# NĚKOLIK POZNÁMEK KE KONZISTENCI M-ODHADŮ

Jana Jurečková, MFF UK Praha

## Úvod

V literatuře bývá doporučována řada M-odhadů parametru polohy a regrese vytvořených různými váhovými funkcemi. Zdůrazňují se některé jejich vlastnosti, hlavně robustnost vzhledem k odchylkám od předpokládaného rozdělení pravděpodobností, ale na druhé straně jejich konzistence bývá považována jaksi za samozřejmost. Že tomu tak nemusí být, ukazují např. Freedman a Diaconis (1982), kteří dokazují, že M-odhad nemusí být konzistentní, pokud příslušná $\rho$-funkce není konvexní a rozdělení chyb není unimodální.

## I. Model polohy

Nechť $X_1, X_2, \ldots$ je posloupnost nezávislých pozorování z populace s distribuční funkcí $F(x-\theta)$; distribuční funkce $F$ je obecně neznámá, pouze předpokládáme, že patří do třídy $\mathcal{F}$ distribučních funkcí symetrických kolem 0. Chceme odhadnout parametr $\theta$ na základě pozorování $X_1, \ldots, X_n$.

Jednu z rozsáhlých tříd robustních odhadů parametru polohy tvoří M-odhady zavedené Huberem (1964). Definujme M-odhad $M_n$ parametru $\theta$ jako globální minimum statistiky

$$(1.1) \qquad R_n(t) = \sum_{i=1}^{n} \rho\left(\frac{X_i - t}{ks_n}\right) = \min$$

vzhledem k $t \in R^1$, kde $s_n = s_n(X_1, \ldots, X_n)$ je škálová statistika vyhovující podmínkám

$$(1.2) \qquad \begin{cases} s_n(X_1 + c, \ldots, X_n + c) = s_n(X_1, \ldots, X_n), \\ s_n(cX_1, \ldots, cX_n) = c\, s_n(X_1, \ldots, X_n), \quad c \in R^1 \end{cases}$$

a $k>0$ je volitelná konstanta. O statistice $s_n$ dále předpokládáme, že

$$(1.3) \qquad s_n \xrightarrow{P} \sigma = \sigma(F) > 0 \quad \text{při} \quad n \to \infty ,$$

kde $\sigma(F)$ je nějaký kladný funkcionál distribuční funkce $F$. Řada autorů doporučuje volit konstantu $k$ tak, aby získaný odhad byl vydatný pro nějakou speciální distribuční funkci $F$, většinou pro normální rozdělení.

Odhad definovaný pomocí (1.1) je **ekvivariantní** vzhledem ke změně měřítka, tj.

(1.4) $M_n(cX_1, \ldots, cX_n) = cM_n(X_1, \ldots, X_n), \quad c>0$ .

Často se též volí $s = 1$; pak posloupnost $s_n$ ovšem nevyhovuje(1.2) a výsledný odhad není ekvivariantní ve smyslu (1.4).Nejčastější robustní volbou $s_n$ je buď **mezikvartilová odchylka**

(1.5) $s_n = X_{n:[3n/4]} - X_{n:[n/4]}$ ,

kde $X_{n:1} \le \ldots X_{n:n}$ jsou pořádkové statistiky příslušné $X_1, \ldots, X_n$ (pak $\sigma(F) = F^{-1}(3/4)-F^{-1}(1/4)$ ) nebo **mediánová absolutní odchylka** (MAD)

(1.6) $s_n = \underset{1\le i\le n}{\mathrm{med}} \left| X_i - \tilde{X}_n \right|$ ,

kde $\tilde{X}_n$ je medián $X_1,\ldots,X_n$ (pak $\sigma(F) = F^{-1}(3/4)$ pro symetrickou F).

Jestliže $\rho$ je konvexní funkce symetrická kolem 0, a tedy $\psi(x) = \frac{d\rho}{dx}$ je neklesající antisymetrická funkce, lze M-odhad určit jednoznačně ve tvaru

$$(1.7)\quad \begin{cases} M_n = M_n^- + M_n^+ , \\ M_n^- = \sup \left\{ t : \sum_{i=1}^{n} \psi\left( \frac{x_i - t}{k\, s_n} \right) > 0 \right\} , \\ M_n^+ = \inf \left\{ t : \sum_{i=1}^{n} \psi\left( \frac{x_i - t}{k\, s_n} \right) > 0 \right\} . \end{cases}$$

