COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE
20,3 (1979)

THE BERRY-ESSEEN THEOREM FOR RANK STATISTICS

Marie HUŠKOVÁ

Abstract: The natural rate of convergence for the distribution function of simple linear rank statistics to the normal one was established for a rather wide class of score-generating functions ([1],[2],[4],[6],[7]), which however, does not include one of the most usual score-generating function - the normal quantile function. The purpose of this paper is to extend the assertion on the rate of convergence to the class of score-generating functions covering the normal case. The null hypothesis is treated in detail while the assertion for the contiguous alternatives is stated without proof.

Key words: Simple linear rank statistics, convergence rate, distribution free tests.

AMS: 60F05, 62G10

---

1. Introduction and Statement of the Theorem under Hypothesis

Let $X_1,\dots,X_N$ be independent identically distributed random variables with common continuous distribution function F. Define a simple linear rank statistics

$$S_N = \sum_{i=1}^{N} c_{iN}\, a_N(R_{iN}), \tag{1.1}$$

where $(c_{1N},\dots,c_{NN})$ are regression constants, $(a_N(1),\dots \dots,a_N(N))$ scores and R_{iN} is the rank of X_i in the sequence

$X_1,\dots,X_N$.

In the previous work on the rate of convergence there were either imposed stronger conditions on the score-generating function or the obtained rate was not natural. If regression constants satisfy the assumption B given below the natural rate means $O(\sum_{i=1}^{N} |c_{iN}|^3)$, i.e. the same as that for the near sum of independent random variables $\sum_{i=1}^{N} c_{iN}\varphi(F(X_i))$. Jurečková and Puri (1975) assumed φ' bounded and obtained the rate $\sum_{i=1}^{N} |c_{iN}|^3 N^{\delta}$, $\delta > 0$. B. von Bahr treating more general statistics showed that the rate is

$$\max_{1\le i\le N} |c_{iN}| \max_{1\le i\le N} |a_N(i)|$$

which coincides with natural rate if

$$\max_{1\le i\le N} |c_{iN}| = O(\sum_{i=1}^{N} |c_{iN}|^3), \quad \max_{1\le i\le N} |a_N(i)| = O(1).$$

Bergström and Puri (1977) treated the problem for the case φ'' bounded and $X_1,\dots,X_N$ independent with continuous distributions $F_1,\dots,F_N$ and obtained the rate $\sum_{i=1}^{N} |c_{iN}|^3 N^{\delta}(\text{var } S_N)^{-3/2}$, $\delta > 0$. Hušková (1977) got the natural rate for the case of square-integrable φ''.

In the present paper we consider the following assumptions:

A. $X_1,\dots,X_N$ be independent identically distributed random variables with common continuous distribution function F.

B. The regression constants satisfy:

$$\sum_{i=1}^{N} c_{iN} = 0, \quad \sum_{i=1}^{N} c_{iN}^2 = 1, \; N = 1,2,\dots$$

(this can be assumed without loss of generality).

C. The scores be either of the forms:

$$(1.2) \qquad a_N(i) = \varphi(i/(N+1)), \ i = 1,\dots,N,$$
$$a_N(i) = \mathrm{E}\,\varphi(U_N^{(i)}), \ i = 1,\dots,N,$$

with $U_N^{(i)}$ being the i-th order statistic from the uniform (0,1) distribution, φ is the score-generating function defined on (0,1).

D. The score-generating function φ defined on (0,1) satisfy :

$$|\varphi(u)| \le K\, r(u)^{1/3-\delta}, \quad \delta > 0 \text{ arbitrary},$$
$$|\varphi'(u)| \le K\, r(u)^{3/4-\delta}, \quad \delta > 0 \text{ arbitrary},$$
$$|\varphi''(u)| \le K\, r(u)^{7/4},$$

where φ' and φ'' denote derivatives, K is a constant not depending on u and

$$(1.3) \qquad r(u) = (u(1-u))^{-1}, \ u \in (0,1).$$

The main assertion is the following:

<u>Theorem:</u> Under assumptions A - D there exists a constant d (not depending on N) such that

$$\sup |P(S_N < x(\mathrm{var}\, S_N)^{-1/2}) - \Phi(x)| \le d \sum_{i=1}^{N} |c_{iN}|^3,$$

where $\Phi(.)$ is the normal distribution (0,1).

The assertion of Corollary concerning the two-sample case in Hušková (1977) remains true under ass. A,C,D.

