KONSISTENCE M-ODHADŮ

Jana Jurečková, MFF UK Praha

I. Model polohy.

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

$$(1.1) \quad R_n(t) = \sum_{i=1}^{n} \varrho\left(\frac{X_i - t}{k s_n}\right) = \min$$

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

$$(1.2) \quad s_n(X_1+c, \ldots, X_n+c) = s_n(X_1, \ldots, X_n)+c$$

$$s_n(cX_1, \ldots, cX_n) = c\, s_n(X_1, \ldots, X_n) \quad , \quad c \in R^1$$

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

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

kde $\sigma(F)$ je nějaký kladný funkcionál d.f.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 <u>ekvivariantní</u> vzhledem ke změně měřítka, tj.

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

Často se též volí $s_n \equiv 1$; pak posloupnost s_n ovšem nevyhovuje (1.2) a výsledný odhad není ekvivariantní ve smyslu (1.4). Nejčastější robustní volbou s_n je buď <u>mezikvartilová odchylka</u>

$$(1.5) \quad s_n = X_{n:[\frac{3}{4}n]} - X_{n:[\frac{1}{4}n]}$$

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

$$(1.6) \quad s_n = \operatorname*{med}_{1 \leq i \leq n} |X_i - \tilde{X}_n|$$

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

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

$$M_n = M_n^- + M_n^+$$

$$(1.7) \quad M_n^- = \sup\left\{t : \sum_{i=1}^{n} \psi\left(\frac{X_i - t}{k s_n}\right) > 0\right\}$$

$$M_n^+ = \inf\left\{t : \sum_{i=1}^{n} \psi\left(\frac{X_i - t}{k s_n}\right) < 0\right\}.$$

Jestliže navíc $\int \varrho(x-t)dF(x)$ má jediné minimum v $t=0$, pak M_n konverguje k θ v pravděpodobnosti i skoro jistě při $n \to \infty$ (Huber (1981)). V téže Huberově knize můžeme najít i obecnější podmínky pro konsistenci posloupnosti M_n. Speciálně, jestliže ρ 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 konsistenci 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 and Diaconis (1981,1982)] .

Nyní uvažujme, co se stane, když funkce

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

nemá globální minimum v bodě t=0. Pak platí následující tvrzení (Freedman and Diaconis (1982)):

Tvrzení 1. Nechť funkce ρ je symetrická kolem 0 a má ohraničenou spojitou derivaci ψ. 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) $R_n(t) = \sum_{i=1}^{n} \rho(X_i - t)$

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 konsistentním odhadem θ.

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

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

kde $F_n(x)$ je empirická distribuční funkce příslušná $X_1,\ldots,X_n$, tj. $F_n(x) = n^{-1}\sum_{i=1}^{n} I\,[X_i \leq x]$.

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

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

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

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

a

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

a tedy

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

pro $|t| \leq \varepsilon$, odkud plyne tvrzení.

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

Uveďme si některé příklady.

(a) $\rho(x) = \log\,(1+x)^2$

$\psi(x) = 2x/(1+x)^2$

(Funkce ρ je věrohodnostní funkcí Cauchyho rozdělení).

(b)

$$\rho(x) = \begin{cases} -(1-x^2)^3 & \ldots \quad |x| \leq 1 \\ 0 & \ldots \quad |x| > 1 \end{cases}$$

$$\psi(x) = \begin{cases} 6x(1-x^2)^2 & \ldots \quad |x| \leq 1 \\ 0 & \ldots \quad |x| > 1 \end{cases}$$

(tzv. Tukey biweight)

(c)

$$\rho(x) = \begin{cases} -(1-x^2)^2 & \ldots \quad |x| \leq 1 \\ 0 & \ldots \quad |x| > 1 \end{cases}$$

$$\psi(x) = \begin{cases} 4x(1-x^2) & \ldots \quad |x| \leq 1 \\ 0 & \ldots \quad |x| > 1 \end{cases}$$

(d)

$$\rho(x) = \begin{cases} x^2 & \ldots \quad |x| \leq 1 \\ 1 & \ldots \quad |x| > 1 \end{cases}$$

$$\psi(x) = \begin{cases} x & \ldots \quad |x| \leq 1 \\ 0 & \ldots \quad |x| > 1. \end{cases}$$

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

(e)

$$\rho(x) = \begin{cases} |x| & \ldots \quad |x| \leq 1 \\ 1 & \ldots \quad |x| > 1 \end{cases}$$

$$\psi(x) = \begin{cases} \operatorname{sign} x & \ldots \quad |x| \leq 1 \\ 0 & \ldots \quad |x| > 1, \end{cases}$$

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. Pocházejí-li pozorování z rozdělení $F(x-\theta)$, znamená to, že globální minimum $R_n(t)$ (1.9) není konsistentním odhadem θ. Někteří autoři doporučují jako odhad parametru θ to řešení rovnice $\sum_{i=1}^{n} \psi(x_i - t) = 0$, které je nejblíže nějakému konsistentnímu počátečnímu odhadu, např. mediánu. V situaci, kterou jsme naznačili, by takový odhad i mohl být konsistentním odhadem, ale konvergujícím k bodu maxima funkce $h(t)$(speciálně, u funkce(a) bychom tedy dostali "minimálně věrohodný odhad".

