ROBUST'98, 179 – 185

POZNÁMKY K CHOVÁNÍ M - ODHADŮ REGRESNÍCH PARAMETRŮ ZA NEOBVYKLÝCH HUSTOT

Jan SVATOŠ

MFF UK, KPMS

Abstract. Asymptotic distribution for M-estimators in regression models is well known for regular types of densities, when $f'(x)$ exists in support of f and when support of f is an open set. My short articlen tries to show both the asymptotic representation and distribution of M-estimators in cases when such regularity conditions are broken. Resulting distribution of $\hat{\boldsymbol{\beta}}$ is no longer normal, when derivative of loss function, ψ, is discontinuous and the discontinuity points are the same as singular points of density. For location case, asymptotics for nonregular density cases were studied at the end of 70. and beginning of 80. years. Results for regression vectors have something in common with results of those studies, but the multidimensionality of $\boldsymbol{\beta}$ complicates the technical matters considerably.

Резюме: В случаях регулярных плотностей асимптотика M естимации для вектора линейной регрессии хорошо знакома и разработана. В работах, которые интересуются етой проблематикой, учитывают, что плотность вектора E в модели линейной регрессии $Y = X\beta + E$ регулярна, что значит, что существует f' и что носитель f связанное множество. Ситуация, в которой не существует $f'(x)$ или $f(x) = 0$ в некоторых изолированных пунктах, была разработана на переломе 70-х и 80-х годов для параметра локации. Моя работа рассматривает асимптотику M естиматоров в таких случаях. В таких нерегулярных случаях можно показать, что не сам $\hat{\boldsymbol{\beta}}$, но его функция следует (асимптотически) распределение $N(\mathbf{0}, \mathbf{V})$ для некоторой матрицы $\mathbf{V}$.

1. Úvod

V příspěvku budou ukázány výsledky vedoucí k asymptotické reprezentaci M-odhadů regresních parametrů za předpokladu porušení jedné z podmínek kladených obvykle na hustotu chybových složek modelu. Jedná se o porušení podmínky nenulovosti hustoty ve speciálních bodech, odpovídajících bodům, ve kterých ztrátová funkce, obvykle označovaná jako ϱ, není hladká. Podobné úlohy byly řešeny v modelu polohy, kdy cílem několika prací bylo získat poznatky o chování výběrového kvantilu $F_n^{-1}(p)$ pro případ, že $f(F^{-1}(p)) = 0$. Výsledky poskytují např. články autorů Ghoshe a Sukhatmeho (1981) nebo, de Haana a Taconis–Haantjese (1979). Podobnou úlohou se zabývala také Jurečková (1983). V případě odhadů parametru polohy je tedy známo, jaké asymptotické rozdělení vzniká za situací, kdy hustota uvažovaného rozdělení porušuje některé předpoklady (např. je porušen předpoklad existence derivace hustoty nebo nosičem hustoty není konvexní množina). Výsledky prací, zabývajících se tímto problémem, ukazují, že ne odhad samotný, ale jeho určitá mocnina, závisející na lokálním chování hustoty, má asymptoticky normální rozdělení.

Příspěvek se bude zabývat rozšířením těchto poznatků na určení chování M odhadů v případě lineární regrese s pevnou maticí $\mathbf{X}$, kdy hustota rozdělení chybových složek porušuje předpoklady způusobem uvedeným výše a který bude dále formalizován.

2. Definice a základní předpoklady

Předpokládejme platnost regresního modelu

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}, \tag{1}$$

kde $\mathbf{Y}$ značí $n \times 1$ vektor pozorování, $\mathbf{X}$ $n \times p$ matici experimentu, $\boldsymbol{\beta}$ je $p \times 1$ neznámý parametr a $\mathbf{E}$ označuje vektor n nezávislých a stejně rozdělených chybových složek E_i. Dále předpokládejme, že distribuční funkce E_1, kterou označíme jako F, je absolutně spojitá a hustotu E_1 označíme $f(e)$. O distribuční funkci rozdělení budeme dále předpokládat, že pro aspoň jeden bod x_0 platí, že

$$\lim_{x \to x_0} \frac{|F(x) - F(x_0)|}{|x - x_0|^{\gamma}} = K \tag{2}$$

pro nějaké $\gamma \neq 1$. Toto je poněkud obecnější možnost, kterou lze rozložit na několik podtříd, z nichž lze uvést například:

- $f(x_0) = 0$, $f'(x_0) = 0$. Tento případ zahrnuje kupříkladu situace, kdy $\gamma = m$, kde m je liché přirozené číslo.
- $f(x_0) = 0$, $f'(x_0)$ neexistuje, ale existují $\lim_{x \to x_0^+} f'(x_0)$ a $\lim_{x \to x_0^-} f'(x_0)$, ať již vlastní, nebo nevlastní.
- $\lim_{x \to x_0} f(x) = \infty$.

