KYBERNETIKA — VOLUME 31 (1995), NUMBER 3, PAGES 239–250

JAROSLAV HÁJEK AND ASYMPTOTIC THEORY OF RANK TESTS

JANA JUREČKOVÁ

We characterize basic Hájek's results in the asymptotic theory of rank tests. As one of many extensions of his ideas, we mention an extension of Hájek's rank score process to the linear model.

1. INTRODUCTION

In the series of papers [1–3, 5, 6, 8, 9, 11, 12] (papers [11] and [12] were written jointly with V. Dupač), Hájek systematically investigated the asymptotic properties of linear rank statistics under null hypotheses, under local (contiguous) and some nonlocal alternatives. Besides that, in [4] he derived the rank test of independence in a bivariate distribution, locally most powerful against specific dependence alternatives. The results published before 1967 were then included, unified and elaborated, in the monograph [10], written jointly with Z. Šidák. Hájek's textbook [7] of rank tests also deserves your attention.

This collection of papers, though not of a great size, represents a substantial contribution to the asymptotic theory of rank tests; it was a starting point of a research of many authors and it is a rich source of ideas even today. Each of these papers not only brings new original results, but these results are proved by new, original methods which were later frequently used also in many other contexts. Let us briefly characterize the main Hájek's asymptotic results on rank tests.

2. LINEAR RANK STATISTICS UNDER HYPOTHESIS H_0

Let $(R_{N1}, \dots, R_{NN})$ be a random vector, uniformly distributed over the set of $N!$ permutations of $\{1, \dots, N\}$ and let $(c_{N1}, \dots, c_{NN})$ and $(a_{N1}, \dots, a_{NN})$ be given triangular arrays of real numbers. Consider the statistic

$$S_N = \sum_{i=1}^{n} c_{Ni} a_N(R_{Ni}), \tag{1}$$

where $a_N(i) = a_{Ni}$, $i = 1, \ldots, N$, and

$$a_N(1) \leq \ldots \leq a_N(N) \tag{2}$$

which holds without loss of generality. Hájek [1] proved a necessary and sufficient condition of the Lindeberg type for the asymptotic normality of S_N as $N \to \infty$; it extends the results of Wald and Wolfowitz [32], Noether [27], Hoeffding [23], Dwass [17, 18], and Motoo [26].

For simplicity, we shall formulate the result under the standardization

$$\bar{c}_N = N^{-1} \sum_{i=1}^{N} c_{Ni} = 0, \qquad \bar{a}_N = N^{-1} \sum_{i=1}^{N} a_{Ni} = 0 \tag{3}$$

$$\sum_{i=1}^{N} c_{Ni}^2 = 1, \qquad \lim_{N\to\infty} \max_{1\leq i\leq N} c_{Ni}^2 = 0 \tag{4}$$

and

$$\sum_{i=1}^{N} a_{Ni}^2 = 1, \qquad \lim_{N\to\infty} \max_{1\leq i\leq n} a_{Ni}^2 = 0. \tag{5}$$

Theorem 2.1. (Permutational CLT) Under (1)–(5),

$$(S_N - ES_N)/(\operatorname{var} S_N)^{1/2} \xrightarrow{D} N(0,1) \tag{6}$$

as $N \to \infty$ if and only if

$$\lim_{N\to\infty} \sum_i \sum_j \sum_{|c_{Ni} a_{Nj}| > \varepsilon} c_{Ni}^2 a_{Ni}^2 = 0 \tag{7}$$

for every $\varepsilon > 0$.

Permutational CLT is applicable not only to linear rank statistics, but also, e. g., in sampling from finite population. If $R_{N1}, \ldots, R_{NN}$ are ranks of $X_{N1}, \ldots, X_{NN}$ where the X_{Ni} are independent, X_{Ni} distributed according to a d.f. F_{Ni}, $i = 1, \ldots, N$, then the theorem implies the asymptotic normality of S_N under the *hypothesis of randomness*

$$H_0: \quad F_{N1} = \ldots = F_{NN} = F, \tag{8}$$

where F is a continuous d.f., otherwise unspecified.

Moreover, under (8), $(R_{N1}, \ldots, R_{NN})$ could be also interpreted as the vector of ranks of the random sample $(U_1, \ldots, U_N)$ from the uniform $R(0,1)$ distribution. The following result of Hájek is an asymptotic representation of S_N by means of a sum of independent summands.

