TESTY ZALOŽENÉ NA REGRESNÍCH POŘADOVÝCH SKÓRECH

Jan PICEK
TU Liberec, KPDM

Abstract: In this paper we construct a class of regression rank scores tests in the linear mixed model where some of the predictors are nonstochastic and some are stochastic. The tests are based on regression rank scores, introduced by Gutenbrunner and Jurečková (1992) as dual variables to the regression quantiles of Koenker and Bassett (1978). Their properties are analogous to those of the corresponding rank tests in location model.

Резюме: В этой статье мы конструируем класс критериев в линейной смешанной модели, т. е., некоторые столбце регрессионной матрицы случайные и некоторые неслучайные. Критерии выведёны на основе регрессионных ранговых меток, которые ввели Ц. Гутенбруннер и Й. Юречкова (1992) как двухственные переменные к регрессыонным квантилам, которые ввели в 1978 году Р. Коенкер и Г. Бассетт. Свойства этих критериев анлогичны свойствам соответствуюших ранговых критериев в модели сдвига.

1. Úvod

Budeme uvažovat následující lineární regresní model

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

kde $\mathbf{Y} = (Y_1, \ldots, Y_n)$ je vektor pozorování, $\mathbf{X}$ je $n \times p$ rozměrná regresní matice vysvětlujících proměnných, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in I\!R^p$ je vektor neznámých parametrů a $\mathbf{E} = (E_1, \ldots, E_n)$ je vektor chyb.

Snaha po zobecnění L-odhadů parametru polohy na regresní model vedla k zavedení pojmu regresní α-kvantil. R. Koenker a G. Basset [11] ho definovali v modelu (1) následovně:
Regresním α-kvantilem $\widehat{\boldsymbol{\beta}}(\alpha)$ $(0 < \alpha < 1)$ v modelu lineární regrese (1) nazýváme každý vektor $\mathbf{t} \in I\!R^p$, který je řešením

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - \mathbf{x}_i'\mathbf{t}) := \min_{\mathbf{t}}, \tag{2}$$

kde

$$\rho_\alpha(x) = x\psi_\alpha(x), \; x \in I\!R^1$$

a

$$\psi_\alpha(x) = \alpha - I_{[x<0]},\ x \in \mathbb{R}^1.$$

Studiem asymptotických vlastností regresních kvantilů a konstrukcí odhadů na nich založených se dále např. zabývali D. Ruppert a R. Carrol [13], Jurečková [7, 8] , Antoch a Jurečková [1], C. Gutenbrunner [2].

Koenker a Basset v [11] charakterizovali regresní α - kvantil $\widehat{\boldsymbol{\beta}}(\alpha)$ jako složku $\widehat{\boldsymbol{\beta}}$ optimálního řešení úlohy parametrického lineárního programování, která má následující tvar:

$$\begin{aligned} \alpha \mathbf{1}_n' \mathbf{u}^+ + (1-\alpha)\mathbf{1}_n' \mathbf{u}^- &:= \min \\ \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{u}^+ - \mathbf{u}^- &= \mathbf{Y} \\ \widehat{\boldsymbol{\beta}} \in \mathbb{R}^p, \mathbf{u}^+, \mathbf{u}^- \in \mathbb{R}_+^n \quad & 0 < \alpha < 1 \end{aligned} \tag{3}$$

Autoři též uvažovali i duální úlohu, ale její řešení využívali jen k výpočtu regresních kvantilů. Optimální řešení $\hat{\mathbf{a}}(\alpha) = (\hat{a}_1(\alpha), \ldots, \hat{a}_n(\alpha))' \in \mathbb{R}^n$ její ekvivalentní verze duální úlohy

$$\begin{aligned} \mathbf{Y}'\hat{\mathbf{a}} &:= \max \\ \mathbf{X}'\hat{\mathbf{a}} &= (1-\alpha)\mathbf{X}'\mathbf{1}_n \\ \hat{\mathbf{a}} &\in [0,1]^n \end{aligned} \tag{4}$$

nazvali Gutenbrunner a Jurečková [3] *regresními pořadovými skóry* a ukázali, že dualitu mezi pořádkovými statistikami a pořadím v modelu polohy lze přirozeně zobecnit na klasický lineární regresní model (1).

