

AD-A037 766

INDIANA UNIV BLOOMINGTON DEPT OF MATHEMATICS F
CONVERGENCE AND REMAINDER TERMS IN LINEAR RANK STATISTICS. (U)
JAN 77 H BERGSTROM, M L PURI AF-AFOSR-2927

F/G 12/1

AF-AFOSR-2927-76

NL

UNCLASSIFIED

AFOSR-TR-77-0168

1 OF 1
AD
A037766

END

DATE
FILMED

4-77

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-77-0168	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. CONVERGENCE AND REMAINDER TERMS IN LINEAR RANK STATISTICS.		5. TYPE OF REPORT & PERIOD COVERED 9 Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Harold Bergstrom and Madan L. Puri	8. CONTRACT OR GRANT NUMBER(s) 15 AF-AFOSR-2927-76	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Indiana University Department of Mathematics Bloomington, Indiana 47401	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5 12 23 p. 16 17	
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332	12. REPORT DATE January 1977	13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 11 Jan 77	15. SECURITY CLASS. (of this report) UNCLASSIFIED	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES COPY AVAILABLE TO DDC DOES NOT PERMIT FULLY LEGIBLE PRODUCTION		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) linear rank statistics, score generating functions, rate of convergence.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A new approach to the asymptotic normality of simple linear rank statistics for the regression case, studied earlier by Hajek (1968; Ann. Math. Statist., 39, 325-346) is provided along with the estimation of the remainder terms in the approximation to normality.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ADA037766

DDC FILE COPY, AU RJ.

DDC
APR 4 1977
RECEIVED

402 522

(4)



Convergence and Remainder Terms
In Linear Rank Statistics

by

Harold Bergström and Madan L. Puri

University of Göteborg, Chalmers University of Technology,
and Indiana University.

A new approach to the asymptotic normality of simple linear rank statistics for the regression case studied earlier by Ha'jek (1968) is provided along with the estimation of the remainder term in the approximation to normality.

1. Introduction and Summary. Let $X_1, \dots, X_n$ be independent random variables having continuous cdfs (cumulative distribution functions) $F_1(x), \dots, F_n(x)$ respectively. Consider a statistic $S_n = s(X_1, \dots, X_n)$ with $ES_n = 0$ and

Work supported by the Air Force Office of Scientific Research, AFSC, USAF, under Grant No. AFOSR 76-2927. Reproduction in whole or in part is permitted for any purpose of the U.S. Government.

AMS 1970 subject classifications. Primary 62E20 ; Secondary 60F05, 60F99 Key words and phrases. Linear rank statistics, score generating functions, rate of convergence.

Approved for public release;
distribution unlimited.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. BLOSE
Technical Information Officer

Approved for public release
Distribution unlimited

$ES_n^2 < \infty$. Then, to prove the asymptotic normality of S_n (as $n \rightarrow \infty$) , Ha'jek (1968) uses the method of projection which gives to the statistic S_n , the approximation of the form

$$(1.1) \quad \hat{S}_n = \sum_{j=1}^n E[S_n | X_j] .$$

Consider now the simple linear rank statistic S_n introduced by Ha'jek (1962, 1968)

$$(1.2) \quad S_n = \sum_{j=1}^n c_j \{ \psi(R_j/n) - E[\psi(R_j/n)] \}$$

where the c 's are known constants, R_j is the rank of X_j among $(X_1, \dots, X_n)$ and $\psi(\cdot)$ is a score generating function defined on $(0,1)$. Ha'jek (1962) [see also Ha'jek-Sida'k (1967)] established the asymptotic normality of S_n in (1.2) under the assumption that the F_i are contiguous, e.g. when $F_i(x) = F(x - \Delta d_{ni})$ where Δ is the unknown parameter and the d 's are the known constants. Later on Ha'jek (1968) studied the asymptotic normality of S_n for the general $F_i(x)$ (the non-contiguous case). Under the set-up of Ha'jek (1962), Jur'ekova' and Puri (1975) , referred to hereafter as JP, studied the problem of determining the rate of convergence of the cdf of S_n to the limiting normal cdf and established it of order $O(N^{-\frac{1}{2}+\delta})$ for $\delta > 0$. In this paper we not only give a new approach to the asymptotic

normality of S_n for the general F_i (i.e. not necessarily contiguous) but improve the results of JP in providing a sharper bound (for the general F_i 's) . In the passing, we may also mention that where as JP requires ψ to have a bounded fourth derivative, here we only require the boundedness of the second derivative. Furthermore whereas this paper gives more explicit error bounds than the JP paper, the later gives more information on the limiting behavior of ES_n and $\text{Var } S_n$.