To prove the theorem we combine several known methods. First, the score-generating function φ is replaced by the function

$$(1.4) \qquad \varphi_N(u) = \varphi(u), \quad [N^{\alpha}]/(N+1) \le u \le 1 - [N^{\alpha}]/(N+1),$$
$$= \varphi([N^{\alpha}]/(N+1)), \quad 0 < u < [N^{\alpha}]/(N+1),$$

$$= \varphi(1-[N^{\alpha}]/(N+1)),\ 1>u>1-[N^{\alpha}]/(N+1),$$

where $0<\alpha<\delta/\delta+8$ and $[N^{\alpha}]$ denotes the largest integer not exceeding N^{α}, then $\varphi_N(R_{iN}/(N+1))$ by

$$\varphi_N(F(X_i))+(R_{iN}/(N+1)-F(X_i))\varphi'_N(F(X_i))+2^{-1}(R_{iN}/(N+1)-F(X_i))^2$$
$$\varphi''_N(\eta_{iN}R_{iN}/(N+1)+(1-\eta_{iN})F(X_i)),\ 0\leq\eta_{iN}\leq 1,$$

for $[N^{\alpha}]\leq R_{iN}\leq N-[N^{\alpha}]$, and by $\varphi_N(F(X_i))$ otherwise, and at last the modified method by Callaert, Janssen (1978) is applied to get the convergence rate for the leading terms S_N^*+ $+T_{1N}+T_{2N}$ defined by (1.9 - 11) below (see the proof of Lemma 5 and Theorem).

Now, we give some notations. Define

$$\langle[N^{\alpha}]/(N+1),1-[N^{\alpha}]/(N+1)\rangle = I_{N,\alpha},$$

$$\{R_{iN}/(N+1)\in I_{N,\alpha}\}= A_i,\ i = 1,\dots,N,$$

$$\{F(X_i)\in I_{N,\alpha}\}= B_i,\ i = 1,\dots,N,$$

the complement of a set A will be denoted by A^c and the characteristic function of A by $I\{A\}$.

The function

$$(1.5)\quad h_N(X_i,X_j) = (u(X_i-X_j)-F(X_i))\varphi'_N(F(X_i)) - E((u(X_i-X_j)-F(X_i))\varphi'_N(F(X_i))|X_i),$$

where $u(x)=1$, if $x\geq 0$ and $u(x)=0$, if $x<0$, has the property

$$(1.6)\quad E(h_N(X_i,X_j)|X_i) = E(h_N(X_i,X_j)|X_j)= 0,\ i\neq j=1,\dots,N.$$

Introduce i.i.d. random variables $Z_{1N},\dots,Z_{NN}$ with the property:

$$(1.7)\quad P(Z_{iN}=1)=D_N/N=1-P(Z_{iN}=0),\ i = 1,\dots,N,$$

where $D_N = N\{\sum_{i=1}^{N} |c_{iN}|^3\}^{-\gamma+1}$, $0 < \gamma < \alpha(1/3-4\delta)$.

To prove the theorem by the mentioned method it is suitable to decompose the statistic S_N (1.1) in the following way:

$$S_N = S_N^* + T_{1N} + T_{2N} + \sum_{m=1}^{6} V_{mN}, \tag{1.8}$$

where

$$S_N^* = \sum_{i=1}^{N} c_{iN}(\varphi_N(F(X_i)) - E\varphi_N(F(X_i))), \tag{1.9}$$

$$T_{1N} = (N+1)^{-1} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} c_{iN}(1-Z_{iN})(1-Z_{jN})h_N(X_i, X_j), \tag{1.10}$$

$$T_{2N} = (N+1)^{-1} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} c_{iN}(Z_i + (1-Z_i)Z_j)h_N(X_i, X_j), \tag{1.11}$$

$$V_{1N} = \sum_{i=1}^{N} c_{iN}(\varphi(R_{iN}/(N+1)) - \varphi_N(R_{iN}/(N+1))) \tag{1.12}$$

$$V_{2N} = \sum_{i=1}^{N} c_{iN} I\{(A_i \cap B_i)^c\}(\varphi_N(R_{iN}/(N+1)) - \varphi_N(F(X_i))) \tag{1.13}$$

$$V_{3N} = -(N+1)^{-1} \sum_{i=1}^{N} c_{iN} \int_0^1 (u(x-F(X_i)) - x)\varphi'_N(x)\,dx \tag{1.14}$$

$$V_{4N} = -2(N+1)^{-1} \sum_{i=1}^{N} c_{iN} F(X_i)\varphi'_N(F(X_i)) \tag{1.15}$$

$$V_{5N} = -\sum_{i=1}^{N} c_{iN}\, I\{(A_i \cap B_i)^c\}(R_{iN}/(N+1) - F(X_i))\varphi'_N(F(X_i)) \tag{1.16}$$

$$V_{6N} = 2^{-1} \sum_{i=1}^{N} c_{iN}\, I\{A_i \cap B_i\}(R_{iN}/(N+1) - F(X_i))^2 \varphi''_N(\eta_{iN}R_{iN}/(N+1) + (1-\eta_{iN})F(X_i)), \tag{1.17}$$

with $0 \leq \eta_{iN} \leq 1$.

