Vedecká Rada Fakulty Matematiky,
Fyziky a Informatiky
Univerzity Komenského v Bratislave

Mgr. Erika Hönschová

Autoreferát dizertačnej práce

L-odhady

na získanie vedecko-akademickej hodnosti philosophiae doctor

v odbore doktorandského štúdia:
9.1.9 **Aplikovaná matematika**

Bratislava 2010

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Katedre aplikovanej matematiky s štatistiky, Fakulty matematiky, fyziky a informatiky (FMFI) Univerzity Komenského (UK) v Bratislave.

Predkladateľ: Mgr. Erika Hönschová
Katedra aplikovanej matematiky a štatistiky
FMFI - UK v Bratislave
Mlynská dolina, 842 48 Bratislava

Školiteľ: doc.RNDr. František Rublík, CSc.
Ústav merania SAV
Dúbravská cesta 9, 841 04 Bratislava

Oponenti:

Autoreferát bol rozoslaný dňa: ....

Obhajoba dizertačnej práce sa koná dňa: .... o .... pred komisiou pre obhajobu dizertačnej práce v odbore doktorandského štúdia 9.1.9 Aplikovaná matematika vymenovanou predsedom spoločnej odborovej komisie dňa .... na Fakulte matematiky, fyziky a informatiky UK v Bratislave, Mlynská dolina, 842 48 Bratislava.

Predseda spoločnej odborovej komisie:
prof. RNDr. Marek Fila, PhD.
Katedra aplikovanej matematiky a štatistiky
FMFI - UK v Bratislave
Mlynská dolina, 842 48 Bratislava

Úvod

L-odhady sú lineárnymi funkciami poriadkových štatistík. Keďže v usporiadanom náhodnom výbere nie je splnená podmienka nezávislosti, nie je vo všeobecnosti jednoduché odvodiť asymptotické rozdelenie týchto štatistík. Tento problém bol už za rôznych podmienok kladených na distribučnú funkciu, z ktorej pochádza náhodný výber a skóre L-odhadu vyriešený viacerými spôsobmi. Väčšina autorov vyžaduje, aby náhodný výber bol z rozdelenia s konečnou strednou hodnotou, alebo aby funkcia, ktorá generuje skóre bola zrezaná, čo znamená, že istý počet najmenších a najväčších poriadkových štatistík zanedbáme. Jednou z možností, ako dokázať asymptotickú normalitu L-odhadov je dokázať, že ich môžeme aproximovať priemerom nezávislých náhodných premenných, čo nazývame asymptotická reprezentácia, alebo aj asymtotická linearita. Pokiaľ je rád konvergencie zvyšku tejto reprezentácie aspoň $o_P(n^{-1/2})$ je asymptotická normalita jej dôsledkom. Asymptotická reprezentácia L-odhadu je užitočná aj v iných prípadoch, napríklad pri výpočte skóre L-odhadu, alebo pri odvádzaní vlastností vektorov, ktorých zložkami sú $L-$odhady poprípade funkcie týchto odhadov. Cieľom dizertačnej práce bolo nájsť podmienky, za ktorých by bola asymptotická reprezentácia $L-$odhadov platná s rádom zvyšku $R_n = \mathcal{O}_P(1/n)$, ak náhodný výber nepochádza z rozdelenia s konečnou strednou hodnotou.

Ďalej sme sa zaoberali robustnými odhadmi parametra škály. Väčšina robustných odhadov je založená na princípe, že z náhodného výberu zanedbáme určitý počet najmenších a najväčších pozorovaní. Známym odhadom je napríklad $\alpha-$useknutý priemer, ktorý je priemerom po odstránení $\lfloor \alpha n/2 \rfloor$ najväčších a najmenších hodnôt. V prípade nesymetrických rozdelení však tento odhad nemusí byť vhodný. Kim vo svojej práci [8] navrhol odhad parametra lokácie pod názvom metricky useknutý priemer, kde zrezávanie závisí od distribučnej funkcie. Je to priemer pozorovaní, ktoré sú v intervale $(\xi(F_n)-\lambda(F_n), \xi(F_n)+\lambda(F_n))$, kde $\xi(F_n)$ je mediánom výberovej distribučnej funkcie F_n a $\lambda(F_n) = \inf\{t : F_n(\xi(F_n)+t) - F_n^-(\xi(F_n)-t) \geq 1-\alpha\}$. My sme podobným spôsobom definovali odhad parametra škály, ktorý nazývame metricky useknutá variancia. Vypočítali sme jeho influenčnú funkciu a dokázali sme asymptotickú normalitu pomocou Hadamardovskej diferencovateľnosti.