We now introduce some notations. We define $\psi(\cdot) = 0$ outside $(0,1)$. Then, we can use the supremum norm

$$(1.3) \quad \|\psi\| = \sup_{t \in (-\infty, \infty)} |\psi(t)|$$

Set

$$(1.4) \quad \rho_i = R_i/n, \quad \rho_{ii} = E[\rho_i | X_i], \quad u(x) = 1 \text{ if } x \geq 0$$

and $u(x) \approx 0$, otherwise.

Then

$$(1.5) \quad R_i = \sum_{j=1}^n u(X_i - X_j) .$$

In this paper, we shall deal with the following approximation of S_n .

ACCESSION for		Write Section	☒
NTIS		Ref Section	☐
DOC			
UNANNOUNCED			
JUSTIFICATION			
BY		DISTRIBUTION AVAILABILITY CODES	
Dist.		Avail. and/or SPECIAL	
A			

$$(1.6) \quad T_n = \sum_{i=1}^n c_i \{ \psi(\rho_{ii}) - E[\psi(\rho_{ii})] + (\rho_i - \rho_{ii}) \psi'(\rho_{ii}) \} ,$$

assuming that ψ' exists on $(0,1)$ and

$$(1.7) \quad \hat{T}_n = \sum_{j=1}^n E[T_n | X_j] .$$

Since $E[(\rho_i - \rho_{ii}) \psi'(\rho_{ii})] = 0$, it follows that

$$(1.8) \quad \hat{T}_n = \sum_{i=1}^n c_i \{ \psi(\rho_{ii}) - E[\psi(\rho_{ii})] + \sum_{j \neq i}^n E[(\rho_i - \rho_{ii}) \psi'(\rho_{ii}) | X_j] \} .$$

Let H_n , G_n and $\hat{G}_n$ be the cdfs of S_n , T_n and $\hat{T}_n$ respectively, and put

$$(1.9) \quad \sigma_n^2 = E[S_n^2] , \quad \hat{\delta}_n^2 = E[\hat{T}_n^2] , \quad \Gamma_{nr}^{2r} = \frac{1}{n} \sum_{i=1}^n c_i^{2r} , \quad \Gamma_{nr} > 0 .$$

Then our theorems are the following:

Theorem 1.1 : If ψ has a derivative on $(0,1)$ then

$$(1.10) \quad \|\hat{G}_n(\hat{\delta}_n) - \Phi(\cdot)\| \leq 4C[2\|\psi\|^3 + \|\psi'\|^3] \sum_{i=1}^n |c_i| \hat{\delta}_n^{-3} ;$$

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^x e^{-t^2/2} dt$$

where C is the constant in Berry-Esseen's inequality

(Zolotarev (1967) gives the approximation 0,9051) . Further

$$(1.11) \quad |\hat{\delta}_n - \sigma_n| \leq C_1 (\|\psi'\| + \|\psi''\|) \Gamma_{n,1}$$

with an absolute constant C_1 , provided ψ'' exists on $(0,1)$.

Theorem 1.2 : If ψ has a second order derivative on $(0,1)$, then for any positive integers n and r such that $n \geq 2r$

$$(1.12) \quad \|H_n(\hat{\delta}_n \cdot) - \phi(\cdot)\| \leq 4C(2\|\psi\|^3 + \|\psi'\|^3) \sum_{i=1}^n |c_i| \hat{\delta}_n^{-3} \\ + C_2 [\hat{\delta}_n^{-1} (\|\psi'\| + \|\psi''\|) r \Gamma_{nr}]^{\frac{2r}{2r+1}}.$$

where C_2 is an absolute constant.

Remark. If the c_i are chosen such that $|c_i| \leq a/\sqrt{n}$ with constant a for all i and n , then

$$\Gamma_{nr} \leq a/\sqrt{n},$$

$$\text{and for } r = [\log n], \quad [r \Gamma_{nr}]^{\frac{2r}{2r+1}} \leq a \sqrt[n]{e} (\log n) n^{-\frac{1}{2}} (1 + O(\frac{1}{\log n}))$$

Note that $\hat{\delta}_n^{-1} c_i$ is invariant and thus also $\hat{\delta}_n^{-1} \Gamma_{nr}$ is invariant under the transformation $c_i \rightarrow \gamma c_i$, $i = 1, 2, \dots$.