Motivací k zavedení regresních pořadových skórů je, že v případě modelu polohy (tj. $\mathbf{X} = \mathbf{1}_n$)

$$\hat{a}_i(\alpha) \equiv a^*(R_i, \alpha) \equiv \begin{cases} 1 & \text{jestliže } \alpha \le (R_i - 1)/n \\ R_i - \alpha n & \text{jestliže } (R_i - 1)/n < \alpha \le R_i/n \\ 0 & \text{jestliže } R_i/n < \alpha \end{cases}, \tag{5}$$

kde R_i je pořadí Y_i mezi $Y_1, \ldots, Y_n$. Funkce $a^*(j, \alpha),\ j = 1, \ldots, n,\ 0 < \alpha < 1$ je přesně táž, kterou zavedl Hájek v r. 1965.

Navrhl tehdy rozšíření Kolmogorov-Smirnovova testu, jehož kritérium je funkcionál procesu $\{T_n(\alpha) = \sum_{i=1}^n c_{ni} a_n^*(R_i, \alpha),\ 0 \le \alpha \le 1\}$ a ukázal, že asymptotické rozdělení tohoto kritéria je shodné s asymptotickým rozdělením obyčejného Kolmogorov-Smirnovova testu. Nejen Kolmogorov-Smirnovův test, ale i standardní (lineární) pořadové testy mohou být vyjádřeny jako funkcionály procesu $T_n(\alpha)$.

Přirozeně se tedy naskytla otázka, zda lze konstruovat testy založené na regresních pořadových skórech jako analogii pořadových testů. První idea se objevila v práci Gutenbrunnera a Jurečkové v [3], teorie však byla aplikovatelná pouze na testy s useknutou skórovou funkcí. Obecná třída testů založených na regresních pořadových skórech byla zkonstruovaná v článku Gutenbrunnera, Jurečkové, Koenkera a Portnoye [4]. Hlavním metodologickým nástrojem pro odvození asymptotických vlastností testů je asymptotická reprezentace regresních kvantilů, stejnoměrná v α. Testy Kolmogorov–Smirnovova typu založených na regresních pořadových skórech odvodila Jurečková [9]. Algoritmus pro výpočet regresních pořadových skórů popsali R. Koenker a V. d'Orey v [12]. V tomto příspěvku se budeme zabývat testy založenými na regresních pořadových skórech ve smíšeném modelu, tj. za předpokladu, že některé sloupce regresní matice jsou náhodné.

2. Vlastnosti regresních pořadových skórů

Z duality mezi $\widehat{\boldsymbol{\beta}}(\alpha)$ a $\hat{\mathbf{a}}(\alpha)$ vyplývá

$$\hat{a}_i(\alpha) = \begin{cases} 1 & \text{jestliže } E_i > \mathbf{x}_i'(\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha)) \\ 0 & \text{jestliže } E_i < \mathbf{x}_i'(\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha)) \end{cases} \tag{6}$$

a

$$\sum_{i \in M_\alpha} \hat{a}_i(\alpha)\mathbf{x}_i = (1-\alpha)\sum_{i=1}^{n} \mathbf{x}_i - \sum_{i=1}^{n} I[E_i > \mathbf{x}_i'(\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha))]\mathbf{x}_i,$$

kde

$$M_\alpha = \{i : E_i = \mathbf{x}_i'(\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha))\}.$$

Z vlastností regresních pořadových skórů při konečném n je patrně nejdůležitější jejich invariance vzhledem k regresi s maticí $\mathbf{X}$, plynoucí přímo z definice (4):

$$\hat{\mathbf{a}}(\alpha, \mathbf{Y} + \mathbf{X}\mathbf{b}) = \hat{\mathbf{a}}(\alpha, \mathbf{Y}), \ \forall\, b \in \mathbb{R}^p, \tag{7}$$

která je rozšířením invariance pořadí (nebo pořadových skórů (5)) vzhledem k posunutí v poloze.

