ROBUST'2002, 240 – 246

KONZISTENCE ODHADU LWS PRO PARAMETR POLOHY

LIBOR MAŠÍČEK

ABSTRACT. The present paper deals with least weighted squares estimator which is robust estimator and it generalize classical least trimmed squares. We will prove $n^{\frac{1}{4}}$-consistency of this estimator for parameter of location. As next result we show asymptotic linearity of the normal equation which gives us under assumption of $n^{\frac{1}{2}}$-consistency asymptotic normality and formula for asymptotic variance.

Резюме. В статье укажем оценку методом наименьших взвешенных квадратов который выражает наименьшие усеченные квадраты. Для оценки параметра положения привидём его $n^{\frac{1}{4}}$-консистенцию и ассимптотическую линеарность нормального уравнения.

1. Úvod

V následujícím textu se budeme zabývat odhadem metodou nejmenších vážených čtverců (least weighted squares, LWS). Jedná se robustní odhad s volitelným bodem selhání, který zobecňuje klasické nejmenší useknuté čtverce (least trimmed squares, LTS).

Pro odhad parametru polohy ukážeme jeho $n^{\frac{1}{4}}$-konzistenci pro nerostoucí váhovou funkci a symetrické jednovrcholové rozdělení náhodných chyb. V důkazu je využito silných výsledků z principů invariance.

Dále odvodíme tvar normální rovnice a dokážeme její asymptotickou linearitu. Asymptotická linearita hraje podstatnou roli při dokazování asymptotické normality. Za předpokladu $\sqrt{n}$-konzistence odhadu dostáváme z asymptotické linearity asymptotickou normalitu odhadu a tvar asymptotického rozptylu.

2. Definice odhadu

Předpokládejme model parametru polohy

$$Y_i = \beta_0 + e_i \qquad \text{pro } i = 1, \ldots, n,$$

kde β_0 je neznámý parametr, e_i jsou nezávislé a stejně rozdělené chyby s nulovou střední hodnotou a Y_i jsou pozorované hodnoty. Vektor chyb a pozorování potom označíme $e = (e_1, \ldots, e_n)$ a $Y = (Y_1, \ldots, Y_n)$.

Pro libovolné $b \in R$ označme (oproti standardnímu značení) absolutní hodnotu reziduí jako $r_i(b) = |Y_i - b|$ a h-tou pořádkovou statistiku absolutních hodnot těchto reziduí jako $r_{(h)}(b)$ (tj. $0 \le r_{(1)}(b) \le r_{(2)}(b) \le \cdots \le r_{(n)}(b)$).

2000 *Mathematics Subject Classification.* 62M10 62F35.

Klíčová slova. Robustní regrese, nejmenší useknuté čtverce, nejmenší vážené čtverce, konzistence, asymptotická linearita.

Tento příspěvek vznikl za přispění grantu GA UK 255/2000 /A EK /FSV a MSM 113200008.

Nyní můžeme definovat *odhad metodou nejmenších vážených čtverců* (least weighted squares, LWS) jako

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \mathrm{MF}(\beta, w, Y) \tag{1}$$

kde

$$\mathrm{MF}(\beta, w, Y) = \sum_{h=1}^{n} w\left(\frac{h}{n}\right) r^2_{(h)}(\beta) \tag{2}$$

a $w : [0,1] \to R$ je daná *váhová funkce*. Tento odhad navrh Víšek v [11].

Vidíme, že pro speciální volbu váhové funkce $w(x) = I\{x < \alpha\}$ dostaneme klasické *nejmenší useknuté čtverce* (viz. [5]).

Náš odhad lze definovat i pro obecný regresní model. Pak budou rezidua ve vztahu (2) v klasickém regresním smyslu.

3. Konzistence odhadu

V následujícím textu budeme chtít odvodit, za jakých podmínek bude náš odhad konzistentní.