2. Same Lemmas.

Lemma 2.1 : For any positive integers r and n , $2r \leq n$, we have

$$(2.1) \quad E[(\rho_i - \rho_{ii})^{2r}] \leq b(r) n^{-r}$$

with

$$(2.2) \quad b(r) \leq n^{-r} \sum_{t=1}^r \binom{n-1}{t} \frac{(2r)!}{(2r-2t)!} t^{2r-2t} \cdot 2^{-3t}$$

$$\text{and for } n^{-1} r^3 \leq \frac{3}{4}$$

$$(2.3) \quad b(r) \leq 2^{-3r} \frac{(2r)!}{r!} [1 + 8n^{-1} r^3]$$

Proof: By (1.4) we obtain

$$\rho_i - \rho_{ii} = \frac{1}{n} \sum_{j \neq i}^n [u(X_i - X_j) F_j(X_i)] .$$

By the polynomial theorem we then get

$$(2.4) \quad E[(\rho_i - \rho_{ii})^{2r}] = n^{-2r} \sum \frac{(2r)!}{s_1! \dots s_n!} E \prod_{j=1}^n [u(X_i - X_j) - F_j(X_i)]^{s_j} .$$

$$s_1 + \dots + s_n = 2r .$$

We claim that any term in this sum is equal to zero if $s_{j_0} = 1$ for some j_0 . Indeed we find that the conditional expectation of the product with respect to all X_j , $j \neq j_0$ is equal to 0 if $s_{j_0} = 1$. Hence we have only to regard terms with $s_j = 0$ or ≥ 2 for any j , and there can be at most $t \leq r$ exponents s_j different from 0. If $s_j \geq 2$, $j = 1, 2, \dots, t$, $s_j = 0$ for $j > t$, $i > t$ we obtain, observing that

$$|u(X_i - X_j) - F_j(X_i)| \leq 1$$

$$\begin{aligned} (2.5) \quad E \left[\prod_{j=1}^t [u(X_i - X_j) - F_j(X_i)]^{s_j} \right] &\leq E \prod_{j=1}^t [u(X_i - X_j) - F_j(X_i)]^2 \\ &= E \left[\prod_{j=1}^t [F_j(X_i) - F_j^2(X_i)] \right] \leq 4^{-t} \end{aligned}$$

This inequality remains true for all permutations of the indices $1, \dots, n$. Put

$$(2.6) \quad \gamma(t) = \sum_{\substack{s_1 + \dots + s_t = 2r \\ s_j \geq 2, j=1, \dots, t}} \frac{(2r)!}{s_1! \dots s_t!} \quad .$$

Since t indices out of $n-1$ indices can be chosen in $\binom{n-1}{t}$ different ways we obtain from (2.4) through (2.6),

$$(2.7) \quad E[(\rho_i - \rho_{ii})^{2r}] \leq -2r \sum_{t=1}^r \binom{n-1}{t} \gamma(t) 4^{-t} \quad .$$

We claim that

$$(2.8) \quad \gamma(t) \leq \frac{(2r)!}{(2r-2t)!} 2^{-t} 2^{r-2t} \quad .$$

Indeed, differentiating the identity

$$\left(\sum_{j=1}^t y_j\right)^{2r} = \sum_{s_1+\dots+s_t=2r} \frac{(2r)!}{s_1! \dots s_t!} \prod_{j=1}^t y_j^{s_j}$$

twice with respect to all y_j and then putting all y_j equal to 1, we obtain

$$\frac{(2r)!}{(2r-2t)!} t^{(2r-2t)} = \sum_{s_1+\dots+s_t=2r} \prod_{j=1}^t s_j (s_j-1) \frac{(2r)!}{s_1! \dots s_t!}$$

$$s_j \geq 2, j = 1 \dots t$$

Now using (2.7) and (2.8), we get (2.1) and (2.2). We now estimate $b(r)$ further, mainly for use when n and r are large. Put $r-t = u$. Then we can write

$$(2.9) \quad b(r) \leq 2^{-3r} \sum_{u=0}^{r-1} k(u)$$

with

$$k(u) = \frac{n^{-u} (2r)! (r-u)^{2u} 2^{3u}}{(r-u)! (2u)!}$$

Particularly

$$k(0) = \frac{(2r)!}{r!}, \quad k(1) < 4n^{-1} r^3 \cdot \frac{(2r)!}{r!}$$

and for $u \geq 1$

$$\frac{k(u+1)}{k(u)} = n^{-1} \left(1 - \frac{1}{r-u}\right)^{2u} \cdot 2^3 \cdot \frac{(r-u)(r-u-1)^2}{(2u+1)(2u+2)}$$

$$< \frac{2}{3} n^{-1} r^3 \leq \frac{1}{2} \quad \text{for} \quad n^{-1} r^3 \leq \frac{3}{4}$$

