного интервала для фиксированного γ.

Замечание 2. Видно, что в (3.4) длина интервала уменьшается с уменьшением σ^2 и с ростом n, стремясь к нулю при $n \to \infty$. Как результат, точность интервальной оценки естественным образом увеличивается с уменьшением дисперсии составляющих выборки и с ростом объёма выборки. При $n \to \infty$ интервал стягивается в точку, совпадающую с точечной оценкой $\bar{\xi}$ параметра μ. Кроме того, чем больше γ, тем больше b_γ. Таким образом, попытка увеличить уровень доверия оценки неизбежно приводит к уменьшению её точности.

Замечание 3. Построенная нами интервальная оценка, конечно, не является единственно возможной. Также нельзя утверждать, что она оптимальна с какой-либо точки зрения. Например, хотя в замечании 1 мы сказали, что оценка (3.4) имеет минимальную длину, это относится только к оценкам, построенным на основе выбранной нами функции $T(\xi;\mu) = \sqrt{n/\sigma^2}(\bar{\xi}-\mu)$. Возможно, оценки, построенные при других функциях $T(\xi;\mu)$, будут более точными. Преимуществами оценки (3.4) является простота её построения, в частности простота решения неравенства $b_1 < T(\xi;\mu) < b_2$ относительно μ. Также в общем случае не являются оптимальными и другие интервальные оценкам, построенные в этом и следующих разделах.

Интервальная оценка дисперсии при известном математического ожидании

Теперь нам нужно выбрать функцию $T(\xi;\sigma^2)$, распределение которой не зависит от σ^2. Для этого воспользуемся свойством **N**3 и рассмотрим функцию

$$\chi_n^2(\xi;\sigma^2) = \chi_n^2 = \frac{1}{\sigma^2}\sum_{k=1}^{n}(\xi_k-\mu)^2 \sim \mathbf{X}_n^2.$$

По заданному $\gamma \in (0,1)$ найдём $0 < c_{\gamma,1}^{(n)} < c_{\gamma,2}^{(n)}$ как корни уравнения

$$P(c_1^{(n)} < \chi_n^2 < c_2^{(n)}) = \gamma. \tag{3.5}$$

Таких решений бесконечно много, чтобы выделить единственное, наложим дополнительное условие

$$P(\chi_n^2 < c_{\gamma,1}^{(n)}) = P(\chi_n^2 > c_{\gamma,2}^{(n)}) = \frac{1-\gamma}{2}. \tag{3.6}$$

Тогда, решая неравенство $c_{\gamma,1}^{(n)} < \chi_n^2 < c_{\gamma,2}^{(n)}$ относительно σ^2, получаем

$$P\left(\frac{\sum_{k=1}^{n}(\xi_k-\mu)^2}{c_{\gamma,2}^{(n)}} < \sigma^2 < \frac{\sum_{k=1}^{n}(\xi_k-\mu)^2}{c_{\gamma,1}^{(n)}}\right) = \gamma. \tag{3.7}$$

Формулы (3.5)–(3.7) задают интервальную оценку параметра σ^2 при известном μ.

Условие (3.6) говорит о том, что наш доверительный интервал «промахивается», падая левее и правее истинного значения параметра σ^2, с одинаковой вероятностью $(1-\gamma)/2$. Можно было бы воспользоваться подходом предыдущей задачи и найти самый короткий интервал (т. е. минимум разности $\frac{1}{c_{\gamma,1}^{(n)}} - \frac{1}{c_{\gamma,2}^{(n)}}$ при фиксированном γ. Эту вариационную задачу можно решить, но не так легко, как удовлетворить условие (3.6)

Мы построили интервальные оценки одного из параметров нормального распределения, считая, что нам известно значение другого параметра. Едва ли можно ожидать, что на практике мы столкнёмся с подобной ситуацией. Скорее всего неизвестны значения обоих параметров, но вполне возможно, что исследователя интересует оценка только для одного из них, например для среднего (математического ожидания). Это усложняет алгоритм оценивания: искомая оценка должна строиться так, чтобы все необходимые вероятности можно было найти, не зная «лишний» параметр (его часто называют мешающим). Идея построения таких оценок для параметров нормального распределения состоит в замене математического ожидания и дисперсии их выборочными значениями.

Интервальная оценка дисперсии при неизвестном математического ожидании

В данном случае заменим неизвестное математическое ожидание μ на выборочное среднее $\bar{\xi}$. Согласно свойству **N**4 мы можем выбрать в качестве $T(\xi;\sigma^2)$ функцию

$$\chi^2_{n-1} = \frac{1}{\sigma^2}\sum_{k=1}^{n}(\xi_k - \bar{\xi}\,)^2 \sim \mathbf{X}^2_{n-1}.$$

В результате интервальная оценка дисперсии при неизвестном математическом ожидании задаётся теми же формулами, что и выше, с заменой n на $n-1$:

$$P\biggl(\frac{\sum_{k=1}^{n}(\xi_k-\mu)^2}{c_2^{(n-1)}} \leqslant \sigma^2 \leqslant \frac{\sum_{k=1}^{n}(\xi_k-\mu)^2}{c_1^{(n-1)}}\biggr) = \gamma, \qquad (3.8)$$

где $c_1^{(n-1)}$, $c_2^{(n-1)}$ определяются уравнениями

$$\begin{gathered} P(c_1^{(n-1)} < \chi^2_{n-1} < c_2^{(n-1)}) = \gamma, \\ P(\chi^2_{n-1} < c_{\gamma,1}^{(n-1)}) = P(\chi^2_{n-1} > c_{\gamma,2}^{(n-1)}) = \frac{1-\gamma}{2}. \end{gathered} \tag{3.9}$$

Интервальная оценка математического ожидания при неизвестной дисперсии

В данном случае заменим неизвестную σ^2 на выборочную дисперсию

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(\xi_k - \bar{\xi}\,)^2.$$

Тогда согласно свойству **N**5 статистика

$$\tau_{n-1} = \frac{1}{\sqrt{n\hat{\sigma}^2}}(\bar{\xi} - \mu) = \frac{\frac{1}{\sqrt{n}}(\bar{\xi} - \mu)}{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(\xi_j - \bar{\xi}\,)^2}} \tag{3.10}$$

имеет распределение Стьюдента $\mathbf{T}_{n-1}$. Плотность вероятности для этого распределения является чётной функцией, поэтому границы интервальной оценки, как и в случае известной σ^2, будем находить из следующего уравнения относительно u_γ:

$$P(-u_\gamma < \tau_{n-1} < u_\gamma) = \gamma, \qquad 0 < \gamma < 1. \tag{3.11}$$

В результате имеем

$$P\left(\bar{\xi} - \sqrt{\frac{\hat{\sigma}^2}{n}}\,u_\gamma < \mu < \bar{\xi} + \sqrt{\frac{\hat{\sigma}^2}{n}}\,u_\gamma\right) = \gamma. \tag{3.12}$$

Мы видим очевидное сходство оценок (3.4) и (3.12) для соответственно известной и неизвестной σ^2. Разница состоит в том, что в оценке (3.12) вместо σ^2 стоит с.в. $\hat{\sigma}^2$, которая в определённом смысле приближает неизвестную дисперсию. Кроме того, решения уравнений (3.2) и (3.11), определяющие границы интервала, выводятся соответственно из стандартного нормального распределения и распределения Стьюдента $\mathbf{T}_{n-1}$.

Если сравнивать точности оценок (3.4) и (3.12) при фиксированных n и γ, считая, что $\hat{\sigma}^2 = \sigma^2$, то, как и следовало ожидать, при неизвестной дисперсии точность оценки ниже, чем в случае известной. Это объясняется поведением решений уравнений (3.2) и (3.11). Поскольку распределение $\mathbf{T}_{n-1}$ шире стандартного нормального, u_γ в уравнении (3.2) больше, чем решение b_γ уравнения (3.11) при том же γ. Точность интервальной оценки (3.12) аналогично оценке (3.4) растёт с ростом n и падает с ростом $\hat\sigma^2$ или γ (см. замечание 2).

3.3. Другие интервальные оценки. Асимптотические интервальные оценки

Рассмотрим интервальные оценки для других распределений.

Пример 1. Дана выборка ξ объёма 1 из экспоненциального распределения $\mathbf{E}(\theta)$, $\theta > 0$, т.е. её плотность имеет вид $p(x) = \theta e^{-\theta x}$ при $x \geqslant 0$, $p(x) = 0$ при $x < 0$. Построим интервальную оценку параметра θ с уровнем доверия γ.

