Kyungpook Math. J.
Volume 23, Number 1
June, 1983

A BOUND ON CHARACTERISTIC FUNCTIONS OF SIGNED LINEAR RANK STATISTICS

By Munsup Seoh

1. Introduction

Let $X_{N1}, X_{N2}, \cdots, X_{NN}$ be independent r.v.'s (random variables) with density functions $f_{N1}, f_{N2}, \cdots, f_{NN}$, respectively, and let R^+_{Nj}, $1 \leq j \leq N$, be the rank of $|X_{Nj}|$ among $\{|X_{Nk}| : 1 \leq k \leq N\}$. We shall consider the signed linear rank statistic

$$(1.1) \qquad T^+_N = \sum_{j=1}^{N} c_{Nj} a_{NR^+_{Nj}} \operatorname{sgn} X_{Nj},$$

where $c_{N1}, c_{N2}, \cdots, c_{NN}$ are arbitrary regression constants; $a_{N1}, a_{N2}, \cdots, a_{NN}$ are scores and sgn $x=1$ or -1 according as $x \geq 0$ or $x < 0$.

The statistics of the type T^+_N are often used (see e.g., Hájek (1962), Hájek and Šidák (1967), chapters V and VI) for testing the hypothesis $H: f_{N1}=f_{N2}=\cdots=f_{NN}=\tilde{f}_N$ and $\tilde{f}_N(x)=\tilde{f}_N(-x)$, against certain classes of alternatives. The case when $c_{N1}=c_{N2}=\cdots=c_{NN}=1$ is well known.

Define for each N, $N=1, 2, \cdots,$

$$(1.2) \qquad T^*_N = \frac{T^+_N - ET^+_N}{\sigma_N}, \quad F^*_N(x) = P(T^*_N \leq x)$$

where σ^2_N is either exact variance of T^+_N or some normalizing constant.

It is well known (see e.g., Hájek and Šidák (1967), Hušková (1970) and Puri and Ralescu (1981)) that under suitable assumptions, T^*_N has asymptotically, as $N \longrightarrow \infty$, the standard normal distribution, i.e., $\lim_{N\to\infty} \Lambda_N = 0$ where $\Lambda_N = \sup_x |F^*_N(x) - \Phi(x)|$ with the standard normal distribution function $\Phi(x)$. However, one needs more precise information than the asymptotic normality can provide. For example, one may try to find (i) suitable order bounds for Λ_N or (ii) some polynomials $Q_j(x)$, $j=1, 2, \cdots, r$, such that

$$(1.3) \qquad \sup_x |F^*_N(x) - \tilde{F}_N(x)| = 0(N^{-r/2}),$$

where, with the standard normal density function $\phi(x)$,

$$\tilde{F}_N(x)=\Phi(x)+\phi(x)\sum_{j=1}^{r} N^{-j/2}Q_j(x). \tag{1.4}$$

Such an expansion $\tilde{F}_N(x)$ is typically called the Edgeworth expansion up to $(r+1)$ terms. Interested readers are referred to Bickel (1974).

An application of Esséen's smoothing lemma (see e.g., Feller (1971), p.538) to our problems (i) and (ii) yields that for any $\gamma>0$ and $\varepsilon\geq 0$

$$\sup_x |F_N^*(x)-\tilde{F}_N(x)|\leq\frac{1}{\pi}\int_{|t|\leq\gamma N^{(r+\varepsilon)/2}} |\phi_N^*(t)-\tilde{\phi}_N(t)|/|t|dt+0(N^{-(r+\varepsilon)/2}) \tag{1.5}$$

where ϕ_N^* is the characteristic function of T_N^*, and $\tilde{\phi}_N$, the Fourier-Stieltjes transform of $\tilde{F}_N$, is of the form

$$\tilde{\phi}_N(t)=e^{-t^2/2}\{1+\sum_{j=1}^{r} N^{-j/2}\tilde{Q}_j(t)\}, \tag{1.6}$$

with some polynomials $\tilde{Q}_j$, $1\leq j\leq r$. To attack above problems (i) and (ii), we shall split the integral in (1.5) into two parts to obtain the following two facts;

$$\int_{|t|\leq\log N} |\phi_N^*(t)-\tilde{\phi}_N(t)|/|t|dt=o(N^{-r/2}), \tag{1.7}$$

$$\int_{\log N\leq|t|\leq\gamma N^{(r+\varepsilon)/2}} |\phi_N^*(t)-\tilde{\phi}_N(t)|/|t|dt=o(N^{-r/2}). \tag{1.8}$$