Jestliže navíc $\int \rho(x-t)\, dF(x)$ má jediné minimum v $t=0$, pak $M_n$ konverguje k $\theta$ v pravděpodobnosti skoro jistě při $n\to\infty$ (Huber(1981)). V téže Huberově knize můžeme najít i obecnější podmínky pro konzistenci posloupnosti $M_n$. Speciálně, jestliže $\rho$ není nutně konvexní, ale je to absolutně spojitá, zdola ohraničená nekonstantní funkce symetrická kolem 0 a taková, že je neklesající na $[0,\infty)$, pak pro konzistenci $M_n$ stačí, že F má silně unimodální hustotu f (tj. f je rostoucí pro $x<0$ a klesající pro $x>0$, viz např. Freedman a Diaconis(1981, 1982)).

Nyní uvažujme, co se stane, když funkce

$$(1.8) \qquad h(t) = \int \rho\,(x-t)\,dF(x)$$

nemá globální minimum v bodě $t=0$.

Pak platí následující tvrzení Freedman a Diaconis (1981):

**Tvrzení 1.** Nechť funkce $\rho$ je symetrická kolem 0 a má ohraničenou spojitou derivaci $\psi$. Nechť $X_1, X_2, \ldots$ jsou nezávislá pozorování se společnou distribuční funkcí $F(x-\theta)$, kde $F$ je spojitá, symetrická kolem 0 a taková, že funkce $h(t)$, definovaná v(1.8), je ohraničená a **nemá** globální minimum v bodě $t=0$. Pak existuje $\varepsilon>0$ tak, že s pravděpodobností 1 statistika

$$(1.9) \qquad R_n(t) = \sum_{i=1}^{n} \rho(x_i-t)$$

pro dostatečně velká $n$ nemá své globální minimum v intervalu$[\theta-\varepsilon, \theta+\varepsilon\,]$. Speciálně, M-odhad definovaný jako globální minimum $R_n(t)$ není konzistentním odhadem $\theta$.

**Důkaz:**

Bez újmy na obecnosti můžeme položit $\theta = 0$. Pak můžeme psát

$$\begin{aligned} n^{-1} R_n(t) &= h(t) + \int \rho(x-t)\, d(\,F_n(x)-F(x)\,) \\ &= h(t) - \int (F_n(x)-F(x))\, \psi(x-t)\, dx \end{aligned}$$

kde $F_n(x)$ je empirická distribuční funkce příslušná $X_1, \ldots, X_n$ , tj.

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I\,[x_i \le x] \quad .$$

Pak tedy k libovolnému $\eta>0$ existuje $n_0$ a pro $n \ge n_o$ platí

$$(1.10) \qquad |\; n^{-1} R_n(t) - h(t) \;| \le \eta \;.$$

Protože $h$ nenabývá globálního minima v bodě $t=0$, existuje $t_0 \ne 0$ tak, že $h(t)>h(t_0)$. Zvolme $\varepsilon>0$ a $\eta>0$ tak, že $h(t) \ge h(t_0)+3\eta$ pro$|t| \le \varepsilon$ . Pak pro $|t| \le \varepsilon$ plyne z (1.10)

$$n^{-1} R_n(t) \ge h(t_0) + 2\eta$$

a

$$n^{-1} R_n(t) \le h(t_0) + \eta$$

a tedy $n^{-1} R_n(t) \ge n^{-1} R_n(t_0) + \eta$ pro $t$ , odkud plyne tvrzení.

Řada autorů (viz např. Hampel. Ronchetti, Rousseeuw a Stahel (1985) ) doporučují užívat tzv. **redescentní M-odhady**, jímž příslušná funkce $\psi$ (= $\rho$ ) je buď rovna 0 pro $|X|\geq c$ nebo konverguje k 0 pro x . Hlavním argumentem pro použití těchto funkcí je, že zcela potlačují vliv odlehlých pozorování na výsledný odhad. Takováto funkce $\psi$ samozřejmě není neklesající a příslušná $\rho$ není konvexní, a mohou tedy vést k M-odhadům, které nejsou konzistentní pro některá rozdělení pravděpodobností.