Выберем с.в. $T(\xi;\theta) = \xi\theta$. Ее функция распределения не зависит от неизвестного параметра θ:

$$F_T(z) = P(\xi\theta < z) = \int_0^{z/\theta} \theta e^{-\theta x}\,dx = 1 - e^{-z}, \qquad z > 0. \tag{3.13}$$

Для $0 < \gamma < 1$ и $0 \leqslant \beta < 1-\gamma$ найдём ε_1 и ε_2 из уравнений

$$P(\xi\theta \leqslant \varepsilon_1) = 1 - e^{-\varepsilon_1} = \beta, \qquad P(\xi\theta \geqslant \varepsilon_2) = e^{-\varepsilon_2} = 1-\gamma-\beta,$$

тогда $P(\varepsilon_1 < \xi\theta < \varepsilon_2) = \gamma$ или

$$P\left(\frac{\varepsilon_1}{\xi} < \theta < \frac{\varepsilon_2}{\xi}\right) = \gamma.$$

Из (3.13) следует, что

$$\varepsilon_1 = -\ln(1-\beta), \qquad \varepsilon_2 = -\ln(1-\gamma-\beta).$$

Интервальная оценка тем точнее, чем меньше длина оценивающего интервала при заданном уровне доверия. В нашем случае длина оценивающего интервала есть

$$\frac{\varepsilon_2 - \varepsilon_1}{\xi} = \frac{\ln(1-\beta) - \ln(1-\gamma-\beta)}{\xi},$$

и это с.в. Но тем не менее мы можем сказать, что наибольшая точность оценивания достигается при $\beta = 0$, так как при фиксированном ξ длина монотонно возрастает на интервале $\beta \in [0, 1-\gamma)$, поскольку

$$\frac{\partial}{\partial \beta}\big(\ln(1-\beta) - \ln(1-\gamma-\beta)\big) = \frac{\gamma}{(1-\beta)(1-\gamma-\beta)} > 0.$$

Итак, искомая оценка имеет вид

$$P\left(0 < \theta < -\frac{\ln(1-\gamma)}{\xi}\right) = \gamma.$$

Пример 2. Пусть теперь дана выборка $\xi = (\xi_1, \ldots, \xi_n)$ объема n из того же экспоненциального распределения, и, как и в предыдущей задаче, требуется построить интервальную оценку параметра θ с уровнем доверия γ.

В этом случае в качестве $T(\xi;\theta)$ можно выбрать функцию

$$\theta\xi_{(n)} = \theta \cdot \max_{1 \leqslant k \leqslant n} \xi_k$$

с не зависящей от θ функцией распределения

$$F_T(z) = P(\xi_{(n)}\theta < z) = (1 - e^{-z})^n, \qquad z > 0.$$

Тогда из уравнения $P(0 < \theta\xi_{(n)} < \varepsilon) = (1 - e^{-\varepsilon})^n = \gamma$ находим $\varepsilon = \ln(1 - \sqrt[n]{\gamma}\,)$ и получаем искомую интервальную оценку

$$P\left(0 < \theta < -\frac{\ln(1 - \sqrt[n]{\gamma}\,)}{\xi_{(n)}}\right) = \gamma.$$

Если в качестве $T(\xi;\theta)$ выбрать функцию

$$\theta\xi_{(1)} = \theta \cdot \min_{1 \leqslant k \leqslant n} \xi_k,$$

то её функция распределения

$$F_T(z) = P(\xi_{(n)}\theta < z) = 1 - e^{-zn}, \qquad z > 0,$$

также не зависит от θ. Тогда, действуя так же, как выше получаем

$$P(0 < \xi_{(1)}\theta < \varepsilon) = 1 - e^{-\varepsilon n} = \gamma, \qquad \varepsilon = -\frac{\ln(1-\gamma)}{n}.$$

В результате имеем интервальную оценку

$$P\left(0 < \theta < -\frac{\ln(1-\gamma)}{n\xi_{(1)}}\right) = \gamma.$$

Пример 3. Пусть теперь $\xi = (\xi_1, \ldots, \xi_n)$ – выборка объема n из равномерного распределения на отрезке $[0, \theta]$. Требуется построить интервальную оценку параметра $\theta > 0$ с уровнем доверия γ.

Выберем в качестве $T(\xi; \theta)$ функцию $\xi_{(n)}/\theta$ с не зависящей от θ функцией распределения

$$F_T(z) = \begin{cases} 0, & z \leqslant 0, \\ z^n, & 0 < z \leqslant 1, \\ 1 & z > 1. \end{cases}$$

Тогда, решая для $0 < \gamma < 1$ уравнение

$$P\left(\varepsilon < \frac{\xi_{(n)}}{\theta} < 1\right) = 1 - \varepsilon^n = \gamma$$

относительно ε, получаем $\varepsilon = \sqrt[n]{1-\gamma}$ и далее, разрешая неравенство под знаком вероятности относительно θ, выводим интервальную оценку с уровнем доверия γ:

$$P\left(\xi_{(n)} < \theta < \frac{\xi_{(n)}}{\sqrt[n]{1-\gamma}}\right) = \gamma.$$

Асимптотические интервальные оценки

Как видно из приведенных выше примеров, интервальные оценки с заданным уровнем доверия можно построить с помощью различных функций $T(\xi; \theta)$, и не всегда выбор этой функции очевиден. Интервальные оценки можно также строить, воспользовавшись предельными теоремами (центральной предельной теоремой, законами больших чисел), так, чтобы вероятность накрыть ими параметр распределения стремилась к γ при стремлении объёма выборки к бесконечности. Такие интервальные оценки (доверительные интервалы) называются асимптотическими. Часто асимптотические оценки оказываются более естественными, чем точные интервальные оценки.

Определение 4. Интервал $\big(\hat{t}_1(\xi_1, \ldots, \xi_n), \hat{t}_2(\xi_1, \ldots, \xi_n)\big)$ со случайными границами называется асимптотической оценкой параметра θ с уровнем доверия γ, если для всех $\theta \in \Theta$

$$\liminf_{n\to\infty} P_\theta\big(\hat{t}_1(\xi_1, \ldots, \xi_n) < \theta < \hat{t}_2(\xi_1, \ldots, \xi_n)\big) = \gamma.$$

В этом определении используется нижний предел, так как последовательность вероятностей может не иметь предела, но нижний предел ограниченной снизу (нулём) числовой последовательности всегда существует.

Опишем общий метод построения асимптотических интервальных оценок.

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ – n-мерная выборка из распределения, зависящего от параметра $\theta \in \Theta$. Для построения асимптотических интервальных оценок параметра θ следует выполнить следующие действия.

• Найти функцию $\widehat{T}(\xi_1, \ldots, \xi_n; \theta)$, по распределению сходящуюся к с.в. η, которая имеет известную функцию распределения $F_\eta(\,\cdot\,)$, не зависящую параметра θ. Необходимо, чтобы при любой фиксированной выборке $(\xi_1, \ldots, \xi_n)$ функция $\widehat{T}(\xi_1, \ldots, \xi_n; \theta)$ была обратима по θ.

• Найти числа ε_1 и ε_2, для которых

$$\liminf_{n\to\infty} P(\varepsilon_1 < \widehat{T}(\xi_1, \ldots, \xi_n; \theta) < \varepsilon_2) = P(\varepsilon_1 < \eta < \varepsilon_2) = \gamma.$$

• Разрешив неравенства $\varepsilon_1 < \widehat{T}(\xi_1, \ldots, \xi_n; \theta) < \varepsilon_2$ относительно θ, получить асимптотически точный доверительный интервал.

Рассмотрим для примера построение асимптотической оценки параметра θ показательного распределения по заданной выборке $\xi_1, \ldots, \xi_n$ объема n. Математическое ожидание этого распределения равно $1/\theta$, дисперсия равна $1/\theta^2$. Согласно центральной предельной теореме последовательность с.в.

$$\frac{1}{\sqrt{n/\theta^2}} \sum_{k=1}^{n} (\xi_k - n/\theta) = \sqrt{n}\,\frac{\bar{\xi} - 1/\theta}{1/\theta} = \sqrt{n}(\theta\bar{\xi} - 1), \quad n = 1, 2 \ldots,$$

сходится по распределению к с.в. $\eta \sim \mathbf{N}(0, 1)$. Тогда при $n \to \infty$

$$P(-\varepsilon < \sqrt{n}(\theta\bar{\xi} - 1) < \varepsilon) \to P(-\varepsilon < \eta < \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon).$$

Для заданного $\gamma \in (0, 1)$ найдём значение $\varepsilon = \varepsilon_\gamma$ как корень уравнения $\Phi(\varepsilon) - \Phi(-\varepsilon) = \gamma$. Тогда

$$P\left(\frac{1}{\bar{\xi}} - \frac{\varepsilon_\gamma}{\sqrt{n}\bar{\xi}} < \theta < \frac{1}{\bar{\xi}} + \frac{\varepsilon_\gamma}{\sqrt{n}\bar{\xi}}\right) =$$

$$= P\left(\frac{n}{\sum_{k=1}^{n} \xi_k} - \frac{\sqrt{n}\varepsilon_\gamma}{\sum_{k=1}^{n} \xi_k} < \theta < \frac{n}{\sum_{k=1}^{n} \xi_k} + \frac{\sqrt{n}\varepsilon_\gamma}{\sum_{k=1}^{n} \xi_k}\right) \xrightarrow[n\to\infty]{} \gamma.$$

Для построения асимптотических интервальных оценок часто бывают полезны предложения 1 и 2 из раздела 1.1.

Для примера, пользуясь указанными утверждениями, построим асимптотическую интервальную оценку параметра распределения Пуассона λ по выборке $\xi_1, \ldots, \xi_n$ объема n.