1 Prehľad doterajších výsledkov

1.1 L-odhady

Shorack [12] uvažoval $L-$odhad tvaru

$$T_n = n^{-1} \sum_{i=1}^{n} c_{ni} g_n(\xi_n^{(i)}),$$

kde $\xi_1, \xi_2, \dots \xi_n$ je náhodný výber z rovnomerného rozdelenie na intervale $(0,1)$. Platnosť tvrdenia

$$\sqrt{n}(T_n - \mu_n) \to N(0, \sigma^2),$$

kde

$$\mu_n = \int_0^1 g_n(t) J_n(t) \mathrm{d}t,\ \sigma^2 = \int_0^1 \int_0^1 (min\{s,t\} - st) J(s) J(t) \mathrm{d}g(s) \mathrm{d}g(t),$$

dokazuje za nasledujúcich predpokladov: Existuje zľava spojitá funkcia g definovaná na intervale $(0,1)$, ktorá má ohraničenú variáciu na $(\theta, 1-\theta)$, $\theta \in (0, 1/2)$ a funkcia J merateľná na $(0,1)$ tak, že

1. s výnimkou množiny bodov, ktorých miera definovaná funkciou $|g|$ je nulová funkcia J je spojitá a $J_n(t)$ konvergujú rovnomerne k $J(t)$ pre nejaké okolie bodu t, kde $J_n(t) = c_{ni}$ pre $(i-1)/n < t \leq i/n$, $1 \leq i \leq n$ a $J_n(0) = c_{n1}$,

2. $|J| \leq B$, $|J_n| \leq B$, $n = 1, 2, \dots$, $|g| \leq D$ a $|g_n| \leq D$, $n = 1, 2, \dots$, kde

   $$B(t) = Mt^{-b_1}(1-t)^{-b_2},$$

   pre nejaké pevné b_1, b_2, M a

   $$D(t) = Mt^{-1/2+b_1+\delta}(1-t)^{-1/2+b_2+\delta},\ \text{pre pevné } \delta > 0,$$

3. $$\lim_{n\to\infty} \int_0^1 B(t(1-t))^{1/2-\delta/2} \mathrm{d}|g_n - g| = 0.$$

Chernoff, Gastwirth a Johns [1] dokázali asymptotickú reprezentáciu štatistiky

$$T_n = n^{-1} \sum_{i=1}^{n} J(i/n+1) X_i^{(n)}$$

v tvare

$$T_n = \mu_n + Q_n + o_p(n^{-1/2}), \tag{1}$$

kde

$$\mu_n = n^{-1} \sum_{i=1}^{n} J(i/(n+1)F^{-1}(i/(n+1))),$$

$$Q_n = n^{-1} \sum_{i=1}^{n} \alpha(i/n+1)(Z_i - 1),$$

$$\alpha(u) = (1-u)^{-1} \int_u^1 J(w)(F^{-1})'(w)(1-w)\mathrm{d}w$$

a $Z_1, Z_2, \ldots, Z_n$ sú nezávislé náhodné premenné s exponenciálnym rozdelením. Podmienky, ktoré kladú na distribučnú funkciu F a funkciu J sú nasledovné:

1. Funkcia F^{-1} je spojitá na intervale (0,1) a spĺňa Lipschitzovu podmienku na každom intervale $(\delta, 1-\delta)$, $\delta \in (0, 1/2)$. $(F^{-1})'$ existuje a je spojitá.

2. Integrály $\int_0^1 J(u)(F^{-1})'(u)[u(1-u)]^{1/2}$, $\int_\delta^{1-\delta} J(u)\mathrm{d}u, \delta \in (0, 1/2)$ konvergujú absolútne.

3. Existuje $\delta_0 \in (0,1)$, také, že buď $J(i/n+1) = 0$ pre $i \leq n\delta_0$, alebo pre každé $K > 0$ existuje konečné M také, že ak $0 < u_1, u_2 < \delta_0$ a $K^{-1} < u_1/u_2 < K$ potom $M^{-1} < (F^{-1})'(u_1)/(F^{-1})'(u_2) < M$. Ďalej buď $J(i/n+1) = 0$ pre $i \geq n(1-\delta_0)$, alebo pre každé $K > 0$ existuje konečné M také, že ak $1-\delta_0 < u_1, u_2 < 1$ a $K^{-1} < (1-u_1)/(1-u_2) < K$ potom $M^{-1} < (F^{-1})'(u_1)/(F^{-1})'(u_2) < M$.

Vo vete 3. (str.63) ukazujú, že hodnotu μ_n v reprezentácii (1) môžeme nahradiť hodnotou $\mu = \int_0^1 J(u)F^{-1}(u)\mathrm{d}u$ ak $\mu_n = \mu + o(n^{-1/2})$.