Theorem 2.2. (Asymptotic representation) Under (1)–(3) and (5),

$$S_N = T_N + r_N, \tag{9}$$

where

$$T_N = \sum_{i=1}^{N} c_{Ni} a_N([NU_i] + 1), \tag{10}$$

$[Nu]$ denotes the integer part of Nu and

$$Er_N^2/\operatorname{var} T_N \to 0 \qquad \text{as } N \to \infty. \tag{11}$$

Among various possible choices of the a_{Ni}, Hájek also considered

$$a_{Ni} = E(\varphi(U_1)|R_{N1} = i) = E\varphi(U_{N:i}), \tag{12}$$

where $\varphi : (0,1) \to \boldsymbol{R}_1$ is a nondecreasing, square- integrable score function and $U_{N:1} \le \ldots \le U_{N:N}$ are the order statistics corresponding to $U_1, \ldots, U_N$. Hájek showed that the conditions of Theorem 2.2 for the scores (12) are guaranteed by the martingale property of $(a_N(R_{N1}), \ldots, a_N(R_{NN}))$.

3. LINEAR RANK STATISTICS UNDER LOCAL ALTERNATIVES

Making use of LeCam's concept of contiguity, Hájek [2] proved the asymptotic normality of S_N under the contiguous alternatives. More precisely, if $R_{N1}, \ldots, R_{NN}$ are the ranks of independent $X_{N1}, \ldots, X_{NN}$, Hájek proved the asymptotic normality of S_N under the model

$$P(X_{Ni} \le x) = F((x - \beta_0 - \beta c_{Ni})/\sigma), \quad i = 1, \ldots, N, \tag{13}$$

where $\beta > 0$ and $\beta_0 \in \boldsymbol{R}_1$, $\sigma \ge 0$ are nuisance parameters, F has an absolutely continuous density f and finite Fisher's information,

$$0 < I(f) = \int_{-\infty}^{\infty} \left(\frac{f'(x)}{f(x)}\right)^2 \mathrm{d}F(x) < \infty \tag{14}$$

and $\{(c_{N1}, \ldots, c_{NN})\}_{N=1}^{\infty}$ satisfy the conditions

$$C_N^2 = I(f) \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N)^2 \to C^2 < \infty \quad \text{as } N \to \infty \tag{15}$$

and

$$\lim_{N\to\infty} \max_{1 \le i \le N} (c_{Ni} - \bar{c}_N)^2 = 0. \tag{16}$$

The linear rank statistic is written in the form

$$S_N = \sum_{i=1}^{N} (c_{Ni} - \bar{c}_N) \varphi\left(\frac{R_{Ni}}{N+1}\right), \tag{17}$$

where $\varphi : (0,1) \longrightarrow \boldsymbol{R}_1$ is assumed being nondecreasing and square-integrable; other choices of scores are also considered.

Hájek [2] showed that the choice of φ

$$\varphi(u) = \varphi(u, f) = -\frac{f'(F^{-1}(u)}{f(F^{-1}(u)}, \qquad 0 < u < 1 \tag{18}$$

leads to an asymptotically efficient test of $H_0 : \beta = 0$ against $K : \beta > 0$ in model (13) in the sense of Pitman efficiency; more precisely, he proved the following theorem:

Theorem 3.1. (Asymptotically most powerful rank test) Let S_N be defined as in (17) with φ given in (18). Then, under (13)–(16), the test with the critical region

$$S_N \geq C_N \tau_\alpha, \qquad \tau_\alpha = \Phi^{-1}(1-\alpha), \qquad 0 < \alpha < 1 \tag{19}$$

has the asymptotic power

$$P_\beta(S_N \geq C_N \tau_\alpha) \cong 1 - \Phi(\tau_\alpha - (\beta/\sigma)\, C_N) \tag{20}$$

and hence it is the asymptotically optimal test of size α for H_0 against K in model (13).

Moreover, under $\beta_0 = 0$ and σ=1, the test is asymptotically equivalent to the *score test* with the criterion

$$M_N = -\sum_{i=1}^{N} (c_{Ni} - \bar{c}_N) \frac{f'(X_{Ni})}{f(X_{Ni})}. \tag{21}$$

A similar treatment is made for the *signed-rank tests* of the hypothesis of symmetry (or the paired comparisons).