В силу центральной предельной теоремы последовательность

$$\frac{\sum\limits_{i=1}^{n}(\xi_i - \lambda)}{\sqrt{n\lambda}} = \sqrt{n}\,\frac{\bar{\xi} - \lambda}{\sqrt{\lambda}}, \qquad n = 1, 2, \ldots,$$

сходится по распределению к $\eta \sim \mathbf{N}(0, 1)$. Из этого следует, что при $n \to \infty$

$$P\left(-\varepsilon < \sqrt{n}\,\frac{\bar{\xi} - \lambda}{\sqrt{\lambda}} < \varepsilon\right) \underset{n\to\infty}{\longrightarrow} P(-\varepsilon < \eta < \varepsilon) = \Phi(\varepsilon) - \Phi(-\varepsilon).$$

Выбираем ε как корень уравнения $\Phi(\varepsilon) - \Phi(-\varepsilon) = \gamma$. Однако мы имеем квадратичное неравенство $-\varepsilon < \sqrt{n}\frac{\bar{\xi}-\lambda}{\sqrt{\lambda}} < \varepsilon$ относительно λ, решение которого весьма громоздкое. Но если заменить в знаменателе дроби $\sqrt{\lambda}$ на $\sqrt{\bar{\xi}}$, то неравенства легко разрешаются.

В данном случае $\bar{\xi} \xrightarrow{\mathrm{P}} \lambda$ с ростом объёма n выборки, функция $\sqrt{\lambda/x}$ непрерывна в $x = \lambda$, значит, $\sqrt{\lambda/\bar{\xi}} \xrightarrow{\mathrm{P}} 1$ при $n \to \infty$. Далее,

$$\sqrt{\frac{\lambda n}{\bar{\xi}}}\,\frac{\bar{\xi} - \lambda}{\sqrt{\lambda}} = \sqrt{n}\,\frac{\bar{\xi} - \lambda}{\sqrt{\bar{\xi}}} \ \underset{n\to\infty}{\xrightarrow{\mathrm{d}}} \ \eta \sim \mathbf{N}(0, 1).$$

Поэтому

$$P\left(-\varepsilon < \sqrt{n}\,\frac{\bar{\xi} - \lambda}{\sqrt{\bar{\xi}}} < \varepsilon\right) \underset{n\to\infty}{\longrightarrow} P(-\varepsilon < \eta < \varepsilon) = \gamma.$$

Решаем неравенство под знаком вероятности относительно λ и получаем асимптотическую интервальную оценку:

$$P\left(\bar{\xi} - \varepsilon\sqrt{\frac{\bar{\xi}}{n}} \leqslant \lambda \leqslant \bar{\xi} + \varepsilon\sqrt{\frac{\bar{\xi}}{n}}\right) \underset{n\to\infty}{\longrightarrow} \gamma.$$

4. Линейная регрессия

На практике часто встречаются задачи, в которых интересующие исследователя параметры не могут наблюдаться непосредственно, измеряться может только некоторая известная функция этих параметров, причём измерения производятся со случайной погрешностью. Задача, состоящая в оценке этих параметров по результату измерения значений этой функции и математической модели погрешности, называется задачей регрессии. Как параметры, так и результаты измерений в этом разделе считаются векторными величинами, принадлежащими вещественным евклидовым пространствам конечной размерности. Наиболее изучены задачи линейной регрессии, в которой связь между неизвестными параметрами и наблюдаемой функцией даётся линейным преобразованием параметров.

Для нахождения оценок параметров нам потребуются дополнительные сведения из линейной алгебры и теории вероятностей, связанные с линейными преобразованиями случайных векторов.

Линейные операторы в евклидовых пространствах

Вещественное евклидово пространство $\mathcal{R}^n$ размерности n – это множество упорядоченных наборов $x = \langle x_1, \dots, x_n \rangle$, где $x_k \in \mathbb{R}$ для всех $k = 1, \dots, n$. Элементы пространства называются векторами. По сути элемент $x \in \mathcal{R}^n$ полностью тождествен вектор-столбцу. Элемент $\langle 0, \dots, 0 \rangle$ (столбец из нулей) называется нулевым элементом пространства, и мы будем обозначать его как 0, т. е. так же, как и число ноль. Это не должно приводить к недоразумениям, поскольку по смыслу формулы должно быть понятно, какова природа символа 0.

Линейные операции для $x = \langle x_1, \dots, x_n \rangle$ и $z = \langle z_1, \dots, z_n \rangle$ вводятся поэлементно:

$$ax + bz = \langle ax_1 + bz_1, \dots, ax_n + bz_n \rangle, \qquad a, b \in \mathbb{R},$$

скалярное произведение и норма задаются как

$$(x,z) = x_1 z_1 + \dots + x_n z_n, \qquad \|x\| = \sqrt{(x,x)} = \sqrt{x_1^2 + \dots + x_n^2}.$$

В любом n-мерном евклидовом пространстве существует ортонормированный базис (кратко ОНБ, мы иногда называем его просто базис), т. е. набор элементов $e_1, \dots, e_n \in \mathcal{R}^n$, таких что $(e_k, e_j) = \delta_{kj}$, где $\delta_{kj} = 1$ при $k = j$ и $\delta_{kj} = 0$ при $k \neq j$ (символ Кронекера). Далее мы обозначаем такой базис как $\{e_k\}$, не указывая его размерность. Любой вектор $x \in \mathcal{R}^n$ можно разложить по ОНБ $\{e_k\}$ по формуле:

$$x = \sum_{k=1}^{n} (x, e_k) e_k \quad \Longrightarrow \quad \|x\|^2 = \sum_{k=1}^{n} (x, e_k)^2. \tag{4.1}$$

Числа (x, e_k), $k = 1, \dots, n$, называются координатами вектора x в базисе $\{e_k\}$. Для любых векторов $x, y \in \mathcal{R}^n$

$$(x, z) = \sum_{k=1}^{n} (x, e_k)(z, e_k).$$

Простейшим примером ОНБ (назовём такой базис естественным) является набор элементов

$$e_k = \langle \underbrace{0, \dots, 0, 1}_{k}, 0, \dots, 0 \rangle, \qquad k = 1, \dots, n. \tag{4.2}$$

Координаты элемента $x = \langle x_1, \dots, x_n \rangle$ в этом базисе совпадают с теми числами, которые образуют набор x, т. е. $x_k = (x, e_k)$ для всех $k = 1, \dots, n$. В $\mathcal{R}^n$ существует бесконечно много ОНБ, более того, любую ортонормированную систему $e_1, \dots, e_r$ с $1 \leqslant r < n$ всегда можно дополнить элементами $e_{r+1}, \dots, e_n$ до ОНБ всего пространства $\mathcal{R}^n$.

Линейное подпространство в $\mathcal{R}^n$ – это множество $\mathcal{L}$, в котором для любых $x, y \in \mathcal{L}$ и любых чисел $a, b \in \mathbb{R}$ элемент $ax + by$ принадлежит $\mathcal{L}$. В линейном подпространстве можно ввести ОНБ $e_1, \dots, e_r$ из элементов, принадлежащих $\mathcal{L}$. При этом $r = \dim \mathcal{L}$ и $0 < r < n$. В случае $r = n$ мы имеем $\mathcal{L} = \mathcal{R}^n$; можно также взять $r = 0$, тогда $\mathcal{L} = \{0\}$, т. е. $\mathcal{L}$ содержит только нулевой элемент.

Для любого $x \in \mathcal{L}$ справедливы формулы (4.1), в которых верхний индекс суммы нужно заменить на r. Множество

$$\mathcal{L}^{\perp} = \{x \in \mathcal{R}^n \colon (x, z) = 0 \text{ для всех } z \in \mathcal{L}\}$$

называется ортогональным дополнением к $\mathcal{L}$. Ортогональное дополнение является линейным подпространством, в нём можно ввести ОНБ, и $\dim \mathcal{L} + \dim \mathcal{L}^{\perp} = \dim \mathcal{R}^n$.

Любой $x \in \mathcal{R}^n$ можно единственным образом разложить как $x = x^{\circ} + x^{\perp}$, где $x^{\circ} \in \mathcal{L}$, $x^{\perp} \in \mathcal{L}^{\perp}$. Это называется разложением пространства в прямую сумму подпространств: $\mathcal{R}^n = \mathcal{L} \oplus \mathcal{L}^{\perp}$. Это разложение линейно в том смысле, что для любых $a, b \in \mathbb{R}$

$$(ax + bz)^{\circ} = ax^{\circ} + bz^{\circ}, \qquad (ax + bz)^{\perp} = ax^{\perp} + bz^{\perp},$$

если выполнены разложения $ax + bz = (ax + bz)^{\circ} + (ax + bz)^{\perp}$ и $x = x^{\circ} + x^{\perp}$, $z = z^{\circ} + z^{\perp}$.

Пусть $\mathcal{R}^m$ и $\mathcal{R}^n$ – два евклидовых пространства. Линейным оператором A, действующим из $\mathcal{R}^m$ в $\mathcal{R}^n$, называется соответствие, которое сопоставляет любому $x \in \mathcal{R}^m$ единственный вектор $y \in \mathcal{R}^n$ так, что $A(ax+bz) = aAx+bAz$ для любых $x, z \in \mathcal{R}^m$ и любых чисел $a, b \in \mathbb{R}$. Мы обозначаем такое соответствие как $A \colon \mathcal{R}^m \to \mathcal{R}^n$.

