

Лекция 3.

Метод максимального правдоподобия

Пусть имеется выборка $\vec{X} = (X_1, \dots, X_n)$ из распределения известного типа, определяемого неизвестными параметрами $\theta = (\theta_1, \dots, \theta_n)$

Идея метода состоит в следующем: подбираем параметры таким образом, чтобы вероятность получения данной выборки при случайном эксперименте была наибольшей.

Если распределение дискретное, то $P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n)$

Def. Функцией правдоподобия $L(\vec{X}, \theta)$ называется функция $L(\vec{X}, \theta) = P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n) = \prod_{i=1}^n P(X_i = x_i)$ при дискретном распределении

и $L(\vec{X}, \theta) = f_\theta(x_1) \cdot \dots \cdot f_\theta(x_n) = \prod_{i=1}^n f_\theta(x_i)$ в абсолютно непрерывном распределении

Def. Логарифмической функцией правдоподобия называется функция $\ln L(\vec{X}, \theta)$

Nota. Так как $y = \ln x$ возрастающая функция, точки максимума совпадают, а такую функцию правдоподобия становится легче дифференцировать

Def. Оценкой максимального правдоподобия $\hat{\theta}$ называется значение θ , при котором функция правдоподобия $L(\vec{X}, \theta)$ достигает наибольшего значения (при фиксированных значениях выборки)

Ex. 1. Пусть $\vec{X} = (X_1, \dots, X_n)$ - выборка из распределения Пуассона Π_λ с неизвестным $\lambda > 0$

Мет. Для распределения Пуассона $P(X = x_i) = \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$

Получаем функцию максимального правдоподобия $L(\vec{X}, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\lambda} =$

$$\frac{\lambda^{n\bar{x}}}{\prod_{i=1}^n x_i!} e^{-n\lambda}$$

$\ln L(\vec{X}, \lambda) = n\bar{x} \ln \lambda - \ln \prod_{i=1}^n x_i! - n\lambda$

$\frac{\partial \ln L}{\partial \lambda} = \frac{n\bar{x}}{\lambda} - n = 0 \implies \hat{\lambda} = \bar{x}$ - оценка максимального правдоподобия

Убедимся, что этот экстремум - максимум: $\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{n\bar{x}}{\lambda^2} < 0 \implies \hat{\lambda} = \bar{x}$ - точка максимума

Ex. 2. Пусть (X_1, \dots, X_n) из $N(a, \sigma^2)$

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

$$L(\vec{X}, a, \sigma^2) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-a)^2}{2\sigma^2}} = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^n (x_i-a)^2}{2\sigma^2}}$$

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

$$\frac{\partial \ln L}{\partial a} = -\frac{1}{2\sigma^2} \sum_{i=1}^n -2(x_i - a) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - a) = \frac{n\bar{x} - na}{\sigma^2}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} - \sum_{i=1}^n (x_i - a)^2 \frac{1}{2} \cdot (-2) \cdot \sigma^{-3} = \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - a)^2 - \frac{n}{\sigma}$$

$$\begin{cases} \frac{n\bar{x} - na}{\sigma^2} = 0 \\ \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - a)^2 - \frac{n}{\sigma} = 0 \end{cases} \implies \begin{cases} \hat{a} = \bar{x} \\ \widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (x_i - a)^2 = D^* \end{cases}$$

Ех. 3. Пусть (X_1, \dots, X_n) из $U(0, \theta)$. Найти оценку θ этого распределения.

Воспользуемся методом моментов:

$$EX = \frac{a+b}{2} = \frac{\theta}{2} \implies \bar{x} = \frac{\theta^*}{2} \implies \theta^* = 2\bar{x}$$

Воспользуемся методом максимального правдоподобия:

$$f_\theta = \begin{cases} 0, & x < 0 \\ \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & x > \theta \end{cases}$$

$$X_{(n)} = \max_i (X_1, \dots, X_n)$$

$$L(\vec{X}, \theta) = \prod_{i=1}^n f_\theta(x_i) = \begin{cases} 0, & \text{если } \theta < X_{(n)} \\ \frac{1}{\theta^n}, & \text{если } \theta \geq X_{(n)} \end{cases}$$

$L(\vec{X}, \theta)$ достигает наибольшего значения при наименьшем значении θ^n , то есть при $\hat{\theta} = X_{(n)}$

Сравним оценки:

$\theta^* = 2\bar{x}$ - несмещенная оценка, так как $E\theta^* = 2E\bar{x} = 2EX = \theta$

$$E(\theta^* - \theta)^2 = D\theta^* = D2\bar{x} = 4D\bar{x} = 4 \frac{D\bar{x}}{n} = \frac{4}{n} \frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

Изучим распределение $X_{(n)}$: $F_{X_{(n)}}(x) = P(X_{(n)} < x) = P(X_1 < x, X_2 < x, \dots, X_n < x) = P(X_1 < x) \cdot \dots \cdot$

$$P(X_n < x) = F_{X_1}(x) \cdot \dots \cdot F_{X_n}(x) = F_{(x_1)}^n(x)$$

$$F_{X_1}(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \leq x \leq \theta \\ 1, & x > \theta \end{cases} \implies F_{X_{(n)}}(x) = \begin{cases} 0, & x < 0 \\ \frac{x^n}{\theta^n}, & 0 \leq x \leq \theta \\ 1, & x > \theta \end{cases} \implies f_{X_{(n)}}(x) = \begin{cases} 0, & x < 0 \\ n \frac{x^{n-1}}{\theta^n}, & 0 \leq x \leq \theta \\ 1, & x > \theta \end{cases}$$

$$EX_{(n)} = \int_0^\theta x \cdot \frac{nx^{n-1}}{\theta^n} dx = \frac{n}{\theta^n} \int_0^\theta x^n dx = \frac{nx^{n+1}}{\theta^n(n+1)} \Big|_0^\theta = \frac{n\theta}{n+1} - \text{смещенная вниз оценка}$$