Na základě reprezentace regresních kvantilů Gutenbrunner a kol. v [4] odvodili asymptotickou reprezentaci a rozdělení procesu regresních pořadových skórů v modelu (1), přičemž uvažovali následující předpoklady.

Chyby $E_1, \ldots, E_n$ jsou nezávislé stejně rozdělené náhodné veličiny s distribuční funkcí F a hustotou f, které splňují tyto podmínky

(A.1) $|F^{-1}(\alpha)| \le c(\alpha(1-\alpha))^{-a}$ pro $0 < \alpha \le \alpha_0$, $1 - \alpha_0 \le \alpha < 1$, kde $0 < a \le 1/4 - \varepsilon$, $\varepsilon > 0$ a $c > 0$.

(A.2) $\dfrac{1}{f(F^{-1}(\alpha))} \le c(\alpha(1-\alpha))^{-1-a}$ pro $0 < \alpha \le \alpha_0,\ 1-\alpha_0 \le \alpha < 1,\ c > 0.$

(A.3) Hustota $f(x)$ je absolutně spojitá, je kladná a omezená na na intervalu (A, B) a klesající pro $x \to A+$ a $x \to B-$, kde

$$-\infty \le A \ \equiv \sup\{x : F(x) = 0\} \ \text{ a } \ \infty \ge B \equiv \inf\{x : F(x) = 1\}.$$

Derivace f' je omezená.

(A.4) $\left|\dfrac{f'(x)}{f(x)}\right| \le c|x|$ pro $|x| \ge K \ \ge 0,\ c > 0.$

Matice $\mathbf{X}$ vyhovuje těmto požadavkům

(B.1.) $x_{i1} = 1,\ i = 1,\ \ldots,\ n.$

(B.2.) $\lim_{n\to\infty} \mathbf{D}_n = \mathbf{D}$, kde matice $\mathbf{D}_n = n^{-1}\mathbf{X}'\mathbf{X}$ a $\mathbf{D}$ je pozitivně definitní $p \times p$ rozměrná matice.

(B.3.) $n^{-1}\sum_{i=1}^n ||x_i||^4 = O(1)$ pro $n \to \infty$.

(B.4.) $\max_{1\le i\le n} ||x_i|| = O\left(n^{(2(b-a)-\delta)/(1+4b)}\right)$ pro nějaké $b > 0$ a $\delta > 0$ takové , že $0 < b - a < \varepsilon/2$.

Poznamenejme, že výše uvedené podmínky splňuje např. normální, logistické a t rozdělení s pěti a více stupni volnosti.

Věta 1 *Nechť* $\mathbf{d}_n = (d_{n1}, \ldots, d_{nn})'$ *je vektor splňující následující podmínky:*

(C.1)

$$\mathbf{X}'_n\mathbf{d}_n = \mathbf{0}, \quad \frac{1}{n}\sum_{i=1}^n d_{ni}^2 \to \Delta^2, \quad 0 < \Delta^2 < \infty$$

(C.2)

$$n^{-1}\sum_{i=1}^n |d_{ni}|^3 = O(1) \ \textit{pro}\ n \to \infty$$

(C.3)

$$\max_{1\le i\le n} |d_{ni}| = O\left(n^{(2(b-a)-\delta)/(1+4b)}\right),$$

kde δ, a, b *jsou dány podmínkou (B.4). Dále nechť jsou v modelu (1) splněny podmínky (A.1)-(A.4) a (B.1)-(B.4). Potom platí* *i)*

$$\sup_{0 \le \alpha \le 1} \left\{ \left| n^{-1/2} \sum_{i=1}^{n} d_{ni}(\hat{a}_{ni}(\alpha) - \tilde{a}_i(\alpha)) \right| \right\} \to 0 \ \textit{pro}\ n \to \infty, \tag{8}$$

kde

$$\tilde{a}_i(\alpha) = I\,[E_i > \alpha] \quad i = 1, \ldots, n. \tag{9}$$

ii)

$$\textit{Proces}\ \left\{ \Delta^{-1} n^{-1/2} \sum_{i=1}^{n} d_{ni}\hat{a}_{ni}(\alpha) : 0 \le \alpha \le 1 \right\} \tag{10}$$

konverguje k Brownovu můstku v Prochorově topologii na $C[0,1]$.

