Содержание

Лекция 1: Задачи и понятия математической статистики. Генеральная и выборочная совокупности

Математическая статистика изучает методы сбора и обработки статистических данных с целью получения теоретических и практических выводов вероятностного характера.

Генеральная и выборочная совокупности

Пусть X — случайная величина, связанная с экспериментом ε, с неизвестным законом распределения.

Генеральная совокупность — множество всех возможных значений, которые может принимать случайная величина X.

Выборочная совокупность значений случайной величины X объема n (или просто выборка объема n) — совокупность n реализаций случайной величины X:

$$x_1, x_2, \ldots, x_n \tag{1}$$

полученных в результате проведения серии из n экспериментов ε.

Задача математической статистики: построение надёжных заключений о генеральной совокупности на основе анализа выборочной совокупности.

Вариационный ряд и распределение выборки

В общем случае выборочные значения (1) могут повторяться:

$$\begin{array}{c|cccc} x_i & x_1 & x_2 & \ldots & x_k \\ \hline n_i & n_1 & n_2 & \ldots & n_k \end{array} \tag{2}$$

где $n_1 + n_2 + \cdots + n_k = n$, $x_1 < x_2 < \cdots < x_k$.

Ряд $x_1, \ldots, x_k$ называется **вариационным рядом**; натуральные числа n_i, $i = 1, 2, \ldots, k$ — **частотами**; числа $w_i = \frac{n_i}{n}$ — **относительными частотами**; $x_1, x_2, \ldots, x_k$ — различные значения выборки; таблица (2) — **распределением выборки**.

Ясно, что $\sum_{i=1}^{k} n_i = n$ и $\sum_{i=1}^{k} w_i = 1$.

Эмпирическая функция распределения

Функция распределения случайной величины X:

$$F(x) = P\{X < x\}, \quad x \in (-\infty, \infty)$$

может быть оценена (исходя из статистического восприятия вероятности) так:

$$P\{X < x\} \approx \frac{\sum\limits_{i:x_i<x} n_i}{n}$$

где $\sum_{i:x_i<x} n_i$ — сумма всех n_i таких, что $x_i < x$.

Обозначим $n_x = \sum_{i:x_i<x} n_i$. По определению, функция

$$F^*(x) = \frac{n_x}{n}$$

называется **эмпирической функцией распределения** случайной величины X, полученной по выборке (2).

Очевидны свойства $F^*(x)$:

1. $0 \leq F^*(x) \leq 1$
2. $F^*(x)$ — неубывающая функция
3. поскольку в (2) x_1 — наименьшее, а x_k — наибольшее значения выборки, $F^*(x) = 0$ при $x \leq x_1$ и $F^*(x) = 1$ при $x \geq x_k$

Полигон частот и относительных частот

Полигон частот — ломаная, отрезки которой соединяют точки $(x_1, n_1), \ldots, (x_k, n_k)$.

Полигон относительных частот — ломаная, отрезки которой соединяют точки $(x_1, w_1), \ldots, (x_k, w_k)$, где $w_1 + w_2 + \cdots + w_k = 1$.

Гистограмма

Разобьем все значения $x_1, x_2, \ldots, x_k$ наблюдаемой случайной величины на некоторое число m частичных интервалов длины h:

$$[x_1, x_1 + h), [x_1 + h, x_1 + 2h), \ldots, [x_{k-1}, x_k]$$

Обозначим:

$$h_j = \sum n_i \quad \text{для которых} \quad x_i \in [x_{j-1}, x_j), \text{ при} \quad j < m \quad \text{или} \quad x_i \in [x_{m-1}, x_m]$$

Гистограммой частот называют фигуру, состоящую из ступенчатых прямоугольников, основаниями которых служат интервалы длиной h, а высоты равны $\frac{h_j}{h}$.

Площадь всех прямоугольников будет равна:

$$\sum_{j=1}^{m} \frac{h_j}{h} \cdot h = \sum_{j=1}^{m} h_j = n$$

— общему числу всех значений случайной величины.

Если высоты прямоугольников сделать равными $\frac{h_j}{nh}$, то соответствующая ступенчатая фигура называется **гистограммой относительных частот**. В этом случае площадь всех прямоугольников будет равна 1, а сама гистограмма даст (в случае непрерывной случайной величины X) представление о плотности распределения случайной величины X.

Некоторые характеристики выборки

Рассмотрим выборочное распределение:

x_i	x_1	x_2	$\ldots$	x_k
n_i	n_1	n_2	$\ldots$	n_k

где $n_1 + n_2 + \cdots + n_k = n$.

Определим следующие характеристики выборки:

Выборочное среднее:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + \cdots + x_k n_k}{n}$$

Выборочная дисперсия:

$$D_x = \frac{(x_1 - \bar{x})^2 n_1 + (x_2 - \bar{x})^2 n_2 + \cdots + (x_k - \bar{x})^2 n_k}{n}$$

Эмпирический начальный момент порядка s:

$$\nu_s = \frac{x_1^s n_1 + x_2^s n_2 + \cdots + x_k^s n_k}{n}$$

Эмпирический центральный момент порядка s:

$$\mu_s = \frac{(x_1 - \bar{x})^s n_1 + (x_2 - \bar{x})^s n_2 + \cdots + (x_k - \bar{x})^s n_k}{n}$$

Как видно, выборочное среднее — это эмпирический начальный момент 1-го порядка, а выборочная дисперсия — это эмпирический центральный момент 2-го порядка.

Двумерная выборка

Если имеется выборочное распределение для системы двух случайных величин X и Y, т.е. таблица

	$Y = y_1$	$Y = y_2$	$\dots$	$Y = y_l$
$X = x_1$	n_{11}	n_{12}	$\dots$	n_{1l}
$X = x_2$	n_{21}	n_{22}	$\dots$	n_{2l}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$X = x_k$	n_{k1}	n_{k2}	$\dots$	n_{kl}

где n_{ij} — частота пары (x_i, y_j), то для случайной величины $Z = f(X, Y)$, где $f(\cdot, \cdot)$ — заданная функция двух переменных, выборочное среднее $\bar{z}$ определяется по формуле:

$$\bar{z} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{l} f(x_i, y_j) n_{ij}}{n}$$

где $n = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij}$.

В частности:

Эмпирическая ковариация:

$$\mathrm{Cov}(X, Y) = \frac{\sum_{i=1}^{k} \sum_{j=1}^{l} (x_i - \bar{x})(y_j - \bar{y}) n_{ij}}{n}$$

Эмпирический коэффициент корреляции:

$$r(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{D_x D_y}}$$

где D_x и D_y — выборочные дисперсии X и Y соответственно.

Лекция 2: Выборка. Статистики. Эмпирическая функция распределения

Пусть X – случайная величина с функцией распределения $F(x) = P\{X < x\}$, $x \in (-\infty, +\infty)$.