Pro libovolné $\alpha \in (0,1)$, $t \in R$ a libovolnou distribuční funkci G definujme následující statistický funkcionál

$$T_\alpha(t, G) = \int_{-u_\alpha(t,G)}^{u_\alpha(t,G)} x^2 \, dG(x+t) = \int_{t-u_\alpha(t,G)}^{t+u_\alpha(t,G)} (x-t)^2 \, dG(x) \tag{3}$$

kde $u_\alpha(t, G)$ je definováno

$$G(t + u_\alpha(t,G)) - G(t - u_\alpha(t,G)) = \int_{-u_\alpha(t,G)}^{u_\alpha(t,G)} 1 \, dG(x+t) = \alpha. \tag{4}$$

Jedná se vlastně o střední kvadratickou odchylku přes symetrický interval kolem bodu t a šířky $2u_\alpha(t,G)$. Pravděpodobnost tohoto intervalu je rovna α. V případě, že distribuční funkce G je nespojitá a rovnice (4) nemá řešení, bereme největší možné $u_\alpha(t,G)$ pro které je intergál menší nebo roven α.

Nyní pomocí tohoto funkcionálu přepíšeme náš odhad. Proto jako F a F_n označíme distribuční funkci a empirickou distribuční funkci veličin e_i, $i = 1, \ldots, n$. Vidíme, že pokud do funkcionálu dosadíme empirickou distribuční funkci dostaneme

$$T_\alpha(t, F_n) = \frac{1}{n} \sum_{h=1}^{[\alpha n]} r^2_{(h)}(\beta_0 + t).$$

Pokud budeme dále předpokládat, že váhová funkce je konstantní na intervalu $[0, \underline{\alpha}]$, nerostoucí na $[\underline{\alpha}, \overline{\alpha}]$ a nulová na $[\overline{\alpha}, 1]$, kde $0 < \underline{\alpha} \le \overline{\alpha} < 1$, můžeme minimalizovanou funkci MF přepsat následovně

$$\mathrm{MF}(\beta, w, Y) = n \cdot \sum_{h=[\underline{\alpha} n]}^{[\overline{\alpha} n]} w_{h,n} \cdot T_{h/n}(\beta - \beta_0, F_n) \tag{5}$$

kde $w_{h,n} = w(h/n) - w((h+1)/n) \ge 0$.

Z nulovosti váhové funkce na $[\overline{\alpha}, 1]$ okamžitě vidíme, že bod selhání tohoto odhadu bude $1 - \overline{\alpha}$ (pro $\overline{\alpha} > \frac{1}{2}$) a tedy náš odhad bude robustní.

Vidíme, že chování našeho odhadu bude úzce spjato s chováním funkcionálu T_α. Konkrétní vztah je uveden v následující větě, která dává postačující podmínku na konzistenci odhadu.

Věta 1. Nechť jsou splněny následující předpoklady

1) existují $0 < \underline{\alpha} \le \overline{\alpha} < 1$ tak, že $w \equiv 1$ na $[0, \underline{\alpha}]$, w je nerostoucí na $[\underline{\alpha}, \overline{\alpha}]$ a $w \equiv 0$ na $[\overline{\alpha}, 1]$,
2) chyby e_i jsou spojité veličiny s hustotou $f(x)$, která je omezená a zdola odražená od nuly na každém omezeném intervalu (tj. $\inf\{f(x), x \in (-K, K)\} > 0$ pro $K > 0$),
3) existuje $K_0 > 0$ a $\delta > 0$ tak, že pro libovolné $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ a $t \neq 0$ platí
$$T_\alpha(t, F) \ge T_\alpha(0, F) + K_0 \cdot \min\{t^2, \delta\},$$

potom je odhad metodou nejmenších vážených čtverců $n^{\frac{1}{4}}$-konzistentní, tj.
$$\sqrt[4]{n}\left(\hat{\beta}^{(LWS,n,w)} - \beta_0\right) = \mathcal{O}_p(1).$$