Stigler vo svojich článkoch [16] a [17] na základe Hájekovej vete o projekcii [4] dokazuje, že štatistika $T_n = \frac{1}{n}\sum_{i=1}^n J(\frac{i}{n+1})X_n^{(i)}$ má rovnaké asymptotické rozdelenie ako $\hat{T}_n = \frac{1}{n}\sum_{i=1}^n J(\frac{i}{n+1})\hat{X}_n^{(i)}$, kde

$$\hat{X}_n^{(i)} = n^{-1} \sum_{k=1}^{n} \int_0^{F(X_k)} \frac{1}{f(F^{-1}(u))} g_{ni}(u)\mathrm{d}u + nE(X_{n-1}^{(i-1)}) - (n-1)E(X_n^{(i)}),$$

a funkcia g_{ni} je hustota $i-$tej poriadkovej štatistiky z rovnomerného rozdelenia na intervale $(0,1)$. Asymptotickú normalitu dokazuje za podmienok, že

$E(X^2) < \infty$, kde X má distribučnú funkciu F a funkcia J je ohraničená a spojitá skoro všade F^{-1}, alebo pre nejaké $\varepsilon > 0$ platí, že $\lim_{x\to\infty} x^{\varepsilon}[1 - F(x) + F(-x)] = 0$, J je ohraničená a spojitá skoro všade F^{-1} a $J(u) = 0$ pre $0 < u < \alpha$ a $1 - \alpha < u < 1$.

Moore [10] dokázal reprezentáciu

$$\frac{1}{n}\sum_{i=1}^{n} J(\frac{i}{n})X_n^{(i)} - \int_0^1 J(u)F^{-1}(u)\mathrm{d}u = \frac{1}{n}\psi(X_i) + R_n,\ R_n = o_P(n^{-1/2}),$$

kde

$$\psi(x) = \int\limits_{-\infty}^{+\infty} J(F(y))F(y)\,\mathrm{d}y - \int\limits_{x}^{+\infty} J(F(y))\,\mathrm{d}y$$

za podmienok, že distribučná funkcia F je spojitá, $E(|X|) = \int_0^1 |F^{-1}(u)|\mathrm{d}u < \infty$, funkcia J je spojitá s výnimkou konečného počtu bodov $a_1, \ldots, a_M$ a J' je spojitá s ohraničenou variáciou na $\langle 0, 1\rangle - \{a_1, \ldots, a_M\}$. Na jeho prácu nadviazal Ghosh [2], ktorý ukázal konvergenciu $R_n = \mathcal{O}_P(n^{-1}\log^2(n))$, za podmienok, že J'' je ohraničená, funkcia F je spojitá a spĺňa podmienku: $\int_0^1 [u(1-u)]\mathrm{d}|F^{-1}(u)| < \infty$, z ktorej vyplýva, ako uvádza na str. 355, $E(|X|) < \infty$. Singh [15] takisto za podmienky ohraničenej druhej derivácie funkcie J dokázal, že $R_n = \mathcal{O}_P(n^{-1}\log(\log(n)))$ ak $E(|X|^{1+\delta}) < \infty$ pre nejaké $\delta > 0$ a $R_n = \mathcal{O}_P(n^{-1}\log(n)^{1+\gamma})$ pre každé $\gamma > 0$ ak $E(|X|) < \infty$.

Podmienku konečnej strednej hodnoty nevyžadujú autori práce [3]. Za podmienok:

- $|F^{-1}(u)| \le K[u(1-u)]^{-1/p}$ pre nejaké $0 < p < \infty$ a $0 < K < \infty$ a
- pre nejaké $0 < \delta < 1/2$

$$|J(u)| \le K[u(1-u)]^{1/p-1/2+\delta},\ |J'(u)| \le K[u(1-u)]^{1/p-3/2+\delta},$$

dokázali asymptotickú reprezentáciu pre $L-$odhad s integrálnym typom skóre a konvergenciou $\sqrt{n}R_n \to 0$ skoro všade.

Asymptotická reprezentácia L-odhadov so zvyškovým členom $R_n = \mathcal{O}_P(\frac{1}{n})$, bola dokázaná v [7] kapitola 4. Pre $L-$odhady v tvare

$$\tilde{L}_n = \sum_{i=1}^{n} \tilde{c}_{ni}X_n^{(i)}, \qquad \tilde{c}_{ni} = \int\limits_{\frac{i-1}{n}}^{\frac{i}{n}} J(u)\,\mathrm{d}u\,. \tag{2}$$

predpokladajú, že ak funkcia $J(u)$ je nenulová na intervale $(0, 1)$, teda všetky poriadkové štatistiky majú nenulové skóre, tak spĺňa Lipschitzovu podmienku rádu $\nu > 1$, ktorá pre prípad nekonštantnej diferencovateľnej funkcie nie je splnená. Pre štatistiky so skóre $c_{ni} = \frac{1}{n} J(\frac{i}{n+1})$ vyžadujú konečnosť strednej hodnoty náhodnej premennej s distribučnou funkciou F.