2. Some Lemmas and the Proof of Theorem

In the proofs of the lemmas the following simple relations will be used repeatedly:

I. $E(U_N^{(i)}-i/(N+1))^{2j} \leq bN^{-j}\ (r(i/(N+1)))^{-j}, j=1,\dots,4$, where $U_N^{(i)}$ is the i-th order statistic from the uniform (0,1) distribution and b is a constant depending neither on i nor on N (see [9]).

II. $\int_{I_{N,\alpha}} (r(u))^{\beta}\, du = O(N^{(1-\alpha)(\beta-1)}),\ \beta > 0,\ 0 < \alpha < 1.$

III. $N^{-1} \sum_{i=[N^{\alpha}]}^{N-[N^{\alpha}]} (r(i/(N+1)))^{\beta} = O(N^{(1-\alpha)(\beta-1)}),\quad \beta > 0,$
$0 < \alpha < 1.$

IV. $\sum_{i=1}^{N} i^{-\gamma} = O(1),\quad \gamma > 1.$

V. $\sum_{i=1}^{N} |c_{iN}|^3 \geq N^{-1/2} \quad \max_{1 \leq i \leq N} |c_{iN}| \leq (\sum_{i=1}^{N} |c_{iN}|^3)^{1/3}.$

<u>Lemma 1</u>. Let assumptions A,B,D be satisfied. If the scores are given by (1.2), then

$$(1.18)\qquad P(|V_{mN}| \geq \sum_{i=1}^{N} |c_{iN}|^3) = O(\sum_{i=1}^{N} |c_{iN}|^3),\ m=1,3,4.$$

Proof. In view of the Chebyshev inequality it suffices to prove that

$$(1.19)\qquad E\,V_{mN}^2 = O((\sum_{i=1}^{N} |c_{iN}|^3)^3).$$

The relation (1.18) for i=3,4, follows easily by direct computations for the random variables V_{3N} and V_{4N} are sums of independent random variables with zero means.

As for V_{1N} we have by Theorem II.3.a in Hájek and Šidák (1967), ass. B and relation IV:

$$E\,V_{1N}^2 \leq (N-1)^{-1} \sum_{i=1}^{N} (\varphi(i/(N+1)) - \varphi_N(i/(N+1)))^2 \leq$$
$$\leq (N-1)^{-1} \sum_{i/(N+1) \notin I_{N,\alpha}} (\varphi(i/(N+1) - \varphi([N^{\alpha}]/(N+i)))^2 \leq$$
$$\leq 2(N-1)^{-1} K \sum_{i=1}^{[N]^{\alpha}} \{i(i-[N^{\alpha}])/(N+1)\}^2\ (r(i/(N+1)))^{-2(-3/4+\delta)}\} \leq$$
$$\leq 2^{5/2-2\delta} N^{-3+2\alpha} (N+1)^{3/2-2\delta} \sum_{i=1}^{[N]^{\alpha}} i^{-3/2+2\delta} = O(N^{-3/2}).$$

Q.E.D.

Lemma 2. Under assumptions of Lemma 1

$$P(|V_{5N}| \geq \sum_{i=1}^{N} |c_{iN}|^3) \doteq O(\sum_{i=1}^{N} |c_{iN}|^3).$$

Proof. The vector of ranks $(R_{1N},\dots,R_{NN})$ and the vector of order statistics are independent under ass. A. Thus one can write (see p. 160 Hájek and Šidák (1967)):

$$(1.20) \quad E\, V_{5N}^2 = E\{E\{(\sum_{i=1}^{N} c_{iN} I\{A_i^c\}(R_{iN}/(N+1) -$$

$$-V^{(R_{iN})})\ \varphi_N'(V^{(R_{iN})}))^2/V^{(\cdot)}\}\} \leq$$

$$\leq (N-1)^{-1} \sum_{i/(N+1) \notin I_{N,\alpha}} E\{(V^{(i)} - i/(N+1))^2 \varphi_N'^2(V^{(i)})\}, \leq$$

$$\leq (N-1)^{-1} \sup_{u \in (0,1)} \varphi'^2(u) \sum_{i/(N+1) \notin I_N} \operatorname{var} V^{(i)}$$

where $V^{(\cdot)} = (V^{(1)},\dots,V^{(N)})$, $V^{(i)}$ denotes the i-th order statistic in the sample $(F(X_1),\dots,F(X_N))$. The last expression in (1.20) is by ass. B and relation I smaller or equal to

$$(N-1)^{-1}(N^{\alpha}/(N+1))^{-3/2+2\delta} \sum_{i/(N+1) \notin I_{N,\alpha}} b\, i(N+1-i)N^{-3} = O(N^{-3/2}).$$

Q.E.D.