Uveďme si některé příklady.

$$\text{(a)} \quad \begin{cases} \rho(x) = \log(1+x^2) \\ \psi(x) = 2x/(1+x^2) \end{cases}$$

(Funkce $\rho$ je věrohodnostní funkcí Cauchyho rozdělení.)

$$\text{(b)} \quad \begin{cases} \rho(x) = \begin{cases} -(1-x^2)^3 & \text{pro} \quad |x| \leq 1 \\ 0 & \phantom{\text{pro}} \quad |x| > 1 \end{cases} \\ \psi(x) = \begin{cases} 6x\,(1-x^2)^2 & \text{pro} \quad |x| \leq 1 \\ 0 & \phantom{\text{pro}} \quad |x| > 1 \end{cases} \end{cases}$$

(tzv. Tukeyho biweight funkce).

$$\text{(c)} \quad \begin{cases} \rho(x) = \begin{cases} -(1-x^2)^2 & \text{pro} \quad |x| \leq 1 \\ 0 & \phantom{\text{pro}} \quad |x| > 1 \end{cases} \\ \psi(x) = \begin{cases} 2x\,(1-x^2) & \text{pro} \quad |x| \leq 1 \\ 0 & \phantom{\text{pro}} \quad |x| > 1 \end{cases} \end{cases}$$

$$\text{(d)} \quad \begin{cases} \rho(x) = \begin{cases} x^2/2 & \text{pro} \quad |x| \leq 1 \\ 1/2 & \phantom{\text{pro}} \quad |x| > 1 \end{cases} \\ \psi(x) = \begin{cases} x & \text{pro} \quad |x| \leq 1 \\ 0 & \phantom{\text{pro}} \quad |x| > 1 \end{cases} \end{cases}$$

(M-odhad, vytvořený těmito funkcemi, se v knize Andrews a kol. (1972) nazývá "skipped mean").

$$(e)\quad \begin{cases} \rho(x) = \begin{cases} |x| & \text{pro } |x| \le 1 \\ 1 & |x| > 1 \end{cases} \\ \psi(x) = \begin{cases} \operatorname{sign}(x) & \text{pro } |x| \le 1 \\ 0 & |x| > 1 \end{cases} \end{cases}$$

(příslušný odhad je v literatuře nazýván "skipped median").

Právě k některé z těchto funkcí může existovat distribuční funkce F, absolutně spojitá a symetrická, ale taková, že funkce(1.8) nabývá v bodě t=0 nikoli minima, ale maxima. Pro populaci s distribuční funkcí $F(x-\theta)$ není globální minimum funkce R(t)

(1.9) obecně konzistentním odhadem $\theta$. Někteří autoři doporučují definovat jako odhad parametru $\theta$ to řešení rovnice $\sum_{i=1}^{n} \psi(X_i - t) = 0$, které je nejblíže nějakému konzistentnímu počátečnímu odhadu,např.mediánu. V situaci, kterou jsme naznačili, by takový odhadmohl být i konzistentním odhadem, ale konvergujícím k bodu maximafunkce h(t) (speciálně, takový odhad by odpovídal "minimálněvěrohodnému odhadu").

Freedman a Diaconis (1981, 1982) ukázali, že takové distribučnífunkce (a hustoty) skutečně mohou existovat. Příklady nalezlimezi multimodálními hustotami s ohraničenými nosičem. Popišmeněkteré z jejich příkladů.

$$(i)\quad \begin{cases} \rho(x) = 1 + x^2, \; x \in R^1 \\ \qquad\qquad \text{(věrohodnostní funkce Cauchyho rozdělení)} \\ \text{a} \quad \rho_k(x) = \rho\left(\frac{x}{k}\right) \end{cases}$$

$$\rho''(x) = (1-x^2)/(1+x^2)^2$$

Položme $x = \sqrt{\sqrt{32} - 5} \approx .81$; zvolme libovolné k z intervalu $(1, 1/x_0)$ a $x_1$ z intervalu $(x_0, 1/k)$. Nechť Z je náhodnáveličina, která nabývá hodnot $\pm kx_1$ a $\pm k\sqrt{3}$, každé s pravděpodobností 1/4. Pak

$$E\,\rho_k(Z) = E\left\{ k^{-2}\rho''(Z/k) \right\} < 0$$

a funkce E nabývá maxima v bodě t=0 .

Nechť nyní Y je náhodná veličina se symetrickou hladkou hustotou g, kladnou na (-1,1) a rovnou 0 jinak. Pak, pro dostatečně velká m, náhodná veličina $Z_m^* = \frac{Z}{m} + \frac{Y}{m}$ má hladkou symetrickou multimodální hustotu na nosiči [-2,2] a funkce $E\left\{\rho_k(Z_m^*-t)\right\}$ nabývá maxima v bodě t=0. Podle tvrzení 1 odtud plyne, že mají-li $X_1-\theta$, $X_2-\theta$, ... stejné rozdělení jako $Z_m$ je M-odhad vytvořený funkcí $\rho_k$ nekonzistentní.