Důkaz. Zvolme $\varepsilon > 0$ libovolné. Potom existuje konstanta $K_1 > 0$ taková, že jev
$$B_n = [\|F_n - F\| < n^{-\frac{1}{2}} K_1]$$
splňuje $P(B_n) > 1 - \varepsilon$ pro všechna n, kde $\|\cdot\|$ je suprémová norma (viz. kapitola 12, [6]). Ukážeme, že potom existuje konstatnta $M > 0$ a n_0 takové, že na B_n je splněno
$$\mathrm{MF}(\beta, w, Y) > \mathrm{MF}(\beta_0, w, Y) \tag{6}$$
pro $|\beta - \beta_0| > n^{-\frac{1}{4}} M$ a $n > n_0$, čímž tvrzení dokážeme. V další části důkazu proto předpokládejme, že pracujeme na B_n, kde je splněno $\|F_n - F\| < n^{-\frac{1}{2}} K_1$.

Nejprve ukážeme, že nerovnost (6) platí pro $|\beta - \beta_0| \le K_2$, kde $K_2 > 0$ je vhodná, ale pevně zvolená konstanta. Z podmínky 1), vztahu (5) a nezápornoti $w_{h,n}$ dostáváme, že stačí ukázat, že na B_n je splněno
$$T_\alpha(t, F_n) > T_\alpha(0, F_n) \tag{7}$$
pro $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, $n^{-\frac{1}{4}} M \le |t| \le K_2$ a $n > n_1$. Uvědomme si, že mezi $w_{h,n}$ je vždy alespoň jedna kladná hodnota, které nám dá ostrou nerovnost v (6).

Přepišme proto rozdíl levé a pravé strany nerovnosti (7) následovně
$$\begin{aligned} &T_\alpha(t, F_n) - T_\alpha(0, F_n) = \\ &\quad = (T_\alpha(t, F_n) - T_\alpha(t, F)) + (T_\alpha(t, F) - T_\alpha(0, F)) + (T_\alpha(0, F) - T_\alpha(0, F_n)) \end{aligned} \tag{8}$$
Z podmínky 3) dostaneme dolní omezení na prostřední člen v (8)
$$T_\alpha(t, F) - T_\alpha(0, F) \ge K_0 \min\{t^2, \delta\} \ge K_0 \min\{n^{-\frac{1}{2}} M^2, \delta\}. \tag{9}$$
Nyní omezíme ostatní dva členy v (8). Přepišme tedy první člen následovně
$$\begin{aligned} T_\alpha(t, F_n) - T_\alpha(t, F) &= \int_{-u_\alpha(t,F_n)}^{u_\alpha(t,F_n)} x^2\, dF_n - \int_{-u_\alpha(t,F)}^{u_\alpha(t,F)} x^2\, dF = \\ &= \int_{-u_\alpha(t,F_n)}^{u_\alpha(t,F_n)} x^2\, d(F_n - F) + \int_{-u_\alpha(t,F_n)}^{-u_\alpha(t,F)} x^2\, dF + \int_{u_\alpha(t,F)}^{u_\alpha(t,F_n)} x^2\, dF. \end{aligned} \tag{10}$$
Pokud ukážeme, že existují konstanty K_3 a K_4 tak, že platí
$$|u_\alpha(t, F) - u_\alpha(t, F_n)| \le K_3 n^{-\frac{1}{2}} \tag{11}$$
$$|u_\alpha(t, F)| \le K_4 \tag{12}$$

pro $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ a $|t| \leq K_2$, dostaneme omezení na první člen v (10)

$$\left| \int_{-u_\alpha(t,F_n)}^{u_\alpha(t,F_n)} x^2 \, d(F_n - F)(x) \right| \leq K_5^2 \cdot 2K_1 n^{-\frac{1}{2}}$$

neboť $|u_\alpha(t, F_n)| \leq K_3 + K_4 = K_5$ a pracujeme na B_n, kde $\|F_n - F\| \leq K_1 n^{-\frac{1}{2}}$. Na třetí člen ve výrazu (10) použijeme následující omezení

$$\left| \int_{u_\alpha(t,F)}^{u_\alpha(t,F_n)} x^2 \, dF(x) \right| \leq K_5^2 \int_{u_\alpha(t,F)}^{u_\alpha(t,F_n)} f(x) \, dx \leq K_5^2 \cdot K_3 n^{-\frac{1}{2}} \max\{f(x)\}.$$

Pro druhý člen v (10) použijeme stejné omezení jako pro třetí. Pokud tato omezení spojíme dostaneme

$$|T_\alpha(t, F_n) - T_\alpha(t, F)| \leq n^{-\frac{1}{2}} K_6 \tag{13}$$

kde $K_6 = 2K_3K_5^2 + 2K_1K_5^3 \max\{f(x)\} < \infty$. Tato nerovnost bude platit i pro $t = 0$, čímž omezíme třetí člen ve výrazu (8).

