ROBUST'2000, 113 – 118

DVOUVÝBĚROVÉ PODMÍNĚNÉ POŘADOVÉ TESTY V ANALÝZE PŘEŽITÍ

LENKA KOBLÍŽKOVÁ

Abstrakt. The present paper deals with conditional rank tests in survival analysis for two sample problem with randomly censored data. Conditional rank tests are exact permutation tests under null hypothesis of randomness if equal censorship is included (restricted null hypothesis). Mainly their asymptotic properties are studied under this hypothesis.

Резюме. В статье изучаются условные ранговые критерии для двухвыборочной проблемы с цензурированием и даны их асымптотические свойства.

1. Úvod

Příspěvek pojednává o některých pořadových testech shody rozdělení dvou cenzorovaných výběrů, které se používají v analýze přežití.

Je zaměřen na testy podmíněné, které jsou založeny na vlastnostech podmíněného rozdělení příslušných statistik při pevné realizaci indikátorových veličin událostí sdruženého výběru. Na základě permutací lze určit přesné hodnoty kvantilů podmíněného rozdělení uvažovaných statistik. Dostáváme tak exaktní testové kritérium. Tento permutační test vyžaduje rovnost rozdělení dob do cenzorování obou uvažovaných výběrů.

Vážené logrankové statistiky patří do třídy zobecněných lineárních pořadových statistik a lze na ně použít již vybudovanou teorii pořadových testů pro necenzorovaná data. S ohledem na tuto skutečnost je odvozeno limitní chování podmíněného rozdělení těchto statistik za platnosti hypotézy náhodnosti a rovnosti rozdělení cenzorování (omezené nulové hypotézy). V tomto případě podmíněné rozdělení nezávisí na podmínce a testovanou hypotézu pak zamítáme nebo nezamítáme na základě kvantilů normovaného normálního rozdělení $\mathrm{N}(0,1)$.

2. Formulace problému a jeho testování

Předpokládejme dvouvýběrový model *náhodného cenzorování*, kde $T_{i1}, T_{i2}, \ldots, T_{in_i}$ je náhodný výběr z nějakého rozdělení s absolutně spojitou distribuční funkcí F_i, $i = 1, 2$. Nechť oba tyto výběry *dob do selhání* jsou na sobě nezávislé.

Nechť $C_{i1}, C_{i2}, \ldots, C_{in_i}$ je náhodný výběr z nějakého rozdělení s absolutně spojitou distribuční funkcí G_i, $i = 1, 2$. Nechť oba tyto výběry *dob do cenzorování* jsou na sobě nezávislé.

Dále předpokládejme, že náhodné veličiny T_{ij}, C_{ij} jsou nezávislé a $S_i = 1 - F_i$ je funkce přežití veličin T_{ij}, $j = 1, 2, \ldots, n_i$, $i = 1, 2$.

Skutečnému pozorování pak odpovídá náhodný vektor $(X_{ij}, \delta_{ij})'$, $j = 1, 2, \ldots, n_i$, $i = 1, 2$, kde

---

2000 *Mathematics Subject Classification.* Primary 62G10; Secondary 62N03.

Klíčová slova. Pořadové testy, analýza přežití, cenzorovaná data.

Tento příspěvek vznikl za přispění grantů GAČR 201/00/0769 a MSM 113200008.

$$X_{ij} = \min(T_{ij}, C_{ij}), \qquad \delta_{ij} = \begin{cases} 1, & T_{ij} \le C_{ij}, \quad X_{ij} \text{ necenzorováno,} \\ 0, & T_{ij} > C_{ij}, \quad X_{ij} \text{ cenzorováno.} \end{cases}$$

Označme $\tilde{X}_{(.)} = (X_{(1)}, X_{(2)}, \dots, X_{(n)})'$ vektor pořádkových statistik příslušný náhodnému vektoru $\tilde{X} = (X_1, X_2, \dots, X_n)' = (X_{11}, \dots, X_{1\,n_1}, X_{21}, \dots, X_{2\,n_2})'$ a nechť $\tilde{\delta} = (\delta_{[1]}, \delta_{[2]}, \dots, \delta_{[n]})'$ je vektor odpovídajících indikátorových veličin událostí

$$\delta_{[j]} = \begin{cases} 1, & X_{(j)} \text{ necenzorováno,} \\ 0, & X_{(j)} \text{ cenzorováno.} \end{cases}$$

Vzhledem k tomu, že distribuční funkce F_1, F_2, G_1, G_2 jsou absolutně spojité, nastává jev $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ s pravděpodobností jedna.