Определение. *Совокупность $\{X_k,\ k = 1, 2, \dots, n\}$ независимых случайных величин, имеющих одинаковую функцию распределения $F_{X_k}(x) = F(x)$, называется **однородной выборкой объёма** n, соответствующей функции распределения $F(x)$, и обозначается Z_n:*

$$Z_n = \{X_1, X_2, \dots, X_n\}.$$

Случайная величина X_k $(k = 1, 2, \dots, n)$ называется k-м элементом выборки.

Определение. ***Реализацией выборки** Z_n называется неслучайный вектор $\mathbf{z}_n = \{x_1, \dots, x_n\}$, компонентами которого являются реализации случайных величин $X_1, \dots, X_n$.*

Пусть $\mathbf{z}_{(n)} = \{x_{(1)}, \dots, x_{(n)}\}$ – вектор, компонентами которого являются упорядоченные по возрастанию числа $x_1, \dots, x_n$, т.е. $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$.

Определение. *Случайная величина $X_{(k)}$, реализацией которой для каждой реализации $\mathbf{z}_{(n)}$ является число $x_{(k)}$, называется k-й **порядковой статистикой**, $k = 1, \dots, n$. При этом случайные величины $X_{(1)}$ и $X_{(n)}$ называются **экстремальными статистиками**.*

Статистики

Если $\varphi(\cdot)$ – некоторая заданная функция от n переменных, то любая случайная величина Y, определяемая как $Y = \varphi(X_1, X_2, \dots, X_n)$, называется **статистикой**.

Примеры статистик:

$$\bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \text{– выборочное среднее,}$$

$$\bar{S}_n^2 = \frac{1}{n}\sum_{k=1}^{n} (X_k - \bar{x}_n)^2 \quad \text{– выборочная дисперсия,}$$

$$\bar{\nu}_r(n) = \frac{1}{n}\sum_{k=1}^{n} X_k^r, \quad r = 1, 2, \dots$$

– выборочный начальный момент r-го порядка,

$$\bar{\mu}_r(n) = \frac{1}{n}\sum_{k=1}^{n} (X_k - \bar{x}_n)^r, \quad r = 1, 2, \dots$$

– выборочный центральный момент r-го порядка.

Заметим, что $\bar{x}_n = \bar{\nu}_1(n)$, $\bar{S}_n^2 = \bar{\mu}_2(n)$.

Теорема (без доказательства)**.** *Пусть распределение $F(x)$ таково, что существуют следующие теоретические моменты любого элемента X_k выборки Z_n:*

$$m_X = \mathrm{M}\{X_k\}, \quad \nu_r = \mathrm{M}\{X_k^r\}, \quad \mu_r = \mathrm{M}\{(X_k - m_X)^r\}, \quad r = 1, 2, \dots.$$

Тогда

$$\bar{\nu}_r(n) \xrightarrow[n\to\infty]{P} \nu_r, \qquad \bar{\mu}_r(n) \xrightarrow[n\to\infty]{P} \mu_r.$$

Эмпирическая функция распределения

Введём функцию-индикатор:

$$I(A) = \begin{cases} 1, & \text{если событие } A \text{ произошло,} \\ 0, & \text{если } A \text{ не произошло.} \end{cases}$$

Для каждого $x \in (-\infty, +\infty)$ определим случайную величину $F_n^*(x)$ по формуле

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i < x).$$

Здесь $\sum_{i=1}^{n} I(X_i < x)$ есть случайное число элементов выборки, меньших x. Эту функцию будем называть **эмпирической функцией распределения выборки**. Каждая реализация выборки даёт свою реализацию случайной функции $F_n^*(x)$ (вид этих реализаций обсуждался в лекции 1).

Нетрудно усмотреть, что при любом фиксированном x случайная величина $\sum_{i=1}^{n} I(X_i < x)$ имеет биномиальное распределение с параметрами n и $p = P\{X_i < x\} = F(x)$, так что

$$P\left\{F_n^*(x) = \frac{k}{n}\right\} = C_n^k [F(x)]^k [1 - F(x)]^{n-k},$$

где $k = 0, 1, \ldots, n$.

Докажем следующую теорему, показывающую, что $F_n^*(x)$ с увеличением n сближается в каждой точке x с теоретической функцией распределения $F(x)$.

Теорема 1. *Для любого $x \in (-\infty; +\infty)$ и любого $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left\{|F_n^*(x) - F(x)| < \varepsilon\right\} = 1,$$

т.е. $F_n^*(x) \xrightarrow{P} F(x)$.

Доказательство. Рассмотрим систему независимых случайных величин

$$\eta_i = \begin{cases} 1, & \text{если } X_i < x, \\ 0, & \text{если } X_i \geq x, \end{cases} \quad i = 1, 2, \ldots, n.$$

Каждая из η_i имеет один и тот же закон распределения

$$\begin{array}{c|cc} \eta_i & 1 & 0 \\ \hline P & p & 1-p \end{array},$$

где $p = P\{X_i < x\} = F(x)$, $\mathrm{M}\{\eta_i\} = p$, $\mathrm{D}\{\eta_i\} = p(1-p)$. Так что выполнены все условия ЗБЧ в форме Чебышёва и

$$\lim_{n\to\infty} P\left\{\left|\frac{\eta_1 + \cdots + \eta_n}{n} - p\right| < \varepsilon\right\} = 1.$$

Но $p = F(x)$, $(\eta_1 + \cdots + \eta_n)/n = F_n^*(x)$. Теорема доказана. □

Замечание 1. *Отметим (без доказательства), что имеет место и более сильный результат:*

$$F_n^*(x) \xrightarrow[n\to\infty]{п.н.} F(x).$$

Эти свойства говорят о поточечной сходимости $F_n^*(x)$ к $F(x)$ при $n \to \infty$. Оказывается, справедливо и более сильное свойство, отражающее факт равномерной сходимости $F_n^*(x)$ к $F(x)$.

Теорема (Гливенко, без доказательства). *Введём последовательность случайных величин*

$$Y_n = \sup_{x\in(-\infty,+\infty)} |F_n^*(x) - F(x)|, \quad n = 1, 2, \dots .$$

Тогда

$$Y_n \xrightarrow[n\to\infty]{п.н.} 0.$$

Таким образом, теоремы 2, 3 говорят о том, что $F_n^*(x)$ при больших значениях n является хорошим приближением для $F(x)$.

Оценка точности приближения

В заключение, получим ещё один результат, позволяющий в вероятностном смысле оценить точность приближения для $F(x)$ с помощью $F_n^*(x)$. Применим ЦПТ к введённой при доказательстве теоремы 2 последовательности независимых случайных величин $\eta_1, \eta_2, \dots$.

