STATISTICKÉ EXPERIMENTY S ROBUSTNOSTÍ A ADAPTIVITOU MODIFIKOVANÝCH D-ODHADŮ

Miloslav Vošvrda
Ústav teorie informace a automatizace ČSAV

1. Úvod

V poslední době se značné úsilí statistiků věnovalo na výzkum robustních a adaptivních odhadů polohy pravděpodobnostního rozdělení. Toto úsilí vedlo k tvorbě značně rozsáhlé třídy odhadů polohy. Každá rozsáhlá třída odhadů vyžaduje zobecňující teorii a srovnávací studii. Předkládaný příspěvek nabízí odhad, který byl vytvořen na základě teorie, která zobecňuje některé doposud známé výsledky z výzkumu robustních odhadů polohy a provádí srovnání s některými dosud známými "dobrými" odhady polohy.

V příspěvku odvodíme tvar modifikovaného D-odhadu pro parametr polohy distribuční funkce. Chování tohoto odhadu budeme srovnávat s chováním některých robustních nebo adaptivních odhadů polohy na třídě diskrétních nebo spojitých distribučních funkcí a to za přítomnosti symetrické nebo asymetrické kontaminace.

2. Modifikovaný D-odhad

Nechť $(\mathcal{X},\mathcal{B})$ je měřitelný prostor a $\mathcal{P}=\{P_\theta,\ \theta\in\Theta\subset R^1\}$ je třída pravděpodobnostních měr na něm. Nechť $P_\theta \ll \lambda$ /kde λ je σ-konečná/ pro každé $\theta\in\Theta$ a tudíž existuje $p_\theta = \frac{dP_\theta}{d\lambda}\ [\lambda]$. Nechť $(\mathcal{X}^\infty,\mathcal{B}^\infty)$ je kartézský součin měřitelných prostorů $(\mathcal{X},\mathcal{B})$ a nechť X_i je i-tá souřadnicová náhodná veličina tj.

$$X_i(x_1,x_2,\dots,x_i,\dots) = x_i \,. \qquad /1/$$

Náhodným výběrem rozsahu n nazveme n-tici

$$(X_1,\dots,X_n)\,. \qquad /2/$$

Pro každý náhodný výběr definujeme pravděpodobnostní míru $P_n\in\mathcal{P}$ následujícím způsobem:

$$P_n(E) = n^{-1}\sum_{i=1}^{n}\chi_E(X_i) \qquad \forall E\in\mathcal{B} \qquad /3/$$

kde

$$\chi_E(X_i) = 1 \quad \text{jestliže } X_i\in E$$
$$= 0 \quad \text{jinak.} \qquad /4/$$

Nyní si zavedeme třídu D-odhadů. Nechť f je spojitá, konvexní funkce na $(0,\infty)$, ryze konvexní v $\underline{1}$ a $f(1)=0$ a pokládáme

$$f(0) = \lim_{\varepsilon\to 0_+} f(\varepsilon)\,, \qquad /5/$$

$$0\cdot f\left(\tfrac{0}{0}\right) = 0\,, \qquad /6/$$

$$0\cdot f\left(\tfrac{z}{0}\right) = \lim_{\varepsilon\to 0_+}\varepsilon f\left(\tfrac{z}{\varepsilon}\right) = z\lim_{u\to\infty}\frac{f(u)}{u} \qquad \text{pro každé } z\in(0,1\rangle\,. \qquad /7/$$

D-odhad $\{T_n^f\}$ definujeme podmínkami:

1/ $$T_n^f\in\Theta \qquad \text{pro každé } n \text{ přirozené}, \qquad /8/$$

2/ $$D_f(P_n, P_{T_n^f}) = \min_{\theta\in\Theta} D_f(P_n,P_\theta)\,, \qquad /9/$$

kde $D_f(\cdot,\cdot)$ je f-divergence /viz Vajda I. /1983//.

Položme

$$f(u) = \alpha^{-1}(1-u^\alpha) \qquad \text{pro } \alpha\in(0,1), \qquad /10/$$

$$= -\ln u \qquad \text{pro } \alpha=0 \text{ a } u\in R^1\,. \qquad /11/$$

Pro takovouto funkci má odhad $\{T_n^f\}$ tvar

$$\min_{\theta\in\Theta} E_{p_n}(1-p_\theta^\alpha) = \max_{\theta\in\Theta} E_{p_n}(p_\theta^\alpha) \qquad \text{pro} \quad \alpha\in(0,1) \qquad /12/$$

a

$$\min_{\theta\in\Theta} E_{p_n}(-\ln p_\theta) = \max_{\theta\in\Theta} E_{p_n}(\ln p_\theta) \qquad \text{pro} \quad \alpha=0 \qquad /13/$$

V článku Vajdy I. /1983/ bylo ukázáno, že odhady tohoto typu jsou konzistentní a asymptoticky normální. V článku Vošvrdy M. /1983/ bylo ukázáno, že v regulárním /podle Rao C.R. /1961//systému hustot podmínka

$$\frac{d^2}{du^2} f(u)\Big|_{u=1} \neq 0 \qquad /14/$$

stačí k tomu, aby odhad $\{T_n^f\}$ byl asymptoticky eficientní.