Hence

$$b(r) \leq 2^{-3r} \cdot \frac{(2r)!}{r!} [1 + 8n^{-1}r^3]$$

$$\text{for } n^{-1}r^3 \leq \frac{3}{4}.$$

Lemma 2.2 : For any positive integers r and n,
 $2r \leq n$, we have

$$(2.10) \quad E(T_n - \hat{T}_n)^{2r} \leq c(r) \|\psi'\|^{2r} \Gamma_{n,r}^{2r}$$

if ψ' exists on $(0,1)$, and if ψ'' exists on $(0,1)$

$$(2.11) \quad E[(S_n - T_n)^{2r}] \leq b(2r) \|\psi''\|^{2r} \Gamma_{n,r}^{2r},$$

$$(2.12) \quad E[(S_n - \hat{T}_n)^{2r}] \leq d(r, \psi) \Gamma_{n,r}^{2r}$$

with

$$b(2r) \leq n^{-2r} \sum_{t=1}^{2r} \binom{n-1}{t} \frac{(4r)!}{(4r-2t)!} t^{4r-2t} \cdot 2^{-3t}$$

$$c(r) \leq 2^{2r} n^{-2r} \sum_{t=1}^{2r} \binom{n}{t} \frac{(4r)!}{(4r-2t)!} t^{4r-2t} \cdot 2^{-t}$$

$$d(r, \psi) \leq [b(2r)]^{\frac{1}{2r}} \|\psi''\| + [c(r)]^{\frac{1}{2r}} \|\psi'\|^{2r}.$$

Further we have the estimates

$$(2.13) \quad b(2r) \leq 2^{-6r} \frac{(4r)!}{(2r)!} [1 + 2^{6n-1} r^3]$$

$$\text{for } 2^{3n-1} r^3 \leq \frac{3}{4},$$

$$(2.14) \quad c(r) \leq \frac{(4r)!}{(2r)!} [1 + 2^{3n-1} r^3] \text{ for } n^{-1} r^3 \leq \frac{3}{8}$$

Remark : By Stirling's approximation of the Γ -function we have

$$\frac{(4r)!}{(2r)!} \leq 2^{6r+\frac{1}{2}} r^{2r} (\exp-2r) \exp \frac{1}{48r}.$$

Proof : By (1.6) and (1.8) we get

$$(2.15) \quad T_n - \hat{T}_n = \sum_{i=1}^n c_i \{ (\rho_i - \rho_{ii}) \psi'(\rho_{ii}) - \sum_{\substack{j=1 \\ j \neq i}}^n E[(\rho_i - \rho_{ii}) \psi'(\rho_{ii}) | X_j] \}$$

and for $j \neq i$

$$(2.16) \quad E[(\rho_i - \rho_{ii}) \psi'(\rho_{ii}) | X_j] = \frac{1}{n} \sum_{k \neq i}^n E\{[u(X_i - X_k) - F_k(X_i)] \psi'(\rho_{ii}) | X_j\} = \frac{1}{n} E[u(X_i - X_j) - F_j(X_i)] \psi'(\rho_{ii}) | X_j]$$

since the conditional expectations in the sum are zero for $j \neq k, i$. Now using the relation

$$(\rho_i - \rho_{ii}) \psi'(\rho_{ii}) = \frac{1}{n} \sum_{j \neq i}^n [u(X_i - X_j) - F_j(X_i)] \psi'(\rho_{ii})$$

and noting that

$$E[(\rho_i - \rho_{ii})\psi'(\rho_{ii}) | X_i] = 0$$

we obtain from (2.15)

$$(2.17) \quad T_n - \hat{T}_n = \frac{1}{n} \sum_{i=1}^n \sum_{j \neq i}^n c_i V_{ij}$$

with

$$(2.18) \quad V_{ij} = [u(X_i - X_j) - F_j(X_i)]\psi'(\rho_{ii}) - E[u(X_i - X_j) - F_j(X_i)]\psi'(\rho_{ii}) | X_j] .$$