Označme Y_{ij} počet objektů z i-té populace, které pozorujeme těsně před událostí v čase $X_{(j)}$, tzn. $Y_{ij} = \sum_{k=1}^{n_i} I(X_{ik} \ge X_{(j)})$. Položme $Y_j = Y_{1j} + Y_{2j} = n - j + 1$. Nechť $Z_j = 1$ (0), jestliže náhodná veličina $X_{(j)}$, $j = 1, 2, \dots, n$, pochází z prvního (druhého) výběru. Položme $p_j = \frac{Y_{1j}}{Y_j}$ a $q_j = 1 - p_j$ pro $1 \le j \le n$.

Předmětem zájmu je testovat platnost *omezené nulové hypotézy*

$$\bar{H}_0: \; F_1 = F_2 = F \text{ (neznámé)}, \quad G_1 = G_2 = G \text{ (neznámé)} \tag{2.1}$$

proti jednostranné alternativě *stochastického uspořádání*

$$K_1: \; F_1(t) \ge F_2(t) \quad \text{pro } \forall\, t, \quad F_1 \ne F_2. \tag{2.2}$$

K testování výše formulované hypotézy (2.1) proti alternativě (2.2) užíváme *váženou logrankovou statistiku* T_n tvaru (viz [5], část 3, popř. viz [3], část 2)

$$T_n = T_n(\tilde{Z}, \tilde{\delta}) = \sum_{j=1}^{n} w_n(j)\, \delta_{[j]}\, (Z_j - p_j), \tag{2.3}$$

kde w_n je nezáporná stochastická váhová funkce. Přitom se omezíme na váhy tvaru

$$w_n(j) = \bar{w}_n(X_{(j)}) = \hat{S}^{\rho}(X_{(j)}-) \left(\frac{Y_j}{n} \right)^{\kappa} = \hat{S}^{\rho}(X_{(j)}-) \left(\frac{n-j+1}{n} \right)^{\kappa}. \tag{2.4}$$

Ve vzorci (2.4) jsou koeficienty $\rho, \kappa \ge 0$ a $\hat{S}(X_{(j)}-)$ značí tzv. *Kaplanův–Meierův odhad* (podrobněji viz [1], kapitola 3) funkce přežití $S(t)$ těsně před okamžikem $X_{(j)}$, tj.

$$\hat{S}(X_{(j)}-) = \prod_{k=1}^{j-1} \left(1 - \frac{\delta_{[k]}}{n-k+1} \right), \text{ kde } \hat{S}\left(X_{(1)}-\right) = 1. \tag{2.5}$$

V praxi se běžně používají statistiky *logranková* ($\rho = 0$, $\kappa = 0$), *Prenticeova–Wilcoxonova* ($\rho = 1$, $\kappa = 0$) a *Gehanovova–Wilcoxonova* ($\rho = 0$, $\kappa = 1$).

Poznámka 2.1. Volba vhodných vah je složitější problém a při jeho řešení se využívá informace o tom, z jakého rozdělení výběr pochází (podrobněji viz [1], oddíl 7.4).

Ze vztahů (2.4) a (2.5) vyplývá, že váhová funkce $w_n(j)$ závisí pouze na indikátorových veličinách $\delta_{[1]}, \delta_{[2]}, \dots, \delta_{[j-1]}$ a p_j, $q_j = 1 - p_j$ závisejí pouze na $Z_1, Z_2, \dots, Z_{j-1}$:

$$p_j = \frac{Y_{1j}}{Y_j} = \frac{\sum_{k=1}^{n_1} I(X_{1k} \ge X_{(j)})}{n-j+1} = \frac{n_1 - \sum_{k=1}^{j-1} Z_k}{n-j+1}. \tag{2.6}$$

Tedy statistika T_n definovaná v (2.3) závisí pouze na vektoru $\tilde{Z} = (Z_1, Z_2, \ldots, Z_n)'$ a vektoru $\tilde{\delta} = (\delta_{[1]}, \delta_{[2]}, \ldots, \delta_{[n]})'$.