Friedman a Diaconis (1981, 1982) ukázali, že takové distribuční funkce (a hustoty) skutečně mohou existovat. Příklady nalezli mezi multimodálními hustotami s ohraničeným nosičem. Popišme některé z jejich příkladů.

(i) $\rho(x) = 1+x^2$, $x \in R^1$(věrohodnostní funkce Cauchyho rozdělení) a $\rho_k(x) = \rho(\frac{x}{k})$.
Pak

$$\psi^{*}(x) = (1-x^2)/(1+x^2)^2$$

Položme $x_0 = \sqrt{\sqrt{32} - 5} \doteq 0.81$; zvolme libovolné k z intervalu $(1, 1/x_0)$ a x_1 z intervalu $(x_0, 1/k)$. Nechť Z je náhodná veličina, která nabývá hodnot $\pm k x_1$ 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 $E\left\{\rho_k(Z-t)\right\}$ nabývá maxima v bodě $t=0$.

Nechť nyní Y je náhodná veličina ze 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 $X_1-\theta$, $X_2-\theta, \ldots$ stejné rozdělení jako Z_m^{*}, je M-odhad vytvořený funkcí ρ_k nekonsistentní. V tomto případě můžeme dokonce dokázat:

a) Rovnice $\sum_{i=1}^{n} \psi_k(X_i - t) = 0$ má tři kořeny, M_{-n} M_{on} M_{+n}. Výraz $R_n(t) = \sum_{i=1}^{n} \rho_k(X_i - t)$ má lokální maximum v $t = M_{on}$ 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 θ. Existuje $\gamma = \gamma(F) > 0$, tak, že $M_{-n} \longrightarrow -\gamma$, $M_{on} \longrightarrow 0$ a $M_{+n} \longrightarrow \gamma$ 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 γ a nemůže být silně konsistentní.

(ii) Uvažujme Tukeyho funkci (b). Pak, podobně jako v předcházejícím příkladě, položíme-li $x_0 = \sqrt{3 - \sqrt{8}} / \sqrt{5} \doteq 0{,}185$ 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 ρ_k je nekonsistentní. Z numerických hodnot se ukazuje, že pro $k > 1/x_0 \doteq 5.41$ je M-odhad patrně již konsistentní (Tukey navrhuje volit $k=6$).

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

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 konsistentní v případě rozdělení s lehkými chvosty, ke kterým typicky patří useknutá rozdělení a 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 k blízkou 1; a funkci ρ z příkladu(b) použijeme jen pro $k \geq 6$.

II. Lineární regresní model

Uvažujme klasický lineární regresní model

$$\underset{\sim}{Y} = \underset{\sim}{X}\beta + \underset{\sim}{e}$$

kde $\underset{\sim}{Y} = (Y_1,\ldots,Y_n)'$ je vektor pozorování, $\underset{\sim}{X} = \underset{\sim}{X}_n$ je daná regresní matice řádu $n \times p$, $\underset{\sim}{\beta} = (\beta_1,\ldots,\beta_p)'$ je parametr a $\underset{\sim}{e} = (e_1,\ldots,e_n)'$ je vektor nezávislých chyb, stejně rozdělených s distribuční funkcí F. M-odhad (nestudentizovaný) definujeme jako řešení minimalizace

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

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

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

Jestliže obecně ϱ není konvexní a funkce $h(t) = \int \varrho(x-t)dF(x)$ 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) nekonsistentním odhadem $\underset{\sim}{\beta}$.

Všimněme si tedy podrobněji situace, kdy funkce ϱ a rozdělení F jsou takové, že ϱ je absolutně spojitá s derivací ψ a platí

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

Za tohoto předpokladu a za některých dalších podmínek regularity na posloupnost $\{\underset{\sim}{X}_n\}$ existuje řešení soustavy (2.2), které je konsistentním odhadem $\underset{\sim}{\beta}$ a pro dostatečně velká n je řešením minimalizace (2.1). Jestliže navíc je ϱ 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 Rheinboldt(1970)). Podstatné pro použití těchto vět je, že levá strana rovnice (2.2) je spojitou funkcí $\underset{\sim}{t}$.

Nevyřešenou dosud zůstává otázka M-odhadu, vytvořeného funkcemi ϱ z příkladů (d) a (e) (skipped mean a skipped median u parametru polohy)a dalšími funkcemi, které nemají absolutně spojitou derivaci. Za předpokladu, že funkce ϱ je zdola ohraničená a funkce $h(t) = \int \varrho(x-t)dF(x)$ 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 ψ ukázala, že minimalizace (2.1) je asymptoticky ekvivalentní minimalizaci konvexní kvadratické funkce, a že tedy existuje M-odhad, konsistentní 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. and Tukey,J.W.(1972). Robust Estimates of Location. Survey and Advances. Princeton University Press.

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

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

Hampel,F.R.,Ronchetti,E.M.,Rousseeuw,P.J. and 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.

Jurečková,J.(1988). Consistency of M-estimators in linear model generated by non-monotone and discontinuous ϱ -functions. Probability and Math.Statist.X.

Ortega,J.M. and Rheinboldt, W.C.(1970). Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York.

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

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