1.2 Robustné odhady a náhodné zrezávanie

Väčšina známych robustných odhadov, je založená na princípe zanedbania určitého počtu najmenších a najväčších hodnôt. Huber [6] str. 112 spomína $\alpha-$useknutú varianciu, ktorá je charakteristikou variability distribučnej funkcie F v tvare:

$$\varsigma^2(F) = \gamma(\alpha) \int_{\alpha/2}^{1-\alpha/2} (F^{-1}(x))^2 \mathrm{d}x \tag{3}$$

a túto hodnotu odhadujeme štatistikou $\varsigma^2(F_n)$. Číslo $\gamma(\alpha)$ je nejaká normujúca konštanta, ktorú môžeme zvoliť tak, aby odhad $\varsigma^2(F_n)$ bol konzistentným odhadom parametra záujmu. Keďže v $\alpha-$useknutej variancii nevystupuje parameter polohy, nebude táto charakteristika vhodná pre všetky typy rozdelení. Predpokladajme, že distribučná funkcia F je rastúca spojitá a symetrická, pod čím rozumieme, že spĺňa podmienku $F(x) = 1 - F(-x),\ x \in \mathbb{R}$. V tomto prípade má táto distribučná funkcia nulový medián a nulovú strednú hodnotu, ak táto existuje, a teda $\alpha-$useknutá variancia je vhodným odhadom parametra škály. Ak distribučná funkcia F nie je symetrická, navrhuje Huber používať upravenú verziu charakteristiky ς^2 v tvare:

$$\tilde{\varsigma}^2(F) = \varsigma^2(\tilde{F}), \tag{4}$$

kde

$$\tilde{F}(x) = \frac{1}{2}(F(x) + 1 - F^-(-x)) \tag{5}$$

je symetrická verzia funkcie F. Ako spomína na str. 113, funkciu $\tilde{F}(x)$ môžeme definovať aj iným spôsobom, napríklad „zosymetrizovaním“ okolo mediánu. Huber vo svojej práci nedokazuje asymptotickú normalitu $\alpha-$useknutej variancie prispôsobenej pre nesymetrické rozdelenia.

Pojem náhodného usekávania (random trimming) spomína Shorack vo svojich článkoch [13] a [14]. V článku [14] dokazuje asymptotickú normalitu $L-$odhadov v tvare

$$T_n = \frac{1}{\sum\limits_{\alpha_n+1}^{\beta_n} c_{ni}} \sum_{\alpha_n+1}^{\beta_n} c_{ni} g(Y_n^{(i)}), \tag{6}$$

kde $Y_n^{(1)}, \ldots Y_n^{(n)}$ je usporiadaný náhodný výber z rovnomerného rozdelenia na intervale $(0,1)$. Shorack predpokladá konvergenciu $\alpha_n/n \xrightarrow{P} a$, $\beta_n/n \xrightarrow{P} b$, kde $0 \le a \le b \le 1$ a symbol $\xrightarrow{P}$ označuje konvergenciu podľa pravdepodobnosti. V prípade, že $a \neq 0$ a $b \neq 1$ dokazuje vo vete 2 asymptotickú normalitu štatistiky T_n za predpokladu $\alpha_n/n - a = o_P(n^{-1/2})$, $\beta_n/n - b = o_P(n^{-1/2})$. Ak sú α_n, β_n výberové kvantily, nebude táto podmienka vo všeobecnosti splnená.

Welsh a Morrison predstavujú vo svojom článku [18] širokú triedu robustných odhadov parametra škály a uvádzajú za akých podmienok sú tieto odhady asymptoticky normálne. Keďže táto trieda nezahrňuje odhady, kde by usekávanie mohlo byť náhodné, metricky useknutá variancia do tejto triedy nepatrí.