Если ввести в $\mathcal{R}^m$ и в $\mathcal{R}^n$ ОНБ $\{e_j\}$ и $\{\tilde{e}_j\}$ соответственно, то с линейным оператором A можно связать набор чисел $A_{ij} = (\tilde{e}_i, Ae_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. Эти числа являются элементами матрицы оператора. Верно и обратное: если задана матрица чисел $\{A_{ij}\}$ с $i = 1, \ldots, n$, $j = 1, \ldots, m$, то можно задать линейный оператор, сопоставляющий элементу $x = \langle x_1, \ldots, x_m \rangle \in \mathcal{R}^m$ элемент $y = \langle y_1, \ldots, y_n \rangle \in \mathcal{R}^n$ по правилу

$$y_i = \sum_{j=1}^{m} A_{ij} x_j, \qquad i = 1, \ldots, n.$$

При этом числа A_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, будут элементами матрицы оператора, если в качестве базиса в $\mathcal{R}^m$ и $\mathcal{R}^n$ мы выберем естественный ОНБ (4.2) соответствующей размерности. Далее мы часто будем отождествлять понятие оператора (как соответствия) и его матрицы (как таблицы чисел). При этом либо ортонормированные базисы будут указаны явно, либо наши утверждения для матриц будут справедливы при любом выборе базисов.

Для любого линейного оператора $A\colon \mathcal{R}^m \to \mathcal{R}^n$ существует линейный оператор $A^*\colon \mathcal{R}^n \to \mathcal{R}^m$, называемый сопряжённым оператором, такой что $(Ax, y) = (x, A^*y)$ для любых $x \in \mathcal{R}^m$, $y \in \mathcal{R}^m$. При любом выборе базисов матрица сопряжённого оператора является транспонированной по отношению к матрице A, т. е. $A^*_{ij} = A_{ji}$ для всех $i = 1, \ldots, n$, $j = 1, \ldots, m$. Понятно, что $(A^*)^* = A$. Иногда мы будем использовать верхний индекс Т для обозначения транспонирования матрицы, но в силу нашего отождествления матрицы и оператора по сути A^{T} – это то же самое, что A^*. Для сопряжённых операторов имеют место следующие хорошо известные равенства:

$$(A^*)^* = A, \qquad (A+B)^* = A^* + B^*, \qquad (AB)^* = B^*A^*, \tag{4.3}$$

которые доказываются, например, переходом к матричному представлению операторов.

Простейшим примером линейного оператора является единичный оператор $I\colon \mathcal{R}^n \to \mathcal{R}^n$, действующий по правилу $Ix = x$ для любого $x \in \mathcal{R}^n$. Матрица единичного оператора в любом ОНБ имеет вид $I = \mathrm{diag}(1, \ldots, 1)$, т. е. все n её диагональных элементов равны 1, а остальные элементы равны 0. Ещё более простым является нулевой оператор, который мы обозначаем символом O. Этот оператор действует по правилу $\mathrm{O}x = 0$ (нулевой вектор) для любого $x \in \mathcal{R}^n$. Все элементы матрицы нулевого оператора в любом ОНБ равны 0. Очевидно, $I = I^*$, $\mathrm{O} = \mathrm{O}^*$.

Важным классом операторов являются ортогональные проекторы. Пусть $\mathcal{L}$ – линейное подпространство в $\mathcal{R}^n$, $\mathcal{L}^\perp$ – его ортогональное дополнение. Тогда в силу единственности и линейности разложения любого $x \in \mathcal{R}^n$ в сумму $x = x^\circ + x^\perp$, где $x^\circ \in \mathcal{L}$, $x^\perp \in \mathcal{L}^\perp$, соответствие $x \mapsto x^\circ \in \mathcal{L}$ задаёт линейный оператор.

Определение 1. Оператор $\Pi\colon \mathcal{R}^n \to \mathcal{R}^n$, такой что $\Pi x = x^\circ$, где $x = x^\circ + x^\perp$ – разложение вектора на $x^\circ \in \mathcal{L}$, $x^\perp \in \mathcal{L}^\perp$, называется *ортогональным проектором* на $\mathcal{L}$.

1. Из единственности разложения элемента в сумму $x = x^\circ + x^\perp$ мы имеем $\Pi x = x$ для любого $x \in \mathcal{L}$ (поскольку тогда $x = x + 0$) и $\Pi x = 0$ для любого $x \in \mathcal{L}^\perp$ (поскольку тогда $x = 0 + x$).

2. Оператор Π самосопряжён, $\Pi = \Pi^*$. Действительно, для любых $x = x^\circ + x^\perp$, $z = z^\circ + z^\perp$, поскольку $(x^\circ, z^\perp) = 0$ и $(x^\perp, z^\circ) = 0$,

$$(\Pi x, z) = (x^\circ, z) = (x^\circ, z^\circ + z^\perp) = (x^\circ + x^\perp, z^\circ) = (x, \Pi z).$$

3. Имеет место равенство $\Pi^2 = \Pi$: вследствие свойства 1

$$\Pi x = x^\circ = \Pi x^\circ = \Pi^2 x, \qquad x \in \mathcal{R}^n.$$

Замечание 1. Очевидно, что если Π – ортогональный проектор на $\mathcal{L}$, то $(I - \Pi)$ – ортогональный проектор на $\mathcal{L}^\perp$. При этом в силу ортогональности слагаемых Πx и $(I - \Pi)x$ для любого $x \in \mathcal{R}^n$

$$\|x\|^2 = \|\Pi x\|^2 + \|(I - \Pi)x\|^2.$$

Легко видеть, что если $\mathcal{L}$ – линейное подпространство размерности $r < n$, проектор на которое равен Π, и ОНБ в $\mathcal{R}^n$ выбран так, что $e_1, \ldots, e_r$ – ОНБ в $\mathcal{L}$, а $e_{r+1}, \ldots, e_n$ – ОНБ в $\mathcal{L}^\perp$, то $\Pi e_k = e_k$ для $k = 1, \ldots, r$ и $\Pi e_k = 0$ для $k = r+1, \ldots, n$. Поэтому для любого $x \in \mathcal{R}^n$

$$\Pi x = \Pi \sum_{k=1}^{n} (x, e_k) e_k = \sum_{k=1}^{r} (x, e_k) e_k.$$

На самом деле верно и обратное: если некоторый линейный оператор Π действует на любой $x \in \mathcal{R}^n$ по указанному выше правилу, то Π – ортогональный проектор.

Линейные преобразования случайных векторов

В этом разделе мы рассматриваем набор с.в. $\xi_1, \ldots, \xi_n$ как случайный вектор ξ со значениями в n-мерном евклидовом пространстве $\mathcal{R}^n$. Другими словами, любая реализация случайного вектора ξ есть $x = \langle x_1, \ldots, x_n \rangle \in \mathcal{R}^n$.

Математическое ожидание случайного вектора ξ зададим как (неслучайный) вектор

$$M\xi = \langle M\xi_1, \ldots, M\xi_n \rangle \in \mathcal{R}^n.$$

Введём матрицу ковариаций этого вектора как матрицу Σ с элементами $\operatorname{cov}(\xi_i, \xi_j)$, $i, j = 1, \ldots, n$. На диагонали этой матрицы, очевидно, стоят дисперсии элементов. С.в. $\xi_1, \ldots, \xi_n$ можно рассматривать как координаты вектора ξ в естественном базисе. Другими словами, $\xi_k = (\xi, e_k)$, где $\{e_k\}$ задан в (4.2). В этом базисе элементы матрицы ковариаций суть матричные элементы линейного оператора $\Sigma\colon \mathcal{R}^n \to \mathcal{R}^n$:

$$(\Sigma e_i, e_j) = M(\xi - M\xi, e_i)(\xi - M\xi, e_j) = \mathrm{cov}(\xi_i, \xi_j), \quad i, j = 1, \dots, n.$$

Абстрактное действие этого оператора на любой вектор $x \in \mathcal{R}^n$ задаётся аналогичным равенством

$$(\Sigma x, y) = M(\xi - M\xi, x)(\xi - M\xi, y), \qquad x, y \in \mathcal{R}^n. \tag{4.4}$$

Далее мы, как и выше, будем отождествлять оператор Σ и его матрицу.

Очевидно, что если ξ и η – случайные векторы в $\mathcal{R}^n$, a и b – неслучайные числа, то

$$M(a\xi + b\eta) = aM\xi + bM\eta. \tag{4.5}$$

Также понятно, что для любого ортонормированного базиса $\{e_k\}$

$$M\|\xi - M\xi\|^2 = M(\xi - M\xi, \xi - M\xi) = M\sum_{k=1}^{n}(\xi - M\xi, e_k)^2 =$$
$$= \sum_{k=1}^{n} M(\xi - M\xi, e_k)(\xi - M\xi, e_k) = \sum_{k=1}^{n}(\Sigma e_k, e_k),$$

где мы воспользовались определением (4.4). Величина в правой части называется следом оператора Σ и обозначается как $\mathrm{Tr}\,\Sigma$. В матричном представлении $\mathrm{Tr}\,\Sigma$ равен сумме диагональных элементов матрицы ковариаций случайного вектора ξ, причём его значение не зависит от выбора базиса, поскольку левая часть равенства

$$M\|\xi - M\xi\|^2 = \mathrm{Tr}\,\Sigma \tag{4.6}$$

не зависит от выбора базиса.