По центральной предельной теореме если $V_n = \eta_1 + \dots + \eta_n$, $\widetilde{V_n} = \frac{V_n - \mathrm{M}\{V_n\}}{\sqrt{\mathrm{D}\{V_n\}}}$, то

$$\lim_{n\to\infty} P\{\alpha \le \widetilde{V_n} \le \beta\} = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\, dt.$$

В нашем случае

$$\begin{aligned} V_n &= \eta_1 + \dots + \eta_n = nF_n^*(x), \\ \mathrm{M}\{V_n\} &= np = nF(x), \\ \mathrm{D}\{V_n\} &= np(1-p) = nF(x)(1-F(x)), \end{aligned}$$

так что

$$\lim_{n\to\infty} P\left\{\alpha \le \frac{nF_n^*(x) - nF(x)}{\sqrt{nF(x)(1-F(x))}} \le \beta\right\} = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\, dt.$$

В частности,

$$\lim_{n\to\infty} P\left\{\sqrt{n}\frac{|F_n^*(x) - F(x)|}{\sqrt{F(x)(1-F(x))}} \le \varepsilon\right\} = 2\Phi_0(\varepsilon),$$

где $\Phi_0(\varepsilon) = \frac{1}{\sqrt{2\pi}} \int_0^\varepsilon e^{-t^2/2}\, dt$ – функция Лапласа.

Курс теории вероятностей, фрагменты

Теорема (Чебышёва, ЗБЧ)**.** *Для последовательности независимых случайных величин $X_1, X_2, \ldots$ с одинаковым математическим ожиданием μ и ограниченными дисперсиями ($\mathrm{D}\{X_i\} \leq C$) выполняется*

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right\} = 1.$$

Теорема (ЦПТ для независимых одинаково распределенных случайных величин)**.** *Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные случайные величины с $\mathrm{M}\{X_i\} = \mu$ и $\mathrm{D}\{X_i\} = \sigma^2 < \infty$. Тогда:*

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

($N(0,1)$ — стандартное нормальное распределение, соответствующая функция распределения $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt$ (не путать с $\Phi_0(x) = \frac{1}{\sqrt{2\pi}}\int_{0}^{x} e^{-t^2/2}dt$, эти функции связаны соотношением $\Phi(x) = 1/2 + \Phi_0(x)$).

Лекция 3: Точечные оценки и их свойства (несмещенность, эффективность, состоятельность)

Пусть θ — неизвестный параметр изучаемого распределения случайной величины X, а $(X_1, X_2, \ldots, X_n)$ – выборка объёма n, порождённая этим распределением. Пусть

$$\hat{\theta}_n = h(X_1, \ldots, X_n)$$

— оценка параметра θ, определяемая некоторой заданной функцией $h(x_1, \ldots, x_n)$ от n переменных по выборке $(X_1, \ldots, X_n)$, так что $\hat{\theta}$ является случайной величиной.

Пример 1. *Пусть $h(x_1, \ldots, x_n) = (x_1+\cdots+x_n)/n$. Тогда оценка $\hat{\theta}_n = (X_1+\cdots+X_n)/n$ может рассматриваться как оценка неизвестного математического ожидания $\mathrm{M}\{X\} = \theta$ изучаемого распределения случайной величины X.*

Определение. *Оценка $\hat{\theta}_n$ называется несмещённой, если её математическое ожидание $\mathrm{M}\{\hat{\theta}_n\}$ совпадает с θ: $\mathrm{M}\{\hat{\theta}_n\} = \theta$.*

Пример 2. *Оценка $\hat{\theta}_n = (X_1 + \cdots + X_n)/n$ является несмещённой оценкой для математического ожидания $\mathrm{M}\{X\} = \theta$, так как*

$$\mathrm{M}\{\hat{\theta}_n\} = \frac{1}{n}\mathrm{M}\{X_1 + \cdots + X_n\} = \frac{\mathrm{M}\{X_1\} + \cdots + \mathrm{M}\{X_n\}}{n} = \frac{\mathrm{M}\{X\} + \cdots + \mathrm{M}\{X\}}{n} = \frac{n\mathrm{M}\{X\}}{n} = \mathrm{M}\{X\}$$

Определение. *Оценка $\hat{\theta}_n = h(X_1, \ldots, X_n)$ называется асимптотически несмещённой, если*

$$\lim_{n\to\infty} \mathrm{M}\{\hat{\theta}_n\} = \theta.$$

Пример 3. *Пусть выборка $(X_1, \ldots, X_n)$ порождена случайной величиной $X \sim R[0, \theta]$, где $\theta > 0$ — неизвестный параметр. Рассмотрим оценку $\hat{\theta}_n = X_{(n)}$, где $X_{(n)}$ — соответствующая порядковая статистика. По условию*

$$F(x) = P\{X < x\} = \begin{cases} 0, & x \le 0, \\ x/\theta, & x \in (0; \theta], \\ 1, & x > \theta. \end{cases}$$

Поэтому

$$F_{(n)}(x) = P\{X_{(n)} < x\} = F^n(x) = \begin{cases} 0, & x \le 0, \\ (x/\theta)^n, & x \in (0; \theta], \\ 1, & x > \theta, \end{cases}$$

и

$$\mathrm{M}\{\hat{\theta}_n\} = \mathrm{M}\{X_{(n)}\} = \int_0^\theta x \cdot n \left(\frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta}\, dx = \frac{n}{\theta^n} \int_0^\theta x^n\, dx = \frac{n}{\theta^n} \cdot \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1} \cdot \theta.$$

Как видно,

$$\lim_{n\to\infty} \mathrm{M}\{\hat{\theta}_n\} = \lim_{n\to\infty} \frac{n}{n+1}\theta = \theta,$$

так что предложенная оценка для θ является асимптотически несмещённой.

Определение. *Несмещённая оценка $\hat\theta_n$ параметра θ называется эффективной, если при заданном объёме n выборки она имеет наименьшую дисперсию.*

Определение. *Оценка $\hat\theta_n$ параметра θ называется состоятельной, если*

$$\forall\varepsilon > 0 \quad \lim_{n\to\infty} P\{|\hat\theta_n - \theta| < \varepsilon\} = 1,$$

т.е. $\hat\theta_n \xrightarrow{P} \theta$ *при* $n \to \infty$.

Теорема 2 (Достаточные условия состоятельности)**.** *Пусть для оценки $\hat\theta_n$ параметра θ выполнены условия:*

1. $\lim_{n\to\infty} \mathrm{M}\{\hat\theta_n\} = \theta$ *(т.е. оценка $\hat\theta_n$ является асимптотически несмещённой);*
2. $\lim_{n\to\infty} \mathrm{D}\{\hat\theta_n\} = 0$*;*

тогда $\hat\theta_n$ — состоятельная оценка для θ.