Známym robustným odhadom parametra škály je medián absolútnych odchýliek (MAD_n) definovaný vzťahom:

$$MAD_n = b \, \mathrm{med}_i |X_i - \mathrm{med}_j X_j|, \tag{7}$$

kde konštantu b môžeme zvoliť tak, aby daný odhad bol konzistentným odhadom parametra záujmu. Rousseeuw a Croux navrhli vo svojom článku [11] dve alternatívy tohto odhadu, štatistiky S_n a Q_n ktoré sú založené na vzdialenostiach medzi prvkami náhodného výberu. Štatistika S_n je definovaná vzťahom:

$$S_n = c \, \mathrm{med}_i \{ \mathrm{med}_j |X_i - X_j| \}, \tag{8}$$

teda pre každé $i = 1, \ldots, n$ nájdeme medián súboru $|X_i - X_j|$, $j = 1, \ldots, n$ a nakoniec určíme medián z týchto mediánov. Inou možnosťou je uvažovať súbor vzdialeností $|X_i - X_j|$, $i < j$, čo je $\binom{n}{2}$ prvkov. Štatistika Q_n/d je $k-$tou poriadkovou štatistikou tohto súboru, kde $k = \binom{\lfloor n/2 \rfloor + 1}{2}$, teda

$$Q_n = d\{|X_i - X_j|, i < j\}^{(k)}, \tag{9}$$

voľba práve tejto poriadkovej štatistiky zaručuje, že odhad Q_n má 50% bod zlomu. Konštanty c a d volíme tak, aby daný odhad bol konzistentným odhadom parametra záujmu. Odhady S_n a Q_n sú podľa autorov vhodnejšie pre nesymetrické rozdelenia ako MAD_n, pretože v nich nevystupuje odhad parametra polohy. Tieto odhady sú jednoduché na výpočet a oba majú 50% bod zlyhania, čo môže spôsobiť, že ich disperzia bude väčšia ako u menej robustných odhadov.

2 Ciele dizertačnej práce

Ciele dizertačnej práce boli nasledovné:

- Nájsť podmienky, za ktorých platí asymptotická reprezentácia $L-$odhadov neuseknutého typu s rádom konvergencie $\mathcal{O}_{\mathcal{P}}(1/n)$, ktoré by zahŕňali rozdelenia bez strednej hodnoty.
- Definovať pomocou náhodného zrezávania charakteristiku variability distribučnej funkcie a ukázať jej asymptotickú normalitu využitím Hadamardovskej diferencovateľnosti.

3 Výsledky dizertačnej práce

3.1 $L-$odhady

Prvý ciel dizertačnej práce sa nám podarilo naplniť dokázaním nasledujúcej vety, ktorá bola publikovaná v článku [5].

Veta 3.1. *Predpokladajme, že podmienky (A 1) - (A 5) sú splnené a definujme funkciu*

$$\psi(x) = \int_{-\infty}^{+\infty} J(F(y))F(y)\,\mathrm{d}y - \int_{x}^{+\infty} J(F(y))\,\mathrm{d}y\,. \tag{10}$$

(I) Nech

$$\tilde{L}_n = \sum_{i=1}^{n} \tilde{c}_{ni} X_n^{(i)}\,, \qquad \tilde{c}_{ni} = \int_{\frac{i-1}{n}}^{\frac{i}{n}} J(u)\,\mathrm{d}u\,. \tag{11}$$

Potom

$$\tilde{L}_n = \mu + \frac{1}{n}\sum_{i=1}^{n} \psi(X_i) + \mathcal{O}_P(\frac{1}{n})\,. \tag{12}$$

(II) Nech

$$L_n^* = \sum_{i=1}^{n} c_{ni} X_n^{(i)}\,, \qquad c_{ni} = \frac{1}{n} J(\frac{i}{n+1})\,. \tag{13}$$

Okrem platnosti podmienok (A 1) - (A 5) predpokladajme, že pre nejaké $t_d\,, t_D \in (d, D)$ *a* $\beta_d > 0$, $\beta_D > 0$ *platia nasledujúce nerovnosti:*

$$\sup\{|x|^{\beta_d} F(x);\, d < x \le t_d\} < +\infty\,, \tag{14}$$

$$\sup\{|x|^{\beta_D}(1 - F(x));\, t_D \le x < D\} < +\infty.$$

Ak

$$\delta_1 = 2 + \gamma_d - \frac{1}{\beta_d} > 0\,, \qquad \delta_2 = 2 + \gamma_D - \frac{1}{\beta_D} > 0\,, \tag{15}$$

potom

$$L_n^* = \mu + \frac{1}{n}\sum_{i=1}^{n}\psi(X_i) + R_n\,, \tag{16}$$

kde

$$R_n = \begin{cases} \mathcal{O}_P\left(\frac{\log n}{n}\right), & \min\{\delta_1, \delta_2\} = 1\,, \\ \mathcal{O}_P\left(\frac{1}{n^{\delta^*}}\right) & \text{inak}\,. \end{cases} \tag{17}$$

Kde

$$\delta^* = \min\{1\,,\,\delta_1\,,\delta_2\}\,. \tag{18}$$

Keďže predpoklady vety zaručujú rovnosť: $\int_{-\infty}^{\infty}\psi(x)\mathrm{d}F(x) = 0$, ak $V = \int_{-\infty}^{+\infty}\psi^2(x)\,\mathrm{d}F(x) < +\infty$, pomocou centrálnej limitnej vety dostávame konvergenciu $\sqrt{n}(\tilde{L}_n - \mu) \to N(0,V)$ podľa rozdelenia. Navyše, ak pre zvyškový člen R_n z rovnice (17) platí rovnosť $R_n = o_P(n^{-1/2})$, potom $\sqrt{n}(L_n^* - \mu) \to N(0,V)$ podľa rozdelenia.