Lemma 3. Under assumptions of Lemma 1

$$P(|V_{2N}| \geq \sum_{i=1}^{N} |c_{iN}|^3) = O(\sum_{i=1}^{N} |c_{iN}|^3).$$

Proof. Decompose V_{2N} as follows

$$V_{2N} = \sum_{i=1}^{N} c_{iN}\ I\{A_i^c\}(\varphi_N(R_{iN}/(N+1) - \varphi_N(F(X_i))) +$$

$$+ \sum_{i=1}^{N} c_{iN}\ I\{F(X_i) < [N^{\alpha}]/(N+1)\}(\varphi_N(R_{iN}/(N+1) - \varphi_N(F(X_i))) +$$

$$+ \sum_{i=1}^{N} c_{iN}\ I\{F(X_i) > N - [N^{\alpha}]/(N+1)\}(\varphi_N(R_{iN}/(N+1) - \varphi_N(F(X_i))) =$$

$$= V_{2N}^{*} + V_{2N}^{**} + V_{2N}^{***}.$$

Similarly as in the proof of Lemma 2 we have

$$E\, V_{2N}^{*2} = E\{E\{\sum_{i=1}^{N} c_{iN}(\varphi_N(R_{iN}/(N+1)) - \varphi_N(F(X_i)))^2\, I\{A_i^c\}/V^{(\cdot)}\}\} \leqq$$

$$\leqq (N-1)^{-1} \sum_{i/(N+1)\notin I_{N_\alpha}} E(\varphi_N(i/(N+1)) - \varphi_N(V^{(i)}))^2 \leqq$$

$$\leqq 4(N-1)^{-1} \sum_{i=1}^{[N^\alpha]} \{(i-[N^\alpha])^2\, (N+1)^{-2} + E(V^{(i)} - i/(N+1))^2\} \cdot$$

$$\cdot \varphi_N'^{\,2}([N^\alpha]/(N+1)) = O(N^{-3/2}).$$

To obtain the assertion on V_{2N}^{**} we define the random variables

$$Y_i = 1 \text{ if } F(X_i) < [N^\alpha]/(N+1),\ i=1,\dots,N,$$
$$=0 \text{ otherwise.}$$

Obviously,

$$P(\sum_{i=1}^{N} Y_i \geq b_N) \leq e^{-b_N}\,(E\, e^{Y_1})^N = e^{-b_N}\,(1-[N^\alpha](1-e)/(N+1))^N \leqq$$

$$\leq \exp\{-b_N + e(1-e)[N^\alpha]\}.$$

Thus choosing $b_N = 5[N^\alpha]+1$ we get

$$P(\sum_{i=1}^{N} Y_i \geq b_N) = o(N^{-1/2}).$$

Now,

$$(1.21)\quad P(|V_{2N}^{**}| \geq \sum_{i=1}^{N} |c_{iN}|^3) \leq P(\sum_{i=1}^{N} Y_i > b_N) +$$

$$+ P(\sum_{i=1}^{N} Y_i \leq b_N,\ |V_{2N}^{**}| \geq \sum_{i=1}^{N} |c_{iN}|^3) \leq o(N^{-1/2}) +$$

$$+ \sum_{i=1}^{b_N} \sum_{(i_1,\dots,i_N)}^{*} P(|V_{2N}^{**}| \geq \sum_{i=1}^{N} |c_{iN}|^3 / G_k(i_1,\dots,i_N))$$

$$P(G_k(i_1,\dots,i_N)),$$

where Σ^* denotes the sum over all permutations of $(1,\dots,N)$ and $G_k(i_1,\dots,i_N) = \{Y_{i_1} = \dots = Y_{i_k} = 1,\ Y_{i_{k+1}} = \dots Y_{i_N} = 0\}$. For the conditional probability in (1.21) one can write

$$(1.22)\quad P(|\sum_{i=1}^{N} c_{iN}\, I\{F(X_i) < [N^\alpha]/(N+1)\}\,(\varphi_N(R_{iN}/(N+1)) -$$

$$- \varphi_N(F(X_i)))| \geq \sum_{i=1}^{N} |c_{iN}|^3 | G_k(i_1,\dots,i_N) =$$

$$= P(|\sum_{m=1}^{k} c_{i_m N}(\varphi_N(R_{i_m N}/(N+1)) - \varphi_N([N^\alpha]/(N+1))) \geq$$

$$\geq \sum_{i=1}^{N} |c_{iN}|^3 / G_k(i_1,\dots,i_N)).$$

Clearly,

$$P(R_{i_m N}=j | G_k(i_1,\dots,i_N)) = k^{-1} \quad \text{for } j,m=1,\dots,k$$
$$=0 \quad \text{otherwise.}$$