3. Lineární regresní pořadová statistika

Protože ve třídě lineárních pořadových testů důležitou roli zaujímají lineární pořadové statistiky, uvažovali Gutenbrunner a kol. [4] analogicky *lineární regresní pořadovou statistiku*

$$S_n = n^{-1/2} \sum_{i=1}^{n} d_{ni}\hat{b}_{ni} \tag{11}$$

se skóry $\hat{\mathbf{b}}_n = (\hat{b}_{n1}, \ldots, \hat{b}_{nn})'$ generovanými neklesající skórovou funkcí

$$\varphi(t) \ : \ (0,1) \to I\!R^1; \quad 0 < \int_0^1 \varphi^2(t)dt < \infty \tag{12}$$

a definovanými jako integrál

$$\hat{b}_{ni} = -\int_0^1 \varphi(t)d\hat{a}_{ni}(t), \quad i = 1, \ldots, n. \tag{13}$$

Odvodili též asymptotickou reprezentaci statistiky S_n (11).

Věta 2 *Nechť $\varphi(t)$ je neklesající funkce daná v (12) taková, že její derivace $\varphi'(t)$ existuje pro $0 < t < \alpha_0,\ 1 - \alpha_0 < t < 1$ a splňuje*

$$\|\varphi'(t)\| \le c(t(1-t))^{-1-\delta^*} \tag{14}$$

pro nějaké $\delta^* \le \delta$, kde δ splňuje (B.4), a pro $t \in (0, \alpha_0) \cup (1 - \alpha_0, 1)$. Jestliže jsou splněny podmínky (A.1)-(A.4), (B.1)-(B.4) a (C.1)-(C.3) potom S_n má reprezentaci

$$S_n = n^{-1/2} \sum_{i=1}^{n} d_{ni}\varphi(F(E_i)) + o_p(1). \tag{15}$$

Tuto větu rozšíříme na situaci, kdy koeficienty v lineární regresní pořadové statistice jsou náhodné. Označíme-li koeficienty jako $z_{1n} \dots z_{nn}$, potom místo podmínek (C.1)-(C.3) budeme uvažovat tyto předpoklady:

(D.1) $\mathbf{z}_n = (z_{1n}, \dots, z_{nn})'$ je nezávislý náhodný výběr z rozdělení s distribuční funkcí G, která má spojitou hustotu g.

(D.2) $\mathbb{E}|z_1|^3 < \infty$.

(D.3) Vektory $\mathbf{z}_n$ a $\mathbf{E}_n$ jsou nezávislé.

Jako H označme sdruženou distribuční funkci $G \cdot F$.

Věta 3 *Nechť $\varphi(t)$ je neklesající funkce daná v (12) taková, že její derivace $\varphi'(t)$ existuje pro $0 < t < \alpha_0,\ 1 - \alpha_0 < t < 1$ a platí pro ní*

$$|\varphi'(t)| \le c(t(1-t))^{-1-\delta^*} \tag{16}$$

pro nějaké $\delta^* < \delta$, kde δ splňuje (B.4), a pro $t \in (0, \alpha_0) \cup (1 - \alpha_0, 1)$. Potom za platnosti předpokladů (A.1)-(A.4), (B.1)-(B.4), (D.1)-(D.3) má statistika S_n asymptotickou reprezentaci

$$S_n = n^{-1/2} \sum_{i=1}^{n} (z_{ni} - \widehat{z}_{ni}) \varphi(F(E_i)) + o_p(1), \tag{17}$$

kde

$$\widehat{z}_{ni} = \mathbf{H}_n z_{ni},\ i = 1, \dots, n \quad a\ \mathbf{H}_n = \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n$$

Důkaz. Pro potřeby důkazu označme $r_i = n^{-1/2}(\widehat{b}_i - \varphi(F(E_i)))$, $i = 1, \dots, n$ a statistiku $S_n^* = n^{-1/2} \sum_{i=1}^n (z_{ni} - \widehat{z}_{ni}) \varphi(F(E_i))$. Cílem je tedy ukázat, že $S_n - S_n^* = o_p(1)$.