As Van Zwet (1982) has mentioned in a different context, proving (1.7) is generally a difficult and highly technical affair, the difficulty lying not so much in finding $\tilde{\phi}_N$ and proving (1.7) but in doing so under reasonably mild assumptions. However, proving (1.8) is a problem of an entirely different nature because it is essentially a smoothness property of the distribution function F_N^* and generally applicable methods for establishing it are not available.

The aim of this paper is to find a sufficient condition for (1.8) in the case of signed linear rank statistic, restricting the Edgeworth expansion valid to three terms (i.e., $r=2$). Because of (1.6), it is sufficient, in order to prove (1.8), to show

$$\int_{\log N\leq|t|\leq\gamma N^{3/2}} |\phi_N^*(t)|/|t|dt=o(N^{-1}). \tag{1.9}$$

The method to prove (1.9) is a follow-up of van Zwet (1982). By accomplishing (1.7) and also using (1.8), Puri and Seoh (1981a, 1981b) have derived Berry-Esséen's bound of order $O(N^{-\frac{1}{2}})$ and an Edgeworth expansion with remainder $o(N^{-1})$, respectively, for a wide class of the signed linear rank statistics including normal scores case. In forthcoming papers Seoh and Puri (1983c, 1983d), Berry-Esséen's theorem and asymptotic expansions shall be established for the statistic under contiguous location alternatives.

2. Assumptions and main theorem

Let for any real numbers $\zeta > 0$, $\theta(a_{N1}, \cdots, a_{NN} : \zeta)$ denote the Lebesgue measure λ of ζ-neighborhood of the set $\{a_{N1}, a_{N2}, \cdots, a_{NN}\}$, thus $\theta(a_{N1}, a_{N2}, \cdots, a_{NN} : \zeta) = \lambda\{x : |x - a_{Nj}| < \zeta \text{ for some } j\}$.

ASSUMPTIONS: Suppose that there exist positive numbers c, C, a, A, δ, a sequence $\tilde{f}_N$ of densities and a sequence $\varepsilon_N \downarrow 0$ such that

(A.1) $$\sum_{j=1}^{N} |c_{Nj}| \geq cN, \quad \sum_{j=1}^{N} c_{Nj}^2 \leq CN$$

(A.2) $$\sum_{j=1}^{N} |a_{Nj}| \geq aN, \quad \sum_{j=1}^{N} a_{Nj}^2 \leq AN$$

(A.3) $$\theta(a_{N1}, a_{N2}, \cdots, a_{NN} : \zeta) \geq \delta N \zeta \text{ for some } \zeta \geq N^{-3/2} \log N$$

(A.4) $$\sum_{j=1}^{N} \int \frac{(f_{Nj}(x) - \tilde{f}_N(x))^2}{\tilde{f}_N(x)} dx \leq N\varepsilon_N$$

(A.5) $$\limsup_{N \to \infty} \int |\tilde{f}_N(x) - \tilde{f}_N(-x)| dx = 0$$

Let $\phi_N(t)$ denote the characteristic function of $N^{-1/2}(T_N^+ - ET_N^+)$, i.e.,

(2.1) $$\phi_N(t) = E \exp\{itN^{-1/2}(T_N^+ - ET_N^+)\}.$$

Then our main theorem reads:

THEOREM 2.1. *Under the assumptions (A.1) to (A.5), there exist positive numbers B, β and γ depending only on c, C, a, A, δ and the sequence ε_N such that for* $\log N \leq |t| \leq \gamma N^{3/2}$

(2.2) $$|\phi_N(t)| \leq BN^{-\beta \log N}.$$

In order to get Berry-Esséen bound of order $O(N^{-1/2})$, we need the conclusion of the Theorem 2.1 for a domain $\log N \leq |t| < \gamma N^{1/2}$. But to get (2.2) on the domain $\log N \leq |t| \leq \gamma N^{1/2}$, the assumption (A.3) is superfluous. Thus we get the following corollary.