Spojením nerovností (9) a (13) dostaneme

$$T_\alpha(t, F_n) - T_\alpha(0, F_n) \geq K_0 \cdot \min\{M^2 n^{-\frac{1}{2}}, \delta\} - 2n^{-\frac{1}{2}} K_6 > 0,$$

kde poslední nerovnost bude platit pro M splňující $K_0 M^2 > 2K_6$ a $n > n_2$ (pro vhodnou volbu n_2). K dokončení důkazu nerovnosti (6) pro $n^{-\frac{1}{4}} M < |\beta - \beta_0| < K_2$ musíme ukázat existenci konstant K_3 a K_4 pro platnost (11) a (12).

Existence K_3 v (11) vyplývá z faktu, že $u_\alpha(t, F)$ a $u_\alpha(t, F_n)$ vyjadřují skutečný a empiricný kvantil veličin $|e_i + t|$. Neboť pracujeme na B_n bude se skutečná a empirická distribuční funkce veličin $|e_i + t|$ lišit nejvýše o $2K_1 n^{-\frac{1}{2}}$. Předpoklad 2) nám dává omezenost $(F^{-1})'(x)$ na libovolném intervalu (a, b), kde $0 < a < b < 1$. Zbytek vyplývá z faktu, že t je omezené a α je odraženo od 1.

Omezenost $u_\alpha(t, F)$ (tj. (12)) pak dostaneme z omezení

$$u_\alpha(t, F) \leq |t| + u_\alpha(0, F) \leq K_2 + u_{\overline{\alpha}}(0, F).$$

K dokončení celého důkazu nám zbývá nalézt vhodnou konstantu K_2 tak, že pro $|\beta - \beta_0| \geq K_2$ bude na B_n splněna nerovnost (6). Předpokládejme tedy, že $|\beta - \beta_0| \geq K_2$ a ukážeme, že pro dostatečně velké K_2 bude nerovnost (6) splněna.

K nalezení K_2 budeme potřebovat, aby velká část pozorování byla omezená. Proto zvolme $\max\{\overline{\alpha}, 1 - \overline{\alpha}\} < \tilde{\alpha} < 1$ libovolně. Analogicky jako u nerovností (11) a (12) dostaneme, že existuje K_7 tak, že pro $h \leq [\tilde{\alpha} n]$ je $r_{(h)}(\beta_0) \leq K_7$ na B_n pro $n > n_3$ (tj. alespoň $[\tilde{\alpha} n]$ z veličin $e_1, \ldots, e_n$ splňuje $|e_i| \leq K_5$). Protože $\overline{\alpha} < \tilde{\alpha}$ dostaneme

$$\mathrm{MF}(\beta_0, w, Y) = \sum_{h=1}^{[\overline{\alpha} n]} w\left(\frac{h}{n}\right) r_{(h)}^2(\beta_0) \leq K_5^2 \sum_{h=1}^{[\overline{\alpha} n]} w\left(\frac{h}{n}\right) \leq K_5^2 \cdot n \int_0^{\overline{\alpha}} w(x) \, dx \tag{14}$$

Nyní chceme nalézt dolní omezení na

$$\mathrm{MF}(\beta, w, Y) = \sum_{h=1}^{[\overline{\alpha} n]} w\left(\frac{h}{n}\right) r_{(h)}^2(\beta). \tag{15}$$