Z podmínky (D.3) plyne, že $\mathbf{r} = (r_1, \dots, r_n)'$ a $(\mathbf{z}_n - \widehat{\mathbf{z}}_n)$ jsou nezávislé. Předpokládejme ještě, že rozptyl $z_{1n} = \sigma^2$.

Nejprve počítejme následující podmíněný moment

$$\begin{aligned}
\mathbb{E}_G \left[(S_n - S_n^*)^2 \middle| E_1, \dots, E_n \right] &= \mathbb{E}_G \left[(\mathbf{z}_n - \widehat{\mathbf{z}}_n)' \, \mathbf{r} \, (\mathbf{z}_n - \widehat{\mathbf{z}}_n)' \, \mathbf{r} \middle| E_1, \dots, E_n \right] \\
&= \mathbf{r}' \, \mathbb{E}_G \left[(\mathbf{I}_n - \mathbf{H}_n) \mathbf{z}_n \mathbf{z}'_n (\mathbf{I}_n - \mathbf{H}_n)' \middle| E_1, \dots, E_n \right] \mathbf{r} \ = \\
&= \ \mathbf{r}' (\mathbf{I}_n - \mathbf{H}_n) \sigma^2 \mathbf{I}_n (\mathbf{I}_n - \mathbf{H}_n) \mathbf{r}
\end{aligned}$$

Stačí si uvědomit, že vzhledem k (B.1) platí $(\mathbf{I}_n - \mathbf{H}_n)\mathbf{1}_n = \mathbf{0}_n$, a tedy $(\mathbf{I}_n - \mathbf{H}_n)\mathbb{E}\mathbf{Z}_n = \mathbf{0}_n$. Dále víme, že projekční matice je idempotentní a symetrická.

Z definice podmíněné střední hodnoty tak získáváme

$$\int_{\mathbf{B}} (S_n - S_n^*)^2 \, dH(E_1), \ldots, dH(E_n) =$$
$$= \sigma^2 \int_{\mathbf{B}} ((\mathbf{I} - \mathbf{H})\mathbf{r})' \, ((\mathbf{I} - \mathbf{H})\mathbf{r}) \, dF(E_1), \ldots, dF(E_n)$$

pro každé $\mathbf{B} \in \mathbb{R}^n$.
Tvrzení potom vyplývá z věty 2 ($(\mathbf{I}_n - \mathbf{H}_n)\mathbf{r} = o_p(1)$). □

4. Testy ve smíšeném modelu

Uvažujme nyní lineární regresní model (1) ve tvaru

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{Z}_n \boldsymbol{\gamma} + \mathbf{E}_n, \tag{18}$$

kde $\boldsymbol{\beta}$ a $\boldsymbol{\gamma}$ jsou r a q rozměrné neznámé parametry, $\mathbf{X}_n$ je pevná $n \times r$ rozměrná pevná matice, $\mathbf{Z}_n$ je $n \times q$ rozměrná náhodná matice, $\mathbf{Y}$ je n rozměrný vektor pozorování a $\mathbf{E}$ je n rozměrný vektor nezávislých stejně rozdělených chyb.

Zajímat nás bude problém testu hypotézy

$$H_0 : \boldsymbol{\gamma} = 0, \qquad \boldsymbol{\beta} \text{ je neurčeno} \tag{19}$$

proti Pitmanově alternativě

$$H_n : \boldsymbol{\gamma} = n^{-1/2} \boldsymbol{\gamma}_0, \qquad (\boldsymbol{\gamma}_0 \in \mathbb{R}^q \text{ pevné}). \tag{20}$$

Předmětem našich úvah budou lineární testy založené na regresních pořadových skórech, které jsou podobně jako obvyklé pořadové testy založeny na lineárních regresních pořadových statistikách. Protože za platnosti H_0 regresní pořadové skóry odpovídají submodelu

$$\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{E}_n \tag{21}$$

pracujeme při jejich výpočtu pouze s maticí $\mathbf{X}_n$, tj. tedy s nenáhodnou maticí. Lze tedy využít výsledků z předcházející části.

