Vlastnosti M-odhadů parametru polohy

- M-odhady jsou ekvivariantní vůči posunutí: $\hat{\theta}(X_1+c, \dots, X_n+c) = \hat{\theta}(X_1, \dots, X_n) + c, \quad \forall c \in \mathbb{R}$

- M-odhady obecně nejsou ekvivariantní vůči měřítku: $\hat{\theta}(cX_1, \dots, cX_n) = c \cdot \hat{\theta}(X_1, \dots, X_n), \quad \forall c > 0$

Řešení:

1.) Zároveň s parametrem polohy odhadujeme i parametr měřítka σ, např. Huber: $(\theta, \sigma)^T$ je řešením soustavy rovnic:

$$\sum_{i=1}^{n} \psi_H\left(\frac{X_i - \theta}{\sigma}\right) = 0$$

ψ_H – Huberova funkce ψ

$$\chi(x) = \psi_H^2(x) - \int_{\mathbb{R}} \psi_H^2(y)\, \varphi(y)\, dy$$, kde $\varphi(y)$ je hustota $N(0,1)$

$$\sum_{i=1}^{n} \chi\left(\frac{X_i - \theta}{\sigma}\right) = 0$$

2.) Studentizované M-odhady

– odhadneme měřítko vhodnou škálovou statistikou $S_n = S_n(X_1, \dots, X_n)$:

a) $S_n(x_1, \dots, x_n) > 0 \quad , x_1, \dots, x_n \in \mathbb{R}$

b) $S_n(x_1+c, \dots, x_n+c) = S_n(x_1, \dots, x_n) \quad , c \in \mathbb{R}, x_1, \dots, x_n \in \mathbb{R}$

c) $S_n(cx_1, \dots, cx_n) = c \cdot S_n(x_1, \dots, x_n) \quad , c > 0, x_1, \dots, x_n \in \mathbb{R}$

d) S_n je konzistentní odhad σ

Studentizovaný M-odhad parametru θ

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{X_i - \theta}{S_n}\right)$$

$$\sum_{i=1}^{n} \psi\left(\frac{X_i - \hat{\theta}}{S_n}\right) = 0$$

$$IF(x, T, P) = \frac{S(P)}{\int_{\mathbb{R}} \psi'\left(\frac{y}{S(P)}\right) dP(y)} \cdot \psi\left(\frac{x - T(P)}{S(P)}\right)$$ je příslušná influenční funkce

Volba S_n:

- výběrová směrodatná odchylka $S_n = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}$ ... není robustní

- mezikvartilová odchylka $S_n = X_{0.75} - X_{0.25}$

- mediánová absolutní odchylka (MAD) $S_n = \text{medián} \{ |X_1 - \tilde{X}_n|, \dots, |X_n - \tilde{X}_n| \}$, kde $\tilde{X}_n = \text{medián} \{ X_1, \dots, X_n \}$.

L-odhady

- jsou lineární kombinací funkcí pořádkových statistik $X_{(1)} \leq \ldots \leq X_{(n)}$

$$T_n = \sum_{i=1}^{n} c_i h(X_{(i)})$$, $c_1, \ldots, c_n$ jsou vhodné konstanty a h je nějaká funkce

Odhady parametru polohy

1.) výběrový průměr $T_n = \overline{X}$, $c_1 = c_2 = \ldots = c_n = \frac{1}{n}$, $h(x) = x$

2.) střed rozpětí $T_n = \frac{X_{(1)} + X_{(n)}}{2}$, $c_1 = c_n = \frac{1}{2}$, $c_2 = \ldots = c_{n-1} = 0$, $h(x) = x$

3.) výběrový medián $T_n = \tilde{X} = X_{(\frac{n+1}{2})}$, $c_1 = \ldots = c_{\frac{n-1}{2}} = c_{\frac{n+3}{2}} = \ldots = c_n = 0, c_{\frac{n+1}{2}} = 1$, $h(x) = x$

pro n liché (pro n sudé analogicky)

4.) α-useknutý průměr $T_n = \overline{X}_\alpha = \frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)}$

5.) α-winsorizovaný průměr $T_n = \overline{X}_\alpha^{w} = \frac{1}{n} \left\{ [n\alpha] X_{([n\alpha]+1)} + \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)} + [n\alpha] X_{(n-[n\alpha])} \right\}$

6.) Jemův vážený průměr

$$T_{n,k} = \binom{n}{2k+1}^{-1} \sum_{i=k+1}^{n-k} \binom{i-1}{k}\binom{n-i}{k} X_{(i)}$$