Thus by Chebyshev inequality the righthand side of (1.22) is smaller or equal to

$$(1.23) \quad (\sum_{i=1}^{N} |c_{iN}|^3)^{-2} \{ k^{-1} \sum_{m=1}^{k} c_{i_m N}^2 (\sum_{i=1}^{k} (\varphi_N(i/(N+1)) - \varphi_N([N^\alpha]/(N+1))))^2 + \sum_{m=1}^{k} \sum_{\substack{v=1 \\ v \neq m}}^{k} c_{i_m N} c_{i_v N} (k(k-1))^{-1} \sum_{i=1}^{k} \sum_{\substack{s=1 \\ s \neq i}}^{k} (\varphi_N(i/(N+1)) - \varphi_N([N^\alpha]/(N+1))(\varphi_N(s/(N+1)) - \varphi_N([N^\alpha]/(N+1)))\}.$$

Notice that

$$(1.24) \quad \Sigma^* \sum_{m=1}^{k} c_{i_m N}^2 \, P(G_k(i_1,\dots,i_N)) = \binom{N-1}{k-1} ([N^\alpha]/(N+1))^k \cdot (1 - [N^\alpha]/(N+1))^{N-k}, \quad k \geq 1,$$

$$(1.25) \quad \Sigma^* \sum_{m=1}^{k} \sum_{\substack{s=1 \\ m \neq s}}^{k} c_{i_m N} c_{i_s N} P(G_k(i_1,\dots,i_N)) = -\binom{N-2}{k-2} ([N^\alpha]/(N+1))^k \cdot (1 - [N^\alpha]/(N+1))^{N-k}, \quad k \geq 2,$$

$$\sum_{k=[N^\alpha]+1}^{b_N} \binom{N-1}{k-1} (k-1)^{-1} \left(\frac{[N^\alpha]}{N+1}\right)^k \left(1 - \frac{[N^\alpha]}{N+1}\right)^{N-k} \leq N^{-1}.$$

By definition of φ_N the expression (1.23) equals zero for $k \leq N^\alpha$. Thus combining the last inequality together with (1.21 - 1.25) and recalling the definition of φ_N we obtain

$$P(|V_{2N}| \geq \sum_{i=1}^{N} |c_{iN}|^3) \leq o(N^{-1/2}) + (\sum_{i=1}^{N} |c_{iN}|^3)^{-2} \sum_{k=[N^\alpha]+1}^{b_N} \Big\{ \binom{N-1}{k-1}$$

$$(k-1)^{-1} \Big(\frac{[N^\alpha]}{N+1}\Big)^k \Big(1 - \frac{[N^\alpha]}{N+1}\Big)^{N-k} \sum_{i=[N^\alpha]+1}^{k} (\varphi_N(i/(N+1)) -$$

$$- \varphi([N^\alpha]/(N+1)))^2 \Big\} \leq o(N^{-1/2}) + (\sum_{i=1}^{N} |c_{iN}|^3)^{-2}\, b_N N^{-1}$$

$$\max_{[N^\alpha] \leq i \leq b_N} (\varphi_N(i/(N+1)) - \varphi_N([N^\alpha]/(N+1)))^2 \leq o(N^{-1/2}) +$$

$$+ (\sum_{i=1}^{N} |c_{iN}|^3)^{-2}\, b_N^3\, N^{-3} (N^{\alpha-1}/2)^{-3/2+2\delta} = o(N^{-1/2}).$$

The member V_{2N}^{***} can be treated in the same way. Q.E.D.

<u>Lemma 4</u>. Under assumptions of Lemma 1

$$P(|V_{6N}| \geq N^{-1/2}) = o(N^{-1/2}).$$

Proof. Since ass. B and the convexity of the function $(r(u))^\gamma$, $\gamma > 0$, one has

$$E\, V_{6N}^2 \leq 4^{-1} \{ E\, (R_{1N}/(N+1) - F(X_1))^4\, \varphi''^2_N (\eta_{1N} R_{1N}/(N+1) + (1 - \eta_{1N})$$

$$F(X_1)) . I\{A_1 \cap B_1\} \} \leq 4^{-1}\, E\{(R_{1N}/(N+1) - F(X_1))^4 .$$

$$. (\eta_{1N} (r(R_{1N}/(N+1)))^{7/2} + (1 - \eta_{1N})(r(F(X_1)))^{7/2}) I\{A_1 \cap B_1\}\} =$$

$$= 4^{-1} (\eta_{1N} D_{n1} + (1 - \eta_{1N}) D_{N2}), \quad 0 \leq \eta_{1N} \leq 1.$$