Clearly

$$(2.19) \quad E[V_{ij} | X_j] = 0, \quad E[V_{ij} | X_i] = 0 .$$

By the polynomial theorem we get

$$(2.20) \quad E[(T_n - \hat{T}_n)^{2r}] = n^{-2r} E\left[\sum_{i=1}^n \sum_{j \neq i}^n c_i V_{ij}\right]^{2r} \\ = n^{-2r} \sum_{\substack{\prod_{i=1}^n \prod_{j \neq i}^n (s_{ij}!)}} \frac{(2r)!}{\prod_{i=1}^n \prod_{j \neq i}^n (s_{ij}!)} E\left\{\prod_{i=1}^n \prod_{j \neq i}^n (c_i V_{ij})^{s_{ij}}\right\}$$

where the sum should be taken over terms corresponding to different vector solutions $\{s_{ij}\}$, $i, j = 1, \dots, n$, $j \neq i$

of the equation

$$(2.21) \quad \sum_{i=1}^n \sum_{j \neq i}^n s_{ij} = 2r .$$

The expectation

$$(2.22) \quad E \left[\prod_{i=1}^n \prod_{j \neq i}^n v_{ij}^{s_{ij}} \right]$$

is equal to 0 for some vector solutions of (2.21) since (2.19) holds, and we have only to regard those solutions for which the expectation (2.22) is not equal to 0 .

We say that s_{ij} gives the contribution $\frac{1}{2} s_{ij}$ to the sum (2.21) from each of the indices i and j . Hence according to this notation an index k gives the contribution

$$(2.23) \quad g(k) = \frac{1}{2} \sum_{j \neq k}^n s_{kj} + \frac{1}{2} \sum_{j \neq k}^n s_{jk}$$

to the sum (2.21). By conditioning with respect to all X_j , $j \neq k$ we easily find that the expectation (2.22) is equal to 0 if k gives the contribution $\frac{1}{2}$ to the sum (2.21), i.e. if $s_{kj} = 1$ for exactly one index $j \neq k$, and $s_{jk} = 0$ for $j \neq k$ or if $s_{jk} = 1$ for exactly one j and $s_{kj} = 0$ for $j \neq k$.

The sum Σ on the right hand side of (2.20) can be divided into partial sums as follows. Let C be a collection of different positive integers belonging to the set

$1, \dots, 2r$, say $C = (1, 2, \dots, t)$. Let Σ_C consist of all terms in (2.20) corresponding to the vector solutions of (2.21) such that

- (a) $s_{ij} = 0$ if not both i and j belong to C ;
- (b) for any $k \in C$ the contribution to the sum (2.21) is larger than $1/2$. Note that C can contain at most $2r$ different integers since every $k \in C$ gives at least the contribution 1 to the sum (2.21). Clearly partial sums Σ_{C_1} and Σ_{C_2} contain no common terms if $C_1 \neq C_2$. Consider now the expectation

$$E \left[\prod_{i=1}^t \prod_{j \neq i}^t (c_i v_{ij})^{s_{ij}} \right]$$

where the i and j belong to the collection C . Note that s_{ij} may be equal to 0 for some pairs (i, j) . By Hölder's inequality we get, using the fact that

$$|v_{ij}| \leq 2 \|\psi'\|$$

$$\begin{aligned} (2.24) \quad |E \prod_{i=1}^t \prod_{j \neq i}^t (c_i v_{ij})^{s_{ij}}| &\leq \prod_{i=1}^t \prod_{j \neq i}^t |c_i|^{s_{ij}} \{E[(v_{ij})^{2r}]\}^{\frac{s_{ij}}{2r}} \\ &\leq 2^{2r \|\psi'\|^{2r}} \prod_{i=1}^t |c_i|^{s_i} \end{aligned}$$

where

$$(2.25) \quad s_i = \sum_{j=1}^t s_{ij}, \quad \sum_{i=1}^t s_i = 2r.$$

The partial sum corresponding to C is then estimated by

$$(2.26) \quad \sum'_C \frac{(2r)!}{t!} (2^{2r} \prod_{i=1}^t |c_i|^{s_i}) \prod_{i=1}^t \prod_{j \neq i} (s_{ij})!$$