$\tilde{\theta} = \frac{n+1}{n} X_{(n)}$ - несмещенная оценка (будем считать, что эффективность не изменилась)

$$E\tilde{\theta}^2 = E\left(\frac{n+1}{n} X_{(n)}\right)^2 = \frac{(n+1)^2}{n^2} EX_{(n)}^2 = \frac{(n+1)^2}{n^2} \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^n} dx = \frac{(n+1)^2 \theta^2}{n(n+2)}$$

$$D\tilde{\theta} = E\tilde{\theta}^2 - (E\tilde{\theta})^2 = \frac{\theta^2}{n(n+2)}$$

$$D\tilde{\theta} = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{3n} = D\theta^*$$

Таким образом, оценка по методу правдоподобия сходится быстрее, чем оценка по методу моментов, поэтому она лучше

Отсюда следует, что при равномерном распределении выборочное среднее не является эффективной оценкой для математического ожидания; вместо нее половина максимального элемента выборки будет лучше

Nota. Эффективной здесь будет несмещенная оценка $\frac{n+1}{2n}X_{(n)}$

В общем случае для $U(a, b)$ будет такая эффективная оценка математического ожидания - $\frac{X_{(1)} + X_{(n)}}{2}$, длины интервала - $\frac{n+1}{n-1}(X_{(n)} - X_{(1)})$

Nota. При методе максимального правдоподобия обычно получаем состоятельные и эффективные оценки, но часто смещенные

Неравенство Рао-Крамера

Пусть $X \in F_\theta$ - семейство распределений с параметром $\theta \in \mathbb{R}$

Def. Носителем семейства распределений F_θ называется множество $C \subset \mathbb{R}$ такое, что $P(X \in C) = 1 \quad \forall X \in F_\theta$

$$f_\theta(x) = \begin{cases} \text{плотность } f_\theta(x) \text{ при непрерывном распределении} \\ P_\theta(X = x) \text{ при дискретном распределении} \end{cases}$$

Def. Информацией Фишера $I(\theta)$ семейства распределений F_θ называется величина $I(\theta) = E \left(\frac{\partial}{\partial \theta} \ln f_\theta(X) \right)^2$ при условии, что она существует

Def. Семейство распределений F_θ называется регулярным, если:

- существует носитель C семейства F_θ такой, что $\forall x \in C$ функция $\ln f_\theta(x)$ непрерывно дифференцируема по θ
- информация Фишера $I(\theta)$ существует и непрерывна по θ

Th. Пусть (X_1, \dots, X_n) - выборка объема n из регулярного семейства F_θ ,
 $\theta^* = \theta^*(X_1, \dots, X_n)$ - несмещенная оценка параметра θ , дисперсия которой $D\theta^*$ ограничена в любой замкнутой ограниченной области параметра θ

Тогда $D\theta^* \geq \frac{1}{nI(\theta)}$

Следствие: если при данных условиях получили $D\theta^* = \frac{1}{nI(\theta)}$, то оценка θ^* является эффективной (то есть дальше улучшать уже некуда)

Ex. Пусть (X_1, \dots, X_n) из $N(a, \sigma^2)$ (то есть $F_a = N(a, \sigma^2)$, σ^2 зафиксируем)

Проверим эффективность $a^* = \bar{x}$

Плотность $f_a(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$, носитель - вся прямая \mathbb{R}

$$\ln f_a(x) = -\ln \sigma - \frac{1}{2} \ln 2\pi - \frac{(x-a)^2}{2\sigma^2}, \quad a \in (-\infty, \infty)$$

$$\frac{\partial}{\partial a} \ln f_a(x) = \frac{1}{2\sigma^2} \cdot 2(x-a) = \frac{x-a}{\sigma^2} - \text{непрерывна для всех } a \in \mathbb{R}$$

$$I(a) = E \left(\frac{\partial}{\partial a} \ln f_a(X) \right)^2 = E \left(\frac{X-a}{\sigma^2} \right)^2 = \frac{1}{\sigma^4} E(X-a)^2 = \frac{E(X-EX)^2}{\sigma^4} = \frac{DX}{\sigma^4} = \frac{1}{\sigma^2} - \text{непрерывна по } a$$

Из этого следует, что $N(a, \sigma^2)$ - регулярное семейство относительно параметра a

$$Da^* = D\bar{x} = \frac{DX}{n} = \frac{\sigma^2}{n} - \text{ограничена по параметру } a$$

По неравенству Рао-Крамера $Da^* = \frac{\sigma^2}{n} = \frac{1}{nI(a)} = \frac{1}{n} \sigma^2$; из этого следует, что a^* - параметра a

Nota. Аналогично можно показать, что S^2 - несмещенная эффективная оценка для параметра σ^2