Pomocou známych metód sme odvodili $L-$odhad parametra škály Burrovho rozdelenia typu XII, ktorého distribučná funkcia má tvar:

$$F(x,\mu,\sigma,k,c) = 1 - \left(1 + \left(\frac{x-\mu}{\sigma}\right)^c\right)^{-k} \quad \text{pre} \quad x \geq \mu, \tag{19}$$

kde $k > 0, c > 0, \sigma > 0,\ \mu \in \mathcal{R}$. Definujme funkciu

$$\phi(x) = \frac{1}{\mathcal{I}(\sigma)}\frac{d}{d\sigma}\ln\left(f(x,\mu,\sigma,k,c)\right)$$

kde $\mathcal{I}(\sigma)$ je Fisherova informácia. Pre $c > 1/2,\ k > 0$ definujme funkciu generujúcu skóre

$$J(u) = \phi^{'}(F^{-1}(u)) = \frac{(k+1)(k+2)}{k}(1-u)^{\frac{2}{k}}((1-u)^{-\frac{1}{k}} - 1)^{\frac{c-1}{c}}\,, \quad u \in (0,1)$$

a odhad parametra σ Burrovho rozdelenia

$$\hat{\sigma}_n = \sum_{i=1}^{n} c_{ni}\left(X_n^{(i)} - \mu\right),\ c_{ni} = \int\limits_{(i-1)/n}^{i/n} J(u)\mathrm{d}u. \tag{20}$$

Pre $c > \frac{1}{2}$, $k > 0$ funkcie $J(u)$ a $F(u)$ spĺňajú podmienky vety 3.1 (I) s $\gamma_d = -\frac{1}{c}$, $\gamma_D = \frac{1}{k} + \frac{1}{kc} - 1$, $\kappa_d = -\frac{1}{c} + \Delta_d$, kde $0 < \Delta_d < 1$, $\kappa_D = \frac{1}{kc} + \Delta_D$, kde $0 < \Delta_D < \frac{1}{k}$, preto asymptotická reprezentácia (12) platí. Ďalej $V = \int_{-\infty}^{\infty} \psi^2(x)\,\mathrm{d}F(x) = \frac{2+k}{kc^2}\sigma^2 = \frac{1}{\mathcal{I}(\sigma)}$. Keďže $\int J(u)F^{-1}(u)\mathrm{d}u = \sigma$, pre štatistiku $\hat{\sigma}_n$ platí konvergencia

$$\sqrt{n}\,(\hat{\sigma}_n - \sigma) \to N(0, V) \tag{21}$$

podľa rozdelenia.

Skóre c_{ni} definované v (20), môžu byť aproximované hodnotami $\tilde{c}_{ni} = \frac{1}{n}J(\frac{i}{n+1})$. Keďže podmienky vety 3.1 (II) sú splnené pre $c > \frac{1}{2}$, $k > 0$ s $\beta_d = -c$, $\beta_D = ck$, reprezentácia (16) platí s $R_n = \mathcal{O}_P(1/n)$.

Ak parameter μ nie je známy, je prirodzené použiť odhad $\hat{\mu} = X_n^{(1)}$ a odhad parametra škály upraviť na tvar: $\hat{\hat{\sigma}}_n = \sum_{i=1}^n c_{ni}(X_n^{(i)} - \hat{\mu})$. Keďže pre $c < 2$ platí $\{X_n^{(1)} - \mu\} = o_p(1/\sqrt{n})$, pre $\frac{1}{2} < c < 2$

$$\sqrt{n}\left(\hat{\hat{\sigma}}_n - \sigma\right) \to N(0, V) \tag{22}$$

podľa rozdelenia, kde $V = \frac{1}{\mathcal{I}(\sigma)} = \frac{2+k}{kc^2}\sigma^2$. V článkoch [2],[10],[15],[17] o asymptotickej reprezentácii $L-$odhadov sa predpokladá, že stredná hodnota rozdelenia, z ktorého pochádza náhodný výber je konečná, v tomto prípade by parametre k a c Burrovho rozdelenia typu XII museli spĺňať podmienku $kc > 1$. Podmienky na asymptotickú reprezentáciu z práce [3] sú splnené, ak $c > 3/2, k > 2/(3c-2)$.