Budeme proto uvažovat, že matice $\mathbf{Z}_n$ splňuje následující podmínky:

(E.1) Matice $\mathbf{Z}_n$ má nezávislé stejně rozdělené řádky $\mathbf{z}_i$, $i = 1, \ldots, n$. Jde tedy o nezávislý náhodný výběr z rozdělení vektoru $\mathbf{z} = (z_1, \ldots, z_q)'$ s q rozměrnou distribuční funkcí G, která má spojitou hustotu g.

(E.2) $\mathbb{E}\|\mathbf{z}\|^3 < \infty$.

(E.3) Matice $\mathbf{Z}_n$ a vektor $\mathbf{E}_n$ jsou nezávislé.

Jako H označme sdruženou distribuční funkci $G \cdot F$, resp. h jako sdruženou hustotu $g \cdot f$.

Zaveďme ještě následující označení

$$\mathbf{Q}_n = n^{-1}(\mathbf{Z}_n - \widehat{\mathbf{Z}}_n)'(\mathbf{Z}_n - \widehat{\mathbf{Z}}_n), \tag{22}$$

kde

$$\widehat{\mathbf{Z}}_n = \mathbf{H}_n \mathbf{Z}_n \text{ a } \mathbf{H}_n = \mathbf{X}_n(\mathbf{X}'{}_n \mathbf{X}_n)^{-1}\mathbf{X}'{}_n. \tag{23}$$

Tedy $\widehat{\mathbf{Z}}_n$ je projekce matice $\mathbf{Z}_n$ do prostoru tvořeného sloupci matice $\mathbf{X}_n$.

Jako testovou statistiku pro test hypotézy (19) uvažujeme

$$T_n = \frac{\mathbf{S}_n' \mathbf{Q}_n^{-1} \mathbf{S}_n}{A^2(\varphi)}, \tag{24}$$

kde

$$A^2(\varphi) = \int_0^1 (\varphi(t) - \overline{\varphi})^2 dt, \quad \overline{\varphi} = \int_0^1 \varphi(t) dt$$

a

$$\mathbf{S}_n = n^{-1/2}(\mathbf{Z}_n - \widehat{\mathbf{Z}}_n)'\hat{\mathbf{b}}_n,$$

přičemž skóry $\mathbf{b}_n = (\hat{b}_{n1}, \ldots, \hat{b}_{nn})'$ jsou vypočteny na základě regresních pořadových skórů odpovídajících submodelu (21).

Test je založen na asymptotickém rozdělení T_n za platnosti H_0 daném následující větou, která zároveň udává asymptotické rozdělení T_n za lokální alternativy H_n.

Věta 4 *Předpokládejme, že matice $\mathbf{X}_n$ splňuje podmínky (B.1)-(B.4) a matice $\mathbf{Z}_n$ podmínky (E.1)-(E.3). Dále předpokládejme, že F splňuje (A.1)-(A.4). Nechť T_n je statistika definovaná v (24), přičemž skórová funkce φ daná v (12) splňuje (16). Předpokládejme, že existuje $\lim_{n\to\infty} \mathbb{E}_G \mathbf{Q}_n = \mathbf{Q}$ a je to pozitivně definitní matice. Potom*

i) za hypotézy H_0 má statistika T_n asymptoticky χ^2 s q stupni volnosti.

ii) Za platnosti alternativy H_n má T_n asymptoticky necentrální χ^2 rozdělení s q stupni volnosti a s parametrem necentrality

$$\eta^2 = \boldsymbol{\beta}_0' \mathbf{Q} \boldsymbol{\beta}_0 \cdot \gamma^2(\varphi, F)/A^2(\varphi), \tag{25}$$

kde

$$\gamma^2(\varphi, F) = -\int_0^1 \varphi(t) df(F^{-1}(t)). \tag{26}$$

Důkaz.(**i**) Z věty 3 plyne, že za nulové hypotézy H_0 má statistika $\mathbf{S}_n$ asymptoticky stejné rozdělení jako statistika

$$\mathbf{S}_n^* = n^{-1/2}(\mathbf{Z}_n - \widehat{\mathbf{Z}}_n)'\varphi(F(\mathbf{E})).$$

Asymptotické rozdělení této statistiky je podle centrální limitní věty q–rozměrné normální s nulovou střední hodnotou (z podmínky B.1) a s varianční maticí $\mathbf{Q} \cdot A^2(\varphi)$. Tedy statistika T_n má asymptoticky χ^2 s q stupni volnosti.