K myšlence podmíněných testů se dostáváme přes následující tvrzení.

Tvrzení 2.1. *Za platnosti omezené nulové hypotézy $\bar{H}_0$ jsou náhodné vektory $\tilde{\delta}$ a $\tilde{Z}$ nezávislé a náhodný vektor $\tilde{Z}$ má rozdělení jako náhodný výběr bez vracení z populace obsahující n_1 jedniček a n_2 nul.*

Důkaz. Tvrzení lze nalézt v [5], str. 1765, lemma 3.1. □

Podmíněný test je sestaven ve dvou krocích:

(1) Na základě pozorování $(x_1, \delta_1^o), \ldots, (x_n, \delta_n^o)$ určíme $\tilde{\delta}^o = (\delta_{[1]}^o, \delta_{[2]}^o, \ldots, \delta_{[n]}^o)'$.

(2) Spočteme hodnotu statistiky T_n pro pozorovaná data podle vzorce (2.3) a užijeme rozhodovacího kritéria pro pevné $\tilde{\delta}^o$:

$$\varphi_{n,\tilde{\delta}^o}(\tilde{z}) = \begin{cases} 1, & T_n(\tilde{z}, \tilde{\delta}^o) > c_n(\alpha, \tilde{\delta}^o), \\ \gamma(\alpha, \tilde{\delta}^o), & T_n(\tilde{z}, \tilde{\delta}^o) = c_n(\alpha, \tilde{\delta}^o), \quad \gamma(\alpha, \tilde{\delta}^o) \in [0,1], \\ 0, & T_n(\tilde{z}, \tilde{\delta}^o) < c_n(\alpha, \tilde{\delta}^o), \end{cases}$$

kde $c_n(\alpha, \tilde{\delta}^o)$ je $(1-\alpha)$-kvantil podmíněného rozdělení $\mathcal{L}(T_n(\tilde{Z}_g, \tilde{\delta}) | \tilde{\delta} = \tilde{\delta}^o)$.

Přičemž $\tilde{Z}_g$ je náhodný vektor, který obsahuje právě n_1 jedniček a n_2 nul a nabývá každé permutace n_1 jedniček a n_2 nul se stejnou pravděpodobností $1/\binom{n}{n_1}$. Z tvrzení 2.1 dostáváme, že za platnosti $\bar{H}_0$ je $\mathcal{L}(\tilde{Z}) = \mathcal{L}(\tilde{Z}_g)$.

Při malých hodnotách n lze stanovit podmíněné rozdělení pravděpodobností statistiky T_n tak, že pro každou hodnotu $T_n = t$ stanovíme počet permutací k_t k ní vedoucích, tzn. $\mathrm{P}_{\bar{H}_0}(T_n = t | \tilde{\delta} = \tilde{\delta}^o) = k_t / \binom{n}{n_1}$. Odtud určíme kvantil $c_n(\alpha, \tilde{\delta}^o)$.

Poznámka 2.2. Podmíněný test $\varphi_{n,\tilde{\delta}^o}$ patří mezi tzv. testy permutační (podrobněji viz [2], str. 42–45).

Výše zmíněný způsob výpočtu kvantilu $c_n(\alpha, \tilde{\delta}^o)$ se stává velmi pracným pro větší rozsahy n_1 a n_2, proto v praxi využíváme simulací, kdy provedeme náhodný výběr ze všech možných permutací o rozsahu m (m dostatečně velké) a určíme kvantil $c_n(\alpha, \tilde{\delta}^o)$ z tohoto výběru. Jiná možnost je sestavit rozhodovací kritérium na základě limitního chování podmíněného rozdělení $\mathcal{L}(T_n(\tilde{Z}, \tilde{\delta}) | \tilde{\delta} = \tilde{\delta}^o)$. K tomu potřebujeme určit podmíněnou střední hodnotu a rozptyl statistiky T_n.