COROLLARY 2.2. *Suppose that the assumptions* $(A.1)$, $(A.2)$, $(A.4)$ *and* $(A.5)$ *are satisfied. Then there exist positive numbers* B, β *and* γ *depending only on* c, C, a, A *and the sequence* ε_N *such that* (2.2) *holds for* $\log N \leq |t| \leq \gamma N^{1/2}$.

To conclude this section we provide a discussion of Theorem 2.1. We shall be brief in the following sequence of remarks because our Theorem 2.1 is par allel to the result of Van Zwet (1982), which deals with the unsigned linear rank statistics, and most of his discussion about his result can be applied to ours.

REMARK 2.1. The standardization of T_N^+ in the theorem is different from the one in section 1. If the normalizing constant σ_N^2 is of exact order N, then the difference is of no importance and (1.9) follows immediately from (2.2).

REMARK 2.2. Assumption (A.1) may be replaced by

$$\Sigma|c_{Nj}|^r \geq c'N, \quad \Sigma|c_{Nj}|^s \leq C'N \tag{2.3}$$

for positive c' and C' and for some $s \geq 2$, $0 < r < s$. For $s=2$, this is equivalent to (A.1) and for $s>2$ it is stronger. The same remark applies to the assumption (A.2).

REMARK 2.3. Assumption (A.3) is well known from previous work of Albers, Bickel and Van Zwet (1976) and Bickel and Van Zwet (1978). Its role is to ensure that the scores $a_{N1}, a_{N2}, \cdots, a_{NN}$ do not cluster too much around too few points.

REMARK 2.4. Assumption (A.4) is satisfied if a sequence $(f_{N1}, f_{N2}, \cdots, f_{NN})$ is contiguous to the hypothesis $H_0: f_{N1}=f_{N2}=\cdots=f_{NN}=\tilde{f}_N$ for some choice of $\tilde{f}_N$, $N=1, 2, \cdots$.

REMARK 2.5. Under the case $X_{N1}, X_{N2}, \cdots, X_{NN}$ are independent and identically distributed with a common density $\tilde{f}_N$, the assumption (A.5) may be replaced by a weaker condition that

$$P\left(\varepsilon < \frac{\tilde{f}_N(X_{N1})}{\tilde{f}_N(X_{N1})+\tilde{f}_N(-X_{N1})} < 1-\varepsilon\right) > 1-\delta_5, \tag{2.4}$$

for some $\varepsilon \in (0, 1/4)$ and δ_5 is defined in (3.1). The condition (2.4) is used in Albers, Bickel and Van Zwet (1976) to prove their Theorem 2.2.

3. Proof of main theorem

The proof of this theorem is a technically complicated affair and we shall split it up in a series of lemmas. To avoid the laborious formulation in each of these lemmas we adopt the following conventions. Whenever we assume that one or more of the assumptions (A. 1)−(A. 5) are satisfied, it will be tacitly understood that the numbers c, C, a, A and δ occurring in these assumptions are indeed positive and that $\varepsilon_N \downarrow 0$. In each lemma where they appear, B_ν and β_ν are positive numbers which may depend on c, C, a, A, δ, $\{\varepsilon_N\}$ and other quantities specified in that lemma. Furthermore constants δ_1, δ_2, δ_3, δ_4 and δ_5 are defined by

$$\delta_1=\frac{9c^2}{16C},\ \delta_2=\frac{\delta}{16}\min\left\{\frac{\delta c}{2c+8},\ 4\delta_1\right\},\tag{3.1}$$
$$\delta_3=\frac{9a^2}{16A},\ \delta_4=\frac{\delta_3}{8}\min\{\delta_1,\ \delta_3\},\ \delta_5=\frac{1}{8}\min\{\delta_2,\ 2\delta_4\}.$$

In the sequel, we will use notations $[x]$ as the integer part of x and $[x]^*$ as the smallest integer greater than or equal to x. Also we should mention that, for convenience of notation, we shall omit indices N in X_{Ni}, R_{Ni}^+, c_{Ni}, a_{Ni}, f_{Ni} and $\tilde{f}_{Ni}$, etc.