Доказательство. Зафиксируем некоторое $\varepsilon > 0$. Тогда существует N такое, что для всех $n \geq N$

$$|\mathrm{M}\{\hat\theta_n\} - \theta| < \frac{\varepsilon}{2}. \tag{*}$$

Заметим, что

$$|\hat\theta_n - \theta| \leq |\hat\theta_n - \mathrm{M}\{\hat\theta_n\}| + |\mathrm{M}\{\hat\theta_n\} - \theta|,$$

откуда

$$|\hat\theta_n - \mathrm{M}\{\hat\theta_n\}| \geq |\hat\theta_n - \theta| - |\mathrm{M}\{\hat\theta_n\} - \theta|. \tag{**}$$

Определим события

$$A_n = \{|\hat\theta_n - \theta| \geq \varepsilon\}, \quad B_n = \{|\hat\theta_n - \mathrm{M}\{\hat\theta_n\}| \geq \varepsilon/2\}.$$

Если $n \geq N$ и произошло событие A_n, то из (*) и (**) следует, что

$$|\hat\theta_n - \mathrm{M}\{\hat\theta_n\}| \geq \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2},$$

т.е. произошло событие B_n. Таким образом, при $n \geq N$ событие A_n влечёт B_n, так что

$$P\{A_n\} \leq P\{B_n\}.$$

По неравенству Чебышёва

$$P\{B_n\} = P\{|\hat\theta_n - \mathrm{M}\{\hat\theta_n\}| \geq \varepsilon/2\} \leq \frac{\mathrm{D}\{\hat\theta_n\}}{(\varepsilon/2)^2},$$

поэтому $P\{A_n\} \leq \dfrac{\mathrm{D}\{\hat\theta_n\}}{(\varepsilon/2)^2} \to 0$ при $n \to \infty$. Следовательно,

$$P\{|\hat\theta_n - \theta| < \varepsilon\} = 1 - P\{|\hat\theta_n - \theta| \geq \varepsilon\} = 1 - P\{A_n\} \to 1,$$

что и требовалось доказать. □

Исследуем на несмещённость и состоятельность выборочные моменты.

Теорема 3. *Выборочный начальный момент k-го порядка*

$$\bar{\nu}_k(n) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

является несмещённой оценкой для соответствующего теоретического момента $\nu_k = \mathrm{M}\{X^k\}$ (если последний существует).

Доказательство.

$$\mathrm{M}\{\bar{\nu}_k(n)\} = \mathrm{M}\{\frac{1}{n}\sum_{i=1}^{n} X_i^k\} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\{X_i^k\} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\{X^k\} = \frac{1}{n}\cdot n\nu_k = \nu_k,$$

что и требовалось доказать. □

Теорема 4. *Если существует теоретический начальный момент ν_{2m}, то выборочный начальный момент $\bar{\nu}_k(n)$ является состоятельной оценкой для ν_k при любом $k \le m$.*

Доказательство. Воспользуемся теоремой 1. Условие 1) выполнено. Проверим выполнение условия 2):

$$\mathrm{D}\{\bar{\nu}_k(n)\} = \mathrm{D}\{\frac{1}{n}\sum_{i=1}^{n} X_i^k\} = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}\{X_i^k\} = \frac{1}{n^2}\cdot n\mathrm{D}\{X^k\} = \frac{1}{n}\mathrm{D}\{X^k\} = \frac{1}{n}\left\{\mathrm{M}\{X^{2k}\} - \left(\mathrm{M}\{X^k\}\right)^2\right\} =$$

при $n \to \infty$, так как $\mathrm{D}\{X^k\} = \nu_{2k} - \nu_k^2$ конечно в силу существования ν_{2m} при $k \le m$. Таким образом, оба условия 1) и 2) теоремы 1 выполнены, так что $\bar{\nu}_k(n)$ есть состоятельная оценка для ν_k. □

Теорема (Слуцкого, вспомогательная техническая теорема)**.** *Пусть имеется k последовательностей случайных величин:*

$$Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_n^{(1)}, \ldots$$

$$Y_1^{(2)}, Y_2^{(2)}, \ldots, Y_n^{(2)}, \ldots$$

$$\vdots$$

$$Y_1^{(k)}, Y_2^{(k)}, \ldots, Y_n^{(k)}, \ldots$$

Пусть при $n \to \infty$

$$Y_n^{(1)} \xrightarrow{P} a_1, \quad \ldots, \quad Y_n^{(k)} \xrightarrow{P} a_k.$$

Пусть функция $f(y_1, y_2, \ldots, y_k)$ от k переменных непрерывна в точке $(a_1, a_2, \ldots, a_k)$. Определим последовательность случайных величин

$$Z_n = f\big(Y_n^{(1)}, Y_n^{(2)}, \ldots, Y_n^{(k)}\big), \quad n = 1, 2, \ldots$$

Тогда

$$Z_n \xrightarrow{P} f(a_1, a_2, \ldots, a_k) \quad \text{при } n \to \infty.$$

Теорема 5. *Если существует теоретический начальный момент ν_{2m}, то выборочный центральный момент*

$$\bar{\mu}_k(n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{\nu}_1(n))^k$$

является состоятельной оценкой при $k \leq m$.

Доказательство. Центральный момент любого порядка μ_k можно представить в виде $\mu_k = f_k(\nu_1, \nu_2, \ldots, \nu_k)$, где f_k — многочлен степени k:

$$\begin{aligned} \mu_2 &= \nu_2 - \nu_1^2, \\ \mu_3 &= \mathrm{M}\{(X-\nu_1)^3\} = \mathrm{M}\{X^3 - 3X^2\nu_1 + 3X\nu_1^2 - \nu_1^3\} \\ &= \mathrm{M}\{X^3\} - 3\nu_1\mathrm{M}\{X^2\} + 3\nu_1^2\mathrm{M}\{X\} - \nu_1^3 \\ &= \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3, \quad \text{и т.д.} \end{aligned}$$

При этом выборочные значения удовлетворяют тем же соотношениям:

$$\bar{\mu}_k(n) = f_k(\bar{\nu}_1(n), \bar{\nu}_2(n), \ldots, \bar{\nu}_k(n)).$$

Поскольку многочлен — непрерывная функция, и по теореме 3

$$\bar{\nu}_j(n) \xrightarrow{P} \nu_j \quad \text{при } n \to \infty \text{ для } j = 1, \ldots, k,$$

то по теореме Слуцкого

$$\bar{\mu}_k(n) = f_k(\bar{\nu}_1(n), \bar{\nu}_2(n), \ldots, \bar{\nu}_k(n)) \xrightarrow{P} f_k(\nu_1, \nu_2, \ldots, \nu_k) = \mu_k,$$

что и требовалось доказать. □

Лекция 4: Точечные оценки для математического ожидания и дисперсии

Оценка математического ожидания

Пусть $(X_1, \dots, X_n)$ — выборка, порождённая случайной величиной X с неизвестным распределением. В качестве оценки неизвестного математического ожидания $\mathrm{M}\{X\}$ рассмотрим статистику

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n},$$

называемую **выборочным средним**.