Vidíme, že pokud $|e_i| \leq K_5$ bude $r_i(\beta) = |e_i - (\beta - \beta_0)| \geq K_2 - K_5$ (neboť $|\beta - \beta_0| \geq K_2$). Protože alespoň $[\tilde{\alpha} n]$ z veličin $e_1, \ldots, e_n$ splňuje $|e_i| \leq K_5$ bude alespoň $[\tilde{\alpha} n] - (n - [\tilde{\alpha} n]) > 0$ reziduí v sumě (15) zdola omezených $K_2 - K_5$ a proto

$$\mathrm{MF}(\beta, w, Y) \geq (K_2 - K_5)^2 \sum_{h=n-[\overline{\alpha} n]}^{[\tilde{\alpha} n]} w\left(\frac{h}{n}\right) \geq (K_2 - K_5)^2 \cdot n \int_{1-\overline{\alpha}}^{\tilde{\alpha}} w(x) \, dx \tag{16}$$

pro $n \geq n_4$ (w je totiž nerostoucí). Neboť $1 - \overline{\alpha} < \tilde{\alpha}$ bude integrál v poslední nerovnosti nenulový a proto pro dostatečně velké K_2 (nezávislé na n) bude pravá strana nerovnosti (14) menší než pravá strana (16) a tedy nerovnost (6) bude na B_n splněna pro $|\beta - \beta_0| \geq K_2$. Tím jsme také dokázali celé tvrzení. □

Druhá a třetí podmínka v předchozí větě nám klade omezení na rozdělení chyb. Druhá podmínka v podstatě říká, že nosič hustoty je celá reálná osa. Třetí podmínka je poněkud neprůhledná. Následující věta však ukazuje její platnost pro jednovrcholová symetrická rozdělení. Pro normální rozdělení je pak odvozen silnější výsledek.

Věta 2. Nechť hustota chyb $f(x)$ je symetrická, spojitě diferencovatelná a klesající na $(0, \infty)$, potom je pro libovolné $\frac{1}{2} \leq \underline{\alpha} \leq \overline{\alpha} < 1$ splněna podmínka 3) z Věty 1. Mají-li navíc náhodné chyby normální rozdělení s nulovou střední hodnotou a kladným rozptylem je podmínka splněna pro libovolné $0 < \underline{\alpha} \leq \overline{\alpha} < 1$.

Důkaz. Zvolme $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ pevné. Budeme chtít vyjádřit $\frac{d}{dt} T_\alpha(t, F)$ a vyšetřit její průběh. Nejprve vyčíslíme $\frac{d}{dt} u_\alpha(t, F)$. Pro jednoduchost budeme v další části důkazu psát $u(t)$ namísto $u_\alpha(t, F)$. Z definice je $u(t)$ dáno vztahem

$$\int_{-u(t)+t}^{u(t)+t} 1 \, dF(x) = \alpha. \tag{17}$$

Zkusme vyčíslit derivační podíl $\frac{1}{h}[u(t+h) - u(t)]$. Proto odečtěme rovnici (17) pro $u(t)$ a $u(t+h)$

$$\begin{aligned} 0 &= \int_{-u(t)+t}^{u(t)+t} 1 \, dF(x) - \int_{-u(t+h)+t+h}^{u(t+h)+t+h} 1 \, dF(x) \\ &= \int_{-u(t)+t}^{-u(t+h)+t+h} f(x) \, dx + \int_{u(t+h)+t+h}^{u(t)+t} f(x) \, dx. \end{aligned}$$

Nyní využijeme nerovnosti $|u(t+h) - u(t)| \leq h$, která plyne přímo z rovnosti (4) pro $u(t)$ a $u(t+h)$ (v opačném případě by byl jeden z integrálů v (4) ostře větší než druhý). Dale již pouze aplikujeme Taylorův rozvoj hustoty $f(x)$ v bodech $-u(t) + t$ a $u(t) + t$ a dostaneme

$$0 = f(-u(t)+t)(u(t) - u(t+h) + h) + f(u(t)+t)(u(t) - u(t+h) - h) + o(h)$$

odkud již přímým výpočtem derivačního podílu a limitou pro $h \to 0$ dostaneme

$$\frac{d}{dt} u(t) = \frac{f(-u(t)+t) - f(u(t)+t)}{f(-u(t)+t) + f(u(t)+t)}. \tag{18}$$