Noting that $f_1, f_2, \cdots, f_N$ denote the densities of $X_1, X_2, \cdots, X_N$ and that $R^+=(R_1^+, R_2^+, \cdots, R_N^+)$ and $Z=(Z_1, Z_2, \cdots, Z_N)$ denote the vectors of the ranks and the order statistics of $|X_1|, |X_2|, \cdots, |X_N|$, we define r.v.'s for $1\leq j\leq N$

$$P_j=\frac{f_j(|X_j|)}{f_j(|X_j|)+f_j(-|X_j|)},\ D_j=c_j a_{R_j^+}.$$

Now we prove the following lemmas which are used in the proof of our main theorem.

LEMMA 3.1. *For any integer N and real t,*

$$|\phi_N(t)|\leq E\{\prod_{j=1}^N [1-2P_j(1-P_j)\{1-\cos(N^{-1/2}2tD_j)\}]^{1/2}\}\tag{3.2}$$

PROOF. Since, conditionally given Z and R^+, $\operatorname{sgn} X_1$, $\operatorname{sgn} X_2, \cdots, \operatorname{sgn} X_N$ are independent with probabilities

$$P_j=P(\operatorname{sgn}X_j=1|Z,\ R^+)=1-P(\operatorname{sgn}X_j=-1|Z,\ R^+),$$

we have

$$|\phi_N(t)| \leq E|E[\exp\{itN^{-1/2}(T_N^+ - E(T_N^+|Z,\ R^+))\}|Z,\ R^+]|$$

$$= E\{\prod_{j=1}^{N} |E[\exp\{itN^{-1/2}D_j(\text{sgn } X_j - (2P_j - 1))\}|Z,\ R^+]|\}$$

$$= E\{\prod_{j=1}^{N} |P_j \exp(itN^{-1/2}D_j(2-2P_j)) + (1-P_j)\exp(itN^{-1/2}D_j(-2P_j))|\}$$

$$= E\{\prod_{j=1}^{N} [1 - 2P_j(1-P_j)\{1-\cos(N^{-1/2}2tD_j)\}]^{1/2}\}.$$

LEMMA 3.2. *If the assumptions* $(A.4)$ *and* $(A.5)$ *hold, then for every* $\varepsilon \in (0, 1/4)$ *and* $\eta \in (0, 1)$,

$$P(\varepsilon \leq P_j \leq 1-\varepsilon \textit{ for at least } [\eta N]^* \textit{ indices } j) \geq 1 - B_1 e^{-\beta_1 N}.$$

PROOF. It follows by $(A.4)$, $(A.5)$, Chebyshev's inequality and Jensen's inequality that

$$\frac{1}{N}\sum_{j=1}^{N} P(|2P_j - 1| > 1 - 2\varepsilon)$$

$$\leq \frac{1}{2(1-2\varepsilon)}\frac{1}{N}\sum_{j=1}^{N}\int_{-\infty}^{\infty}(2|f_j(x) - \tilde{f}(x)| + |\tilde{f}(x) - \tilde{f}(-x)|)dx$$

$$\leq \frac{1}{1-2\varepsilon}\left\{\frac{1}{N}\sum_{j=1}^{N}\int_{-\infty}^{\infty}\frac{(f_j(x) - \tilde{f}(x))^2}{\tilde{f}(x)}dx\right\}^{1/2} + \frac{1}{2(1-2\varepsilon)}\int_{-\infty}^{\infty}|\tilde{f}(x) - \tilde{f}(-x)|dx$$

$$\leq 2\,\varepsilon_N^{\frac{1}{2}} + \int_{-\infty}^{\infty}|\tilde{f}(x) - \tilde{f}(-x)|dx.$$