Note that $\frac{(2r)!}{t!} \prod_{i=1}^t \prod_{j \neq i} (s_{ij})!$ is an integer. Hence

$$\text{we have } N(t) = \sum'_C \frac{(2r)!}{t!} \prod_{i=1}^t \prod_{j \neq i} (s_{ij})!$$

terms in the class C which are estimated by (2.24). Let C_t be the set of all terms

$$\sum_{i=1}^n \prod_{j \neq i} (c_i v_{ij})^{s_{ij}}$$

in (2.20) which belong to some class C containing exactly t indices. Let $(s_1, s_2, \dots, s_t)$ in (2.26) be given, $0 \leq s_1 \leq s_2 \leq \dots \leq s_t$, $\sum_{i=1}^t s_i = 2r$. Then according to the symmetry the set C_t contains a sum of terms, each estimated by

$$(2.27) \quad 2^{2r} \prod_{i=1}^t |c_{k_i}|^{s_i}$$

where $(k_1, \dots, k_t)$ is any combination of numbers $1, 2, \dots, n$ to the t^{th} class and in any order within this class. Let the number of terms in C_t for a fixed vector

$(s_1, s_2, \dots, s_t)$ as above be $n(t)$ and the sum of terms (2.27) belonging to $(s_1, s_2, \dots, s_t)$ be $A(s_1, s_2, \dots, s_t)$. (Note that $n(t)$ depends on $s_1, \dots, s_t$). Then, since $A(s_1, \dots, s_t)$ is a symmetrical function

$$(2.28) \quad A(s_1, s_2, \dots, s_t) = \frac{n(t)}{n!} \sum' 2^{2r} \|\psi'\|^{2r} \prod_{i=1}^t |c_{k_i}|^{s_i}$$

where $\sum'$ is the sum all terms belonging to all permutations of the numbers $1, 2, \dots, n$. By Hölder's inequality we get observing that

$$(2.29) \quad \sum' \prod_{i=1}^t |c_{k_i}|^{s_i} \leq \prod_{i=1}^t (\sum' c_{k_i}^{2r})^{\frac{s_i}{2r}},$$

$$|c_{k_i}|^{s_i} = [c_{k_i}^{2r}]^{\frac{s_i}{2r}}, \quad \sum_{i=1}^t \frac{s_i}{2r} = 1,$$

and here

$$\sum' c_{k_i}^{2r} = \frac{n!}{n} \sum_{i=1}^n c_i^{2r}.$$

Hence we obtain by (2.28) and (2.29)

$$A(s_1, s_2, \dots, s_t) \leq 2^{2r} \|\psi'\|^{2r} \cdot n(t) \cdot \frac{1}{n} \sum_{i=1}^n c_i^{2r}.$$

Since C_t contains $\binom{n}{t} N(t)$ terms we then find that C_t gives at most the contribution

$$n^{-2r} 2^{2r} \|\psi'\|^{2r} \binom{n}{t} N(t) \cdot \frac{1}{n} \sum_{i=1}^n c_i^{2r}$$

to the right hand side of (2.20). Putting

$$\Gamma_{nr}^{2r} = \frac{1}{n} \sum_{i=1}^n c_i^{2r}, \quad \Gamma_{nr} \geq 0,$$

and regarding the sets c_t for $t = 1, 2, \dots, 2r$, we obtain from (2.20) that

$$(2.30) \quad E[(T_n - \hat{T}_n)^{2r}] \leq 2^{2r} n^{-2r} \|\psi'\|^{2r} \Gamma_{nr}^{2r} \cdot \sum_{t=1}^{2r} \binom{n}{t} N(t).$$

We estimate $N(t)$ in the following way. Consider the identity

$$(2.31) \quad \left(\sum_{i=1}^t \sum_{j=i}^t x_i x_j \right)^{2r} = \sum_{\substack{\prod_{i=1}^t \prod_{j=1}^t (s_{ij}) \\ i=1, j=1}} \frac{(2r)!}{t^t} \prod_{i=1}^t \prod_{j \neq i}^t (x_i x_j)^{s_{ij}}.$$

If an index k gives the contribution ≥ 1 to the sum (2.21), i.e. to the sum

$$\sum_{i=1}^t \sum_{j \neq i}^t s_{ij} = 2r$$

then the double product

$$\prod_{j=1}^t \prod_{j \neq i}^t (x_i x_j)^{s_{ij}}$$

contains x_k as factor at least in the power 2. Hence differentiating the identity twice with respect to each x_k , $k = 1, 2, \dots, t$ and then putting all x_n equal to 1 we get the inequality

$$(2.32) \quad 2^t N(t) \leq \left\{ \prod_{k=1}^t \frac{\partial^2}{\partial x_k^2} \left(\sum_{i=1}^t \sum_{j \neq i}^t x_i x_j \right)^{2r} \right\}_{x_k=1, k=1, 2, \dots, t}.$$