Vzhledem k tomu, že hustota $f(x)$ je symetrická a klesající na $(0, \infty)$ (a ze symetrie také rostoucí na $(-\infty, 0)$), dostaneme, že výraz

$$f(-u(t)+t) - f(u(t)+t)$$

je kladný pro $t > 0$, záporný pro $t < 0$ a nulový pro $t = 0$. Ze vztahu (18) pak vyplývá, za totéž platí i pro $\frac{d}{dt} u(t)$ a proto $u(t)$ nabývá minima pro $t = 0$.

Ze vzorce (18) pak přímým výpočtem obdržíme první a druhou derivaci $T_\alpha(t, F)$

$$\frac{d}{dt} T_\alpha(t, F) = 2\alpha t - 2 \int_{-u(t)+t}^{u(t)+t} x \, dF(x)$$

$$\frac{d^2}{dt^2} T_\alpha(t, F) = 2\alpha - 4u(t) \frac{f(-u(t)+t) f(u(t)+t)}{f(-u(t)+t) + f(u(t)+t)}. \tag{19}$$

Vidíme, že první derivace je pro $t = 0$ nulová. Pokud nalezneme vhodné $K_0 > 0$ tak, aby $\frac{d^2}{dt^2} T_\alpha(t, F) \geq K_0$ pro $\underline{\alpha} \leq \alpha \leq \overline{\alpha}$ a $t \in R$, dostaneme omezení

$$T_\alpha(t, F) \geq T_\alpha(0, F) + K_0 t^2$$

čímž také celé tvrzení dokážeme.

Omezení za předpokladu $\underline{\alpha} \geq \frac{1}{2}$ nalezneme s využitím vztahu (17) a nerovností $t - u(t) \leq 0 \leq t + u(t)$. Pro $t \geq 0$ tedy dostaneme

$$\begin{aligned} \frac{d^2}{dt^2} T_\alpha(t, F) &= 2 \int_{-u(t)+t}^{u(t)+t} \left[f(x) - \frac{f(-u(t)+t)}{f(-u(t)+t) + f(u(t)+t)} f(u(t)+t) \right] dx \\ &\geq 2 \int_0^{u(t)+t} [f(x) - f(u(t)+t)]\, dx \\ &\geq 2 \int_0^{u_{\underline{\alpha}}(0,F)} \left[f(x) - f(u_{\underline{\alpha}}(0, F)) \right] dx = K_0 > 0 \end{aligned}$$

kde druhá nerovnost plyne z faktu, že $u(t) = u_\alpha(t, F)$ nabývá minima pro $t = 0$ a $\alpha = \underline{\alpha}$. Pro $t \leq 0$ odvodíme analogicky stejné omezení.

Omezení za předpokladu normality pak dostaneme opětovným derivováním vzorce (19) a dosazením hustoty normálního rozdělení. Vyšetřením průběhu derivace zjistíme, že $\frac{d^2}{dt^2} T_\alpha(t, F)$ je konvexní funkce s minimem v nule, které je kladné. □

4. Asymptotická linearita normální rovnice

V další části se budeme zabývat normální rovnicí našeho odhadu a její asymptotickou linearitou. Tato hraje podstatnou roli při dokazování asymptotické normality odhadu a zejména v odvozování tvaru asymptotického rozptylu. Pro důkaz asymptotické normality však kromě asymptotické linearity potřebujeme také $\sqrt{n}$-konzistenci, proto v další části ukážeme pouze asymptotickou linearitu.