V tomto případě můžeme dokonce dokázat:

(i) Rovnice $\sum_{i=1}^{n} \psi_k(X_i-t) = 0$ má tři kořeny, $M_{-n}$, $M_{0n}$, $M_{+n}$. Výraz $R_n(t) = \sum_{i=1}^{n} \rho_k(X_i-t)$ má lokální maximum v $t=M_{0m}$ a lokální minima v bodě $t=M_{+n}$, $M_{-n}$. Jeden z bodů $M_{+n}$, $M_{-n}$ je zároveň globálním minimem, a tedy M-odhadem $\theta$. Existuje $\gamma = \gamma(F) > 0$ tak, že $M_{-n} \to -\gamma$, $M_{0n} \to 0$ a $M_{+n} \to \gamma$ s pravděpodobností 1 při $n \to \infty$. Dále existují podposloupnosti $n_{+j}$ a $n_{-j}$ přirozených čísel tak, že $R_{n_{-j}}(t)$ má globální minimum v $t=M_{+n_{+j}}$ a $R_{n_{-j}}(t)$ má globální minimum v $t=M_{-n_{-j}}$. Tedy M-odhad nekonečně mnohokrát osciluje mezi $-\gamma$ a $\gamma$ a nemůže být silně konzistentní.

(ii) Uvažujme Tukeyho funkci (b). Pak, podobně jako v předcházejícím příkladě, položíme-li $x_0 = \sqrt{3 - \sqrt{8}} \cong 0.185$ a zvolíme k v intervalu $(\sqrt{5/3}), 1/x_0)$, můžeme najít hladkou symetrickou multimodální hustotu s ohraničeným nosičem tak, že odhad vytvořený funkcí $\rho_k$ je nekonzistentní. Z numerických hodnot se ukazuje, že pro $k>1/x_0 \cong 5.41$ je M-odhad patrně jižkonzistentní (Tukey navrhuje volit k=6).

(iii) Uvažujme funkci $\rho$ z příkladu (c) a libovolné $k>1$. Pakexistuje symetrická hladká hustota na ohraničeném nosiči taková,že řídí-li se pozorování touto hustotou, je M-odhad nekonzistentní.

Z těchto konstrukcí nemůžeme vyslovit jednoznačný závěr, spíše několik **poznámek**:

M-odhad je konstruovaný tak, aby omezil vliv odlehlých pozorování a rozdělení s těžkými chvosty. Použijeme-li však redescentních M-odhadů, riskujeme, že odhad nebude konzistentní v případě rozdělení s lehkými chvosty, ke kterým patří useknutá rozdělenía jiná rozdělení s ohraničeným nosičem. Pokud z povahy měření dat víme, že rozdělení je useknuté, jsme opatrní při použití redescentních M-odhadů. Vzhledem k uvedeným protipříkladům je vhodné nejprve porovnat výběrové rozpětí s výběrovou mediánovou odchylkou: pokud je rozpětí menší než čtyřnásobek mediánové odchylky, nepoužijeme Cauchyho věrohodnostní funkci s hodnotou blízkou 1; a funkci $\rho$ z příkladu (b) použijeme jen pro $k \geq 6$.

## II. Lineární regresní model

Uvažujme klasický lineární regresní model

$$\underset{\sim}{y} = \underset{\sim}{X}\,\underset{\sim}{\beta} + \underset{\sim}{e}$$

kde $\underset{\sim}{y} = (Y_1,\dots,Y_n)'$ je vektor pozorování,

$\underset{\sim}{X} = \underset{\sim}{X}_n$ je daná regresní matice řádu n.p.

$\underset{\sim}{\beta} = (\beta_1,\dots,\beta_n)'$ je parametr

a $\underset{\sim}{e} = (e_1,\dots,e_n)'$ je vektor nezávislých chyb, stejně rozdělených s distribuční funkcí F.