Утверждение 1. *$\overline{X}_n$ есть несмещённая оценка для $\mathrm{M}\{X\}$.*

Доказательство.

$$\mathrm{M}\{\overline{X}_n\} = \mathrm{M}\{\frac{X_1 + \dots + X_n}{n}\} = \frac{1}{n}\big(\mathrm{M}\{X_1\} + \dots + \mathrm{M}\{X_n\}\big) = \frac{1}{n} \cdot n\mathrm{M}\{X\} = \mathrm{M}\{X\}.$$

□

Утверждение 2. *$\overline{X}_n$ есть состоятельная оценка для $\mathrm{M}\{X\}$.*

Доказательство. Из утверждения 1 следует $\lim\limits_{n\to\infty} \mathrm{M}\{\overline{X}_n\} = \mathrm{M}\{X\}$. Далее, дисперсия

$$\mathrm{D}\{\overline{X}_n\} = \mathrm{D}\{\frac{X_1 + \dots + X_n}{n}\} = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{D}\{X_i\} = \frac{1}{n^2} \cdot n\mathrm{D}\{X\} = \frac{\mathrm{D}\{X\}}{n} \to 0 \quad (n \to \infty).$$

Таким образом, выполнены достаточные условия состоятельности (см. лекцию 3), следовательно, $\overline{X}_n$ — состоятельная оценка $\mathrm{M}\{X\}$. □

Утверждение 3. *$\overline{X}_n$ является эффективной оценкой для $\mathrm{M}\{X\}$ в классе всех линейных несмещённых оценок.*

Доказательство. Рассмотрим произвольную линейную оценку $\hat{X}_n = c_1X_1 + \dots + c_nX_n$, где $c_1, \dots, c_n$ — некоторые константы. Её математическое ожидание: $\mathrm{M}\{\hat{X}_n\} = (c_1 + \dots + c_n)\mathrm{M}\{X\}$. Для несмещённости необходимо $c_1 + \dots + c_n = 1$. Дисперсия (в предположении существования $\mathrm{D}\{X\}$): $\mathrm{D}\{\hat{X}_n\} = (c_1^2 + \dots + c_n^2)\mathrm{D}\{X\}$. Требуется минимизировать $c_1^2 + \dots + c_n^2$ при условии $c_1 + \dots + c_n = 1$, $c_i \in (-\infty, +\infty)$. Задача на условный экстремум решается методом множителей Лагранжа; минимум достигается при $c_1 = c_2 = \dots = c_n = \frac{1}{n}$. Следовательно, минимальное значение дисперсии равно $\mathrm{D}\{X\}/n$, что совпадает с дисперсией $\overline{X}_n$. Таким образом, $\overline{X}_n$ — эффективная оценка в указанном классе. □

Оценки дисперсии

Случай известного математического ожидания

Предположим, что математическое ожидание $\mathrm{M}\{X\} = m$ известно. В качестве оценки для $\mathrm{D}\{X\}$ рассмотрим статистику

$$S_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2.$$

Утверждение 4. *S_n является несмещённой оценкой* $\mathrm{D}\{X\}$.

Доказательство.

$$\mathrm{M}\{S_n\} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\{(X_i - m)^2\} = \frac{1}{n}\cdot n\mathrm{D}\{X\} = \mathrm{D}\{X\}.$$

□

Утверждение 5. *S_n есть состоятельная оценка* $\mathrm{D}\{X\}$.

Доказательство. Достаточно показать, что $\mathrm{D}\{S_n\} \to 0$ при $n \to \infty$ (см. лекцию 3). Имеем:

$$\mathrm{D}\{S_n\} = \mathrm{D}\{\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2\} = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}\{(X_i - m)^2\} = \frac{1}{n}\mathrm{D}\{(X - m)^2\}.$$

Если $\mathrm{D}\{(X - m)^2\}$ существует (т.е. конечен четвёртый центральный момент), то $\mathrm{D}\{S_n\} \to 0$. □

Случай неизвестного математического ожидания

Пусть теперь $\mathrm{M}\{X\}$ неизвестно. В качестве оценки дисперсии $\mathrm{D}\{X\}$ рассмотрим статистику

$$S_{\text{new}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

где $\overline{X}_n$ — выборочное среднее.

Вычислим математическое ожидание S_{new}. Заметим, что все слагаемые $\mathrm{M}\{(X_i - \overline{X}_n)^2\}$ одинаковы для $i = 1, \dots, n$. Найдём $\mathrm{M}\{(X_1 - \overline{X}_n)^2\}$:

$$\begin{aligned}\mathrm{M}\{(X_1 - \overline{X}_n)^2\} &= \mathrm{M}\{X_1^2 - 2X_1\overline{X}_n + \overline{X}_n^2\} \\ &= \mathrm{M}\{X_1^2\} - 2\mathrm{M}\{X_1\overline{X}_n\} + \mathrm{M}\{\overline{X}_n^2\}.\end{aligned}$$

Имеем:

$$\mathrm{M}\{X_1^2\} = \mathrm{D}\{X\} + (\mathrm{M}\{X\})^2.$$

Далее,

$$\mathrm{M}\{X_1\overline{X}_n\} = \mathrm{M}\{X_1\frac{X_1 + \dots + X_n}{n}\} = \frac{1}{n}\big(\mathrm{M}\{X_1^2\} + (n-1)(\mathrm{M}\{X\})^2\big).$$

И

$$\mathrm{M}\{\overline{X}_n^2\} = \mathrm{D}\{\overline{X}_n\} + (\mathrm{M}\{\overline{X}_n\})^2 = \frac{\mathrm{D}\{X\}}{n} + (\mathrm{M}\{X\})^2.$$

Подставляя:

$$\begin{aligned}\mathrm{M}\{(X_1 - \overline{X}_n)^2\} &= \mathrm{D}\{X\} + (\mathrm{M}\{X\})^2 - \frac{2}{n}\Big(\mathrm{D}\{X\} + (\mathrm{M}\{X\})^2 + (n-1)(\mathrm{M}\{X\})^2\Big) + \frac{\mathrm{D}\{X\}}{n} + (\mathrm{M}\{X\}) \\ &= \mathrm{D}\{X\} + (\mathrm{M}\{X\})^2 - \frac{2}{n}\big(\mathrm{D}\{X\} + n(\mathrm{M}\{X\})^2\big) + \frac{\mathrm{D}\{X\}}{n} + (\mathrm{M}\{X\})^2 \\ &= \mathrm{D}\{X\} + (\mathrm{M}\{X\})^2 - \frac{2}{n}\mathrm{D}\{X\} - 2(\mathrm{M}\{X\})^2 + \frac{\mathrm{D}\{X\}}{n} + (\mathrm{M}\{X\})^2 \\ &= \mathrm{D}\{X\} - \frac{\mathrm{D}\{X\}}{n} = \frac{n-1}{n}\mathrm{D}\{X\}.\end{aligned}$$