In the same paper [2], Hájek constructed a histogram-type estimator $\hat{\varphi}(u)$ of the optimal score function $\varphi(u, f)$ and showed that the rank test based on $\hat{\varphi}(u)$ is uniformly asymptotically efficient (notice that the paper appeared only in 1962 !).

4. LINEAR RANK STATISTICS UNDER GENERAL ALTERNATIVES

Chernoff and Savage [13] and Govindarajulu, LeCam and Raghavachari [19] proved the asymptotic normality of two-sample linear rank statistics under some non-local alternatives and for some classes of score-generating functions.

Hájek [6] gave a far-reaching extension of these results. Typically for him, he developed new pioneering methods to prove these results, and these methods were later used by many authors in various contexts: He derived a general *variance inequality* for linear rank statistics and proved their asymptotic normality by means of L_2-*projection* of S_N on the space of sums of N independent summands.

Theorem 4.1. (Variance inequality) Let $X_1, \ldots, X_N$ be independent random variables with the ranks $R_1, \ldots, R_N$ and arbitrary continuous distribution functions $F_1, \ldots, F_N$. Let $(c_1, \ldots, c_N)$ and $(a_1, \ldots, a_N)$ be arbitrary vectors, $a_1 \le \ldots \le a_N$. Then

$$\operatorname{var}\left[\sum_{i=1}^{N} c_i a(R_i)\right] \le 21 \max_{1 \le i \le N} (c_i - \bar{c}) \sum_{i=1}^{N} (a_i - \bar{a})^2, \tag{22}$$

where $\bar{c} = N^{-1} \sum_{i=1}^{N} c_i$, $\bar{a} = N^{-1} \sum_{i=1}^{N} a_i$, $a(i) = a_i$, $i = 1, \ldots, N$.

The variance inequality enables to study the asymptotic behavior of the statistics $S_N = \sum_{i=1}^{N} c_{Ni} a_N(R_{Ni})$ with the scores of the form

$$a_N(i) = \varphi\left(\frac{i}{N+1}\right) \tag{23}$$

or

$$a_N(i) = E\varphi(U_{N:i}), \tag{24}$$

$i = 1, \ldots, N$, generated by a nondecreasing, square-integrable, possibly unbounded function $\varphi : (0,1) \to \boldsymbol{R}_1$. On the other hand, the L_2-projection applies to scores generated by possibly non-monotone function φ which has a bounded second derivative in $(0,1)$; this leads to the following approximation of S_N:

Theorem 4.2. (Projection approximation of S_N) Let $\varphi : (0,1) \to \boldsymbol{R}_1$ have a bounded second derivative in $(0,1)$. Then there exists a constant $M = M(\varphi)$ such that for any N, $(c_1, \ldots, c_N)$ and continuous $F_1, \ldots, F_N$,

$$E\left(S_N - ES_N - \sum_{i=1}^{N} Z_i\right)^2 \le MN^{-1} \sum_{i=1}^{N} (c_i - \bar{c})^2 \tag{25}$$

and

$$E(S_N - \mu_N)^2 \le MN^{-1} \sum_{i=1}^{N} c_i^2, \tag{26}$$

where

$$Z_i = N^{-1} \sum_{j=1}^{N} (c_j - c_i) \int_{-\infty}^{\infty} (I[X_i \le x] - F_i(x))\, \varphi'(H(x))\, \mathrm{d}F_j(x), \qquad i = 1, \ldots, N \tag{27}$$

$$\mu_N = \sum_{i=1}^{N} c_i \int_{-\infty}^{\infty} \varphi(H(x))\, \mathrm{d}F_i(x) \tag{28}$$

and

$$H(x) = H_N(x) = N^{-1} \sum_{i=1}^{N} F_i(x). \tag{29}$$

Using the fact that, to any function φ being a difference of two nondecreasing, square-integrable functions, absolutely continuous inside $(0,1)$, and to any $\alpha > 0$, there exists a decomposition

$$\varphi(t) = \psi(t) + \varphi_1(t) - \varphi_2(t), \qquad 0 < t < 1, \tag{30}$$

where ψ is a polynomial and φ_1, φ_2 are nondecreeasing functions satisfying

$$\int_0^1 \varphi_1^2(t)\,\mathrm{d}t + \int_0^1 \varphi_2^2(t)\,\mathrm{d}t < \alpha, \tag{31}$$

a combination of Theorems 1 and 2 leads to the following final result:

Theorem 4.3. (Asymptotic normality of S_N under general alternatives)
Assume that the scores of $S_N = \sum_{i=1}^N c_{Ni} a_N(i)$ are generated by φ, being a difference of two square-integrable functions, absolutely continuous inside $(0,1)$, either by (23) or by (24). Then to every $\varepsilon > 0$, $\eta > 0$, there exist N_0 and $\delta > 0$ such that, for $N > N_0$ and for any $(c_1, \dots, c_N)$, $F_1, \dots, F_N$ satisfying

$$\sum_{i=1}^N (c_{Ni} - \bar{c}_N)^2 > N\eta \max_{1 \le i \le N} (c_{Ni} - \bar{c}_N)^2 \tag{32}$$

and

$$\sup_{x \in R_1} |F_i(x) - F_j(x)| < \delta, \qquad i,j = 1, \dots, N \tag{33}$$

it holds

$$\sup_{x \in R_1} |P(S_N - ES_N < x\sigma) - \Phi(x)| < \varepsilon \tag{34}$$

with

$$\sigma^2 = \sum_{i=1}^N (c_{Ni} - \bar{c}_N)^2 \int_0^1 (\varphi(t) - \overline{\varphi})^2\,\mathrm{d}t, \tag{35}$$

$\overline{\varphi} = \int_0^1 \varphi(t)\,\mathrm{d}t$ and Φ being the distribution function of $N(0,1)$.

Hájek and Dupač [11], using the projection method and a more elaborated treatment of the residual variance, extended the above results to possibly discontinuous score functions, under slightly more restrictive conditions on the distributions. The same authors then in [12] specialized the results to the two-sample Wilcoxon statistic under various alternatives.

5. NONLINEAR RANK TESTS

In paper [3] Hájek extended the Kolmogorov–Smirnov test to verify the hypothesis of randomness against the regression alternative stating that the vector $\boldsymbol{X}_N = (X_{N1}, \dots, X_{NN})$ is distributed according to the density

$$H_\beta^+ : \quad q_\beta(x_1, \dots, x_N) = \prod_{i-1}^N f(x_i - c_{Ni}\beta), \quad \beta > 0, \tag{36}$$

where f is an arbitrary one-dimensional density.

Hájek attacked this problem using the weak convergence of empirical processes which was a pioneering method in 1965. He considered the *rank-scores process*

$$\mathcal{X}_N = \left\{ X_N(t) = \sum_{i=1}^{N} c_{Ni} \hat{a}_{Ni}(t),\ 0 \le t \le 1 \right\}, \tag{37}$$

where the scores $\hat{a}_{Ni}(t)$ depend on the ranks $R_{N1}, \ldots, R_{NN}$ of $X_{N1}, \ldots, X_{NN}$ in the following way:

$$\hat{a}_{Ni}(t) = \begin{cases} 1 & \ldots 0 \le t \le (R_{Ni} - 1)/N \\ R_{Ni} - Nt & \ldots (R_{Ni} - 1)/N < t \le R_{Ni}/N \\ 0 & \ldots R_{Ni}/N < t \le 1, \end{cases} \tag{38}$$

$i = 1, \ldots, n$. $\mathcal{X}_N$ is a process with trajectories in $C[0,1]$. Hájek proved that, under $H_0 : \beta = 0$ and under the standardization $\sum_{i=1}^{N} c_{Ni} = 0$, $\sum_{i=1}^{N} c_{Ni}^2 = 1$, $\max_{1 \le i \le N} |c_{Ni}| = o(1)$, $\mathcal{X}_N$ converges to the Brownian bridge in the Prochorov topology on $C[0,1]$. To prove the tightness of the sequence of distributions of $\{\mathcal{X}_N\}$, Hájek extended the Kolmogorov inequality to dependent summands $Y_1, \ldots, Y_n$ which are a realization of a simple random sampling of size n without replacement from the population $\{c_1, \ldots, c_N\}$; this inequality has an interest of its own. The Kolmogorov–Smirnov type test criterion of H_0 against H_β^+ is then defined as

$$K_N^+ = \sup\{X_N(t) : 0 \le t \le 1\} \tag{39}$$

and hence it is a continuous functional of $\mathcal{X}_N$. The tests against two-sided alternatives $\beta \ne 0$ are based on the criterion

$$K_N^+ = \sup\{|X_N(t)| : 0 \le t \le 1\}. \tag{40}$$

The weak convergence and the Prochorov theorem imply that the asymptotic distributions of K_N^+ and K_N under H_0 coincide with those of the classical Kolmogorov–Smirnov criteria.