Define, for $j = 1, 2, \cdots, N$, $Y_j = 1$ if $|2P_j - 1| > 1 - 2\varepsilon$ and $Y_j = 0$ otherwise. Then Bernstein's inequality (see e.g. Hoeffding (1963)) yields

$$P(\varepsilon \leq P_j \leq 1-\varepsilon \text{ at least } [\eta N]^* \text{ indices } j)$$

$$= 1 - P(Y_j = 1 \text{ at least } N - [\eta N] \text{ indices } j) \geq 1 - B_1 e^{-\beta_1 N}.$$

The proof is complete.

A straightforward computation shows that the assumption (A.1) implies that

$$(3.5) \qquad |c_j| \geq \frac{c}{4} \text{ for at least } [\delta_1 N]^* \text{ indices } j$$

and the assumption $(A.2)$ implies that

$$(3.6) \qquad |a_j| \geq \frac{a}{4} \text{ for at least } [\delta_3 N]^* \text{ indices } j.$$

Because the assumptions as well as the conclusion of the theorem are invariant under simultaneous permutation of the c_j, X_j and f_j, $j = 1, 2, \cdots, N$, in the sequel, we may, without loss of generality, assume that

(3.7) $$|c_j| \geq \frac{c}{4} \text{ for } j=1,\ 2,\ \cdots,\ [\delta_1 N]^*.$$

Under the model we are discussing, $X_1, X_2, \cdots, X_N$ are independent with densities $f_1, f_2, \cdots, f_N$ and probabilities and expectations under this model are indicated by P and E. We now introduce an auxiliary model under which X_1, $X_2, \cdots, X_N$ are independent and identically distributed with a common density f and we shall write P_0 and E_0 for probabilities and expectations under this model.

By comparing (3.2) and (2.19) of Van Zwet (1982), our plan of attack is easily recognized. We shall show that, with large probability, there is a large set J of indices, $J \subset \{1, 2, \cdots, N\}$, such that the sequences P_j and D_j for $j \in J$ satisfy the conditions (2.17) and (2.18) of Van Zwet (1982). To do that we already proved a necessary lemma (Lemma 3.2) and now need

LEMMA 3.3. *Suppose that assumptions (A.3), (A.4) and the condition (3.7) are satisfied and let δ_2 be defined by (3.1). Then for some $\zeta \geq N^{-3/2} \log N$*

$$P(\theta(D_1, D_2, \cdots, D_N : \zeta) \geq \delta_2 N\zeta) < 1 - B_2 e^{-\beta_2 N}.$$

PROOF. By Lemma 2.5 of van Zwet (1982), it is sufficient to show that for some positive numbers B and β

(3.8) $$P_0(\theta(D_1, D_2, \cdots, D_N : \zeta) < \delta_2 N\zeta) \leq Be^{-\beta N}.$$

Under the model P_0, $(R_1^+, R_2^+, \cdots, R_N^+)$ equals each permutation of $(1, 2, \cdots, N)$ with probability $1/N!$. Take ζ as in the assumption (A.3) and define

(3.9) $$r = \left[\min\left(\frac{\delta c N}{4(2c+8)},\ \delta_1 N\right)\right]^*.$$

We build up $\theta(D_1, D_2, \cdots, D_r : \zeta)$ in r steps by successively choosing R_1^+, R_2^+, $\cdots, R_r^+$ at random without replacement from $\{1, 2, \cdots, N\}$ and running through the sequence $\theta(D_1 : \zeta)$, $\theta(D_1, D_2 : \zeta)$, $\cdots$, $\theta(D_1, D_2, \cdots, D_r : \zeta)$. If we choose R_k^+ in such a way that D_k is not contained in the 2ζ-neighborhood of $\{D_1, D_2, \cdots, D_{k-1}\}$, then $\theta(D_1, D_2, \cdots, D_k : \zeta) = \theta(D_1, D_2, \cdots, D_{k-1} : \zeta) + 2\zeta$. This is the case unless $|D_k - D_j| < 2\zeta$ for some $j=1, 2, \cdots, k-1$, i.e.,