Пусть $A\colon \mathcal{R}^n \to \mathcal{R}^m$ есть линейный оператор. Рассмотрим случайный вектор $A\xi$ со значениями в $\mathcal{R}^m$. Введём в $\mathcal{R}^n$ и в $\mathcal{R}^m$ какие-либо ортонормированные базисы $\{e_k\}$ и $\{\tilde{e}_j\}$ соответственно. Тогда для любого $j = 1, \dots, m$

$$M(A\xi, \tilde{e}_j) = M\left(A\sum_{k=1}^{n}(\xi, e_k)e_k, \tilde{e}_j\right) = M\left(\sum_{k=1}^{n}(\xi, e_k)(Ae_k, \tilde{e}_j)\right) =$$
$$= \sum_{k=1}^{n} M(\xi, e_k)(Ae_k, \tilde{e}_j) = \sum_{k=1}^{n}(Ae_k, \tilde{e}_j)M(\xi, e_k).$$

Мы получили поэлементную запись произведения матрицы A на вектор $M\xi$. Таким образом, для линейного оператора $A\colon \mathcal{R}^n \to \mathcal{R}^m$

$$MA\xi = A\,M\xi. \tag{4.7}$$

В правой части этого равенства мы видим стандартное действие линейного оператора A на неслучайный вектор $M\xi$. Кроме того, для любого неслучайного вектора $y \in \mathcal{R}^m$

$$M(y, A\xi) = M\sum_{j=1}^{m}(y, \tilde{e}_j)(A\xi, \tilde{e}_j) = \sum_{j=1}^{m}(y, \tilde{e}_j)M(A\xi, \tilde{e}_j),$$

следовательно,

$$M(y, A\xi) = (y, A\,M\xi), \tag{4.8}$$

поскольку $M(A\xi, \tilde{e}_j)$ – это математическое ожидание j-й координаты вектора $A\xi$, которое по определению равно j-й координате $(MA\xi, \tilde{e}_j)$ вектора $MA\xi = A\,M\xi$.

Теперь положим $M\xi = 0$ (нулевой вектор) и найдём $M\|A\xi\|^2$:

$$M\|A\xi\|^2 = M\sum_{j=1}^{m}(A\xi, \tilde{e}_j)(A\xi, \tilde{e}_j) = \sum_{j=1}^{m} M(\xi, A^*\tilde{e}_j)(\xi, A^*\tilde{e}_j) =$$
$$= \sum_{j=1}^{m}(\Sigma A^*\tilde{e}_j, A^*\tilde{e}_j) = \sum_{j=1}^{m}(A\Sigma A^*\tilde{e}_j, \tilde{e}_j).$$

Отсюда получаем, что для вектора ξ с нулевым математическим ожиданием

$$M\|A\xi\|^2 = \operatorname{Tr} A\Sigma A^*. \tag{4.9}$$

Воспользуемся этим свойствами для построения оценок векторных параметров.

4.1. Задачи линейной регрессии

Пусть требуется оценить вектор параметров $f = \langle f_1, \ldots, f_m\rangle$, однако измерены, причём с ошибками, могут быть лишь линейные комбинации его координат

$$a_{i1}f_1 + \cdots + a_{im}f_m, \qquad i = 1, \ldots, n,$$

где число измеряемых параметров не меньше, чем число неизвестных параметров, $n \geqslant m$. Результат ξ_i измерения каждой такой линейной комбинации сопровождается случайной аддитивной погрешностью ν_i, так что

$$\xi_i = a_{i1} f_1 + \dots + a_{im} f_m + \nu_i, \qquad i = 1, \dots, n; \tag{4.10}$$

соотношение (4.10) называется *линейной схемой измерения*. Эту схему можно записать в векторном виде, введя n-мерные векторы

$$\xi = \langle \xi_1, \dots, \xi_n \rangle, \quad \nu = \langle \nu_1, \dots, \nu_n \rangle \quad \text{и} \quad \mathsf{a}_j = \langle a_{1j}, \dots, a_{nj} \rangle$$

для $j = 1, \dots, m$. Тогда вектор ξ задаётся как линейная комбинация векторов $\mathsf{a}_1, \dots, \mathsf{a}_m$ плюс вектор ν:

$$\xi = f_1 \mathsf{a}_1 + \dots + f_m \mathsf{a}_m + \nu. \tag{4.11}$$

Если считать a_{ij}, $i = 1, \dots, n$, $j = 1, \dots, m$, матричными элементами линейного оператора $A\colon \mathcal{R}^m \to \mathcal{R}^n$ и рассматривать $f \in \mathcal{R}^m$ как вектор-столбец с элементами $f_1, \dots, f_m$, то мы приходим к ещё одной записи линейной модели измерения:

$$\xi = Af + \nu. \tag{4.12}$$

Приведем примеры измерений, выполненных по схеме (4.10).

Пример 1. Пусть в заданные моменты времени $t_1, \dots, t_n$ измеряются координаты материальной точки, движущейся прямолинейно и равноускоренно. Неизвестными параметрами являются начальное положение x_0, начальная скорость v_0 и ускорение a. Измерение координаты в каждый момент времени t_i, $i = 1, \dots, n$, сопровождается случайной погрешностью ν_i, такой что $M\nu_i = 0$, $D\nu_i = \sigma^2$, и погрешности измерений координаты в разные моменты времени некоррелированы: $M\nu_i\nu_k = 0$ при $i \neq k$ для всех $i, k = 1, \dots, n$. Поскольку закон равноускоренного движения имеет вид

$$S(t) = x_0 + v_0 t + \frac{at^2}{2},$$

схему измерения можно записать в виде

$$\xi_i = x_0 + v_0 t_i + \frac{at_i^2}{2} + \nu_i, \qquad i = 1, \dots, n,$$

или, положив $f_1 = x_0$, $f_2 = v_0$, $f_3 = a$, – в виде (4.11):

$$\xi = \sum_{j=1}^{3} \mathbf{a}_j f_j + \nu, \qquad \begin{cases} \mathbf{a}_1 = \langle 1, \dots 1 \rangle \\ \mathbf{a}_2 = \langle t_1, \dots, t_n \rangle, \\ \mathbf{a}_3 = \langle t_1^2/2, \dots, t_n^2/2 \rangle, \end{cases}$$

при этом случайный вектор ν имеет нулевое математическое ожидание и матрицу ковариаций $\sigma^2 I$.

Пример 2. Пусть измеряется энергетический спектр излучения $f(E)$, $E \geqslant 0$. Излучение со спектральным составом $f(E)$ поступает на вход спектрометра с аппаратной функцией $a(E, E')$, выходной сигнал спектрометра, измеряющего интенсивность излучения с энергией E'_i, $i = 1, \dots, n$, даётся формулой

$$g(E'_i) = \int_0^\infty a(E'_i, E) f(E)\, dE. \tag{4.13}$$

Измерение выходного сигнала спектрометра сопровождается аддитивной погрешностью, так что результат измерения можно записать в виде $\xi_i = g(E'_i) + \nu_i$, $i = 1, \dots, n$. Записав интеграл в (4.13) по формуле численного интегрирования в виде

$$\int_0^\infty a(E'_i, E) f(E)\, dE = \sum_{j=1}^{m} a_{ij} f(E_j) + \Delta_j,$$

где E_j, $j = 1, \dots, m$ – узлы сетки, Δ_j – остаточный член, и выбрав численный метод интегрирования так, чтобы Δ_j был пренебрежимо мал по сравнению с погрешностью измерения, получим схему измерения в виде $\xi = Af + \nu$, где матрица A имеет элементы a_{ij}, координаты вектора $f \in \mathcal{R}^m$ равны $f(E_j)$, $i = 1, \dots, n$, $j = 1, \dots, m$.

Математическая модель линейной схемы измерений считается заданной, если известны значения коэффициентов a_{ij}, $i = 1, \dots, n$, $j = 1, \dots, m$, математические ожидания измерительных погрешностей $M\nu_i = 0$, $i = 1, \dots, n$, и задана матрица ковариаций случайного вектора ν: её матричные элементы определяются равенствами $\Sigma_{ik} = \operatorname{cov}(\nu_i, \nu_k)$ при $i \neq k$, $\Sigma_{ii} = D\nu_i$, $i, k = 1, \dots, n$. Равенство $M\nu = 0$ оправданно, так как систематическая погрешность, как правило, может быть учтена.

Кроме того, далее для простоты будем считать, что погрешности измерений некоррелированы и имеют одинаковую дисперсию $D\nu_i = \sigma^2$. Тогда ковариационная матрица вектора погрешностей имеет вид $\Sigma = \sigma^2 I$. Кроме того, будем считать, что коэффициенты a_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m$, таковы, что столбцы $\mathbf{a}_j$ матрицы A линейно независимы. Это, в частности, означает, что матрица $A^* A \colon \mathcal{R}^m \to \mathcal{R}^m$ имеет обратную.