2.1. **Podmíněná střední hodnota a rozptyl statistiky.** Pro následující výpočet je třeba si uvědomit toto: $\mathrm{E}(Z_j | Z_1, \ldots, Z_{j-1}) = p_j$. Standardním výpočtem pak odvodíme (podrobněji viz [4], str. 31–32):

$$\mathrm{E}(T_n | \tilde{\delta}) = 0 \quad s.\,j.,$$

$$\text{(2.7)} \qquad \operatorname{var}(T_n | \tilde{\delta}) = \sum_{j=1}^{n} w_n^2(j)\, \delta_{[j]}\, \frac{n_1 n_2}{n(n-1)}\, \frac{n-j}{n-j+1} = \sum_{j=1}^{n} w_n^2(j)\, \delta_{[j]}\, \mathrm{E} p_j q_j \quad s.\,j.$$

Je užitečné si uvědomit souvislost s pořadovými statistikami pro necenzorovaná data. Statistiku $T_n(\tilde{Z}, \tilde{\delta}^o)$ definovanou vzorcem (2.3) lze upravit následovně

$$\text{(2.8)} \qquad T_n(\tilde{Z}, \tilde{\delta}) = \sum_{j=1}^{n} w_n(j)\, \delta_{[j]}\, (Z_j - p_j) = \sum_{j=1}^{n} Z_j\, a_j^*,$$

kde skóry jsou určeny vztahem

$$a_j^* = w_n(j)\,\delta_{[j]} - \sum_{i=1}^{j} w_n(i)\,\frac{\delta_{[i]}}{n-i+1}, \quad j = 1, 2, \ldots, n. \tag{2.9}$$

Jedná se tedy o *zobecněnou lineární pořadovou statistiku.*

Poznámka 2.3. Výše definované skóry a_j^* závisejí na $\delta_{[1]}, \delta_{[2]}, \ldots, \delta_{[j]}$, a tudíž jsou funkcí náhodného vektoru $\tilde{\delta}$, což kvůli zbytečně složitému značení nebudeme explicitně vyjadřovat.

Pro skóry typu (2.9) platí (viz [4], str. 35)

$$\sum_{j=1}^{n} a_j^* = 0, \qquad \sum_{j=1}^{n} (a_j^*)^2 = \sum_{j=1}^{n} w_n^2(j)\,\delta_{[j]}\,\frac{n-j}{n-j+1} = \frac{n(n-1)}{n_1 n_2}\,\mathrm{var}(T_n|\tilde{\delta}). \tag{2.10}$$

3. Asymptotické vlastnosti testu

Tvrzení 3.1. Nechť existuje limita $\lim_{n\to\infty} n_i/n = \eta_i \in (0,1)$, $i = 1, 2$. Pak za platnosti omezené nulové hypotézy $\bar{H}_0$ skóry a_j^* definované v (2.9) s vahami tvaru (2.4) splňují podmínku

$$\frac{\max_{1\le j\le n} (a_j^*)^2}{\sum_{j=1}^{n} (a_j^*)^2} \xrightarrow{\mathrm{P}} 0, \quad n \to \infty. \tag{3.1}$$

Důkaz. Skóry a_j^* definované v (2.9) lze omezit *s. j.*:

$$\max_{1\le j\le n} (a_j^*)^2 \le \left(\sum_{k=1}^{n} \frac{1}{k} \right)^2 = s_n^2.$$

Odtud a z (2.10) obdržíme

$$0 \le \frac{\max_{1\le j\le n} (a_j^*)^2}{\sum_{j=1}^{n} (a_j^*)^2} \le \frac{\frac{n_1}{n}\,\frac{n_2}{n}\,\frac{s_n^2}{n-1}}{\frac{n_1}{n}\,\frac{n_2}{n}\,\frac{1}{n-1}\sum_{j=1}^{n} (a_j^*)^2} = \frac{\frac{n_1}{n}\,\frac{n_2}{n}\,\frac{s_n^2}{n-1}}{\frac{1}{n}\,\mathrm{var}(T_n|\tilde{\delta})}. \tag{3.2}$$