Nejprve odvodíme tvar normální rovnice. Protože náš odhad je vlastně klasická vážená regrese aplikovaná na pozorování s nejmenšími rezidui vidíme, že za předpokladu $w(1) = 0$ bude mít normální rovnice pro tento odhad tvar

$$\begin{aligned} 0 = \mathrm{NR}_n(\beta, w) &= \sum_{h=1}^n w\left(\frac{h}{n}\right) \sum_{i=1}^n (Y_i - \beta) I\left\{ r_i^2(\beta) = r_{(h)}^2(\beta) \right\} \\ &= \sum_{h=1}^n w_{h,n} \sum_{i=1}^n (Y_i - \beta) I\left\{ r_i^2(\beta) \leq r_{(h)}^2(\beta) \right\}. \end{aligned} \tag{20}$$

Při asymptotické linearitě sledujeme chování normální rovnice v okolí β_0. Proto definujme $\beta_n = \beta_0 - n^{-\frac{1}{2}} t$ kde $t \in [-T, T]$ a sledujme rozdíl normální rovnice v bodech β_0 a β_n

$$\begin{aligned} S_n(t, w) &= \mathrm{NR}_n(\beta_n, w) - \mathrm{NR}_n(\beta_0, w) \\ &= \sum_{h=1}^n w_{h,n} \sum_{i=1}^n \left[(Y_i - \beta_n) I\left\{ r_i(\beta_n) \leq r_{(h)}(\beta_n) \right\} \right. \\ &\qquad \left. - (Y_i - \beta_0) I\left\{ r_i(\beta_0) \leq r_{(h)}(\beta_0) \right\} \right]. \end{aligned}$$

O tomto rozdílu budeme chtít ukázat, že je asymptoticky lineární ve smyslu

$$n^{-\frac{1}{4}} \sup_{t \in T_M} \left| S_n(t,w) - n^{\frac{1}{2}} R_w \right| = \mathcal{O}(1) \qquad \text{pro } n \to \infty \tag{21}$$

kde

$$R_w = t \int_0^{\alpha} [x - 2u_x f(u_x)]\, dw(x). \tag{22}$$

Asymptotickou linearitu ukážeme pro po částech konstantní váhovou funkci.

Věta 3. Nechť jsou splněny následující předpoklady

1) váhová funkce je tvaru $w(x) = \sum_{k=1}^{N} c_k I\{x < \alpha_k\}$ kde $\alpha_1, \ldots, \alpha_N \in (0,1)$ a $c_1, \ldots, c_N \in (0, \infty)$,
2) chyby e_i jsou spojité veličiny s hustotou $f(x)$, která je symetrická, omezená a zdola odražená od nuly na každém omezeném intervalu,

potom je normální rovnice asymptoticky lineární ve smyslu rovnosti (21).

Důkaz Věty 3 je uveden v preprintu [4].

Literatura

[1] Billingsley, P.: Convergence of Probability and Measures, New York, 1968.
[2] Boček, P., Lachout, P.: Linear programming approach to LMS-estimation, Memorial volume of Computational Statistics & Data Analysis 19, 1995, 129–134.
[3] Jurečková, J.: Asymptotic representation of M-estimators of location, Math. Operationsforsch. Statist., Ser. Statistics, Vol. 11, 1980, No.1, 61-73.
[4] Mašíček, L.: Konzistence odhadu LWS pro parametr polohy, KPMS Preprint 25, Department of Probability and Mathematical Statistics, Faculty of Mathemetics and Physics, Charles University, Prague, 2002.
[5] Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection, J.Wiley & Sons, New York, 1987.
[6] Shorack, R. G.: Probability for Statisticians, Springer-Verlag NY, 2000.
[7] Štěpán, J.: Teorie pravděpodobnosti, Academia, Praha, 1987.
[8] Víšek, J.Á.: A cautionary note on the method of least median squares reconsidered, Transactions of the Twelfth Prague Conference of Information Theory, Statistical Decision Functions and Random Processes, Prague, 1994, 254–259.
[9] Víšek, J.Á.: On high breakdown point estimation, Computational Statistics 11, 1995, 137–146.
[10] Víšek, J.Á.: On the diversity of estimates, Computational Statistics & Data Analysis 34, 2000, 67–89.
[11] Víšek, J.Á.: Regression with high breakdown point, ROBUST, 2000, 324–356.

UK MFF, Sokolovská 83, Praha 8