Задача линейной регрессии

Задача линейной регрессии для нашей модели состоит в следующем. Рассмотрим класс линейных оценок, т. е. оценок вида $\hat{f} = R\xi$, где R – линейный оператор из $\mathcal{R}^n$ в $\mathcal{R}^m$. Зададим с.к. погрешность такой оценки:

$$h(R) = M\|R\xi - f\|^2.$$

Чем меньше $h(R)$, тем точнее $R\xi$ оценивает f. Потребуем несмещённость оценки, т. е. наложим условие $M\hat{f}_j = f_j$ для любого $j = 1, \ldots, m$ и для любого значения $f_j \in \mathbb{R}$. Поскольку $\hat{f} = R\xi$, мы можем переписать условие несмещённости как цепочку векторных равенств, используя свойства (4.5)–(4.9):

$$MR\xi = MR(Af + \nu) = M\,RAf + MR\nu = RAf + MR\nu = RAf.$$

Здесь мы учли, что RAf – неслучайный вектор и $MR\nu = 0$. Условие должно быть выполнено для любого $f \in \mathcal{R}^n$, что приводит к операторному равенству

$$RA = I. \tag{4.14}$$

В результате мы приходим к следующей задаче: требуется найти оператор $R_0 \colon \mathcal{R}^n \to \mathcal{R}^m$, такой что

$$M\|R_0\xi - f\|^2 = \min_{R\colon RA=I} M\|R\xi - f\|^2. \tag{4.15}$$

Эта задача называется задачей линейной регрессии. Если рассматривать скалярную величину $M\|R\xi - f\|^2$ как аналог дисперсии для векторной несмещённой оценки $R\xi$, то можно сказать, что задача линейной регрессии – это задача поиска НОМД для вектора $f \in \mathcal{R}^m$ в классе линейных оценок, построенных по вектору $\xi \in \mathcal{R}^n$.

Заметим, что требование несмещённости совпадает с необходимым и достаточным условием того, что с.к. погрешность конечна при любом $f \in \mathcal{R}^m$. Действительно,

$$\begin{aligned} M\|R\xi - f\|^2 &= M\|R(Af+\nu) - f\|^2 = \\ &= M\big(\|(RA-I)f\|^2 + 2((RA-I)f, R\nu) + \|R\nu\|^2\big) = \\ &= \|(RA-I)f\|^2 + 2((RA-I)f, MR\nu) + M\|R\nu\|^2 = \\ &= \|(RA-I)f\|^2 + M\|R\nu\|^2, \end{aligned}$$

где мы учли неслучайность вектора $(RA-I)f$, формулу (4.8) вместе с условием $M\nu = 0$. Если не выполнено равенство (4.14), то с.к. погрешность оценки зависит от неизвестного вектора f, а поскольку f – любой вектор из $\mathcal{R}^m$, погрешность может принимать любое, в том числе и сколь угодно большое, значение. Наоборот, при выполнении (4.14) с.к. погрешность оценки $\hat{f}_j$ оказывается конечной и равной

$$M\|R\nu\|^2 = \operatorname{Tr} R\Sigma R^* = \sigma^2 \operatorname{Tr} RR^*,$$

где мы воспользовались формулой (4.9) и тем, что $\Sigma = \sigma^2 I$. Таким образом, задача линейной регрессии сводится к поиску оператора R_0, который удовлетворяет условию $R_0 A = I$ и доставляет минимум $\operatorname{Tr} RR^*$ среди всех операторов R, таких что $RA = I$.

Теорема 1. *Пусть в линейной схеме $\xi = Af + \nu$ погрешность ν имеет нулевое среднее и ковариационный оператор $\Sigma = \sigma^2 I$, а оператор A^*A невырожден. Тогда оператор $R_0 = (A^*A)^{-1}A^*$ является единственным решением задачи линейной регрессии. С.к. погрешность наилучшей оценки равна $h(R_0) = \sigma^2 \operatorname{Tr}(A^*A)^{-1}$.*

Доказательство. Прежде всего заметим, что $R_0\xi$ – несмещённая оценка для f, поскольку $R_0 A = (A^*A)^{-1}A^*A = I$.

Пусть R – другой оператор, для которого $RA = I$ и $Z = R - R_0$, тогда $ZA = \mathrm{O}$ (нулевой оператор из $\mathcal{R}^m$ в $\mathcal{R}^m$). Отсюда, очевидно, $A^*Z^* = (ZA)^* = \mathrm{O}$.

Для оператора $R = R_0 + Z$ погрешность оценивания $h(R)$ равна $M\|R\nu\|^2 = \sigma^2 \operatorname{Tr} RR^*$ и

$$RR^* = (R_0 + Z)(R_0 + Z)^* = R_0 R_0^* + R_0 Z^* + Z R_0^* + ZZ^*.$$

При этом $R_0 Z^* = (A^*A)^{-1}A^*Z^* = (A^*A)^{-1}(ZA)^* = \mathrm{O}$, следовательно, $ZR_0^* = (R_0Z^*)^* = \mathrm{O}$. Итак, для любого $R = R_0 + Z$, такого что $RA = I$, мы имеем

$$\mathrm{Tr}(RR^*) = \mathrm{Tr}(R_0R_0^*) + \mathrm{Tr}(ZZ^*).$$

При этом, если записать матрицу оператора Z в каком-либо базисе, обозначив её элементы как Z_{ji}, $j = 1, \ldots, m$, $i = 1, \ldots, n$, то

$$\begin{aligned}\mathrm{Tr}(ZZ^*) = \sum_{j=1}^{m}(ZZ^*)_{jj} = \sum_{j=1}^{m}\sum_{i=1}^{n} Z_{ji}Z_{ij}^{\mathrm{T}} = \\ = \sum_{j=1}^{m}\sum_{i=1}^{n} Z_{ji}Z_{ji} = \sum_{j=1}^{m}\sum_{i=1}^{n}(Z_{ji})^2 \geqslant 0,\end{aligned}$$

откуда следует, что $\mathrm{Tr}(RR^*) \geqslant \mathrm{Tr}(R_0R_0^*)$, причём равенство возможно, только если $Z_{ji} = 0$ для всех $j = 1, \ldots, m$, $i = 1, \ldots, n$. После умножения на числовой множитель σ^2 получаем

$$h(R) \geqslant h(R_0) = \sigma^2\,\mathrm{Tr}\big((A^*A)^{-1}A^* \cdot A(A^*A)^{-1}\big) = \sigma^2\,\mathrm{Tr}(A^*A)^{-1},$$

причём неравенство превращается в равенство тогда и только тогда, когда $R = R_0$. Теорема доказана.

Оценки параметров линейной регрессии методом наименьших квадратов

Другой метод получения оценок параметров линейной схемы измерения состоит в подборе такого вектора f, чтобы предсказанное моделью значение Af как можно меньше отличалось от результата ξ измерения (4.12), критерием отличия является квадрат нормы разности вектора измерения ξ и предсказанного моделью вектора Af. Оценка вектора f, минимизирующая этот квадрат нормы, называется оценкой метода наименьших квадратов (МНК). Эта оценка определяется как решение экстремальной задачи

$$\min_{f\in\mathcal{R}^m} \|\xi - Af\|^2. \tag{4.16}$$

Квадрат нормы $\|\xi - Af\|^2$ называется невязкой.

Теорема 2. *Пусть оператор A^*A невырожден. Тогда*

$$f_0 = (A^*A)^{-1}A^*\xi \in \mathcal{R}^m$$

доставляет единственный минимум в (4.16).

Доказательство. Для любого $f \in \mathcal{R}^m$ положим $h = f - f_0$. Тогда $f = f_0 + h$ и

$$\begin{aligned}\|\xi - Af\|^2 &= \|\xi - Af_0 - Ah\|^2 = \\ &= \|\xi - Af_0\|^2 - 2(\xi - Af_0, Ah) + \|Ah\|^2 = \\ &= \|\xi - Af_0\|^2 - 2(A^*\xi, h) + 2(A^*Af_0, h) + \|Ah\|^2.\end{aligned}$$

При этом $A^*Af_0 = A^*A(A^*A)^{-1}A^*\xi = A^*\xi$, поэтому $(A^*Af_0, h) = (A^*\xi, h)$, и линейные по h члены сокращаются. Отсюда

$$\|\xi - Af\|^2 = \|\xi - Af_0\|^2 + \|Ah\|^2 \geqslant \|\xi - Af_0\|^2,$$

причём неравенство превращается в равенство тогда и только тогда, когда $Ah = 0$. Отсюда следует, что $A^*Ah = 0$. Если, с другой стороны, $A^*Ah = 0$, то

$$\|Ah\|^2 = (Ah, Ah) = (A^*Ah, h) = 0 \quad \Longrightarrow \quad Ah = 0.$$

Таким образом, $Ah = 0$ равносильно $A^*Ah = 0$, а последнее равенство возможно только при $h = 0$ в силу невырожденности A^*A (мы получаем $h = (A^*A)^{-1}0 = 0$). Теорема доказана.

Сформулируем полученные результаты в форме теоремы, которую называют теоремой Гаусса–Маркова.