The right hand side, however, is at most equal to

$$(2.33) \quad \left\{ \prod_{k=1}^t \frac{\partial^2}{\partial x_k^2} \left(\sum_{i=1}^t x_i \right)^{4r} \right\}_{x_k=1, k=1, 2, \dots, t} = \frac{(4r)!}{(4r-2t)!} t^{4r-2t}.$$

Combining (2.30), (2.32) and (2.33), we get

$$E[(T_n - \hat{T}_n)^{2r}] \leq c(r) \|\psi'\|^{2r} \Gamma_{nr}^{2r}$$

with

$$c(r) = 2^{2r} n^{-2r} \sum_{t=1}^{2r} \binom{n}{t} \frac{(4r)!}{(4r-2t)!} t^{4r-2t} \cdot 2^{-t}$$

$$\Gamma_{nr}^{2r} = \frac{1}{n} \sum_{i=1}^n |c_i|^{2r}.$$

We estimate $c(r)$ exactly in the same way as we have estimated $b(r)$ in Lemma 2.1 and then obtain for $u = 2r - t$

$$c(r) \leq \sum_{u=0}^{r-1} k(u)$$

with

$$k(u) = n^{-u} \frac{(4r)!}{(2u)!(2r-u)!} (2r-u)^{2u} \cdot 2^u .$$

Hence

$$k(0) = \frac{(4r)!}{(2r)!} , \quad k(1) < n^{-1} \cdot (2r)^3 \frac{(4r)!}{(2r)!}$$

and for $u \geq 1$

$$\frac{k(u+1)}{k(u)} \leq \frac{4}{3} n^{-1} r^3 \leq \frac{1}{2} \quad \text{for} \quad n^{-1} r^3 \leq \frac{3}{8} .$$

Hence for $n^{-1} r^3 \leq \frac{3}{8}$

$$c(r) \leq \frac{(4r)!}{(2r)!} [1 + 8n^{-1} r^3] .$$

Thus we have proved (2.13) and (2.14) of the lemma.

It follows by the definition of T_n that

$$S_n - T_n = \sum_{i=1}^n c_i [\xi_i - E(\xi_i)]$$

with

$$|\xi_i| \leq \frac{1}{2}(\rho_i - \rho_{ii})^2 \|\psi''\| .$$

Hence

$$E[(S_n - T_n)^{2r}] \leq n^{2r-1} \sum_{i=1}^n c_i^{2r} E[(\xi_i - E\xi_i)^{2r}]$$

and by Lemma 2.1

$$\begin{aligned} E[(\xi_i - E(\xi_i))^{2r}] &\leq 2^{2r} E[\xi_i^{2r}] \\ &\leq \|\psi''\|^{2r} E[(\rho_i - \rho_{ii})^{4r}] \leq n^{-2r} b(2r) \|\psi''\|^{2r} . \end{aligned}$$

Thus we get (2.11)

$$E[(S_n - T_n)^{2r}] \leq b(2r) \Gamma_{nr} .$$

By Minkovski's inequality we obtain (2.12) from (2.10) and (2.11)

$$E \frac{1}{2r} [(S_n - \hat{T}_n)^{2r}] \leq E \frac{1}{2r} [(S_n - T_n)^{2r}] + E \frac{1}{2r} [(T_n - \hat{T}_n)^{2r}] .$$

Lemma 2.3: $\hat{T}_n = \sum_{j=1}^n \hat{T}_n^{(j)}$ with independent random variables

$$(i) \quad \hat{T}_n^{(j)} = c_j \{ \psi(\rho_{jj}) - E[\psi(\rho_{jj})] \} \\ + \frac{1}{n} \sum_{i \neq j}^n c_i [E(u(X_i - X_j) - F_j(X_i)) \psi'(\rho_{ii}) | X_j] .$$

Further

$$(ii) \quad \sum_{j=1}^n [E|\hat{T}_n^{(j)}|^3] \leq 4[2\|\psi\|^3 + \|\psi'\|^3] \sum_{j=1}^n |c_j|^3 .$$