Nyní takový odhad zkonstruujeme vzhledem k normální $\mathcal{N}(\theta,\sigma^2)$ rodině hustot tj.

$$p_\theta(x) = (2\pi\sigma^2)^{-1/2} \exp\{-(2\sigma^2)^{-1}(x-\theta)^2\} \qquad /15/$$

Podle /12/ má tento odhad tvar

$$T_n^{\mathcal{N}} = \max_{\theta\in\Theta} E_{p_n}(p_\theta^\alpha) = \max_{\theta\in\Theta} n^{-1}\sum p_\theta^\alpha(x_i) =$$

$$\begin{cases} = \max\limits_{\theta\in\Theta} n^{-1}\sum\limits_{i=1}^{n} \exp\{-\alpha(x_i-\theta)^2(2\sigma^2)^{-1}\} & \text{pro} \quad \alpha\in(0,1), \qquad /16/ \\ = \min\limits_{\theta\in\Theta} n^{-1}\sum\limits_{i=1}^{n}(x_i-\theta)^2 & \text{pro} \quad \alpha=0 \qquad /17/ \end{cases}$$

Jak volit hodnotu α ? Odpověď můžeme nalézti v chování influenční křivky konstruovaného odhadu. V článku Vajdy I. /1983/ je ukázáno chování influenční křivky odhadu uvedeného v /16/, /17/; influenční křivka je definována podle Hubera /1972/, kde ψ -funkce má zde tvar

$$\psi_\theta(x_i) = \frac{d}{d\theta} f(p_\theta(x)), \qquad /18/$$

a odhad budeme nazývat robustní, jestliže má omezenou influenční křivku.
Na obr. 1 je znázorněno chování influenční křivky odhadu uvedeného v /16/, /17/ s parametry $\theta=0$, $\sigma^2=1$ pro různá α . Z obr. 1 je vidět, že existuje $\alpha\in(0,1)$ takové, že odhad v /16/, /17/ je robustní. Vybereme-li $\alpha=0.1$ získáme odhad jehož influenční křivka je podobná influenční křivce odhadu, který zavedl Hampel F. /1974/. Též je zajímavá souvislost s α -uřezaným průměrem, kde Hogg V. /1974/ uvádí, že po mnoha experimentálních studiích došli k závěru, že nejvhodnější hodnota je 0.1. Uvidíme později jak 0.1 - uřezaný průměr má blízko k 0.1 - minimálně divergentnímu odhadu /D-odhad/.

Jak volit odhad parametru σ ? Tuto otázku lze řešit dvěma způsoby. Buďto použít v f -divergenci vhodné váhové funkce, které vedou k odhadu, jež má potom tvar

$$\min_{\theta,\sigma^2} n^{-1}\sum_{i=1}^{n}\left((x_i-\theta)^2-\sigma^2\right)^\alpha \qquad \text{pro} \quad \alpha\in(0,1). \qquad /19/$$

Nebo parametr σ můžeme odhadovat následujícím způsobem:
Nechť $X_{(1)}, \ldots\ldots, X_{(n)}$ je uspořádaný náhodný výběr. Sestrojíme následující náhodné veličiny

$$W_{(i)} = X_{(n+1-i)} - X_{(i)} \qquad \text{pro } 2 \leq i \leq \left[\tfrac{n}{2}\right] \tag{20}$$

kde $\left[\frac{n}{2}\right]$ znamená celou část čísla $\frac{n}{2}$. Pro odhad parametru σ bychom mohli použít nestranný odhad Downtona F. /1966/ tvaru

$$\hat{\sigma} = 2(\pi)^{1/2} (n(n-1))^{-1} \sum_{i=1}^{n} (i - 2^{-1}(n+1))\, X_{(i)} \, . \tag{21}$$

Použijeme-li rovnost

$$\sum_{i=1}^{n} (i - 2^{-1}(n+1))\, X_{(i)} = 4^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |X_i - X_j| \tag{22}$$

získáme odhad

$$\hat{\sigma} = 2^{-1} (\pi)^{1/2} (n(n-1))^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |X_i - X_j| \, , \tag{23}$$

kde
$$G = (n(n-1))^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} |X_i - X_j| \tag{24}$$

je odhad variability podle Giniho C. /1912/. Po jednoduché úpravě výrazu /24/ získáme následující odhad variability:

$$S_M = \sum_{i=1}^{\left[\frac{n}{2}\right]} \frac{n+1-2i}{n(n-1)} W_{(i)} \, . \tag{25}$$

Výraz /25/ dosadíme do výrazu /16/ a získáme následující odhad:
/modifikovaný D-odhad/

$$T_n^{N} = \max_{\theta \in \Theta} n^{-1} \sum_{i=1}^{n} \exp\left\{-\alpha \frac{(x_i - \theta)^2}{2 S_M^2}\right\} \, . \tag{26}$$

Tímto jsme získali odhad polohy pravděpodobnostního rozdělení, který budeme v dalším nazývat α -odhadem. Chování ve vychýlenosti a ve varianci tohoto odhadu budeme srovnávat s chováním níže uvedených odhadů polohy pravděpodobnostního rozdělení. Srovnání bude prováděno jak na diskrétních, tak na spojitých pravděpodobnostních modelech. Též bude prováděna kontaminace a to jak symetrická, tak asymetrická.