(**ii**) Posloupnost lokálních alternativ H_n je kontiguitní vzhledem k posloupnosti rozdělení s hustotami $\{\prod_{i=1}^n h(Y_i - \mathbf{x}_i'\boldsymbol{\beta})\}$ (rozdělení za nulové hypotézy). Potom věta 3 platí též za H_n a statistika $\mathbf{S}_n$ má tedy asymptoticky stejné rozdělení jako statistika $\mathbf{S}_n^*$ za H_n. Z faktu, že asymptotické rozdělení $\mathbf{S}_n^*$ je za H_n normální se střední hodnotou $\gamma(\varphi, F)\mathbf{Q}\boldsymbol{\beta}_0$ a s varianční maticí $\mathbf{Q}A^2(\varphi)$, plyne požadované tvrzení. □

Jako příklad volby φ, můžeme vzhledem ke klasickým pořadovým testům uvést $\varphi(t) = t - 1/2,\ 0 < t < 1$ (což odpovídá Wilcoxonovu testu). Potom skóry $\hat{b}_{ni} = \int \hat{a}_{ni}(t)\mathrm{d}t - 1/2$ a $A^2(\varphi) = 1/12$.

References

[1] Antoch, J. and Jurečková, J. (1985). *Trimmed LSE resistant to leverage points.* Comp. Statist. Quarterly, 4, 329-339.

[2] Gutenbrunner, C. (1986). *Zur Asymptotik von Regressionquantileprozessen und daraus abgeleiten Statistiken.* Ph.D. disertace, Universität Freiburg.

[3] Gutenbrunner, C. a Jurečková, J. (1992). *Regression rank-scores and regression quantiles.* Ann. Statist., 20, 305-330

[4] Gutenbrunner, C., Jurečková, J., Koenker, R. and Portnoy, S. (1993). *Tests of linear hypotheses based on regression rank scores.* Nonparametric Statistics, 2, 307-331.

[5] Hájek, J. (1965). *Extension of the Kolmogorov-Smirnov Test to the Regression Alternatives.* Bernoulli-Bayes-Laplace, Proc. Intern. Research Seminar (J. Neyman and L. Le Cam, eds.), 45-60. Springer-Verlag., Berlin.

[6] Hájek, J. a Šidák, Z. (1967). *Theory of Rank Tests*, Academia, Praha.

[7] Jurečková, J. (1983). *Trimmed Polynomial Regression.* Commentationes Mathematicae Universitatis Carolinae, 24, 4, 579-607.

[8] Jurečková, J. (1984). *Regression quantiles and trimmed least squares estimator under a general design.* Kybernetika, 20, 345-357.

[9] Jurečková, J. (1991). *Tests of Kolmogorov-Smirnov type based on regression rank scores.* Transactions of the 11th Prague Conf. on Information Theory, Statist. Decis. Functions and Random Processes, (J. Á. Víšek, ed.), pp. 41-49. Academia, Prague.

[10] Jurečková, J. (1996). *Regression rank scores tests applied to heavy-tailed distributions.*, v recenzním řízení.

[11] Koenker, R. a Bassett, G. (1978). *Regression quantiles.* Econometrica, 46, 33-50.

[12] Koenker, R. a d'Orey, V. (1994). *Remark on algorithm 229.* Applied Statistics, 43, 410-414.

[13] Ruppert, D. a Carroll, R. J. (1980) *Trimmed least squares estimation in the linear model.* J. Amer. Statist. Assoc., 75, 828-838