Definice $M-$odhadu
$M-$odhadem regresního vektoru $\boldsymbol{\beta}$ nazveme vektor minimalizující součet

$$\sum_{i=1}^{n} \varrho(y_i - \mathbf{x}_i' \mathbf{b}) \tag{3}$$

vzhledem k $\mathbf{b}$. V případě, že existuje $\psi = \varrho'$, je možno definici $M-$odhadu vyjádřit i jako řešení úlohy

$$\sum_{i=1}^{n} \mathbf{x}_i \psi(y_i - \mathbf{x}_i \mathbf{b}) = 0. \tag{4}$$

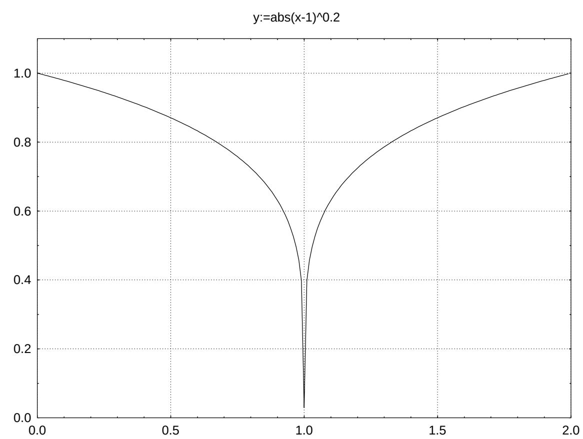
Navíc, pokud je ϱ striktně konvexní a ψ hladká, existuje jediné $\mathbf{b}$ řešící rovnici (4). Jelikož na chování odhadu a existenci řešení (4) má zásadní vliv skutečnost, zda ψ je absolutně spojitá nebo zda obsahuje body skoku, provedeme následující rozklad: $\psi = \psi_1 + \psi_2$, kde ψ_1 je absolutně spojitá, ψ_2 skoková a monotonní.

3. MOTIVACE

Motivací k hledání výsledků pro případ rozdělení odpovídajícímu podmínce (2) jsou jednak výsledky dosažené pro podobný typ rozdělení při odhadování parametrů polohy, jednak výsledky dosažené pro odhady regresního vektoru pro „regulární“ hustoty. V prvním případě se jedná o změnu řádu konzistence z obvyklého $n^{-1/2}$ na řád odpovídající γ. Ve druhém případě se jedná o již klasické výsledky, které uvádí monografie Jurečkové a Sena (1996). Pro odhady kvantilů distribuční funkce za podmínky (2) platí, pokud $\{p_n\}$ splňuje podmínku, že $n^{1/2}(p_n - p)$ je omezená posloupnost (viz Ghosh a Sukhatme, nebo de Haan a Taconis–Haantjes):

$$K|X_{p_n} - F^{-1}(p)|^\gamma \operatorname{sgn}(X_{p_n} - F^{-1}(p)) = p - F_n(F^{-1}(p)) + R_n, \tag{5}$$

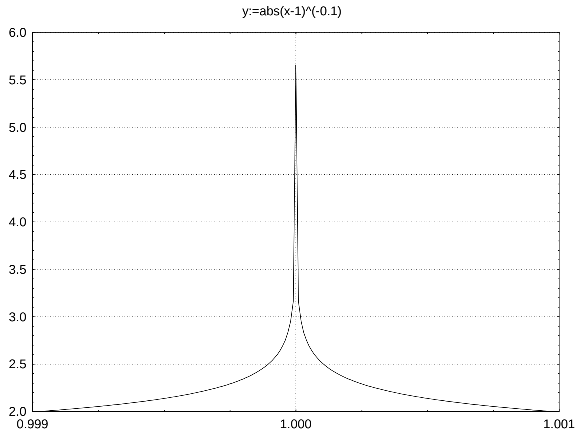
kde $n^{1/2}R_n \to 0$ v pravděpodobnosti a K je limita z (2). Otázkou je, za jakých podmínek lze tento výsledek rozšířit na odhady regresních parametrů.