Not only the Kolmogorov–Smirnov test, but many other rank tests, linear and non-linear, can be expressed as functionals of $\mathcal{X}_N$. Monograph [10] also describes the tests of Cramér–von Mises and of Rényi types. The linear rank statistics can be expressed as the following functionals of $\mathcal{X}$:

$$S_N = -\int_0^1 \varphi(t)\,\mathrm{d}X_N(t) = \sum_{i=1}^{N} c_{Ni} a_N(R_{Ni}) \tag{41}$$

with the scores $a_N(i) = N \int_{(i-1)/N}^{i/N} \varphi(t)\,\mathrm{d}t$, $i = 1, \ldots, N$, representing another alternative form of the scores.

6. FURTHER ASYMPTOTIC PROPERTIES OF RANKS

Hájek [9] demonstrated that not only the best Pitman efficiency but also the best exact Bahadur slope is attainable by rank statistics; otherwise speaking, that the vector of ranks is *sufficient in the Bahadur sense.*

Consider the two-sample model with independent samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ with the respective densities and d.f.'s f, g, F, G; let $\lim_{N\to\infty} \frac{n}{m+n} = \lambda \in (0,1)$ and denote

$$H(x) = \lambda F(x) + (1-\lambda)\, G(x), \quad x \in \boldsymbol{R}_1, \tag{42}$$

$$\overline{f}(u) = \frac{\mathrm{d}}{\mathrm{d}u} F(H^{-1}(u)), \quad \overline{g}(u) = \frac{\mathrm{d}}{\mathrm{d}u} G(H^{-1}(u)), \quad 0 < u < 1. \tag{43}$$

Let $R_1, \dots, R_{n+m}$ be the ranks of $(Z_1, \dots, Z_{n+m}) = (X_1, \dots, X_n, Y_1, \dots, Y_m)$ and

$$S_N = \sum_{i=1}^{n} a_N(R_i), \quad N = m+n \tag{44}$$

with the scores $a_N(i)$ generated by an integrable function $\varphi : (0,1) \to \boldsymbol{R}_1$. Then:

(i) S_N satisfies the law of large numbers, i.e.

$$N^{-1} S_N \to \lambda \int_0^1 \varphi(u)\, \overline{f}(u)\, \mathrm{d}u \quad \text{a.s.} \quad \text{as} \quad N \to \infty \tag{45}$$

and

(ii) the score function

$$\varphi(u) = \log \frac{\overline{f}(u)}{\overline{g}(u)}, \quad 0 < u < 1 \tag{46}$$

is optimal in the Bahadur sense; the limit in (45) is equal to $\lambda K(\overline{f}, \overline{g})$ (*Kullback–Leibler information number*) and, under the hypothesis of randomness $H_0 : F \equiv G$,

$$\lim_{N\to\infty} \log P\{N^{-1} S_N > \lambda K(\overline{f}, \overline{g})\} = -\overline{J}(f, g, \lambda), \tag{47}$$

where

$$2\overline{J}(f, g, \lambda) = \int_0^1 (\lambda \overline{f} \log \overline{f} + (1-\lambda)\overline{g} \log \overline{g})\, \mathrm{d}u \tag{48}$$

is the best attainable exact slope.

(iii) The best Bahadur slope is also attainable by the Neyman–Pearson rank test with the criterion

$$N! Q_N\{(R_1, \dots, R_N) = (r_1, \dots, r_N)\} \tag{49}$$

where Q_N is the distribution of the vector of ranks under the alternative.

Among various other Hájek's results concerning the ranks, let us mention partially adaptive procedures which Hájek proposed in [8]. The procedures select one of a finite set of scores functions and hence one of the corresponding rank tests by means of a decision rule depending on the ranks of observations or of residuals.

7. EXTENSION OF RANK–SCORES PROCESS TO REGRESSION MODEL

Hájek's results and methods were used and extended by a host of statisticians; it is impossible to characterize all this work as a whole. Among many possible extensions, let us briefly describe a recent extension of Hájek's rank-scores process to linear regression model, which in turn has further interesting applications.