Теорема 3. *Пусть случайный вектор* $\xi \in \mathcal{R}^n$ *получен как результат измерения в линейной схеме* (4.12), *в которой матрица* $A\colon \mathcal{R}^m \to \mathcal{R}^n$, $m \leqslant n$, *известна, а вектор случайных погрешностей измерений имеет нулевое математическое ожидание и некоррелированные координаты с одинаковыми дисперсиями* σ^2. *Пусть матрица* A^*A *невырожденна. Тогда линейные НОМД координат вектора* $f \in \mathcal{R}^m$, *являющиеся решением задачи линейной регрессии* (4.15), *даются вектором* $\hat{f} = R_0\xi = (A^*A)^{-1}A^*\xi$ *и совпадают с оценкой* f_0 *МНК* (4.16). *Матрица ковариаций оценок равна* $\sigma^2(A^*A)^{-1}$.

Дисперсия оценки j-й координаты вектора f равна $\sigma^2(A^*A)_{jj}$, внедиагональные элементы матрицы $\sigma^2(A^*A)$ задают ковариации оценок.

4.2. Множества, оценивающие параметры линейной схемы при нормальном распределении погрешности измерений

Пусть в (4.12) случайный вектор ν погрешности измерений имеет независимые координаты $\nu_i \sim \mathbf{N}(0, \sigma^2)$, $i = 1, \ldots, n$. Тогда мы пишем $\nu \sim \mathbf{N}(0, \sigma^2 I)$, имея в виду, что его математическое ожидание $M\nu = 0$ (нулевой вектор) и ковариационный оператор $\Sigma = \sigma^2 I$.

При нормальной погрешности в линейной модели (4.12) по результату измерения ξ можно построить случайные множества, накрывающие истинные значения неизвестных параметров f с заданной вероятностью, в том числе если неизвестна дисперсия σ^2 погрешности измерений.

Воспользуемся свойствами $\mathbf{N}3$, $\mathbf{N}1$ из раздела 3.2: мы знаем, что если с.в. $\nu_k \sim \mathbf{N}(0, \sigma^2)$, $k = 1, \ldots, n$, независимы, то

$$\frac{1}{\sigma^2} \sum_{k=1}^{n} \nu_k^2 \sim \mathbf{X}_n^2.$$

Тогда для вектора $\nu \sim \mathbf{N}(0, \sigma^2 I)$ и любого ОНБ $\{e_i\}$

$$\sum_{k=1}^{n} \nu_k^2 = \sum_{i=1}^{n} (\nu, e_i)^2 = \|\nu\|^2 \sim \sigma^2 \mathbf{X}_n^2.$$

Здесь мы обозначили $\sigma^2 \mathbf{X}_n^2$ распределение с.в. $\sigma^2 \chi_n^2$, где $\chi_n^2 \sim \mathbf{X}_n^2$.

Это свойство является частным случаем другого известного свойства нормального распределения.

Свойство N6. *Пусть случайный вектор $\nu \sim \mathbf{N}(0, \sigma^2 I)$, тогда для любого ОНБ $\{e_i\}$ и любого $r = 1, \ldots, n$ с.в. (ν, e_i), $i = 1, \ldots, r$, имеют распределение $\mathbf{N}(0, \sigma^2)$ и независимы.*

Отсюда следует, что если Π – ортогональный проектор на подпространство $\mathcal{L}$ размерности r, $1 \leqslant r < n$, то с.в.

$$\|\Pi\nu\|^2 \sim \sigma^2 \mathbf{X}_r^2, \qquad \|(I - \Pi)\nu\|^2 \sim \sigma^2 \mathbf{X}_{n-r}^2$$

и независимы. Чтобы это показать, достаточно выбрать в $\mathcal{R}^n$ ОНБ $e_1, \ldots, e_r, e_{r+1}, \ldots, e_n$ так, чтобы первые r векторов лежали в $\mathcal{L}$, а остальные – в $\mathcal{L}^\perp$.

Перейдём к построению оценивающих множеств. Рассмотрим с.в. $s^2 = \|\xi - Af\|^2$. Поскольку $\xi - Af = \nu \sim \mathbf{N}(0, \sigma^2 I)$, имеем $s^2 \sim \sigma^2 \mathbf{X}_n$, .

Введём пространство значений $\mathcal{R}(A)$ оператора A:

$$\mathcal{R}(A) = \{u \in \mathcal{R}^n \colon u = Ax, x \in \mathcal{R}^m\}.$$

В матричном представлении оно состоит из всех линейных комбинаций векторов-столбцов $\mathtt{a}_j = \langle a_{1j}, \ldots, a_{nj} \rangle$, $j = 1, \ldots, m$, матрицы оператора A. Как мы отмечали выше, если матрица A^*A невырожденна, то условие $Ax = 0$ влечёт $x = 0$. Это означает, что столбцы $\mathtt{a}_j$, $j = 1, \ldots, m$, линейно независимы. В самом деле, условие $Ax = 0$ для $x = \langle x_1, \ldots, x_m \rangle$ можно записать как

$$x_1 \mathtt{a}_1 + \cdots + x_m \mathtt{a}_m = 0$$

(в правой части равенства стоит нулевой столбец), а по нашему условию это влечёт $x_1 = \cdots = x_m = 0$. Поэтому $\dim \mathcal{R}(A) = m$. Предположим, что $m < n$, где n – размерность пространства измерений ξ.

Также введём ортогональное дополнение к $\mathcal{R}(A)$

$$\mathcal{R}^{\perp}(A) = \{v \in \mathcal{R}^n \colon (v, u) = 0 \text{ для всех } u \in \mathcal{R}(A)\}.$$

Тогда любой вектор y пространства $\mathcal{R}^n$ можно единственным образом представить как сумму $y = u + v$, где $u \in \mathcal{R}(A)$, $v \in \mathcal{R}^{\perp}(A)$. Найдём проектор Π на $\mathcal{R}(A)$. Он задаётся, как известно, равенством $\Pi y = u$. Покажем, что $\Pi = A(A^*A)^{-1}A^*$. Если $y = Ax \in \mathcal{R}(A)$, то

$$\Pi y = A(A^*A)^{-1}A^* y = A(A^*A)^{-1}A^*Ax = Ax = y.$$

Если $y \in \mathcal{R}^{\perp}(A)$, т. е. $(y, Ax) = 0$ для любого $x \in \mathcal{R}^m$, то, последовательно перенося сопряжённые операторы, имеем

$$\begin{aligned}
\|A(A^*A)^{-1}A^*y\|^2 &= (A(A^*A)^{-1}A^*y, A(A^*A)^{-1}A^*y) = \\
&= ((A^*A)^{-1}A^*y, A^*A(A^*A)^{-1}A^*y) = \\
&= ((A^*A)^{-1}A^*y, A^*y) = (A^*y, (A^*A)^{-1}A^*y) = \\
&= (y, A(A^*A)^{-1}A^*y).
\end{aligned}$$

В этой цепочке равенств мы учли, что $[(A^*A)^{-1}]^* = (A^*A)^{-1}$. Положим $z = (A^*A)^{-1}A^*y$, тогда в силу $y \in \mathcal{R}^{\perp}(A)$

$$\|A(A^*A)^{-1}A^*y\|^2 = (y, Az) = 0, \qquad A(A^*A)^{-1}A^*y = 0.$$

Таким образом, $\Pi = A(A^*A)^{-1}A^*$ проецирует на линейное подпространство $\mathcal{R}(A)$ размерности m, и $I - \Pi$ есть проектор на линейное подпространство $\mathcal{R}^{\perp}(A)$ размерности $n - m$.

Рассмотрим линейную схему измерения $\xi = Af + \nu$. Поскольку $Af \in \mathcal{R}(A)$, имеем $(I - \Pi)Af = 0$, следовательно,

$$(I - \Pi)\xi = (I - \Pi)Af + (I - \Pi)\nu = (I - \Pi)\nu.$$

Далее заметим, что $Af \in \mathcal{R}(A)$, поэтому

$$\Pi\nu = \Pi(\xi - Af) = \Pi\xi - \Pi Af = \Pi\xi - Af.$$

При этом

$$\Pi\xi = A(A^*A)^{-1}A^*\xi = AR_0\xi = A\hat{f},$$

где R_0 – оператор, дающий решение задачи линейной регрессии, а $\hat{f}$ – наилучшая в с.к. смысле линейная несмещённая оценка вектора f.

Таким образом, по свойству ортогональных проекций нормально распределённого вектора ν с.в.

$$S_2^2 = \|\Pi\xi - Af\|^2 = \|A\hat{f} - Af\|^2 = \|\Pi\nu\|^2 \sim \sigma^2\mathbf{X}_m \tag{4.17}$$

и не зависит от с.в.

$$S_1^2 = \|(I - \Pi)\xi\|^2 = \|(I - \Pi)\nu\|^2 \sim \sigma^2\mathbf{X}_{n-m}, \tag{4.18}$$

так как случайные векторы $\Pi\nu$ и $(I - \Pi)\nu$ лежат в ортогональных подпространствах $\mathcal{R}(A)$ и $\mathcal{R}^{\perp}(A)$.

Основываясь на свойствах распределений этих проекций, построим множество в $\mathcal{R}^m$, оценивающее одновременно все координаты вектора f, интервальные оценки каждой его координаты, а в случае, когда дисперсия σ^2 погрешности измерений неизвестна, получим оценку и для неё.