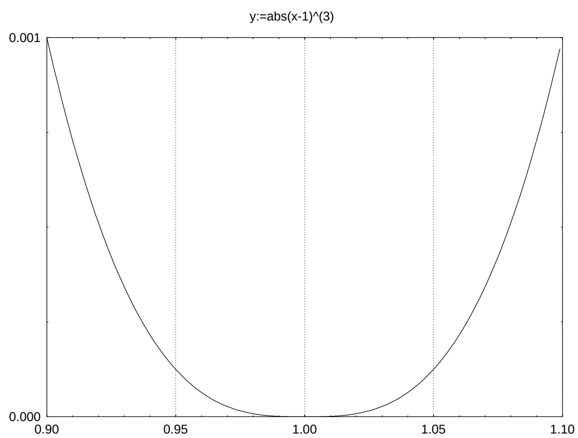
OBRÁZEK 1. $f(x)$ v případě $\gamma = 1,2$

Pro $\gamma < 1$ roste $f(x)$ v okolí x_0 nade všechny meze a intuitivní představa je ta, že řád konzistence bude vyšší.

V případě $\gamma > 1$ klesá $f(x)$ v okolí x_0 k nule a intuitivní představa napovídá, že řád konzistence bude nižší. V následující části bude ukázáno, jak intuitivní představy v tomto případě odpovídají nebo neodpovídají skutečnosti.



OBRÁZEK 2. $f(x)$ v případě $\gamma = 0.9$



OBRÁZEK 3. $f(x)$ v případě $\gamma = 3$

4. VÝSLEDKY

Nejprve je nutno stanovit nutné podmínky pro odvozování asymptotických vlastností $M-$odhadu. Tyto podmínky jsou:

- $0 < \int \psi^2(x) dF(x) < \infty$.
- Pokud $\psi = \psi_1$, pak $\int \psi'(x) > 0$.

- $\frac{1}{n}\mathbf{X}'\mathbf{X} \to \mathbf{Q} > 0$, případně $E\mathbf{x}\mathbf{x}' \to \mathbf{Q} > 0$, pokud matice $\mathbf{X}$ je určena náhodnými vektory $\mathbf{x}_i$.

Věta 1. Asymptotická linearizace vhodné funkce. *Nechť* $\psi = \operatorname{sgn}(x)$. *Pak za výše uvedených podmínek platí, že vhodnou funkci $M-$odhadu regresního vektoru lze vyjádřit vztahem*

$$n^{-\frac{1}{2}-\frac{\gamma}{2}} \sum \mathbf{x}_i |\mathbf{x}_i'(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))|^{\gamma} \operatorname{sgn}(\mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) = n^{-1/2} \sum \mathbf{x}_i \psi(E_i) \frac{1}{K} + o_p(1) \tag{6}$$

Tato reprezentace má za následek asymptotické rozdělení příslušné funkce:
Věta 2. Rozdělení vhodné funkce $\hat{\boldsymbol{\beta}}$. *Jsou-li splněny předpoklady předchozí věty, platí:*

$$n^{-\frac{1}{2}-\frac{\gamma}{2}} \sum \mathbf{x}_i |\mathbf{x}_i'(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))|^{\gamma} \operatorname{sgn}(\mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \sim N_p(\mathbf{0}, \mathbf{Q}\frac{1}{K^2} \operatorname{var} \psi(E_1)) \tag{7}$$

Odůvodnění volby $\psi(x) = \operatorname{sgn}(x)$ odpovídající L_1 odhad a diskuse obecnějších případů jsou tématy poslední části (Diskuse a důsledky).