Consider the linear regression model

$$Y_i = \boldsymbol{x}_i'\beta + E_i, \quad i = 1, \dots, n \tag{50}$$

with $\boldsymbol{x}_i \in \boldsymbol{R}_p$, $x_{i1} = 1, i = 1, \dots, p$ and with independent errors $E_1, \dots, E_n$. Koenker and Bassett [25] introduced the *α-regression quantile* $\hat{\beta}(\alpha)$ $(0 < \alpha < 1)$ for model (50) as a solution of the minimization

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - \boldsymbol{x}_i'\boldsymbol{b}) := \min, \quad \boldsymbol{b} \in \boldsymbol{R}_p, \tag{51}$$

where

$$\rho_\alpha(x) = x(\alpha - I[x < 0]), \quad x \in \boldsymbol{R}_1. \tag{52}$$

Koenker and Bassett [25] and Ruppert and Carroll [31] showed that the asymptotic properties of regression quantiles are in correspondence with those of the sample quantiles in the location model with $\boldsymbol{x}_i = 1$, $i = 1, \dots, n$. More precisely, the latter authors proved, under some regularity conditions on the matrix $\boldsymbol{X}_n = (\boldsymbol{x}_1', \dots, \boldsymbol{x}_n')'$ and on the joint d.f. F of the errors $E_1, \dots, E_n$, the Bahadur-type representation of regression quantiles,

$$n^{1/2}(\hat{\beta}(\alpha) - \beta(\alpha)) = n^{-1/2}[f(F^{-1}(\alpha))]^{-1}\boldsymbol{Q}_n^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i\varphi_\alpha(E_{i\alpha}) + o_p(1) \tag{53}$$

with

$$\beta(\alpha) = (\beta_1 + F^{-1}(\alpha), \beta_2, \dots, \beta_p)'$$

$$\boldsymbol{Q}_n = n^{-1}\sum_{i-1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'$$

$$\varphi_\alpha(x) = \alpha - I[x < 0], \quad x \in \boldsymbol{R}_1$$

and

$$E_{i\alpha} = E_i - F^{-1}(\alpha), \quad i = 1, \dots, n.$$

(53) immediately implies that

$$n^{1/2}\boldsymbol{Q}_n^{1/2}(\hat{\beta}(\alpha) - \beta(\alpha)) \xrightarrow{D} \mathcal{N}_p\left(\boldsymbol{0}, \frac{\alpha(1-\alpha)}{f^2(F^{-1}(\alpha))}\boldsymbol{I}_p\right) \tag{54}$$

which is in the correspondence with the asymptotic distribution of location sample quantiles.

Koenker and Bassett [25] characterized $\hat{\beta}(\alpha)$ as the component $\hat{\beta}$ of the optimal solution $(\hat{\beta}, \boldsymbol{r}^+, \boldsymbol{r}^-)$ of the linear programming problem

$$\alpha \sum_{i=1}^{n} r_i^+ + (1-\alpha) \sum_{i=1}^{n} r_i^- = \min$$

$$\sum_{j=1}^{p} x_{ij}\beta_j + r_i^+ - r_i^- = Y_i, \quad i = 1,\ldots,n \tag{55}$$

$$\beta_j \in \boldsymbol{R}_1,\ j = 1,\ldots,p;\ r_i^+ \geq 0,\ r_i^- \geq 0,\ i = 1,\ldots,n;\ 0 < \alpha < 1.$$

The dual program to (55) can be written as follows

$$\sum_{i=1}^{n} Y_i a_i := \max$$

$$\sum_{i=1}^{n} x_{ij}(a_i - (1-\alpha)) = 0, \quad j = 1,\ldots p \tag{56}$$

$$0 \leq a_i \leq 1,\ i = 1,\ldots,n;\ 0 < \alpha < 1.$$

By the duality of (55) and (56), the optimal solution of (56),

$$\hat{\boldsymbol{a}}_n(\alpha) = (\hat{a}_{n1}(\alpha),\ldots,\hat{a}_{nn}(\alpha))'$$

satisfies the inequalities

$$\hat{a}_{ni}(\alpha)) = \begin{cases} 1 & \ldots Y_i > \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}(\alpha) \\ 0 & \ldots Y_i < \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}(\alpha), \end{cases} \tag{57}$$