3. Srovnávací odhady

A. Skupina obsahující robustní odhady

$_1T$ α -odhad, který byl výše popsán

$_2T$ Maximálně věrohodný odhad polohy pravděpodobnostního rozdělení pro rozdělení Bernoulliho, binomické, Poissonovo, geometrické, Johnsonovo a normální je aritmetický průměr

$_3T$ Hodges - Lehmannův odhad má tvar:

$$T_n^{HL} = \mathrm{med}\left\{\tfrac{1}{2}(X_{(i)} + X_{(j)})\right\} \qquad \text{pro } i, j = 1, 2, \ldots, n. \tag{27}$$

$_4T$ α - uřezaný průměr má tvar:

$$T_n^{\alpha} = \frac{1}{n-2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i)} \qquad \text{kde } \alpha = 0.1 \qquad /28/$$

$_5T$ Hampelův odhad polohy má tvar:

T_n^H je řešením rovnice

$$\sum_{i=1}^{n} \varphi\left(\frac{X_{(i)} - T_n^H}{s_1}\right) = 0 \qquad /29/$$

kde

$$\varphi(u) = \operatorname{sgn} u \begin{cases} |x| & \text{jestliže } 0 \le |x| \le 2 \\ 2\,\dfrac{5.5-|x|}{3.5} & 2 \le |x| \le 5.5 \\ 0 & 5.5 \le |x| \end{cases}$$

$$s_1 = med\{\,|X_i - med\{X_i\}|\,\} \qquad \text{pro } i=1,\dots,n$$

B. Skupina obsahující adaptivní odhady

$_6T$ Takeuchiho odhad má následující tvar:

$$T_n^{TK} = \frac{\sum_{\alpha}\sum_{\beta} \tilde{s}^{\alpha\beta} T_{\alpha}}{\sum_{\alpha}\sum_{\beta} \tilde{s}^{\alpha\beta}} \qquad /30/$$

kde

$$T_{\alpha} = \sum_{i=1}^{n} p_{\alpha}^{i} X_{(i)}$$

$$p_{\alpha}^{i} = \frac{\binom{i-1}{\alpha-1}\binom{n-i}{k-\alpha}}{\binom{n}{k}} \qquad \alpha < i < n-k+\alpha$$

$$= 0 \qquad \text{jinak},$$

$$s_{\alpha\beta} = \sum_{i,j} p_{\alpha\beta}^{ij} X_{(i)} X_{(j)} - T_{\alpha} T_{\beta} \qquad \alpha < \beta$$

$$p_{\alpha\beta}^{ij} = \frac{\binom{i-1}{\alpha-1}\binom{j-i-1}{\beta-\alpha-1}\binom{n-j}{k-\beta}}{\binom{n}{k}} \qquad \alpha < i,\ \beta \le n-j+k$$

$$= 0 \qquad \text{jinak}$$

$$\tilde{s}_{\alpha\beta} = 2^{-1}(s_{\alpha\beta} + s_{\beta\alpha}), \quad \tilde{s}^{\alpha\beta} = (\tilde{s}_{\alpha\beta})^{-1}, \quad k = \left[\left(\frac{37n}{20}\right)^{1/2}\right]$$

$_7T$ Hogg' 69 má následující tvar:
Vybírá se mezi čtyřmi odhady podle hodnoty výběrové špičatosti

$$k = \frac{n \sum_{i=1}^{n} (X_i - \bar{X})^4}{\left(\sum_{i=1}^{n} (X_i - \bar{X})^2\right)^2} \qquad /31/$$

Jestliže $K < 1.9$ užije se aritmetický průměr z horního a dolního výběrového kvartilu vypočteného z dat

$1.9 \le K < 3.1$ aritmetický průměr

$3.1 < k \leq 4.5$ uřiznutý aritmetický průměr s $\alpha = 0.25$

$4.5 < k$ $T_n^{GA}(x) = 0.3X_{([\frac{n}{3}]+1)} + 0.4X_{([\frac{n}{2}])} + 0.3X_{(n-[\frac{n}{3}])}$

/Gastwirthův odhad/

${}_8T$ Hogg' 67 má podobnou konstrukci jako Hogg' 69 s tím, že výběr mezi odhady je prováděn následovně:

jestliže $k < 2$ užije se aritmetický průměr z horního a dolního výběrového kvartilu vypočteného z dat

