Robustní odhady jednorozměrného parametru

M-odhady

$X_1, \dots, X_n$ náhodný výběr z rozdělení $P = P_\theta$, $\theta \in \Theta$ je neznámý parametr

$\theta = T(P)$ ... neznámý parametr

$\hat{\theta} = T(P_n)$ ... jeho odhad

M-odhad parametru θ definujeme jako řešení minimalizace $\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{i=1}^{n} \rho(X_i, \theta)$, kde ρ je nějaká vhodně zvolená funkce.

Existuje-li její derivace $\psi(\cdot, \theta) = \dfrac{\partial \rho(\cdot, \theta)}{\partial \theta}$ a je spojitá, pak $\hat{\theta}$ je (jedním z) řešení rovnice $\sum_{i=1}^{n} \psi(X_i, \theta) = 0$.

Příklad (maximálně věrohodný odhad

nechť X_i má hustotu $f(x_i, \theta)$, pak zvolme $\rho(x, \theta) = -\log f(x, \theta)$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} -\log f(X_i, \theta) = \underset{\theta}{\operatorname{argmax}} \underbrace{\sum_{i=1}^{n} \log f(X_i, \theta)}_{\text{logaritmická věr. funkce}}$$ je maximálně věrohodný odhad parametru θ.

Poznámka

$T(P_n) = \underset{\theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \theta) = \underset{\theta}{\operatorname{argmin}}\, E\rho(X, \theta)$, kde $X \sim P_n$ ... příslušný stat. funkcionál $T(P) = \underset{\theta = T(P)}{\operatorname{argmin}}\, E\rho(X, \theta)$, kde $X \sim P_\theta$.

ex.-li spojitá fce $\psi(\cdot, \theta)$

$T(P_n)$ je řešení $\frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \theta) = 0 \iff E\psi(X, T(P_n)) = 0$, kde $X \sim P_n$ ... stat. funkcionál $T(P)$ řeší $E\psi(X, T(P)) = 0 = \int_{\mathbb{R}} \psi(x, T(P))\, dP(x)$, kde $X \sim P$.

influenční funkce M-odhadu:

nechť $\psi(\cdot,\theta)$ je absolutně spojitá v θ a odhad $T(P_n)$ je určen jednoznačně

$$P_t = (1-t)\cdot P + t\cdot \delta_x$$

$T(P_t)$ je řešení rovnice $\displaystyle\int_{\mathbb{R}} \psi(y, T(P_t))\, d\big((1-t)P(y) + t\cdot\delta_x(y)\big) = 0.$

$$(1-t)\int_{\mathbb{R}} \psi(y, T(P_t))\, dP(y) + t\cdot \psi(x, T(P_t)) = 0. \qquad \Big|\ \text{derivace podle } t$$

$$-\int_{\mathbb{R}} \psi(y, T(P_t))\, dP(y) + (1-t)\cdot \int_{\mathbb{R}} \frac{\partial \psi(y,\theta)}{\partial\theta}\Big|_{\theta=T(P_t)} \cdot \frac{dT(P_t)}{dt}\, dP(y) + \psi(x, T(P_t)) + t\cdot \frac{\partial\psi(x,\theta)}{\partial\theta}\Big|_{\theta=T(P_t)}\cdot\frac{dT(P_t)}{dt} = 0.$$

dosadíme $t=0$ (resp. počítejme $\lim\limits_{t\to 0+}$):

$$\overbrace{-\int_{\mathbb{R}} \psi(y, T(P))\, dP(y)}^{=0} + IF(x,T,P)\cdot \int_{\mathbb{R}} \frac{\partial\psi(y,\theta)}{\partial\theta}\Big|_{\theta=T(P)} dP(y) + \psi(x, T(P)) = 0$$

$$IF(x,T,P) = -\frac{\psi(x, T(P))}{\underbrace{\int_{\mathbb{R}} \frac{\partial\psi(y,\theta)}{\partial\theta}\Big|_{\theta=T(P)} dP(y)}_{\text{není funkce } x}}$$

má-li být odhad robustní, má mít omezenou IF $\Rightarrow$ má-li být M-odhad robustní, pak by ψ měla být omezená.