5. Myšlenky důkazů, technické poznámky

Jelikož podrobné rozepsání důkazů překračuje možnost krátkého článku, uvedu pouze základní myšlenky a postupy vedoucí k výsledku, tedy k asymptotické linearitě odpovídající funkce rozdílu $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$. [1] Postup se skládá z následujících kroků:

- Výpočet vhodné aproximace střední hodnoty veličiny $\sum \mathbf{x}_i(\psi(E_i) - \psi(E_i - n^{-\alpha}\mathbf{x}_i'\mathbf{t}))$ jako funkce $\mathbf{t}$ pro (nejprve obecné) α. Pro $\psi(x) = \operatorname{sgn}(x)$ lze jednotlivé příspěvky do součtu spočíst jako $\mathbf{x}_i P[0 < E_i < n^{-\alpha}\mathbf{x}_i'\mathbf{t}]$ pro případ $\mathbf{x}_i'\mathbf{t} > 0$, v opačném případě samozřejmě $\mathbf{x}_i P[0 > E_i > n^{-\alpha}\mathbf{x}_i'\mathbf{t}]$. V prvním případě tedy hraje klíčovou roli výraz $\int_0^{n^{-\alpha}\mathbf{x}_i'\mathbf{t}} f(e)de$. Aproximací tohoto výrazu (ať již pomocí Taylorova rozvoje, nebo, pro γ celé liché, rekurzívním vyjadřováním $\int_0^A f^{(i)}(e_i)de_i = \int_0^A f^{(i)}(e_i) - f^{(i)}(0)de_i = \int_0^A \int_0^{e_i} f^{(i+1)}(e_{i+1})de_{i+1}de_i$ pro $i =, \ldots, \gamma - 1$) dostáváme $K(n^{-\alpha}\mathbf{x}_i'\mathbf{t})^{\gamma}$, samozřejmě při splnění nutných podmínek na $\mathbf{x}_i$.
- Stejnoměrnost konvergence vyšetřovaného součtu ke střední hodnotě (pro vhodné α). Vhodné α vyhovuje rovnici $\alpha\gamma = 1/2$, tedy $\alpha = 1/(2\gamma)$. Pro ukázání stejnoměrnosti lze využít princip Skorochodova vnoření. Jednotlivé příspěvky závisí na hodnotě výrazu $\psi(E_i) - \psi(E_i - n^{-\alpha}\mathbf{x}_i'\mathbf{t})$. Tento výraz, vzhledem k tomu, že $\psi(x) = \operatorname{sgn}(x)$,

[1] Některé technické prvky jsou ukázány v příspěvku, který jsem přednesl na Robustu'96

může nabývat jedné z hodnot -1,0,+1, přičemž pokud $\mathbf{x}_i'\mathbf{t} < 0$, omezuje se množina hodnot na 0,+1, v případě opačného znaménka na 0,-1. Tyto náhodné veličiny opravíme o střední hodnotu, tedy $K(n^{-\alpha}\mathbf{x}_i'\mathbf{t})^\gamma$. Tyto výsledné veličiny označíme c_i. Na takto vzniklé náhodné veličiny lze nahlížet jako na hodnoty Wienerova procesu v čase prvního výstupu z intervalů $(-a_i, b_i)$, kde hodnoty $-a_i$, b_i jsou určeny možnými hodnotami c_i. Příslušné časy označme T_i. Platí $ET_i = a_i b_i$. Vyšetření ET_i jako funkce $\mathbf{t}$ pak dává výsledek zaručující stejnoměrnou konvergenci $\mathbf{S}_n = \sum_i^n \mathbf{x}_i(\psi(E_i) - \psi(E_i - n^{-1/2\gamma}\mathbf{x}_i'\mathbf{t}))$ ke své střední hodnotě, $||\mathbf{t}|| \leq C$.

- Vyjádření reprezentace. Jelikož pro $\gamma \neq 1$ nelze obecně vyjádřit reprezentaci pro $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$, je uveden tvar pro $|\mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))|^\gamma \operatorname{sgn}(\mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))$.
- Asymptotické rozdělení se získá z reprezentace jednoduchým výpočtem za předpokladu $\lim_{n\to\infty} \frac{1}{n}\mathbf{X}'\mathbf{X} = \mathbf{Q} > 0$.

6. Diskuse a důsledky

Výsledky se zabývají pouze případem skokových ψ funkcí, tedy funkcí definujících L_1 odhad, regresní kvantily a jejich konvexní lineární kombinace. V případě ψ funkcí spojitých, odpovídajících například L_2 odhadu nebo Huberově odhadu, se řád konzistence nemění a platí klasické výsledky. Ke zdůvodnění tohoto faktu lze využít prostou skutečnost, že $E(\psi(E_i) - \psi(E_i - n^{-\alpha}\mathbf{x}_i'\mathbf{t}))$ je pro spojitou ψ funkcí lineární v $n^{-\alpha}\mathbf{x}_i'\mathbf{t}$, například pro L_2 odhad je přímo $(\psi(E_i) - \psi(E_i - n^{-\alpha}\mathbf{x}_i'\mathbf{t})) = n^{-\alpha}\mathbf{x}_i'\mathbf{t}$, což vede na $\alpha = 1/2$, neboť hodnota výrazu nezávisí na lokálním chování hustoty.