$2 \leq k \leq 4$ aritmetický průměr

$4 < k \leq 5.5$ uřiznutý aritmetický průměr s $\alpha = 0.25$

$5.5 < k$ medián

${}_9T$ Bickelův odhad má následující tvar:

$$T^B_{n,\alpha} = \frac{1}{n-2[n\tilde{\alpha}]} \sum_{i=2}^{n-1} X_{([n\tilde{\alpha}]+i)} + \frac{(1+[\tilde{\alpha}n]-\tilde{\alpha}n)}{n-2[n\tilde{\alpha}]} \left(X_{([n\tilde{\alpha}]+1)} + X_{(n-[n\tilde{\alpha}])} \right) \qquad /32/$$

kde $\tilde{\alpha}$ je takové $\alpha \in \langle 0, 1/4 \rangle$ pro něž je výraz

$$\hat{B}(\alpha) = \frac{1}{(1-2\alpha)^2} \left[\sum_{j=[n\alpha]+2}^{n-[n\alpha]-1} (X_{(j)})^2 + \alpha \left(X^2_{([n\alpha]+1)} + X^2_{(n-[n\alpha])} \right) \right] \qquad /33/$$

minimální. Náhodné veličiny $X_{(\cdot)}$ mají distribuční funkci $\tilde{F}(\cdot)$ jenž je symetrizovanou verzí distribuční funkce F t.j.

$$\tilde{F}(t) = 2^{-1} \{ F(t - F^{-1}(1/2)) + \bar{F}(t - F^{-1}(1/2)) \}$$

kde $\bar{F}(t) = 1 - F(-t)$.

${}_{10}T$ Symetrizovaný adaptivní uřezaný průměr je Bickelův odhad u něhož odhad variance je založen na pseudovýběrech rozsahu $2n$ v každém kroku.

${}_{11}T$ Jaeckelův odhad má následující tvar:

$$T^J_{n,\tilde{\alpha}} = \frac{1}{n-2[n\tilde{\alpha}]} \sum_{i=2}^{n-1} X_{([n\tilde{\alpha}]+i)} + \frac{(1+[n\tilde{\alpha}]-n\tilde{\alpha})}{n-2[n\tilde{\alpha}]} \left(X_{([n\tilde{\alpha}]+1)} + X_{(n-[n\tilde{\alpha}])} \right) \qquad /34/$$

kde $\tilde{\alpha}$ je takové $\alpha \in \langle 0, 1/4 \rangle$ pro něž je výraz

$$\hat{A}(\alpha) = \frac{1}{(1-2\alpha)^2} \left[\sum_{j=[n\alpha]+2}^{n-[n\alpha]-1} (X_{(j)} - T^J_{n,\alpha})^2 + \alpha \{ (X_{([n\alpha]+1)} - T^J_{n,\alpha})^2 + (X_{(n-[n\alpha])} - T^J_{n,\alpha})^2 \} \right] \qquad /35/$$

minimální.

${}_{12}T$ Jaeckel-Bickelův odhad spočívá ve výběru takového Jaeckelova odhadu pro nějž je $\hat{A}(\cdot)$, kde $\hat{A}$ je počítáno z $\hat{A}([\frac{n}{4}]/n)$ a $\hat{A}([\frac{n}{4}]/n)$ minimální.

${}_{13}T$ Lineární kombinace dvou uřezaných průměrů má následující tvar:

$$T_\alpha^{Lk} = c\,\frac{[0.05n]}{n} + (1-c)\,\frac{[0.25n]}{n} \qquad /36/$$

kde $c = \frac{V_{12} - V_{22}}{V_{11} - 2V_{12} + V_{22}}$, $V_{11} = \mathring{A}\left(\frac{[0.05n]}{n}\right)$, $V_{12} = \mathring{B}\left(\frac{[0.05n]}{n}, \frac{[0.25n]}{n}\right)$, $V_{22} = \mathring{A}\left(\frac{[0.25n]}{n}\right)$ /37/

pro $n = 2k$, $k = 1, 2, \dots$,

$$\mathring{A}(\alpha) = \frac{1}{(1-2\alpha)^2} \left\{ \sum_{j=k+1}^{[n(1-\alpha)]} \left(X_{(j)} - X_{(n-j+1)}\right)^2 + \alpha \left(X_{(n-[n\alpha])} - X_{([n\alpha]+1)}\right)^2 \right\} \quad /38/$$

$$\mathring{B}(\alpha_1, \alpha_2) = \frac{1}{(1-2\alpha_1)(1-2\alpha_2)} \left\{ \sum_{j=k+1}^{[n(1-\alpha_2)]} \left(X_{(j)} - X_{(n-j+1)}\right)^2 + \sum_{j=[n(1-\alpha_2)]}^{[n(1-\alpha_1)]} \left(X_{(j)} - X_{(n-j+1)}\right) \cdot \left(X_{([n(1-\alpha_2)])} - X_{([n\alpha_2]+1)}\right)^2 + \alpha_1 \left(X_{([n(1-\alpha_2)])} - X_{([n\alpha_2]+1)}\right)\left(X_{([n(1-\alpha_2)])} - X_{([n\alpha_2]+1)}\right) \right\} \quad /39/$$

s $\alpha_1 < \alpha_2$ a vybírá se takové α, α_1, α_2 aby asymptotická variance odhadu byla minimální.