Přičemž užijeme vlastnosti částečného součtu harmonické řady $s_n = \sum_{k=1}^{n} \frac{1}{k}$ a vlastnosti přirozeného logaritmu $\ln(n)$ (viz [6], str. 331–332, bod 6, a str. 365–366, bod 7)

$$\lim_{n\to\infty} \frac{\ln^{\alpha}(n)}{n^{\beta}} = 0, \quad \alpha > 0,\ \beta > 0, \qquad\qquad \lim_{n\to\infty} (s_n - \ln(n)) = c, \tag{3.3}$$

kde $c \doteq 0,577215665$ je tzv. *Eulerova konstanta.* Opakovaným použitím (3.3) dostaneme, že $\lim_{n\to\infty} s_n^2/n = 0$. Tedy čitatel výrazu na pravé straně v (3.2) konverguje k nule pro $n \to \infty$. Pokud jmenovatel uvažovaného zlomku bude konvergovat v pravděpodobnosti ke kladné konstantě pro $n \to \infty$, což nyní ověříme, podmínka (3.1) bude splněna. Jinak řečeno, chceme, aby za $\bar{H}_0$ platilo:

$$\frac{1}{n}\,\mathrm{var}(T_n|\tilde{\delta}) \xrightarrow{\mathrm{P}} \mathit{const} > 0, \quad n \to \infty. \tag{3.4}$$

Označme $V_n = \frac{n}{n_1\,n_2} \sum_{j=1}^{n} w_n^2(j)\,\delta_{[j]}\,p_j\,q_j$. Pro V_n s vahami tvaru (2.4) za $\bar{H}_0$ platí (viz [5], oddíl 2.2, podrobněji viz [1], oddíl 7.2)

$$V_n \xrightarrow{\mathrm{P}} \sigma^2, \quad n \to \infty, \tag{3.5}$$

kde σ^2 je asymptotický rozptyl statistiky $\left(\frac{n}{n_1\,n_2}\right)^{1/2} T_n$. Pro naše potřeby stačí, že se jedná o kladnou konstantu.

Abychom ověřili (3.4), stačí dokázat tvrzení, že za hypotézy $\bar{H}_0$

$$\frac{1}{n}\Big(\frac{n_1\, n_2}{n}\, V_n - \mathrm{var}(T_n|\tilde{\delta})\Big) \xrightarrow{\mathrm{P}} 0, \quad n \to \infty,$$

$$\text{tj. z (2.7)} \quad \frac{1}{n}\sum_{j=1}^{n} w_n^2(j)\, \delta_{[j]}\, (p_j q_j - \mathrm{E} p_j q_j) \xrightarrow{\mathrm{P}} 0, \quad n \to \infty, \tag{3.6}$$

neboť z (3.5) vyplývá, že $\frac{n_1\, n_2}{n^2}\, V_n \xrightarrow{\mathrm{P}} \eta_1\, \eta_2\, \sigma^2$ při $n \to \infty$.

Zvolme libovolně malé pevné $\varepsilon \in (0,1)$ a využijme vlastnost vah $|w_n(j)| \leq 1$ pro $1 \leq j \leq n$, pak

$$\Big|\frac{1}{n}\sum_{1\leq j<n\varepsilon} w_n^2(j)\, \delta_{[j]}\, (p_j q_j - \mathrm{E} p_j q_j)\Big| \leq \frac{1}{n}\sum_{1\leq j<n\varepsilon} |p_j q_j - \mathrm{E} p_j q_j| \leq \frac{1}{n}\sum_{1\leq j<n\varepsilon} 1 < \varepsilon \;\; s.\, j.$$

Stejnou nerovnost dostaneme i pro součet přes všechna j, $n(1-\varepsilon) < j \leq n$, poněvadž ho lze převést na předchozí případ úpravou $k = n - j$.

Z výše uvedeného vyplývá, že stačí vyšetřovat konvergenci podle pravděpodobnosti pro součet přes všechna j splňující nerovnost $n\varepsilon \leq j \leq (1-\varepsilon)n$:

$$\begin{aligned}\Big|\frac{1}{n}\sum_{n\varepsilon\leq j\leq n(1-\varepsilon)} w_n^2(j)\, \delta_{[j]}\, (p_j q_j - \mathrm{E} p_j q_j)\Big| &\leq \frac{1}{n}\sum_{n\varepsilon\leq j\leq n(1-\varepsilon)} |p_j q_j - \mathrm{E} p_j q_j| \\ &\leq \frac{1}{n}\sum_{n\varepsilon\leq j\leq n(1-\varepsilon)} \big(|p_j - \mathrm{E} p_j|\, |1 - (p_j + \mathrm{E} p_j)| + \mathrm{var}\, p_j\big) \quad s.\, j.\end{aligned}$$