4 Metricky useknutá variancia

Variabilitu distribučnej funkcie F založenú na náhodnom zrezávaní sme navrhli charakterizovať hodnotou $S(F)$, kde

$$S(F) = \frac{1}{1-\alpha} \int\limits_{F(\xi(F)-\lambda(F))}^{F(\xi(F)+\lambda(F))} (F^{-1}(x))^2 \mathrm{d}x - (T(F))^2, \tag{23}$$

$$\xi(F) = F^{-1}(1/2) \tag{24}$$

$$\lambda(F) = K_F^{-1}(1-\alpha),\ K_F(x) = F(\xi(F)+x) - F^{-}(\xi(F)-x) \tag{25}$$

a

$$T(F) = \frac{1}{1-\alpha} \int\limits_{F(\xi(F)-\lambda(F))}^{F(\xi(F)+\lambda(F))} F^{-1}(x)\mathrm{d}x. \tag{26}$$

Pomocou Hadamarodvskej diferencovateľnosti sme dokázali nasledujúce tvrdenie:

Veta 4.1. *Nech náhodná premenná s distribučnou funkciou F má diferencovateľnú hustotu, ktorej derivácia je ohraničená. Predpokladajme ďalej, že derivácia funkcie $F^{-1}(x)$ je ohraničená na intervale $\langle \varepsilon_1, 1-\varepsilon_2 \rangle$, kde $\varepsilon_1 < F(\xi(F) - \lambda(F))$, ak $F(\xi(F) - \lambda(F)) \neq 0$, inak $\varepsilon_1 = 0$ a $1 - \varepsilon_2 > F(\xi(F) + \lambda(F))$, ak $F(\xi(F)+\lambda(F)) \neq 1$, inak $\varepsilon_2 = 0$. Potom, ak pre influenčnú funkciu $IF(x, F, S)$ platí $0 < \int_{-\infty}^{\infty} IF^2(x, F, S)\mathrm{d}F(x)$,*

$$n^{1/2}(S(F_n) - S(F)) \to N(0, V) \tag{27}$$

podľa rozdelenia, kde $V = \int_{-\infty}^{\infty} IF^2(x, F, S)\mathrm{d}F(x)$.

Aby bolo možné vypočítať asymptotickú varianciu z predchádzajúcej vety, odvodili sme explicitný vzorec pre influenčnú funkciu. Za predpokladov vety 4.1 má táto funkcia tvar:

$$IF(x, F, S) = \frac{1}{1-\alpha}[(\xi+\lambda)^2 f(\xi+\lambda)(\xi'+\lambda')- \tag{28}$$

$$-(\xi-\lambda)^2 f(\xi-\lambda)(\xi'-\lambda') + x^2\ I_{\{\xi-\lambda<x<\xi+\lambda\}}] - S(F)-$$

$$-2T(F)\frac{1}{1-\alpha}[(\xi+\lambda)f(\xi+\lambda)(\xi'+\lambda')-$$

$$-(\xi-\lambda)f(\xi-\lambda)(\xi'-\lambda') + x\ I_{\{\xi-\lambda<x<\xi+\lambda\}}] + T(F)^2,$$

kde

$$\xi' = \frac{sign(x-\xi)}{2f(\xi)}, \tag{29}$$

$$\lambda' = \frac{(1-\alpha) - (f(\alpha_2) - f(\alpha_1))\xi' - I[\alpha_1 < x < \alpha_2]}{f(\alpha_2) + f(\alpha_1)}. \tag{30}$$

Táto funkcia je ohraničená, čo zaručuje konečnosť asymptotickej variancie.

V práci je simulačné porovnanie s inými typmi odhadov parametra škály. Výberová metricky useknutá variancia vykazovala v niektorých prípadoch lepšie vlastnosti, najmä pre malé rozsahy výberov.

5 Práce autora

1. HoRu Hönschová, E. a Rublík, F. (2009), *On asymptotic linearity of L-estimates.* Math. Slovaca, 59 , No. 6., 667-678.