Следовательно,
$$\mathrm{M}\{S_{\text{new}}\} = \frac{1}{n}\cdot n\cdot\frac{n-1}{n}\mathrm{D}\{X\} = \frac{n-1}{n}\mathrm{D}\{X\}.$$
Таким образом, S_{new} — **смещённая** оценка, однако она является *асимптотически несмещённой*, поскольку
$$\lim_{n\to\infty}\mathrm{M}\{S_{\text{new}}\} = \lim_{n\to\infty}\frac{n-1}{n}\mathrm{D}\{X\} = \mathrm{D}\{X\}.$$

Если ввести ещё одну оценку
$$S_{\text{испр}} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{n}{n-1}S_{\text{new}},$$
то эта оценка будет **несмещённой**:
$$\mathrm{M}\{S_{\text{испр}}\} = \frac{n}{n-1}\mathrm{M}\{S_{\text{new}}\} = \mathrm{D}\{X\}.$$

Оценку S_{new} называют **выборочной дисперсией**, а оценку $S_{\text{испр}}$ — **исправленной выборочной дисперсией**.

Утверждение 6. *Оценка $S_{\text{испр}}$ является состоятельной оценкой для* $\mathrm{D}\{X\}$*.*

Доказательство. Можно проверить, что
$$\mathrm{D}\{S_{\text{испр}}\} = \frac{1}{n}\Big(\mathrm{M}\{(X-\mathrm{M}\{X\})^4\} - \frac{n-3}{n-1}\big(\mathrm{D}\{X\}\big)^2\Big),$$
откуда $\mathrm{D}\{S_{\text{испр}}\}\to 0$ при $n\to\infty$ (при условии существования четвёртого центрального момента). Следовательно, достаточные условия состоятельности выполнены. $\square$

Заметим, что S_{new} также является состоятельной оценкой для $\mathrm{D}\{X\}$, поскольку она отличается от $S_{\text{испр}}$ лишь множителем $\frac{n-1}{n}\to 1$.

Лекция 5: Метод моментов

Пусть вид (закон) распределения случайной величины X известен с точностью до неизвестных параметров $\theta_1, \ldots, \theta_k$, $k \in \mathbb{N}$. Тогда все начальные моменты $\nu_1, \ldots, \nu_k$ этой случайной величины X являются известными функциями этих параметров:

$$\begin{aligned} \nu_1 &= \mathrm{M}\{X\} = g_1(\theta_1, \ldots, \theta_k), \\ &\vdots \\ \nu_k &= \mathrm{M}\{X^k\} = g_k(\theta_1, \ldots, \theta_k). \end{aligned} \tag{1}$$

Если X — непрерывная случайная величина с плотностью распределения $f(x, \theta_1, \ldots, \theta_k)$, то

$$g_1(\theta_1, \ldots, \theta_k) = \int\limits_{-\infty}^{+\infty} x\, f(x, \theta_1, \ldots, \theta_k)\, dx,$$

$$g_k(\theta_1, \ldots, \theta_k) = \int\limits_{-\infty}^{+\infty} x^k f(x, \theta_1, \ldots, \theta_k)\, dx.$$

Если X — дискретная случайная величина, определяемая законом распределения

$$P\{X = x_m\} = p(x_m, \theta_1, \ldots, \theta_k), \quad m = 1, 2, \ldots,$$

то

$$g_1(\theta_1, \ldots, \theta_k) = \sum_m x_m p(x_m, \theta_1, \ldots, \theta_k),$$

$$g_k(\theta_1, \ldots, \theta_k) = \sum_m x_m^k p(x_m, \theta_1, \ldots, \theta_k),$$

где $\sum$ — либо конечная сумма, если X — дискретная случайная величина с конечным числом возможных значений, либо сумма ряда, если X — дискретная случайная величина с бесконечным числом возможных значений.

Предполагаем, что моменты $\nu_1, \ldots, \nu_k$ существуют.

Предположим, что систему (1) можно разрешить относительно параметров $\theta_1, \ldots, \theta_k$:

$$\theta_1 = h_1(\nu_1, \ldots, \nu_k), \quad \ldots, \quad \theta_k = h_k(\nu_1, \ldots, \nu_k),$$

где $h_1, \ldots, h_k$ — непрерывные функции.

Определение. *Оценки параметров $\theta_1, \ldots, \theta_k$, заданные формулами*

$$\hat{\theta}_1 = h_1(\bar{\nu}_1, \ldots, \bar{\nu}_k), \quad \ldots, \quad \hat{\theta}_k = h_k(\bar{\nu}_1, \ldots, \bar{\nu}_k),$$

называются оценками метода моментов. *Здесь $\bar{\nu}_1, \ldots, \bar{\nu}_k$ — соответствующие выборочные начальные моменты, полученные по выборке $\{X_1, \ldots, X_n\}$, порождённой случайной величиной X:*

$$\bar{\nu}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j, \qquad j = 1, \ldots, k.$$

Теорема (о состоятельности оценок метода моментов)**.** *Если распределение зависит от параметров $\theta_1, \ldots, \theta_k$, и при любом допустимом наборе их значений распределение имеет начальные моменты до порядка $2k$ включительно, то оценки $\hat{\theta}_1, \ldots, \hat{\theta}_k$ параметров $\theta_1, \ldots, \theta_k$, полученные по методу моментов, являются состоятельными.*

Рассмотрим несколько примеров на получение оценок методом моментов.

Пример 1

Пусть выборка $\{X_1, \ldots, X_n\}$ порождена случайной величиной X, имеющей равномерное распределение на отрезке $[\theta_1, \theta_2]$. Найти методом моментов оценки параметров θ_1, θ_2.