K dalšímu potřebujeme odhad rozptylu $\mathrm{var}\, p_j$, $n\varepsilon \leq j \leq (1-\varepsilon)n$, (viz [4], str. 41):

$$0 \leq \mathrm{var}\, p_j \leq \frac{n_1\, n_2}{n^2}\, \frac{1}{n-j+1} \leq \frac{n_1\, n_2}{n^2}\, \frac{1}{n\varepsilon+1}. \tag{3.7}$$

Vezmeme-li v úvahu, že $|1-(p_j+\mathrm{E} p_j)| \leq 1$ $s.\, j.$ pro $1 \leq j \leq n$ spolu s odhadem (3.7), pak

$$\Big|\frac{1}{n}\sum_{n\varepsilon\leq j\leq n(1-\varepsilon)} w_n^2(j)\, \delta_{[j]}\, (p_j q_j - \mathrm{E} p_j q_j)\Big| \leq \max_{n\varepsilon\leq j\leq n(1-\varepsilon)} |p_j - \mathrm{E} p_j| + \frac{n_1 n_2}{n^2}\, \frac{1}{n\varepsilon+1} \;\; s.\, j.$$

Přičemž výraz na pravé straně bude konvergovat podle pravděpodobnosti k nule pro $n \to \infty$, pokud

$$\max_{n\varepsilon\leq j\leq n(1-\varepsilon)} |p_j - \mathrm{E} p_j| \xrightarrow{\mathrm{P}} 0, \quad n \to \infty. \tag{3.8}$$

Tuto zbývající vlastnost dokážeme:

Pro p_j, viz (2.6), platí $p_j = \frac{n_1}{n-j+1}\Big(1 - \hat{H}_{n_1}(X_{(j)}-)\Big)$, kde $\hat{H}_{n_1}(x)$ je empirická distribuční funkce posloupnosti náhodných veličin $X_1, X_2, \ldots, X_{n_1}$. Označme $\hat{H}_n(x)$ empirickou distribuční funkci posloupnosti náhodných veličin $X_1, X_2, \ldots, X_n$. Dále nechť H_i značí distribuční funkci veličin X_{ij}, $j = 1, 2, \ldots, n_i$, $i = 1, 2$. Za platnosti $\bar{H}_0$ je $H_1(x) = H_2(x) = H(x)$ pro $\forall\, x$.

K odvození vlastnosti (3.8) užijeme *Glivenkovu větu*, tedy za platnosti $\bar{H}_0$

$$\sup_{x\in\mathrm{R}} |\hat{H}_{n_1}(x) - H(x)| \xrightarrow{\mathrm{P}} 0, \quad n_1 \to \infty, \tag{3.9}$$

$$\sup_{x\in\mathrm{R}} |\hat{H}_n(x) - H(x)| \xrightarrow{\mathrm{P}} 0, \quad n \to \infty. \tag{3.10}$$

Dále využijeme

$$\frac{n_1}{n-j+1} \le \frac{n_1}{n\varepsilon+1} \overset{n\to\infty}{\longrightarrow} \frac{\eta_1}{\varepsilon} \quad \text{pro } n\varepsilon \le j \le n(1-\varepsilon). \tag{3.11}$$

Rozdíl $p_j - \mathrm{E}p_j$ upravíme přičtením a odečtením vhodných výrazů

$$p_j - \mathrm{E}p_j = \frac{n_1}{n-j+1}\left\{\left[1-\hat{H}_{n_1}(X_{(j)}-) - \big(1-H(X_{(j)}-)\big)\right] + \right.$$
$$\left. + \left[1-H(X_{(j)}) - \Big(1-\hat{H}_n(X_{(j)})\Big)\right] + \left[1-\hat{H}_n(X_{(j)}) - \frac{n-j+1}{n_1}\,\mathrm{E}p_j\right]\right\}.$$