2. Robust Hönschová, E. (2008), *Estimation of the scale parameter in Burr type XII distribution.* Robust 2008.

6 Zhrnutie

Výsledky práce môžeme zhrnúť v nasledujúcich bodoch

- V práci je dokázaná veta o asymptotickej reprezentácii L-odhadov za podmienok, ktoré nevyžadujú konečnú strednú hodnotu rozdelenia, z ktorého pochádza náhodný výber. Reprezentácia je pre prípad L-odhadov s integrálnym skóre dokázaná s rádom zvyšku $R_n = \mathcal{O}_P(\frac{1}{n})$ a prípad $L-$odhadov s bodovým skóre je tiež zahrnutý.
- Odvodili sme $L-$odhad parametra škály Burrovho rozdelenia typu XII, a dokázali sme jeho asymptotickú normalitu a to aj pre prípad, keď parametre k a c tohto rozdelenia nespĺňajú rovnosť $kc > 1$, ktorá zaručuje konečnú strednú hodnotu.
- Prezentovali sme novú charakteristiku variability distribučnej funkcie, založenú na náhodnom zrezávaní a dokázali sme jej asymptotickú normalitu.

7 Summary

The main results of the thesis can be summarized as follows:

- A theorem on asymptotic linearity of L-estimates is proved under the general set of regularity conditions, allowing the sampled distribution to be non-integrable. The main result is the improvement in the order of the remainder term in the formula for asymptotic linearity of the L-statistic. It is shown that in the case of the integral coefficients this term $R_n = \mathcal{O}_P(\frac{1}{n})$ and the case of functional coefficients is also covered.
- An L-estimate of the scale parameter of the Burr distribution is presented and its asymptotic normality is proved. The parameters k and c of this distribution are assumed to fulfill the inequality $c > \frac{1}{2}$, hence the estimate is defined and asymptotically normal, covering the case of infinite mean of this distribution.
- A robust scale estimate, a generalized version of the α-trimmed variance mentioned in [6], is presented. Since the trimming depends on the distance of the data from the sample median, this estimate is useful also for asymetric distributions. The asymptotic normality of the estimate is proved using the theory on Hadamard differentiable functionals.

Referencie

[1] Chernoff, H., Gastwirth, J. L. a Johns, M. V. Jr. (1967), *Asymptotic distribution of linear combinations of order statistics with applications to estimation.* Ann. Math. Statist, 38, 52-72.

[2] Ghosh, M. (1972), *On the representation of linear functions of order statistics.* Sankhyã: The Indian Journal of Statistics, 34, No.4, 349-356.

[3] Govindarajulu,Z. a Mason, D.M. (1983), *A Strong Representation for Linear Combinations of Order Statistics with Applications to Fixed-width Confidence Intervals for Location and Scale Parameters.* Scan. J. Statist, 10, No. 2, 97-115.

[4] Hájek, J. (1968), *Asymptotic normality of simple linear rank statistics under alternatives.* Ann. Math. Statist, 39, 325-346.

[5] Hönschová, E. a Rublík, F. (2009), *On asymptotic linearity of L-estimates.* Math. Slovaca, 59 , No. 6., 667-678.

[6] Huber, P.J. (1981), *Robust statistics.* John Wiley & Sons, Inc.

[7] Jurečková, J. a Sen, P. K. (1996), *Robust Statistical Procedures. Asymptotic and Interrelations.* John Wiley & Sons, New York.

[8] Kim, S. (1992), *The metrically trimmed mean as a robust estimator of location.* The Annals of Statistics, 20, No. 3, 1534-1547.

[9] Kulldorf, G. a Vännman, K. (1973), *Estimation of the Location and Scale Parameters of a Pareto Distribution by Linear Functions of Order Statistics.* Journal of the American Statistical Association, 68, 218-227.

[10] Moore, D. S. (1968), *An elementary proof of asymptotic normality of linear functions of order statistics.* Ann. Math. Statist, 39, 263-265.

[11] Rousseeuw, P.J., Croux, C. (1993), *Alternatives to the median Absolute Deviation.* Journal of the American Statistical Association, 88, No. 424.

[12] Shorack, G.R. (1972), *Functions of Order Statistics.* Ann. Math. Statist, 43,412-427.

[13] Shorack, G.R. (1974), *Random means.* Ann. Math. Statist, 2, 661-675.

[14] Shorack, G.R. (1989), *Randomly trimmed L-statistics.* Journal of Statistical Planning and Inference, 31, 293-304.

[15] Singh, K. (1981), *On asymptotic representation and approximation to normality of L-statistics-I.* Sankhyã: The Indian Journal of Statistics, 34, 67-83.

[16] Stigler, S. M. (1969), *Linear functions of order statistics.* Ann. Math. Statist, 40, No. 3, 770-788.

[17] Stigler, S. M. (1974), *Linear functions of order statistics with smooth weight functions.* Ann. Math. Statist, 2, No. 4, 676-693.

[18] Welsh, A.H., Morrison, H.L. (1990), *Robust L Estimation of Scale With an Application in Astronomy.* Journal of the American Statistical Association, 85, No. 411.