M-odhad (nestudentizovaný) definujeme jako řešení minimalizace

$$(2.1) \qquad \sum_{i=1}^{n} \rho(Y_i - \underset{\sim}{x}_i'\,\underset{\sim}{t}) \longrightarrow \min, \quad \underset{\sim}{t} \in R^p ,$$

kde $\underset{\sim}{x}_i'$ je i-tý řádek matice $\underset{\sim}{X}$. Jestliže $\rho$ je konvexní funkce, je (2.1) ekvivalentní řešení soustavy rovnic

$$(2.2) \qquad \sum_{i=1}^{n} \underset{\sim}{x}_i'\,\psi(Y_i - \underset{\sim}{x}_i'\,\underset{\sim}{t}) = 0, \qquad \psi = \rho' .$$

Jestliže obecně $\rho$ není konvexní a funkce $h(t)=\int\rho(x-t)dF(x)$ nemá minimum v bodě $t=0$, můžeme vyslovit tytéž závěry jako u modelu polohy, a tedy pro některá rozdělení chyb může být řešení (2.1) nekonzistentním odhadem $\underset{\sim}{\beta}$.

Všimněme si tedy podrobněji situace, kdy funkce $\rho$ a rozdělení F jsou takové, že $\rho$ je absolutně spojitá s derivací $\psi$ a platí

$$(2.3) \qquad \gamma = \int \psi'(x)\, dF(x) > 0 .$$

Za tohoto předpokladu a za některých dalších podmínek regularity na posloupnost $\{X\}$ existuje řešení soustavy (2.2), které je konzistentním odhadem $\underset{\sim}{\beta}$ a pro dostečné velká n ,je řešením minimalizace (2.1). Jestliže navíc je $\rho$ konvexní funkcí, je toto řešení určeno jednoznačně. Tímto případem se zabýval Portnoy (1984) a za více omezujících podmínek Yohai a Maronna (1979); tito autoři dokonce připouštěli situaci, že $p_n \to \infty$ při $n \to \infty$. Portnoy ve svém důkazu zajímavě použil vět z obecné teorie nelineárních rovnic, založených na větách o pevném bodě (viz Ortega a Rheinholdt (1970)). Podstatné pro použití těchto vět je, že levá strana rovnice (2.2) je spojitou funkcí t.

Nevyřešenou dosud zůstává otázka M-odhadu, vytvořeného funkcemi $\rho$ z příkladu (d) a (e) (skipped mean a skipped median u parametru polohy) a dalšími funkcemi, které nemají absolutně spojitou derivaci. Za předpokladu, že funkce $\rho$ je zdola ohraničená a funkce $h(t) = \int \rho\,(x-t)\,dF(x)$ je ryze konvexní v bodě t=0, studovala tuto otázku Jurečková (1988). Za určitých podmínek regularity na posloupnost $\{ \underset{\sim}{X}_n \}$ a na chování distribuční funkce v okolí bodů nespojitosti funkce $\psi$ ukázala, že minimalizace (2.1) je asymptoticky ekvivalentní mimimalizaci konvexní kvadratické funkce, a že tedy existuje M-odhad, konzistentní vzhledem ke konvergenci v pravděpodobnosti a asymptoticky normální. Podstatný je však předpoklad, že h(t) je ryze konvexní v bodě t=0 který, stejně jako v části I, nelze zaručit při neznámé distribuční funkci F.

**Literatura**

Andrews,D.F.-Bickel,P.J.-Hampel,F.R.-Huber,P.J.-Rogers, W.H.-Tukey,J.W.(1972): Robust Estimates of Location. Survey and Advances. Princeton University Press.

Freedman,D.A.-Diaconis,P.(1981): On inconsistent M-estimates. Technical Report No.170, Department of Statistics, Stanford University.

Freedman,D.A.-Diaconis,P.(1982): On inconsistent M-estimators. Ann.Statist.10, 454-461.

Hampel,F.R.-Ronchetti,E.M.-Rousseeuw,P.J.-Stahel,W.A.(1985): Robust Statistics. The Approach based on Influence Functions. J.Wiley, New York.

Huber,P.J.(1964): Robust estimation of a location parameter. Ann.Math.Statist.35,73-101.

Huber,P.J.(1981): Robust Statistics. J.Wiley, New York.

Jureckova,J.(1989): Consistency of M-estimators in linear model generated by non-monotone and discontinuous $\rho$-functions. Ann. Math.Statist. 10, 1-10.

Ortega,J.M.-Rheinboldt,W.C.(1970): Iterative Solution of Nonlinear equations in Several Variables. Academic Press, new York.

Portnoy,S.(1984): Asymptotic behavior if M-estimators of p regression parameters when $p^2/n$ is large. I. Inconsistency. Ann.Statist.12, 1298-1309.

Yohai,V.J.-Maronna,R.A.(1979): Asyptotic behavior of M-estimators for the linear model. Ann.Statist.7, 258-268.