We estimate D_{N1} and D_{N2} separately. The independence of $(R_{1N}, \dots, R_{NN})$ and $V^{(\cdot)}$ and relations I and II imply

$$D_{N1} \leq E\{E((R_{1N}/(N+1)) - V^{(R_{1N})})^4 (r(R_{1N}/(N+1)))^{7/2}\, I\{A_1\}) | V^{(\cdot)}\} =$$

$$= N^{-1} \sum_{i=[N^\alpha]+1}^{N-[N^\alpha]} E(V^{(i)} - i/(N+1))^4 (r(i/(N+1)))^{7/2} \leq$$

$$\leq b\, N^{-3} \sum_{i=1+[N^\alpha]}^{N-[N^\alpha]} r(i/(N+1))^{3/2} = o(N^{-3/2}).$$

Clearly,

$$E\{(R_{1N}-E(R_{1N}|X_1))^4/X_1\} \le (r(F(X_1)))^{-1}N+(r(F(X_1)))^{-2}N^2.$$

(R_{1N} forms for given X_1 the sum of independent random variables.) Thus we have

$$D_{N2} \le 2^3(N+1)^{-4}\ E\{(R_{1N}-E(R_{1N}|X_1))^4(r(F(X_1)))^{7/2}\ I\{B_1\}\} +$$

$$+E\{(r(F(X_1)))^{7/2}(1-2F(X_1))^4\ I\{B_1\}\} \le$$

$$\le 2^3(N+1))^{-3}\ E\{(r(F(X_1)))^{5/2}\ I\{B_1\}\} + (N-2)\ E\{(r(F(X_1)))^{3/2}$$

$$I\{B_1\}\} + o(N^{-3/2}).$$

Q.E.D.

Lemma 5. Under assumptions of Lemma 1

$$P(|T_{2N}| \ge \sum_{i=1}^{N} |c_{iN}|^3) = O(\sum_{i=1}^{N} |c_{iN}|^3).$$

Proof. Decompose T_{2N} into three summands as follows:

$$T_{2N}=(N+1)^{-1}\sum_{i<j} c_{iN}Z_{iN}h_N(X_i,X_j) +$$

$$+ (N+1)^{-1}\sum_{i>j} c_{iN}Z_{iN}h_N(X_i,X_j) +$$

$$+ (N+1)^{-1}\sum_{i}\sum_{j} c_{iN}Z_{jN}(1-Z_{iN})h_N(X_i,X_j)=Q_{1N}+Q_{2N}+Q_{3N}.$$

For Q_{1N} we can write

$$Q_{1N}=\sum_{i=1}^{N} q_{iN}=\sum_{i=1}^{N}(\sum_{j=i+1}^{N} c_{iN}Z_{iN}h_N(X_i,X_j)(N+1)^{-1}),$$

where q_{iN} are martingale summands $(E(q_{iN}|X_{i+1},\dots,X_N,Z_N)=0)$ and, for given X_i and Z_N, q_{iN} is the sum of independent random variables with zero means. Thus applying the theorem of Dharmadhikary, Fabian and Jodgeo (1968) to Q_{1N} with $\nu=3$ and then to q_{iN} we obtain

$$(1.26)\qquad E\{E|Q_{1N}|^3/Z_N\} \leq (N+1)^{-3}E(\sum_{i=1}^{N} Z_{iN})^{1/2}\sum_{v=1}^{N} Z_{vN}\cdot$$

$$\cdot E|\sum_{j=1+v}^{N} c_{vN}h_N(X_v,X_j)|^3 \leq$$

$$\leq (N+1)^{-1}E(\sum_{i=1}^{N} Z_{iN})^{1/2}\sum_{v=1}^{N} Z_{vN}|c_{vN}|^3N^{3/2}E|h_N(X_v,X_j)|^3.$$

Obviously,

$$E|h_N(X_i,X_j)|^3 \leq k\int_0^1 (u(1-u))|\varphi'_N(u)|^3\,du = O(N^{(1-\alpha)(1/4-3\delta)}),$$

$$E(\sum_{i=1}^{N} Z_{iN})^{1/2}Z_{jN} = O(D_N^{3/2}N^{-1}).$$

Combining the last two inequalities with (1.26) we get

$$E|Q_{1N}|^3 = o(N^{-3/4}(\sum_{i=1}^{N}|c_{iN}|^3)^{5/2}).$$

Similarly,

$$E|Q_{2N}|^3 = o(N^{-3/4}(\sum_{i=1}^{N}|c_{iN}|^3)^{5/2}),$$

$$E|Q_{3N}|^3 = o(N^{-5/2}(\sum_{i=1}^{N}|c_{iN}|^3)^{5/2}).$$

The assertion of our Lemma now follows from the Chebyshev inequality.

Q.E.D.

<u>Proof of Theorem</u>. It suffices to consider the scores (1.2) (see Lemma 2.4 in Hušková (1977)). In view of the proved lemmas and the decomposition (1.8) it suffices to show:

$$(1.27)\qquad \sup_x |P(S_N^*+T_{1N} < x(\mathrm{var}\ S_N^*)^{1/2}) - \Phi(x)| = O(\sum_{i=1}^{N}|c_{iN}|^3).$$

We will show it using the Berry-Esseen argument in a similar way as in [3].