M-odhady parametru posunutí (polohy)

$X_i \sim F(x-\theta)$, ekvivalentně $X_i = \theta + \varepsilon_i$, kde $\varepsilon_i \sim F$, která je symetrická kolem 0 => θ je střed symetrie

zvolíme $\rho(x,\theta) = \rho(x-\theta)$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \rho(X_i - \theta)$$, resp. $\hat{\theta}$ řeší rovnici $\sum_{i=1}^{n} \psi(X_i - \hat{\theta}) = 0.$ $\psi(y) = \rho'(y)$ je spojitá funkce

$$IF(x,T,P) = \frac{\psi(x - T(P))}{\int_{\mathbb{R}} \psi'(y)\, dP(y)}.$$

Poznámky

Jak volit funkci ρ, resp. ψ?

ρ symetrická kolem 0 => ψ je lichá funkce

ρ ryze konvexní => odhad je určen jednoznačně

je-li ρ na nějakém intervalu lineární => ψ je na tomto intervalu konstantní, pak rovnice $\sum_{i=1}^{n} \psi(X_i - \theta) = 0$ může mít více řešení.

$$\hat{\theta} = \frac{1}{2}(\hat{\theta}^+ + \hat{\theta}^-)$$, kde $$\hat{\theta}^+ = \inf\{t : \sum_{i=1}^{n} \psi(X_i - t) < 0\}$$

- stejným způsobem můžeme definovat M-odhad pro ψ neklesající, nespojitou se skoky.

$$\hat{\theta}^- = \sup\{t : \sum_{i=1}^{n} \psi(X_i - t) > 0\}$$

Bod zhroucení M-odhadu parametru polohy:

$$\varepsilon^* = \begin{cases} 0, & \text{je-li } \psi \text{ neomezená} \\ \frac{1}{2}, & \text{je-li } \psi \text{ omezená a lichá} \end{cases}$$

<u>Příklady</u>

a) $\rho(x) = x^2$

$\psi(x) = 2x$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} (X_i - \theta)^2$$

$\hat{\theta}$ je řešením rovnice $\sum_{i=1}^{n} 2(X_i - \hat{\theta}) = 0$

$$\sum_{i=1}^{n} X_i - n\hat{\theta} = 0$$

$\hat{\theta} = \overline{X}$ ... výběrový průměr (není robustní)

$$IF(x, T, P) = \frac{2\cdot(x - EX)}{\int 2\,dP} = x - EX.$$

b) $\rho(x) = |x|$ $\quad \hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} |X_i - \theta| = \tilde{X}_n$ je výběrový medián

$$IF(x, T, P) = \frac{\operatorname{sign}(x - \tilde{X})}{2f(\tilde{X})}$$

c) Huberův odhad

$$\psi_H(x) = \begin{cases} x, & -k \le x \le k \\ k \cdot \operatorname{sign} x, & |x| > k \end{cases}$$

$k > 0$ je pevně zvolená konstanta

d) Tukeyho biweight

$$\psi_T(x) = \begin{cases} x\left[1-\left(\frac{x}{k}\right)^2\right] & , |x| \le k \\ 0 & , |x| > k \end{cases}$$

e) věr. funkce Cauchyho rozdělení

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{.. hustota Cauchyho rozdělení}$$

$$f'(x) = -\frac{2x}{\pi(1+x^2)^2}$$

$$\psi_C(x) = -\frac{f'(x)}{f(x)} = \frac{2x}{1+x^2} \quad , x \in \mathbb{R}$$

f) Andrewsova sinusová funkce

$$\psi_A(x) = \begin{cases} \sin\frac{x}{k} & , |x| \le k\pi \\ 0 & , |x| > k\pi \end{cases}$$

g) Hampel

$$\psi_H(x) = \begin{cases} |x| \cdot \operatorname{sign} x & , |x| < a \\ a \cdot \operatorname{sign} x & , a \le |x| < b \\ \frac{c-|x|}{c-b}\, a \operatorname{sign} x & , b \le |x| < c \\ 0 & , |x| > c \end{cases}$$

h) skipped mean

$$\psi^*(x) = \begin{cases} x & , |x| \le k \\ 0 & , |x| > k \end{cases}$$

i) skipped median

$$\tilde{\psi}(x) = \begin{cases} -1 & , -k \le x < 0 \\ 0 & , |x| > k \\ 1 & , 0 \le x \le k \end{cases}$$