$_{14}T$ Johnsův odhad má následující tvar:

$$T_n^{JH} = c_1 T_{n,1} + c_2 T_{n,2} \quad /40/$$

kde

$$T_{n,1} = \frac{1}{2s} \sum_{j=r+1}^{r+s} \left(X_{(j)} + X_{(n-j+1)}\right) \quad /41/$$

$$T_{n,2} = \frac{1}{n-2r-2s} \sum_{j=r+s+1}^{[\frac{n}{2}]} \left(X_{(j)} + X_{(n-j+1)}\right) \quad /42/$$

s

$$c_1 + c_2 = 1 \quad /43/$$

a

$$c_1 \sim \frac{1}{D_1^2}\left(\frac{2s(n+r+2s}{(2r+s)(n-2r)}\right) - \frac{2}{D_1 D_2} \frac{n-2r-2s}{n-2r} \quad /44/$$

$$c_2 \sim \frac{1}{D_2^2}\left(\frac{2(n-2r-2s}{n-2r}\right) - \frac{4}{D_1 D_2}\left(\frac{s}{n-2r}\right) \quad /45/$$

D_i jsou proporcionální k délkám vstupních proměnných v $T_{n,i}$ pro $i = 1, 2, n < 20$.

$_{15}T$ Adaptivní forma přeskakování má následující tvar:

Jestliže	potom užijeme následujícího odhadu	
body leží za bodem s	1/10 (4.medián + 2.(1/4(3/4) - kvantilu) + + 1/8(7/8)-kvantilu)	/46/
body leží za bodem s , ale před bodem c	$T_n^{TR} = \frac{1}{4}\left(h_1 + 2X_{([\frac{n}{2}])} + h_2\right)$	/47/
body leží za bodem c	jednoduchý C - přeskakující medián	/48/

kde
$$h_1 = \begin{cases} X_{([\frac{n}{4}])} & \text{jestliže } n \text{ není násobek 4} \\ \frac{1}{2}\left(X_{([\frac{n}{4}])} + X_{([\frac{n}{4}]+1)}\right) & \text{jestliže } n \text{ je násobek 4} \end{cases} \quad /49/$$

$$h_2 = \begin{cases} X_{(n+1-[\frac{n+3}{4}])} & \text{jestliže } n \text{ není násobek 4} \\ \frac{1}{2}\left(X_{(n+1-[\frac{n}{4}])} + X_{(n-[\frac{n}{4}])}\right) & \text{jestliže } n \text{ je násobek 4} \end{cases}$$ /50/

$$h = (h_2 - h_1) \qquad (\text{hinges})$$

$$c_1 = h_1 - 2(h_2 - h_1) \qquad c_2 = h_2 + 2(h_2 - h_1)$$

$$c = c_2 - c_1$$

$$t_1 = h_1 - 1.5(h_2 - h_1) \qquad t_2 = h_2 + 1.5(h_2 - h_1)$$

$$t = t_2 - t_1$$

$$z_1 = h_1 - 1(h_2 - h_1) \qquad z_2 = h_2 + 1(h_2 - h_1)$$

$$s = z_2 - z_1$$

/přeskakování je procedura, která zahrnuje vymazávání těch pozorování ležících "vně" c_1, t_1 nebo z_1, tedy máme c- (t-, s-) přeskakování/.

$_{16}T$ Andrews a kol. odhad je M-odhad pro který

$$\varphi(x) = \begin{cases} \sin \frac{x}{2.1} & \text{jestliže } |x| < 2.1\pi \\ 0 & \text{jinak.} \end{cases}$$ /51/

Rovnice tohoto M-odhadu T_n^A má tvar

$$\sum_{i=1}^{n} \varphi\left(\frac{X_{(i)} - T_n^A}{s_1}\right) = 0$$ /52/

4. Použité pravděpodobnostní modely

Srovnávání bude prováděno na následujících pravděpodobnostních modelech:

a/ Bernoulliho rozdělení

$$p_\theta(x) = \theta^x (1-\theta)^{1-x} \qquad \text{pro } x=0,1$$ /53/

$$\theta = 0.3$$

b/ Binomické rozdělení

$$p_\theta(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \qquad \text{pro } x = 0,1,\dots,n$$ /54/

$$\theta = 0.3$$

c/ Poissonovo rozdělení

$$p_\theta(x) = e^{-\theta} \frac{\theta^x}{x!} \qquad \text{pro } x = 0,1,\dots$$ /55/

$$\theta = 3$$

d/ Geometrické rozdělení

$$p_\theta(x) = \theta(1-\theta)^{x} \qquad \text{pro } x=0,1,2,\ldots \tag{56}$$