(3.10) $$c_k a_{R_k^+} \notin \bigcup_{j=1}^{k-1} (D_j - 2\zeta,\ D_j + 2\zeta).$$

Since $k \leq r \leq [\delta_1 N]^*$, (3.7) ensures that $|c_k| \geq c/4 > 0$ and hence (3.10) restricts

$a_{R_k^+}$ to a set A_k which is the union of $(k-1)$ intervals of length$\leq 16\zeta/c$. The set of a_j in A_k has ζ-neighborhood of Lebesgue measure at most $(k-1)(16\zeta/c+2\zeta)$, so the assumption (A.3) implies that the number of j for which $a_j \notin A_k$ equals at least

$$\left[\frac{1}{2\zeta}\{\delta N\zeta-2\zeta(k-1)\left(\frac{8}{c}+1\right)\}\right]^{*}=\left[\frac{\delta N}{2}-(k-1)\left(1+\frac{8}{c}\right)\right]^{*}.$$

Subtracting the $(k-1)$ indices R_1^+, R_2^+, $\cdots$, R_{k-1}^+ chosen before R_k^+ and for which the corresponding a_j may well be outside A_k, we find that the conditional probability that $a_{R_k^+} \notin A_k$, given R_1^+, $\cdots$, R_{k-1}^+, equals at least

$$\frac{\delta N/2-(k-1)(1+8/c)-(k-1)}{N-(k-1)}\geq \delta/2-\frac{r-1}{N}\left(2+\frac{8}{c}\right)\geq\frac{\delta}{4}$$

in view of (3.9). As $a_{R_k^+}\notin A_k$ implies that 2ζ is added to θ at the k-th step, we see that $\theta(D_1, D_2, \cdots, D_r : \zeta)/2\zeta$ is stochastically larger than a binomial random variable with parameters r and $\delta/4$. Since $\theta(D_1, D_2, \cdots, D_N : \zeta)\geq\theta(D_1, D_2, \cdots, D_r : \zeta)$ and $r\delta/4\geq\delta_2 N$, Bernstein's inequality (see e.g. Hoeffding (1963)) ensures (3.8) for some positive B and β and the proof is complete.

Finally we need

LEMMA 3.4. *If assumptions* (A.2), (A.4) *and the condition* (3.7) *are satisfied and* δ_4 *is given by* (3.1), *then*

$$P(|D_j|\geq\frac{ac}{16}\textit{ for at least } [\delta_4 N]^{*} \textit{ indices } j)\geq 1-B_3 e^{-\beta_3 N}.$$

PROOF. Take $r=[N\ \min(\delta_1,\ \delta_3/2)]^{*}$ and let $j=1, 2, \cdots, r$. Under P_0 and given R_1^+, R_2^+, $\cdots$, R_{j-1}^+, the conditional probability that $|a_{R_j^+}|\geq a/4$ is at least

$$\frac{\delta_3 N-(j-1)}{N-(j-1)}\geq\delta_3-\frac{r-1}{N}\geq\frac{\delta_3}{2}$$

because $j-1\leq r-1\leq\delta_3 N/2$ and because of (3.6).

It follows that the number of indices $j\leq r$ for which $|a_{R_j^+}|\geq\frac{a}{4}$ is stochastically larger, under P_0, than a binomial random variable with parameter r and $\delta_3/2$. Since $r\delta_3/2\geq 2\delta_4 N$, Bernstein's inequality yields positive B and β such that

(3.11) $\quad P_0(|a_{R_j^+}|\geq a/4$ for at least $[\delta_4 N]^{*}$ indices $j\leq r)\geq 1-Be^{-\beta N}$.

But if $j\leq r$, then $j\leq[\delta_1 N]^{*}$. Thus $|c_j|\geq c/4$ by (3.7), which, together with

(3.11), implies

$$P_0(|D_j|\geq ac/16 \text{ for at least } [\delta_4 N]^* \text{ indices } j)\geq 1-Be^{-\beta N}.$$

Again the proof is complete by Lemma 2.5 of Van Zwet (1982).