$T_{n,0} = \overline{X}$

$T_{n,k} = \tilde{X}_n$ pro $k = \begin{cases} \frac{n-1}{2} & n \text{ liché} \\ \frac{n}{2}-1 & n \text{ sudé} \end{cases}$

L-odhady parametru měřítka

1.) $T_n = X_{(n)} - X_{(1)}$ výběrové rozpětí

2.) Giniho průměrná diference

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} |X_i - X_j| = \frac{2}{n(n-1)} \sum\sum_{i<j} |X_i - X_j| = \frac{2}{n(n-1)} \sum_{i=1}^{n} (2i-n-1) X_{(i)}$$

R-odhady parametru polohy

jsou inverzí jednovýběrových pořadových testů

vsuvka z teorie pořadových testů:

$X_1, \dots, X_n$ je náh. výběr z rozdělení s dist. fcí $F(x-\theta)$, kde $\theta \in \mathbb{R}$ je neznámý parametr (funkce F je symetrická: $F(-y) = 1 - F(y)$, $\forall y$)

$X_i = \theta + e_i$, kde $e_i \sim F$ — rozdělení X_i je symetrické kolem bodu θ (medián)

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

definujeme rozdíly $|X_1-\theta_0|, |X_2-\theta_0|, \dots, |X_n-\theta_0|$ a určíme jejich pořadí mezi $|X_1-\theta_0|, \dots, |X_n-\theta_0|$

$R_1^+(\theta_0)$ $\quad R_2^+(\theta_0)$ $\quad R_n^+(\theta_0)$

testová statistika: $S_n(\theta_0)=\sum_{i=1}^{n} \operatorname{sign}(X_i-\theta_0)\cdot a_n(R_i^+(\theta_0))$, kde $a_n: \{1,2,\dots,n\}\mapsto \mathbb{R}$ je nějaká funkce pořadí

Příklad

$a_n(i)=i \qquad S_n(\theta_0)=\sum_{i=1}^{n} \operatorname{sign}(X_i-\theta_0)\cdot R_i^+(\theta_0)$ ... jednovýběrový Wilcoxonův test

Za platnosti $H_0(\theta=\theta_0)$ jsou $\operatorname{sign}(X_i-\theta_0)$ a $R_i^+(\theta_0)$ nezávislé $\Rightarrow ES_n(\theta_0)=0$.

My chceme odhadnout skutečnou hodnotu středu symetrie (mediánu) θ_0. Budeme jej tedy hledat jako řešení rovnice $S_n(\theta_0)=0$.

$S_n(\theta_0)=\sum_{i=1}^{n} \operatorname{sign}(X_i-\theta_0)\cdot a_n(R_i^+(\theta_0))=0$ – nerostoucí, schodovitá funkce v θ_0 $\Rightarrow$ řešení nemusí existovat, nebo není jednoznačné

Příslušný R-odhad: $T_n=\frac{1}{2}(T_n^-+T_n^+)$, kde $T_n^-=\sup\{t: S_n(t)>0\}$
$T_n^+=\inf\{t: S_n(t)<0\}$

Příklady

1) $a_n(i)=1 \ \forall i=1,\dots,n \qquad S_n(\theta_0)=\sum_{i=1}^{n} \operatorname{sign}(X_i-\theta_0)=0$ ... $T_n=\tilde{X}_n$ je výběrový medián

2) $a_n(i) = i, \ \forall i = 1, \dots, n$ $\quad S_n(\theta_0) = \sum_{i=1}^{n} \mathrm{sign}(X_i - \theta_0) \cdot R_i^+(\theta_0) = 0$ Wilcoxonův jednovýběrový test

$\varepsilon^* \doteq 0,293$

$T_n = \text{medián} \left\{ \frac{X_i + X_j}{2}, \ 1 \le i \le j \le n \right\}$ ... Hodgesův-Lehmannův odhad

3) $a_n(i) = \Phi^{-1}\left(\frac{i}{2(n+1)} + \frac{1}{2}\right)$ $\quad S_n(\theta_0) = \sum_{i=1}^{n} \mathrm{sign}(X_i - \theta_0)\, \Phi^{-1}\left(\frac{R_i^+(\theta_0)}{2(n+1)} + \frac{1}{2}\right) = 0$ van der Waerdenův test

numerické řešení

<u>Vlastnosti:</u>

R-odhady jsou ekvivariantní vůči posunutí i měřítku.