$\theta = 0.75$

e/ Johnsonovo rozdělení

$$p_\theta(x) = \frac{2}{(2\pi)^{1/2}} \frac{1}{(x-(\theta-\frac{1}{2}))(1-x+\theta-\frac{1}{2})} \exp\left\{-\frac{1}{2}\left(2\ln\frac{x-(\theta-\frac{1}{2})}{1-x+\theta-\frac{1}{2}}\right)^2\right\} \tag{57}$$

pro $\langle 0,1\rangle$

$\theta = 0.5$

f/ Normální rozdělení

$$p_\theta(x) = \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(x-\theta)^2\right\} \qquad \text{pro } x \in R^1 \tag{58}$$

$\theta = 0.3$

5. Forma experimentu

Experiment má následující formu:

1/ rozsah náhodného výběru bude $n = 20$

2/ provádí se vždy 30 opakování experimentu

3/ vypočítáme aritmetický průměr z 30 hodnot odhadů a aritmetický průměr z 30 hodnot výběrových variancí

4/ kontaminace je prováděna následovně:

symetrická:

$$X_{(1)}, \ldots, X_{(n-[n\varepsilon])}, Y_{(n-[n\varepsilon]+1)}, \ldots, Y_{(n)} \tag{59}$$

kde X je náhodná veličina s pravděpodobnostním rozdělením postupně typu a, b, c, d, e, f,

Y je náhodná veličina s rovnoměrným rozdělením na intervalu $(-10+\theta\,,\ 10+\theta\,)$, /60/

$$\varepsilon = 0.0, (= 0.05, = 0.1, = 0.25, = 0.5) \tag{61}$$

asymetrická:

$$X_{(1)}, \ldots, X_{(n-[n\varepsilon])}, Y_{(n-[n\varepsilon]+1)}, \ldots, Y_{(n)}$$

kde Y je náhodná veličina s pravděpodobnostním rozdělením na intervalu $(0,5)$, /62/

jinak je situace obdobná jako v případě symetrické kontaminace.

6. Vyhodnocení experimentu

Vyhodnocení experimentu budeme provádět následovně:

Označme si O_{ij} hodnoty odhadu ${}_iT$ za j-tého pravděpodobnostního modelu, kde i=1,...,16, j=1,...,6. Provedeme normování a vypočteme relativní chybu odhadu RE_T

$$RE_T(i) = \frac{1}{6} \sum_{j=1}^{6} \frac{|O_{ij} - \theta_j|}{\frac{1}{k} \sum_{l=1}^{k} |O_{lj} - \theta_j|}$$ i=1,2,3,4,5 pro skupinu A /k=5/ a i=1,2,6,7,...,16 pro skupinu B /k=13/. /63/

Dále si označme V_{ij} hodnoty výběrové variance odhadu $_iT$ za j-tého pravděpodobnostního modelu. Vypočteme si relativní varianci odhadu RV_T

$$RV_T(i) = \frac{1}{6} \sum_{j=1}^{6} \frac{V_{ij}}{\frac{1}{k} \sum_{l=1}^{k} V_{lj}}$$ /64/

i=1,2,...5 pro skupinu A /k=5/ a i=1,2,6,....,16 pro skupinu B /k=13/.

Označme si

$$RE_T^{(1)} = \{RE_T(1), \ldots, RE_T(5)\}$$ /65/

$$RV_T^{(1)} = \{RV_T(1), \ldots, RV_T(5)\}$$ /66/

$$RE_T^{(2)} = \{RE_T(1), RE_T(2), RE_T(6), \ldots, RE_T(16)\}$$ /67/

$$RV_T^{(2)} = \{RV_T(1), RV_T(2), RV_T(6), \ldots, RV_T(16)\}$$ /68/

a $r(\cdot)$ bude značit pořadí v posloupnosti. Množiny /65/ - /68/ postupně uspořádáme v nerostoucí posloupnosti. Každému prvku přiřadíme jeho pořadí. V případě, že prvky jsou si rovny, přiřazujeme každému aritmetický průměr jedné z permutací jejich pořadí. Odhadu $_iT$ dáváme přednost před odhadem $_lT$ pro $i \neq l$, jestliže

$$r(RE_T(i)) + r(RV_T(i)) > r(RE_T(l)) + r(RV_T(l))$$ /69/

Odhad $_iT$ je rovnocenný odhad $_lT$ pro $i \neq l$ jestliže

$$r(RE_T(i)) + r(RV_T(i)) = r(RE_T(l)) + r(RV_T(l)).$$ /70/

7. Závěr

Výsledky experimentů jsou uvedeny v tab. 1 pro robustní odhady a asymetrickou kontaminaci /pro symetrickou kontaminaci viz Simonoff K.S. /1983//. Z výsledků uvedených v tab.1 vidíme, že modifikovaný D-odhad patří do nejlepší třídy jako maximálně věrohodný odhad pro 0%-kontaminaci a v této nejlepší třídě se drží až do 25%-kontaminace. Ve třídě robustních odhadů můžeme tvrdit, že ve zkoumané třídě distribučních funkcí je použití modifikovaného D-odhadu polohy až do 25%-kontaminace spolehlivé.