Without loss of generality we may assume that

$$\mathrm{var}\ S_N^* = 1.$$

Put

$$p_N = \sum_{i=1}^{N} |c_{iN}|^3.$$

By Berry-Esseen lemma the righthand side of (1.27) is smaller or equal to

$$(1.28) \quad \int_{|t|<p_N^{-1}\varepsilon} |E \exp\{it(S_N^*+T_{1N})\} - E \exp\{it\, S_N^*\}| |t^{-1}| dt +$$

$$+ \int_{|t|<p_N^{-1}\varepsilon} |E \exp\{it\, S_N^*\} - \exp\{-t^2/2\}| |t^{-1}| dt + O(p_N)$$

for any arbitrary fixed $\varepsilon > 0$.

Since S_N^* is the sum of independent random variables with

$$E|\varphi_N(F(X_i))|^3 < +\infty$$

there exists $\varepsilon_1 > 0$ such that

$$\int_{|t|<p_N^{-1}\varepsilon_1} |E \exp\{it\, S_N^*\} - \exp\{-t^2/2\}| . |t^{-1}| dt = O(p_N).$$

As for the first member in (1.28) one can easily get

$$E(E(T_{1N}^2/Z_N)) = (N+1)^{-2} \sum_{i=1}^{N} \sum_{\substack{j=1\\ j\neq i}}^{N} \{c_{iN}^2 E(1-Z_{iN})(1-Z_{jN}) E\, h_N^2(X_i,X_j) +$$

$$+ c_{iN}c_{jN} E(1-Z_{iN})(1-Z_{jN}) h_N(X_i,X_j) h_N(X_j,X_i)\} = O(N^{-1}),$$

where $Z_N = (Z_{1N},\dots,Z_{NN})$. Further, by direct computations (as in Lemma 2.7 in [6])

$$|E\, T_{1N} \exp\{it\, S_N^*\}| = O(p_N^{3/2}).$$

Thus regarding

$$E(\exp\{it(S_N^*+T_{1N})\} - \exp\{it\, S_N^*\}) = E\, T_{1N} \exp\{it\, S_N\} + 2^{-1} E T_{1N}^2 \gamma_N,$$

$|\gamma_N| \leq 1$, one has

$$(1.29) \quad \int_{|t|<p_N^{-1/2}} |t^{-1}| . |E \exp\{it(S_N^*+T_{1N})\} - E \exp\{it\, S_N^*\}| dt =$$

$$= O(p_N).$$

Put

$$S^*_{1N}=\sum_{i=1}^{N} c_{iN}Z_{iN}(\varphi_N(F(X_i))-E\,\varphi_N(F(X_i))).$$

For the characteristic function of the summands of S^*_{1N} the relation

$$E(\exp\{itc_{jN}(\varphi_N(F(X_j))-E\,\varphi_N(F(X_j)))Z_{jN}\}|Z_{jN}) =$$

$$= 1-t^2c^2_{jN}Z_{jN}/2+ \eta_{jN}|t|^3|c^3_{jN}|Z_{jN}E|\varphi_N(F(X_j))-E\,\varphi_N(F(X_j))|^3/3! \ ,$$

$$|\eta_{jN}|\le 1,$$

holds. Then recalling the definition of Z_{jN} we observe that for $|t|\le p_N^{-1}(E|\varphi_N(F(X_1))-E\,\varphi_N(F(X_1))|^3)^{-1}3/2$

$$|E((E\exp\{it(S^*_N+T_{1N})\}-E\exp\{it\ S_N\})/Z_N)| =$$

$$=|E((E\exp\{it\ S^*_{1N}\}.E\exp\{it(T_{1N}+S^*_N-S^*_{1N})\})/Z_N)|\le$$

$$\le |E(E\exp\{it\ S^*_{1N}\}|Z_N) = \prod_{j=1}^{N} E(1-\frac{t^2}{2}c^2_{jN}Z_{jN}+\eta_{jN}|t^3c_{jN}{}^3|Z_{jN}$$

$$E|\varphi_N(F(X_j))-E\,\varphi_N(F(X_j))|^3/3!)\le \exp\{-t^2(1-$$

$$-\sum_{j=1}^{N}\eta_{jN}|c_{jN}|^3|t|E|\varphi_N(F(X_j))-E\,\varphi_N(F(X_j))|^3/3)D_N/2N\}\le$$

$$\le \exp\{-t^2D_N/(4N)\}.$$

Consequently,

$$\int\limits_{p_N^{-1/2}<|t|\le p_N^{-1}\varepsilon_2} |E\exp\{it(S^*_N+T_{1N})\}-E\exp\{it\ S^*_N\}|.|t^{-1}|dt\le$$

$$\le \exp\{-p_N^{-1}p_N^{1-\gamma}/4\}\log(\varepsilon_2 p_N^{-1})=o(p_N),$$

where $\varepsilon_2=3/2\ E|\varphi_N(F(X_1))-\varphi_N(F(X_1))|^3$.