$i = 1,\ldots,n$. Moreover, the $\hat{a}_{ni}(\alpha)$ are continuous piecewise linear functions of α, $\hat{a}_{ni}(0) = 1$, $\hat{a}_{ni}(1) = 0$. In the location model with $\boldsymbol{x}_i = 1$, $i = 1,\ldots,n$, the $\hat{a}_{ni}$ coincide with Hájek's rank-scores (38). Hence, the linear programming duality of (55) and (56) also extends the stochastic duality of order statistics and ranks from the location to the linear regression model. This gives us a justification to call $\hat{a}_{n1}(\alpha),\ldots,\hat{a}_{nn}(\alpha)$, $0 \leq \alpha \leq 1$, the *regression rank scores* of the model (50). Their most interesting property is the *invariance to the regression* under the model (50), i.e.,

$$\hat{\boldsymbol{a}}_n(\alpha, \boldsymbol{Y} + \boldsymbol{X}\boldsymbol{b}) = \hat{\boldsymbol{a}}_n(\alpha, \boldsymbol{Y}) \quad \forall \boldsymbol{b} \in \boldsymbol{R}_p \tag{58}$$

which is in correspondence with the fact that Hájek's scores (38) (and the ranks) are invariant to the translation. Naturally, $\hat{\boldsymbol{a}}_n(\alpha)$ is also scale-invariant.

The regression rank scores process

$$\mathcal{X}_n^* = \left\{ X_n^*(t) = \sum_{i=1}^{n} c_{ni}\hat{a}_{ni}(t) : 0 \leq t \leq 1 \right\} \tag{59}$$

is a natural extension of the rank-scores process (37) and, under some regularity conditions, it is asymptotically equivalent to (37) with the R_{ni} in (38) being the (unobservable) ranks of the errors $E_1,\ldots,E_n$. The properties of the regression rank scores process are studied in Gutenbrunner and Jurečková [20], Gutenbrunner, Jurečková, Koenker and Portnoy [21] and Jurečková [23]. [21] and [24] construct the linear and nonlinear tests of the hypothesis $H : \boldsymbol{\delta} = \mathbf{0}$ in the extended linear model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\delta} + \boldsymbol{E} \tag{60}$$

with $\boldsymbol{X}$ of order $(n \times p)$, $x_{i1} = 1$, $i = 1, \ldots, n$, $\boldsymbol{Z}$ of order $(n \times q)$ and where $\boldsymbol{\beta}$ is considered as a nuisance parameter. Tests based on regression rank scores calculated via (56) under the hypothesis H, i.e. under $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{E}$, are invariant to the $\boldsymbol{X}$-regression and therefore invariant to the nuisance $\boldsymbol{\beta}$. Their structure is analogous to that of ordinary rank tests, and so is their Pitman efficiency. More details could be found in the papers mentioned above where other papers, also concerning the pertaining computional algorithms, are cited. The research is still in the progress; our ultimate goal is to establish the asymptotics of regression rank-scores tests under the weakest possible regularity conditions, keeping in mind that their counterpart tests, based solely on the ordinary ranks, are practically universal.

ACKNOWLEDGEMENT

The research was supported by the Grant Agency of the Czech Republic under Grant No. 2168. The paper was partially written while the author was visiting in Université Bordeaux 2, Laboratoire de Mathématiques Stochastiques under the C.N.R.S. support (JF-91).

(Received October 26, 1994.)

REFERENCES

[1] J. Hájek: Some extensions of the Wald–Wofowitz–Noether theorem. Ann. Math. Statist. *32* (1961), 506–523.
[2] J. Hájek: Asymptotically most powerful rank order tests. Ann. Math. Statist. *33* (1962), 1124–1147.
[3] J. Hájek: Extension of the Kolmogorov–Smirnov test to the regression alternatives. Bernoulli–Bayes–Laplace. In: Proc. Internat. Research Seminar (J. Neyman and L. LeCam, eds.), Springer-Verlag, Berlin 1965, pp. 45–60.
[4] J. Hájek: Locally most powerful tests of independence. In: Studies in Math. Statist. (K. Sarkadi and I. Vincze, eds.), Akadémiai Kiadó, Budapest 1968, pp. 45–51.
[5] J. Hájek: Some new results in the theory of rank tests. In: Studies in Math. Statist. (K. Sarkadi and I. Vincze, eds.), Akadémiai Kiadó, Budapest 1968, pp. 53–55.
[6] J. Hájek: Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. *39* (1968), 325–346.
[7] J. Hájek: A Course in Nonparametric Statistics. Holden-Day, San Francisco 1969.
[8] J. Hájek: Miscellaneous problems of rank test theory. In: Nonparam. Techniques in Statist. Inference (M. L. Puri, ed.), Cambridge Univ. Press 1970, pp. 3–19.
[9] J. Hájek: Asymptotic sufficiency of the vector of ranks in the Bahadur sense. Ann. Statist. *2* (1974), 75–83.
[10] J. Hájek and Z. Šidák: Theory of Rank Tests. Academia, Prague and Academic Press, New York 1967. (Russian translation: Nauka, Moscow 1971.)
[11] J. Hájek and V. Dupač: Asymptotic normality of simple linear rank statistics under alternatives II. Ann. Math. Statist. *40* (1969), 1992–2017.