Vzhledem k tomu, že $\mathrm{E}p_j = \frac{n_1}{n}$ a $\hat{H}_n(X_{(j)}) = \frac{j}{n}$, máme

$$p_j - \mathrm{E}p_j = \frac{n_1}{n-j+1}\left\{\left[H(X_{(j)}-) - \hat{H}_{n_1}(X_{(j)}-)\right] + \left[\hat{H}_n(X_{(j)}) - H(X_{(j)})\right] - \frac{1}{n}\right\}.$$

Za platnosti $\bar{H}_0$ lze náhodnou veličinu $\max_{n\varepsilon\le j\le n(1-\varepsilon)} |p_j - \mathrm{E}p_j|$ omezit *s. j.* následovně:

$$\max_{n\varepsilon\le j\le n(1-\varepsilon)} |p_j - \mathrm{E}p_j| \le \frac{n_1}{n\varepsilon+1}\left(\sup_{x\in\mathrm{R}}\left|\hat{H}_{n_1}(x) - H(x)\right| + \sup_{x\in\mathrm{R}}\left|\hat{H}_n(x) - H(x)\right| + \frac{1}{n}\right).$$

Z vlastností (3.9), (3.10) a (3.11) plyne vlastnost (3.8). Tím jsme dokončili důkaz (3.6), a tedy i tvrzení 3.1. □

Z tvrzení 3.1 vyplývá, že za platnosti omezené nulové hypotézy $\bar{H}_0$ standardizovaná statistika $\frac{T_n}{\sqrt{\mathrm{var}(T_n|\tilde{\delta})}}$, kde T_n je tvaru (2.8), má asymptoticky podmíněně při daném $\tilde{\delta}$ normované normální rozdělení $\mathrm{N}(0,1)$ (viz [2], str. 194–195, dodatky 4 a 8), tj.

$$\lim_{n\to\infty} \mathrm{P}\left\{\sup_{x\in\mathrm{R}}\left|\mathrm{P}\left(\frac{T_n}{\sqrt{\mathrm{var}(T_n|\tilde{\delta})}} \le x \,\Big|\, \tilde{\delta}\right) - \Phi(x)\right| > \varepsilon\right\} = 0, \quad \forall\, \varepsilon > 0.$$

Poznámka 3.1. Vzhledem k této vlastnosti standardizovaná statistika $\frac{T_n}{\sqrt{\mathrm{var}(T_n|\tilde{\delta})}}$ má i asymptoticky (nepodmíněně) normované normální rozdělení $\mathrm{N}(0,1)$ (viz [2], str. 195, dodatek 5).

Na základě získaných poznatků stanovíme asymptotické kritérium podmíněného pořadového testu v případě velkých hodnot n_1 a n_2:

$$\varphi_{n,\tilde{\delta}^o} = \begin{cases} 1, & T_n\left(\mathrm{var}(T_n|\tilde{\delta} = \tilde{\delta}^o)\right)^{-1/2} > u_{1-\alpha}, \quad \text{zamítáme hypotézu } \bar{H}_0, \\ 0, & T_n\left(\mathrm{var}(T_n|\tilde{\delta} = \tilde{\delta}^o)\right)^{-1/2} \le u_{1-\alpha}, \quad \text{nezamítáme hypotézu } \bar{H}_0, \end{cases}$$

kde $u_{1-\alpha}$ je $(1-\alpha)$-kvantil normovaného normálního rozdělení $\mathrm{N}(0,1)$.

Literatura

[1] Fleming T. R., Harrington D. P. (1991): Counting Processes and Survival Analysis. John Wiley & Sons, Inc., New York.

[2] Hájek J., Šidák Z. (1967): Theory of Rank Tests. Academia, Praha.

[3] Janssen A. (1991): Conditional Rank Tests for Randomly Censored Data. *The Annals of Statistics* **Vol. 19, No. 3**, 1434 – 1456.

[4] Koblížková L. (2000): Pořadové testy a odhady v analýze přežití. Diplomová práce MFF UK.

[5] Neuhaus G. (1993): Conditional Rank Tests for the Two–Sample Problem Under Random Censorship. *The Annals of Statistics* **Vol. 21, No. 4**, 1760 – 1779.

[6] Rektorys K. a spolupracovníci (1995): Přehled užité matematiky I. Prometheus, Praha.