Poněkud jinou situaci nám nabízejí výsledky uvedené v tab.2. Pro 0%-kontaminaci patří do nejlepší třídy $_2T$, $_6T$, $_{12}T$. Pro jakoukoli 5% kontaminaci je spolehlivé použít odhad polohy $_{12}T$ a pro jakoukoli 10% kontaminaci je spolehlivé použít odhad polohy $_6T$. Pro symetrickou kontaminaci až 10% je spolehlivé použít odhad polohy $_6T$ a pro asymetrickou kontaminaci až do 25% je spolehlivé použít odhad $_{13}T$. Z uvedeného je vidět, že modifikovaný D-odhad není konkurence schopný s některými adaptivními odhady. Kvalitativní analýza je uvedena v tab. 3 a 4.

KATEGORIE NEJLEPŠÍ BE /10 - 8/ DOBRÝ GO /7 - 5/ ŠPATNÝ BA /4 - 2/

	0%	5%	10%	25%	50%
BE	1^T 2^T	1^T 4^T	1^T 4^T	1^T 4^T	
GO		2^T		5^T	1^T 2^T 3^T 4^T 5^T
BA	3^T 4^T 5^T	3^T 5^T	2^T 3^T 5^T	2^T 3^T	

Tab. 3 Asymetrická kontaminace /robustní odhady/

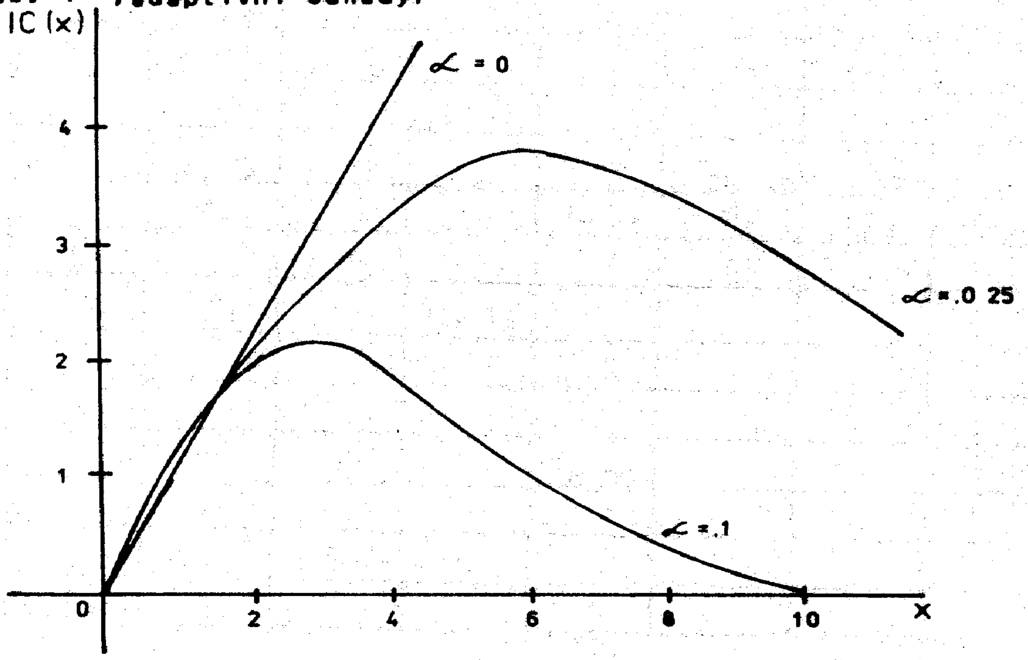
KATEGORIE BE BT GO WR BA

NEJLEPŠÍ LEPŠÍ DOBRÝ HORŠÍ ŠPATNÝ

/26 - 22/ /21 - 17/ /16 - 12/ /11 - 7/ /6 - 2/

	Symetrická kontaminace 0%	5%	10%	25%	50%	Asymetrická kontaminace 0%	5%	10%	25%	50%
BE	2^T 6^T 12^T	12^T	6^T	—	—	2^T 6^T 12^T	12^T 13^T	6^T 13^T	13^T	—
BT	1^T 13^T	1^T 6^T 10^T 13^T	13^T	6^T 13^T	1^T 2^T 8^T	1^T 13^T	6^T 11^T	1^T 10^T	1^T 9^T 10^T 16^T	1^T 2^T 6^T 10^T 11^T 16^T
GO	7^T 10^T 11^T	7^T 8^T	1^T 7^T 8^T 9^T 10^T 12^T 14^T 16^T	1^T 2^T 7^T 8^T 9^T 10^T 11^T 14^T 16^T	6^T 7^T 9^T 10^T 12^T 14^T	7^T 10^T 11^T	1^T 8^T 10^T 14^T 16^T	7^T 12^T 14^T	2^T 11^T 14^T	9^T 12^T
WR	8^T 9^T 14^T	9^T 14^T	2^T 11^T	12^T 15^T	11^T 13^T 15^T 16^T	8^T 9^T 14^T	7^T 9^T 15^T	8^T 9^T 16^T	6^T 7^T 8^T 12^T	7^T 8^T 13^T 15^T
BA	15^T 16^T	2^T 15^T	15^T			15^T 16^T	2^T	2^T 15^T	15^T	14^T

Tab. 4 /adaptivní odhady/



Obr. 1

Tab. 1 Asymetrická kontaminace /robusní odhady/

POŘADÍ	0 %	5 %	10 %	25 %	50 %
10	2^T	1^T	4^T		
9					
8	1^T	4^T	1^T	1^T 4^T	
7				5^T	
6		2^T			1^T 2^T 3^T 4^T 5^T
5					
4	4^T		2^T	2^T	
3	3^T	3^T 5^T	5^T	3^T	
2			3^T		

	Symetrická kontaminace					Asymetrická kontaminace				
POŘADÍ	U (<– 10 + θ, 10 + θ >)					U (< 0,5 >)				
	0 %	5 %	10 %	25 %	50 %	0 %	5 %	10 %	25 %	50 %
26		12^T					12^T			
25								13^T		
24	2^T 6^T					2^T 6^T	13^T	6^T		
23	12^T		6^T			12^T			13^T	