Имеем

$$\nu_1 = \mathrm{M}\{X\} = \int\limits_{\theta_1}^{\theta_2} \frac{x}{\theta_2 - \theta_1}\, dx = \frac{\theta_1 + \theta_2}{2},$$

$$\nu_2 = \mathrm{M}\{X^2\} = \int\limits_{\theta_1}^{\theta_2} \frac{x^2}{\theta_2 - \theta_1}\, dx = \frac{\theta_2^3 - \theta_1^3}{3(\theta_2 - \theta_1)} = \frac{\theta_2^2 + \theta_1\theta_2 + \theta_1^2}{3}.$$

Или, что равноценно,

$$\mathrm{M}\{X\} = \nu_1 = \frac{\theta_1 + \theta_2}{2}, \qquad \mathrm{D}\{X\} = \nu_2 - \nu_1^2 = \frac{(\theta_2 - \theta_1)^2}{12}.$$

Разрешаем эту систему относительно θ_1 и θ_2:

$$\theta_1 + \theta_2 = 2\nu_1, \qquad \theta_2 - \theta_1 = 2\sqrt{3(\nu_2 - \nu_1^2)}.$$

Заменяя в правых частях последней системы ν_1 на $\bar{\nu}_1$, ν_2 на $\bar{\nu}_2$, получим оценки

$$\hat{\theta}_1 = \bar{\nu}_1 - \sqrt{3(\bar{\nu}_2 - \bar{\nu}_1^2)}, \qquad \hat{\theta}_2 = \bar{\nu}_1 + \sqrt{3(\bar{\nu}_2 - \bar{\nu}_1^2)},$$

где $\bar{\nu}_1 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ — выборочное среднее, $\bar{\nu}_2 = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$, а $\sqrt{\bar{\nu}_2 - \bar{\nu}_1^2} = \sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2}$ — корень из выборочной дисперсии.

Пример 2

Пусть выборка $\{X_1, \ldots, X_n\}$ порождена случайной величиной X, имеющей нормальное распределение с плотностью

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in (-\infty, +\infty),$$

a, σ — неизвестные параметры. Найти оценки a и σ методом моментов.

Имеем

$$\nu_1 = \mathrm{M}\{X\} = a, \qquad \nu_2 = \mathrm{M}\{X^2\} = a^2 + \sigma^2,$$

или

$$a = \nu_1, \qquad \sigma^2 = \nu_2 - \nu_1^2.$$

Таким образом, оценки метода моментов имеют вид

$$\hat{a} = \bar{\nu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \hat{\sigma} = \sqrt{\bar{\nu}_2 - \bar{\nu}_1^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

— корень из выборочной дисперсии.

Пример 3

Пусть известно, что случайная величина X имеет распределение, задаваемое формулой

$$P\{X=k\}=\theta(1-\theta)^k, \quad k=0,1,2,\ldots,$$

т.е. X — дискретная случайная величина, которая может принимать только целые неотрицательные значения. Параметр $\theta\in(0,1)$ является неизвестным. В результате трёх измерений (экспериментов) случайной величины X она приняла значения 0, 1 и 2. Предложить численную оценку для θ.

Решение. Воспользуемся методом моментов. Найдём общую оценку по выборке $\{X_1,\ldots,X_n\}$. Имеем

$$\nu_1=\mathrm{M}\{X\}=\sum_{k=0}^{\infty}k\cdot P\{X=k\}=\sum_{k=0}^{\infty}k\,\theta(1-\theta)^k.$$

Обозначим для краткости $q=1-\theta$. Тогда

$$\nu_1=\theta\sum_{k=0}^{\infty}kq^k.$$

Вычислим сумму $S=\sum\limits_{k=0}^{\infty}kq^k$ (при $|q|<1$). Заметим, что при $k=0$ слагаемое равно нулю, поэтому $S=\sum_{k=1}^{\infty}kq^k$. Воспользуемся тем, что степенные ряды можно почленно дифференцировать в круге сходимости. Рассмотрим геометрическую прогрессию $\sum_{k=0}^{\infty}q^k=\frac{1}{1-q}$. Дифференцируя обе части по q, получаем

$$\sum_{k=1}^{\infty}kq^{k-1}=\frac{1}{(1-q)^2}.$$

Умножая обе части на q, находим

$$\sum_{k=1}^{\infty}kq^k=\frac{q}{(1-q)^2}.$$

Следовательно,

$$\nu_1=\theta\cdot\frac{q}{(1-q)^2}=\theta\cdot\frac{1-\theta}{\theta^2}=\frac{1-\theta}{\theta}.$$

Откуда

$$\theta=\frac{1}{1+\nu_1}.$$

Таким образом, оценка метода моментов:

$$\hat{\theta}=\frac{1}{1+\bar{\nu}_1},$$

где $\bar{\nu}_1=\frac{1}{n}\sum_{i=1}^{n}X_i$.

В нашем случае $n=3$ и реализацией выборочного момента $\bar{\nu}_1$ является число $\frac{0+1+2}{3}=1$. Следовательно, искомая численная оценка θ есть

$$\hat{\theta}=\frac{1}{1+1}=\frac{1}{2}.$$

Лекция 6: Метод максимального правдоподобия

Пусть закон распределения случайной величины X известен с точностью до неизвестных параметров $\theta_1, \ldots, \theta_k$. Если случайная величина X, порождающая выборку $(X_1, \ldots, X_n)$, является дискретной, то пусть

$$p(x; \theta_1, \ldots, \theta_k) = P\{X = x\}, \quad x \in \mathcal{X}, \tag{1}$$

где $\mathcal{X}$ — множество всех возможных значений случайной величины X. Если X является непрерывной и имеет плотность распределения $f(x)$, то пусть

$$p(x; \theta_1, \ldots, \theta_k) = f(x), \quad x \in \mathcal{X}. \tag{2}$$

(Правые части равенств (1) и (2), разумеется, также зависят от $\theta_1, \ldots, \theta_k$.)

Определение. Функцией правдоподобия *выборки* $(X_1, \ldots, X_n)$ *называется функция*

$$L_n(\theta_1, \ldots, \theta_k; X_1, \ldots, X_n) = \prod_{i=1}^{n} p(X_i; \theta_1, \ldots, \theta_k), \tag{3}$$

которая при фиксированных $X_1 = x_1, \ldots, X_n = x_n$ *и фиксированных* $\theta_1, \ldots, \theta_k$ *в непрерывном случае равна значению совместной плотности распределения системы независимых случайных величин* $X_1, \ldots, X_n$, *а в дискретном случае — вероятности события* $\{X_1 = x_1, \ldots, X_n = x_n\}$.

Определение. *Пусть* $(X_1, \ldots, X_n)$ *— выборка из распределения, зависящего от параметров* $\theta_1, \ldots, \theta_k$. *Пусть* $h(x_1, \ldots, x_n)$ *— точка глобального максимума функции правдоподобия (3) (на множестве возможных значений* $\theta_1, \ldots, \theta_k$*) при фиксированных* $X_1 = x_1, \ldots, X_n = x_n$*:*

$$h(x_1, \ldots, x_n) = \begin{pmatrix} h_1(x_1, \ldots, x_n) \\ \vdots \\ h_k(x_1, \ldots, x_n) \end{pmatrix}$$

— k*-мерная точка в пространстве параметров* $\theta_1, \ldots, \theta_k$. *Тогда оценки*

$$\hat{\theta}_1 = h_1(X_1, \ldots, X_n), \quad \ldots, \quad \hat{\theta}_k = h_k(X_1, \ldots, X_n)$$

называются **оценками максимального правдоподобия** *параметров* $\theta_1, \ldots, \theta_k$.