1. Пусть сначала дисперсия σ^2 погрешности измерений известна. Тогда, чтобы построить оценивающее множество для вектора f, можно использовать с.в. S_2^2, заданную в (4.17).

Выбрав уровень доверия γ, можем записать

$$P(\sigma^{-2}\|A(f-\hat{f}\,)\|^2<\varepsilon_\gamma)=P(\sigma^{-2}\cdot S_2^2<\varepsilon_\gamma)=\gamma, \tag{4.19}$$

где ε_γ выбирается из условия $P(\chi_m^2<\varepsilon_\gamma)=\gamma$. Множество

$$\mathcal{E}_\gamma=\{f\in\mathcal{R}^m\colon \|A(f-\hat{f}\,)\|^2<\sigma^2\varepsilon_\gamma\}$$

представляет собой m-мерный эллипсоид со случайным центром в точке $\hat{f}=(A^*A)^{-1}A^*\xi$, который накрывает истинное значение $f\in\mathcal{R}^m$ с вероятностью γ.

2. Если дисперсия σ^2 неизвестна, то для её оценки можно использовать статистику $S_1^2\sim\sigma^2\mathbf{X}_{n-m}^2$ из (4.18). По определению с.в. χ_r^2 как суммы $\sum_{k=1}^r\alpha_k^2$ квадратов независимых с.в. α_k с распределением $\mathbf{N}(0,1)$ имеем

$$M\chi_r^2=M\sum_{k=1}^r\alpha_k^2=\sum_{k=1}^r M\alpha_k^2=r.$$

Как известно, $D\chi_r^2=2r$. Таким образом,

$$\hat{\sigma}^2=\frac{\|(I-\Pi)\xi\|^2}{n-m} \tag{4.20}$$

есть несмещённая оценка для σ^2. Дисперсия $D\hat{\sigma}^2=2\sigma^4/(n-m)$ этой оценки зависит от неизвестной величины σ^2, тем не менее полученное соотношение позволяет утверждать, что чем выше размерность вектора ξ (т. е. чем больше проведено измерений по схеме (4.10)), тем точнее в с.к. можно оценить значение σ^2. Если распределение ν отлично от нормального, то несмещённость оценки $\hat{\sigma}^2$ сохраняется, а для вычисления дисперсии этой оценки необходимы дополнительные сведения.

3. Построим оценивающее множество для f при неизвестной дисперсии σ^2. Для этого нам потребуется новое распределение.

Определение 2. С.в. $\varphi_{s,m}$ имеет *распределение Снедекора–Фишера* с r и s степенями свободы, и мы пишем $\varphi_{r,s}\sim\mathbf{\Phi}_{r,s}$, если

$$\varphi_{r,s}=\frac{\chi_r^2/r}{\chi_s^2/s},$$

где с.в. χ_r^2 и χ_s^2 независимы и имеют распределения хи-квадрат с s и m степенями свободы. Также можно записать

$$\varphi_{r,s} = \frac{\frac{1}{r}\sum_{k=1}^{r}\xi_k^2}{\frac{1}{s}\sum\limits_{k=r+1}^{r+s}\xi_k^2},$$

где с.в. $\xi_1, \ldots, \xi_r, \xi_{r+1}, \ldots, \xi_{r+s}$ имеют распределение $\mathbf{N}(0,1)$ и независимы.

Поставим в (4.19) вместо σ^2 её несмещённую оценку по формуле (4.20). В силу (4.17) и (4.18) получившаяся дробь имеет распределение Снедекора–Фишера с m и $n-m$ степенями свободы:

$$\frac{\frac{1}{m}\|A(f-\hat{f}\,)\|^2}{\frac{1}{n-m}\|(I-\Pi)\xi\|^2} \sim \mathbf{\Phi}_{m,n-m}.$$

Выбрав уровень доверия γ, в этом случае также можем записать

$$P\left(\frac{\frac{1}{m}\|A(f-\hat{f}\,)\|^2}{\frac{1}{n-m}\|(I-\Pi)\xi\|^2} \leqslant \varepsilon\right) = \gamma, \tag{4.21}$$

где $\hat{f} = (A^*A)^{-1}A^*\xi$, а ε выбирается из условия

$$P(\varphi^2_{m,(n-m)} < \varepsilon) = \gamma.$$

Получаем, что истинное значение вектора $f \in \mathcal{R}^m$ с вероятностью γ накрывается случайным эллипсоидом

$$\left\{ f \in \mathcal{R}^m \colon \|A(f-\hat{f}\,)\|^2 < \varepsilon\frac{m}{n-m}\|(I-\Pi)\xi\|^2 \right\}$$

с центром в точке $\hat{f} = (A^*A)^{-1}A^*\xi$.

4. Построим теперь интервальную оценку j-й координаты вектора f при известной дисперсии σ^2. Поскольку

$$\hat{f} = R_0\xi = R_0(Af+\nu) = f + R_0\nu,$$

оценка $\hat{f}_j = (R_0\xi)_j$ координаты f_j является линейной комбинацией нормальных с.в. и, следовательно, имеет нормальное распределение (см. свойство $\mathbf{N}2$ в разделе 3.2). Эта оценка несмещённая, а её дисперсия равна j-му диагональному элементу матрицы $(A^*A)^{-1}$, умноженному на σ^2. Обозначим эту дисперсию как $\sigma^2 D_{jj}$. Следовательно,

$$\frac{\hat{f}_j - f_j}{\sqrt{\sigma^2 D_{jj}}} \sim \mathbf{N}(0,1), \qquad P\left(\frac{|\hat{f}_j - f_j|}{\sqrt{\sigma^2 D_{jj}}} < \varepsilon\right) = \gamma, \tag{4.22}$$

если ε выбрать из равенства $P(|\nu_*| \leqslant \varepsilon) = \gamma$ для стандартной нормальной с.в. ν_*. Интервальная оценка для f_j даётся равенством

$$P\big(|f_j - \hat{f}_j| < \varepsilon\sqrt{\sigma^2 D_{jj}}\,\big) = \gamma. \tag{4.23}$$

5. Пусть дисперсия погрешности измерений неизвестна. Как и при построении доверительного эллипсоида, заменим в формуле (4.22) дисперсию σ^2 на её оценку (4.20). Заметим, что случайный вектор $\eta = A(\hat{f} - f) \in \mathcal{R}(A)$ не зависит от $(I - \Pi)\xi \in \mathcal{R}^{\perp}(A)$, поэтому $(\hat{f} - f) \sim \mathbf{N}(0, \sigma^2 D_{jj})$ как вектор-функция от η не зависит от $\|(I - \Pi)\xi\|^2 \in \sigma^2\mathbf{X}_{n-m}$. Отсюда получаем, что дробь

$$\frac{\frac{1}{\sigma^2\sqrt{D_{jj}}}(\hat{f}_j - f_j)}{\frac{1}{\sigma^2\sqrt{n-m}}\|(I - \Pi)\xi\|} = \sqrt{\frac{n-m}{D_{jj}}}\,\frac{\hat{f}_j - f_j}{\|(I - \Pi)\xi\|}$$

имеет распределение Стьюдента с $n - m$ степенями свободы.

Выбирая по заданному уровню доверия γ значение ε из равенства $P(|t_{n-m}| \leqslant \varepsilon) = \gamma$, получим

$$P(|t_{n-m}| \leqslant \varepsilon) = P\left(\sqrt{\frac{n-m}{D_{jj}}}\,\frac{|\hat{f}_j - f_j|}{\|(I - \Pi)\xi\|} < \varepsilon\right) = \gamma \tag{4.24}$$

и найдем интервальную оценку для f_j в виде интервала случайной длины со случайным центром $\hat{f}_j$, разрешая неравенство под знаком вероятности относительно $|f_j - \hat{f}_j|$:

$$P\left(|f_j - \hat{f}_j| < \varepsilon\sqrt{\frac{D_{jj}}{n-m}}\|(I - \Pi)\xi\|\right) = \gamma. \tag{4.25}$$

Заметим, что длина интервала увеличивается с ростом дисперсии D_{jj} оценки $\hat{f}_j$ и уменьшается с ростом $n - m$, т. е. когда количество измерений увеличивается по сравнению с количеством оцениваемых параметров.

Список литературы

1. Пытьев Ю. П., Шишмарев И. А. Теория вероятностей, математическая статистика и элементы теории возможностей для физиков. — Издательство МГУ, Москва, 2023. —410 с.

2. Боровков А. А. Математическая статистика — Издательство «Лань», СанктПетербург, 2010. — 704 с.

3. Чернова Н. И. Математическая статистика: Учебное пособие — Издательство НГУ, Новосибирск, 2007. — 148 с.

4. Коршунов Д. А., Чернова Н. И. Сборник задач и упражнений по математической статистике. Учебное пособие. 2-е изд., испр. — Издательство во Института математики СО РАН, Новосибирск, 2004. — 128 с.

5. Горяинов В. Б., Павлов И. В., Цветкова Г. М. Математическая статистика: Учебник для вузов. — Издательство МГТУ им. Н. Э. Баумана, Москва, 2001. — 424 с..