22										
21		13^T	13^T				6^T			10^T 11^T
20	1^T	6^T			1^T				10^T	
19						1^T		10^T 11^T		
18	13^T	10^T		13^T	2^T	13^T	11^T		1^T 16^T	16^T
17	10^T	1^T		6^T	8^T				9^T	1^T 6^T 2^T
16			10^T		6^T					9^T
15			1^T 7^T	16^T	7^T 9^T 10^T	10^T	1^T	1^T	11^T 14^T	
14			12^T 16^T	1^T 2^T 8^T 9^T 11^T 14^T		11^T		7^T	2^T	
13		7^T 8^T 11^T	8^T 9^T	7^T 10^T	12^T			12^T 14^T		
12	11^T	16^T	14^T		14^T		8^T 10^T 14^T 16^T			
11	9^T 14^T	9^T	11^T	12^T	13^T 15^T			16^T	6^T 12^T	7^T 12^T 8^T 15^T
10			2^T	15^T	16^T	9^T 14^T	7^T	9^T	7^T	
9	7^T 8^T				11^T			8^T		13^T
8							9^T 15^T		8^T	
7		14^T				7^T 8^T				
6		2^T								
5	15^T 16^T	15^T	15^T			15^T 16^T		2^T		
4										
3							2^T	15^T		
2									15^T	14^T

Tab.2 Symetrická a asymetrická kontaminace /adaptivní odhady/

8. Odkazy

|1| Andrews D.F. et al. /1972/, Robust Estimates of Location, Princeton, New Jersey, Princeton University Press

|2| David H.A. /1968/, Order Statistics, New York, Wiley

|3| Downton F. /1966/, Linear Estimates with Polynomial Coefficients, Biometrika 53, 129-41

|4| Gini C. /1912/, Variabilita e Mutabilita, Anno 3, Part 2, pp. 88, Studi Economico-Giaridici della R. Universita de Cagliari

|5| Hampel F.R. /1974/, The Influence Curve and its Role in Robust Estimation, JASA 69, pp. 383-93

|6| Hogg V. /1974/, Adaptive Robust Procedures: A partial Review and some Suggestions for Future,Applications and Theory, JASA, 1969, 909-23

|7| Huber P.I. /1972/, Robust Statistics: A review, Ann. Math. Statist. 43, 1041-67

|8| Johnson N.L. /1949/, Systems of Frequency Curve Generated by Methods of Translation, Biometrika 36, 149

|9| Rao C.R. /1961/, Asymptotic Efficiency and Limiting Information, Proc. Fourth Berkeley Symposium, University of California, 1, 531-46

|10| Simonoff J.S. /1983/,A comparison of robust methods and detection of outliers techniques when estimating a location parameter, Proc. 14th Symp. Interface, New York

|11| Vajda I. /1983/,Minimum Divergence Principle in Statistical Estimation, Statistics and Decisions, 1

|12| Vošvrda M. /1983/, On Second Order Efficiency of Minimum Divergence Estimators, Trans. 9th Prague Conf.,Prague