Таким образом, технически нахождение оценки максимального правдоподобия сводится к нахождению точки глобального максимума функции k переменных $L_n(\theta_1, \ldots, \theta_k; X_1, \ldots,$ по всем возможным значениям параметров

$$L_n(\hat{\theta}_1, \ldots, \hat{\theta}_k; X_1, \ldots, X_n) = \max_{\text{по всем возможным значениям } \theta_1, \ldots, \theta_k} L_n(\theta_1, \ldots, \theta_k; X_1, \ldots, X_n)$$

Пример 1. Распределение Бернулли

Методом максимального правдоподобия найти оценку $\hat{\theta}$ параметра θ распределения Бернулли

x	0	1
$P\{X = x\}$	$1 - \theta$	θ

Отрезок $[0,1]$ — множество допустимых значений параметра.

Имеем:

$$p(x;\theta) = P\{X = x\} = \begin{cases} \theta, & x = 1, \\ 1-\theta, & x = 0. \end{cases}$$

Функция правдоподобия

$$L_n(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i;\theta) = \theta^k (1-\theta)^{n-k},$$

где $k = \sum_{i=1}^n x_i$ — число единиц среди $x_1, \ldots, x_n$.

Найдём максимум L_n при фиксированных x_i. Вычислим производную по θ:

$$\begin{aligned} \frac{dL_n}{d\theta} &= k\theta^{k-1}(1-\theta)^{n-k} - (n-k)\theta^k(1-\theta)^{n-k-1} \\ &= \theta^{k-1}(1-\theta)^{n-k-1}\big(k(1-\theta) - (n-k)\theta\big) \\ &= \theta^{k-1}(1-\theta)^{n-k-1}\big(k - n\theta\big). \end{aligned}$$

При $\theta \in (0,1)$ выражение в скобках равно нулю при

$$k - n\theta = 0 \quad \Longrightarrow \quad \theta = \frac{k}{n}.$$

Поскольку $L_n = 0$ при $\theta = 0$ или $\theta = 1$, а на интервале $(0,1)$ функция положительна и имеет единственную критическую точку, то в этой точке достигается максимум. Учитывая, что $k = \sum_{i=1}^n x_i$, окончательно

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n$$

— выборочное среднее.

Пример 2. Показательное распределение

Методом максимального правдоподобия найти оценку параметра λ показательного распределения с плотностью

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Промежуток $(0,+\infty)$ — множество возможных значений λ.

Выписываем функцию правдоподобия:

$$L_n(\lambda; x_1, \ldots, x_n) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}.$$

Поскольку $L_n > 0$ на области определения, вместо максимума L_n можно искать максимум её логарифма:

$$\ell(\lambda) = \ln L_n(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i.$$

Производная:

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

Приравнивая её к нулю, находим

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

Это точка максимума (проверяется знаком второй производной). Таким образом, искомая оценка

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}_n}.$$

Пример 3. Равномерное распределение на отрезке $[\theta_1, \theta_2]$

Выборка $(X_1, \ldots, X_n)$ порождена случайной величиной X, равномерно распределённой на отрезке $[\theta_1, \theta_2]$, т.е. плотность распределения

$$f(x; \theta_1, \theta_2) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1}, & x \in [\theta_1, \theta_2], \\ 0, & x \notin [\theta_1, \theta_2]. \end{cases}$$

Методом максимального правдоподобия найти оценки неизвестных параметров θ_1 и θ_2, $\theta_1 < \theta_2$.

Выпишем функцию правдоподобия (см. (3)):

$$L_n(\theta_1, \theta_2; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2) = \begin{cases} \dfrac{1}{(\theta_2 - \theta_1)^n}, & \text{если все } x_i \in [\theta_1, \theta_2], \\ 0, & \text{если хотя бы одно } x_i \notin [\theta_1, \theta_2]. \end{cases}$$

Обозначим через $X_{(1)} = \min\{X_1, \ldots, X_n\}$ и $X_{(n)} = \max\{X_1, \ldots, X_n\}$ соответствующие порядковые статистики, а через $x_{(1)} = \min\{x_1, \ldots, x_n\}$ и $x_{(n)} = \max\{x_1, \ldots, x_n\}$ соответствующие реализации. Тогда условие $x_i \in [\theta_1, \theta_2]$ для всех i равносильно $\theta_1 \leq x_{(1)}$ и $\theta_2 \geq x_{(n)}$. Следовательно,

$$L_n(\theta_1, \theta_2; x_1, \ldots, x_n) \leq \left(\frac{1}{x_{(n)} - x_{(1)}} \right)^n,$$

обозначим правую часть $L_{\max}$.

Теперь заметим, что

$$L_n(\theta_1, \theta_2; x_1, \ldots, x_n)\Big|_{\theta_2 = x_{(n)}}^{\theta_1 = x_{(1)}} = L_{\max}.$$

Другими словами

$$L_n(\theta_1, \theta_2) = \frac{1}{(\theta_2 - \theta_1)^n} \quad \text{при} \quad \theta_1 \leq x_{(1)},\ \theta_2 \geq x_{(n)},$$

и $L_n = 0$ вне этой области.

Чтобы максимизировать L_n, нужно минимизировать знаменатель $(\theta_2 - \theta_1)^n$ при ограничениях $\theta_1 \leq x_{(1)}$ и $\theta_2 \geq x_{(n)}$. Очевидно, минимум разности $\theta_2 - \theta_1$ достигается при $\theta_1 = x_{(1)}$ и $\theta_2 = x_{(n)}$. Таким образом, оценками максимального правдоподобия являются

$$\hat{\theta}_1 = X_{(1)}, \qquad \hat{\theta}_2 = X_{(n)},$$

где $X_{(1)} = \min\{X_1, \dots, X_n\}$, $X_{(n)} = \max\{X_1, \dots, X_n\}$ — соответствующие порядковые статистики.

Если реализацией выборки $(X_1, \dots, X_n)$ является последовательность чисел $(x_1, \dots, x_n)$, то реализацией оценок $\hat{\theta}_1, \hat{\theta}_2$ будут числа $x_{(1)} = \min\{x_1, \dots, x_n\}$ и $x_{(n)} = \max\{x_1, \dots, x_n\}$.

В заключение отметим (без доказательства), что при определённых условиях гладкости, налагаемых на функцию правдоподобия, оценки максимального правдоподобия являются асимптотически несмещёнными и состоятельными.