[12] J. Hájek and V. Dupač: Asymptotic normality of the Wilcoxon statistic under divergent alternatives. Zastos. Mat. *10* (1969), 171–178.
[13] H. Chernoff and I. R. Savage: Asymptotic normality and efficiency of certain nonparametric test statistics. Ann. Math. Statist. *29* (1958), 972–994.
[14] M. D. Donsker: Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorems. Ann. Math. Statist. *23* (1952), 277–281.
[15] J. L. Doob: Heuristic approach to the Kolmogorov–Smirnov theorems. Ann. Math. Statist. *20* (1949), 393–403.
[16] M. Driml: Convergence of compact measures on metric spaces. In: Trans. 2nd Prague Conference on Inform. Theory, NČSAV, Prague 1959, pp. 71–92.
[17] M. Dwass: On the asymptotic normality of certain rank order statistics. Ann. Math. Statist. *24* (1953), 303–306.
[18] M. Dwass: On the asymptotic normality of some statistics used in non-parametric setup. Ann. Math. Statist. *26* (1955), 334–339.
[19] Z. Govindarajulu, L. LeCam and M. Raghavachari: Generalizations of theorems of Chernoff and Savage on the asymptotic normality of test statistics. In: Proc. 5th Berkeley Symp. Math. Statist. Probab. *1* (1966), pp. 609–638.
[20] C. Gutenbrunner and J. Jurečková: Regression rank scores and regression quantiles. Ann. Statist. *20* (1992), 305–330.
[21] C. Gutenbrunner, J. Jurečková, R. Koenker and S. Portnoy: Tests of linear hypotheses based on regression rank scores. J. Nonpar. Statist. *2* (1993), 307–331.
[22] W. Hoeffding: A combinatorial central limit theorem. Ann. Math. Statist. *22* (1951) 558–566.
[23] J. Jurečková: Uniform asymptotic linearity of regression rank scores process. In: Nonpar. Statist. Rel. Topics (A. K. Md. E. Saleh, ed.), Elsevier Sci. Publ. 1992, pp. 217–228.
[24] J. Jurečková: Tests of Kolmogorov–Smirnov type based on regression rank scores. In: Trans. 11th Prague Conference on Inform. Theory (J.Á.Víšek, ed.), Academia, Prague and Kluwer Acad. Publ. 1992, pp. 41–49.
[25] R. Koenker and G. Bassett: Regression quantiles. Econometrica *46* (1978), 33–50.
[26] M. Motoo: On the Hoeffding's combinatorial central limit theorem. Ann. Inst. Statist. Math. *8* (1957), 145–154.
[27] G. E. Noether: On a theorem of Wald and Wolfowitz. Ann. Math. Statist. *20* (1949), 455–458.
[28] Ju. V. Prochorov: Convergence of stochastic processes and limiting theorems of probability theory. Theory Probab. *1* (1956), 177–238.
[29] R. Pyke and G. R. Shorack: Weak convergence of a two-sample empirical process and a new approach to Chernoff–Savage theorems. Ann. Math. Statist. *39* (1968), 755–771.
[30] M. Raghavachari: On a theorem of Bahadur on the rate of convergence of tests statistics. Ann. Math. Statist. *41* (1970), 1695–1699.
[31] D. Ruppert and R. J. Carroll: Trimmed least squares estimation in the linear model. J. Amer. Statist. Assoc. *75* (1980), 828–838.
[32] A. Wald and J. Wolfowitz: Statistical tests based on permutations of the observations. Ann. Math. Statist. *15* (1944), 358–372.
[33] G. Woodworth: Large deviations and Bahadur efficiency in linear rank statistic. Ann. Math. Statist. *41* (1970), 251–283.

Prof. RNDr. Jana Jurečková, DrSc., Matematicko-fyzikální fakulta Univerzity Karlovy (Faculty of Mathematics and Physics – Charles University), Sokolovská 83, 186 00 Praha 8. Czech Republic.