Možná zobecnění: Předpokládejme, že výraz v (2) má různé (konečné a nenulové) limity pro $x \to x_0^+$ a $x \to x_0^-$. Pak výsledná reprezentace se mění následovně:

Důsledek 1 *Označme*

$$\lim_{x\to x_0^+} \frac{|F(x) - F(x_0)|}{|x - x_0|^\gamma} = K^+, \lim_{x\to x_0^-} \frac{|F(x) - F(x_0)|}{|x - x_0|^\gamma} = K^-.$$

Jsou-li splněny podmínky Věty 1, pak platí:

$$\begin{aligned} &n^{-\frac{1}{2}-\frac{\gamma}{2}} \sum \mathbf{x}_i |\mathbf{x}_i'(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))|^\gamma \operatorname{sgn}(\mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \\ &= n^{-1/2} \sum \mathbf{x}_i \psi(E_i) \frac{K^+ + K^-}{(K^+)^2 + (K^-)^2} + o_p(1) \end{aligned} \tag{8}$$

Pro důkaz tohoto tvrzení si stačí uvědomit, že část vpravo přispívá hodnotou $K^+\frac{K^+}{K^+ + K^-}$, část vlevo pak hodnotou $K^-\frac{K^-}{K^+ + K^-}$.

Předpokládejme, že podmínka (2) platí s $K^+ = K^- = K$, ale že hustota rozdělení náhodné veličiny E_1 obsahuje obecně k bodů $s_1, \ldots, s_k$ splňujících (2) pro $\gamma_j \neq 1 \, \forall \, j = 1, \ldots, k$ a nějaká $K_1, \cdots, K_k$. Pak lze z věty 1 vyvodit

Důsledek 2 *Označme* $\gamma_0 = \min\{\gamma_1, \ldots, \gamma_k\}$, $K_0 = \{j : \gamma_j = \gamma_0\}$. *Nechť* $\psi(x)$ *je po částech konstantní neklesající funkce se skoky o velikostech* $(a_j - a_{j-1})$ *v bodech* s_j, $j = 1, \ldots, k$, *přičemž* $a_0 = \lim_{x \to -\infty} \psi(x)$. *Pak, jsou-li splněny předpoklady Věty 1, platí:*

$$
\begin{aligned}
&n^{-\frac{1}{2}-\frac{\gamma}{2}} \sum_i \mathbf{x}_i \sum_{j \in K_0} (a_j - a_{j-1}) K_j |\mathbf{x}_i'(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))|^{\gamma_0} \operatorname{sgn}(\mathbf{x}_i'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \\
&\qquad = n^{-1/2} \sum \mathbf{x}_i \psi(E_i) + o_p(1) \qquad (9)
\end{aligned}
$$

Toto zobecnění vyplývá ze skutečnosti, že příspěvky členů s $\gamma_j > \gamma_0$ budou, pro $n \to \infty$, v součtu $\sum_j K_j(a_j - a_{j-1}) |\mathbf{x}_i'(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))|^{\gamma_j}$ zanedbatelné vůči příspěvkům členů s $\gamma_j = \gamma_0$.

LITERATURA

[1] Ghosh M. a Sukhatme S., *On Bahadur's representations of quantiles in nonregular cases*, Communication in Statistics, Theory and Methods, 1981, **A10(3)**, 269 – 282.

[2] de Haan L. a Taconis-Haantjes E., *On Bahadur's representation of sample quantiles*, Annals of Inst. Statist. Math., 1979, **31 A**, 299 – 308.

[3] Jurečková J. a Sen P. K., *Robust Statistical Procedures: Asymptotic and Interrelations*, John Wiley, New York, 1996.

[4] Jurečková J., *Asymptotic behavior of $M-$estimators of location in nonregular cases*, 1983, Statistics and Decisions **1**, 323 – 340.

[5] Svatoš J., *$M-$odhady za předpokladu neregulární hustoty*, Robust'96, Antoch J. a Dohnal G. (eds.), JČMF, 1996, 211 – 216.