Proof: We get the representation (i) by (2.16). Using well-known inequalities

$$|(a+b)^3| \leq 4[|a|^3 + |b|^3] , \quad |(\sum_{i=1}^n a_i)^3| \leq n^2 \sum_{i=1}^n |a_i|^3$$

we obtain

$$E[|\hat{T}_n^{(j)}|^3] \leq 4|c_j|^3 E[|\psi(\rho_{jj}) - E\psi(\rho_{jj})|^3] \\ + \frac{4}{n} \sum_{i \neq j}^n |c_i|^3 \|\psi'\|^3 .$$

Here

$$E[|\psi(\rho_{jj}) - E\psi(\rho_{jj})|^3] \leq 2\|\psi\| E(\psi(\rho_{jj}) - E(\psi(\rho_{jj})))^2 .$$

Thus we get (ii).

3. Proofs of the theorems.

(a) Proof of Theorem 1.1 : (1.10) follows from Berry-Esséen's inequality and Lemma 2.3 and (1.11) from Lemma 2.2 (2.12).

(b) Proof of Theorem 1.2. For $h > 0$ we get

$$(3.1) \quad P[S_n \leq \hat{\delta}_n x] \leq P(S_n \leq \hat{\delta}_n x, |S_n - \hat{T}_n| < h \hat{\delta}_n) \\ + P[|S_n - \hat{T}_n| \geq h \hat{\delta}_n] \leq P[\hat{T}_n \leq \hat{\delta}_n(x+h)] + P[|S_n - \hat{T}_n| \geq h \hat{\delta}_n].$$

Applying Theorem 1.1 we get

$$(3.2) \quad P[\hat{T}_n \leq \hat{\delta}_n(x+h)] \leq \Phi(x+h) + 4C(2\|\psi\|^3 + \|\psi'\|^3) \cdot \sum_{i=1}^n |c_i|^3 \hat{\delta}_n^{-3}.$$

Here

$$(3.3) \quad \Phi(x+h) \leq \Phi(x) + h \|\Phi'\| = \Phi(x) + \frac{h}{\sqrt{2\pi}}.$$

By Chebychev's inequality and the inequality (2.12) of Lemma 2.2 we get

$$(3.4) \quad P[|S_n - \hat{T}_n| \geq h \hat{\delta}_n] \leq d(r, \psi) \Gamma_{nr}^{2r} (h \hat{\delta}_n)^{-2r}.$$

Now we choose n such that

$$\frac{h}{\sqrt{2\pi}} = d(r, \psi) \Gamma_{nr}^{2r} (h \hat{\delta}_n)^{-2r}$$

i.e.

$$(3.5) \quad h = [(2\pi)^{\frac{1}{2}} d(r, \psi) \hat{\delta}_n^{-2r} \Gamma_{nr}^{2r}]^{\frac{1}{2r+1}}.$$

It follows by Lemma 2.2 (2.12), (2.13) and (2.14) and the remark made there in Lemma 2.2 that for $n^{-1}r^3 \leq 3/8$

$$[d(r, \psi)]^{\frac{1}{2r}} \leq c'r(\|\psi'\| + \|\psi''\|)$$

with an absolute constant c' . Then it follows by (3.4) and (3.5) that

$$\begin{aligned} & \frac{h}{\sqrt{2\pi}} + d(r, \psi) \Gamma_{nr}^{2r} (h \hat{\delta}_n)^{-2r} \\ & \leq c_2 [\hat{\delta}_n^{-1} (\|\psi'\| + \|\psi''\|) r \Gamma_{nr}]^{\frac{2r}{2r+1}}. \end{aligned}$$

By (3.1) - (3.6) we get the inequality (1.12) in one direction. It follows for the other direction in the same way.

4. Acknowledgements. The authors would like to express their sincere appreciation to Professor I.R. Savage, the associate editor, and the referees for the critical examination of the original draft. Their constructive criticism,

detailed comments and suggestions for improvements are gratefully acknowledged.

REFERENCES

- [1] Jurečková', J. and Puri, Madan L. (1975). Order of normal approximation for rank test statistics distribution. Ann. Prob. 3 526-533.
- [2] Ha'jek, J. (1962). Asymptotically most powerful rank order tests. Ann. Math. Statist. 33 1124-1147.
- [3] Ha'jek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. 39 325-346.
- [4] Ha'jek, J. and Sidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.
- [5] Zolotarev, V.M. (1967). A sharpening of the inequality of Berry-Esseen. Z. Wahrscheinlich-keitstheorie und Verw. Gebiete. 8 332-342.

FILM
4