Combining the last inequality together with (1.28 - 1.29) we can conclude that the first member in (1.28) is $O(p_N)$ and thus

(1.28) is $O(p_N)$ for $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$. Q.E.D.

4. Statement Theorem under contiguous alternatives

We will assume that

E. $X_{1N},\dots,X_{NN}$ be independent random variables, X_{jN} have a density $f(x, \theta_{jN}) \in \mathcal{F}$, where θ_{jN} are unknown parameters and $\mathcal{F}$ is a family of densities $f(x,\theta)$, $\theta \in J$ (J is an open interval containing zero) satisfying

a. $f(x,\theta)$ is absolutely continuous in θ;

b. the limit

$$\dot{f}(x,0) = \lim_{\theta\to 0} \theta^{-1}(f(x,\theta)-f(x,0))$$

exists for almost all x;

c. there exist θ_0 and a constant C such that for all $|\theta| \leq \theta_0$

$$\int \frac{(\dot{f}(x,0))^2}{f(x,0)}\, dx \leq C;$$

F. Unknown parameters $\theta_{1N},\dots,\theta_{NN}$ satisfy:

$$\sum_{j=1}^{N} \theta_{jN}^2 = 1, \qquad \sum_{j=1}^{N} \theta_{jN} = 0.$$

G. The score-generating function defined on (0,1) satisfy:

$$|\varphi(u)| \leq Kr(u)^{+1/4-\delta}, \quad \delta > 0 \text{ arbitrary},$$
$$|\varphi'(u)| \leq Kr(u)^{2/3-\delta} \quad \delta > 0 \text{ arbitrary},$$
$$|\varphi''(u)| \leq Kr(u)^{3/2},$$

where r(u) is defined by (1.3) and K is a constant.

The main assertion of this section:

Theorem. Consider the statistic S_N given by (1.1). Then under assumptions B,C,E,F,G there exist constants A and θ_0 (both not depending on N) such that for $\max_{1\leq j\leq N} |\theta_{jN}| \leq \theta_0$

$$\sup_x |P(S_N - (\mu_N < x(\mathrm{var}\ S_N)^{-1/2}) - \Phi(x)| \leq A \sum_{j=1}^{N} (|c_{jN}| + |\theta_{jN}|)^3,$$

where $\mu_N = \sum_{j=1}^{N} c_{jN} \int_{f(x_j, 0) \neq 0} \varphi'(F(x_j, 0)) \dot{f}(x_j, \theta_{jN}) dx_j$.

<u>Remark</u>. There is some flexibility in ass. E and G in the following sense: admitting milder conditions on the distribution of $X_{1N}, \ldots, X_{NN}$ we must impose stronger conditions on the score-generating function φ and on the contrary.

The proof of the Theorem is omitted for it is very closed to that under the hypothesis (it is easy to prove analogous lemmas utilizing Lemma 3.5 and 3.6 in Hušková (1977)).

References

[1] von BAHR B.(1976): Remainder Term Estimate in a Combinatorial Limit Theorem, Z. Wahrscheinlichkeitstheorie verw. Gebiete 35, 131-139.

[2] BERGSTRÖM H. and PURI M.L.(1977): Convergence and Remainder Terms in Linear Rank Statistics, Ann. Statist. 5, 671-680.

[3] CALLAERT H. and JANSSEN P.(1978): The Berry-Esseen Theorem for U-statistics, 6, 417-421.

[4] ERICKSON R.V. and KOUL H.(1976), Rates of Convergence for Linear Rank Statistics, Ann. Statist. 4, 771-774.

[5] HÁJEK J. and ŠIDÁK Z.(1967): Theory of Rank Tests, Academia, Prague.

[6] HUŠKOVÁ M.(1977): The Rate of Convergence of Simple Linear Rank Statistics under Hypothesis and Alternatives, Ann. Statist. 4, 658-670.

[7] JUREČKOVÁ J. and PURI M.L.(1975): Order of Normal Approximation for Rank Test Statistics Distribution, Ann. Probability 3, 526-533.

[8] DHARMADHIKARI S.W., FABIAN V. and JOGDEO K.(1968): Bounds

on the Moments of Martingales, Ann. Math. Statist., 39, 1719-1723.

[9] WELLNER J.(1974): Convergence of the sequential uniform empirical processes with bounds for centered beta rv's and a log-log law, Technical report no. 31, Univ. of Washington.

Matematicko-fyzikální fakulta

Universita Karlova

Sokolovská 183, 18600 Praha 8

Československo

(Oblatum 4.3.1979)