PROOF OF THEOREM 2.1. Recall that, in addition to the assumptions (A.1)–(A.5), we may assume that (3.7) is satisfied for positive c, C, a, A, δ, $\tilde{f}$ and $\varepsilon_N \downarrow 0$. Also note that δ_1, δ_2, δ_3, δ_4 and δ_5 are defined by (3.1).

Choose $\varepsilon\in\left(0, \frac{1}{4}\right)$ and define

$$d=2^{-10}\varepsilon(1-\varepsilon)a^2c^2\delta_4,\quad D=((A+C)/\delta_5)^4. \tag{3.12}$$

Let $J\subset\{1, 2, \cdots, N\}$ be the random set of indices j for which $|D_j|\leq D^{\frac{1}{4}}$ and let M be the cardinality of J thus

$$J=\{j: |D_j|\leq D^{\frac{1}{4}}\},\quad M=\|J\|.$$

Because of (A.1) and (A.2), the sets $\{j: |c_j|>D^{1/8}\}$ and $\{j: |a_j|>D^{1/8}\}$ have cardinalities at most $CND^{-\frac{1}{4}}$ and $AND^{-\frac{1}{4}}$ respectively, and thus we have

$$N\geq M\geq N-(A+C)ND^{-\frac{1}{4}}=N-\delta_5 N, \tag{3.13}$$

with probability one. Since $\delta_5\leq 3/4$, (3.13) implies that

$$N/4<M\leq N. \tag{3.14}$$

Take $\zeta=N^{-\frac{3}{2}}\log N$ and define the event F by

$$\begin{aligned} F=&\{\varepsilon\leq P_j\leq 1-\varepsilon \text{ for at least } [(1-\delta_5)N]^* \text{ indices } j\}\\ &\cap\{\theta(D_1, D_2, \cdots, D_N:\zeta)\geq\delta_2 N\zeta\}\\ &\cap\{|D_j|\geq ac/16 \text{ for at least } [\delta_4 N]^* \text{ indices } j\}. \end{aligned}$$

Then Lemma 3.2, Lemma 3.3 and Lemma 3.4 yield

$$P(F)\geq 1-B_4 e^{-\beta_4 N}, \tag{3.15}$$

where $B_4=B_1+B_2+B_3$ and $\beta_4=\min(\beta_1, \beta_2, \beta_3)$ are positive numbers depending only on c, C, a, A, δ and the sequence ε_N.

Now we need a notation. Consider real numbers d_1, d_2, $\cdots$, d_m and p_1, p_2, $\cdots$, p_m with $0\leq p_j\leq 1$ for $j=1, 2, \cdots, m$. For $\zeta>0$ and $0<\varepsilon<\frac{1}{4}$, let $\theta(d_1, \cdots, d_m; p_1, \cdots, p_m; \zeta, \varepsilon)$ denote the Lebesgue measure λ of the ζ-neighborhood of the set of those d_j for which the corresponding p_j satisfy $\varepsilon\leq p_j\leq 1-\varepsilon$, thus

$$\theta(d_1, \cdots, d_m ; p_1, \cdots, p_m ; \zeta, \varepsilon)=\lambda\{x : |x-d_j| < \varepsilon, \ \varepsilon \le p_j \le 1-\varepsilon \text{ for some } j\}.$$

On the the set F in our sample space, the number of indices $j \in \{1, 2, \cdots, N\}$ for which $\varepsilon \le P_j \le 1-\varepsilon$ as well as $|D_j| \ge ac/16$ equals at least $(\delta_4-\delta_5)N$. Because of (3.13), $(\delta_4-2\delta_5)N$ of these indices must also belong to J. Combining this with (3.12) and (3.14), we find that

$$\text{(3.16)} \qquad \begin{aligned} \sum_{j\in J} P_j(1-P_j)D_j^2 &\ge \varepsilon(1-\varepsilon)(ac/16)^2(\delta_4-2\delta_5)N \\ &\ge 2^{-10}\varepsilon(1-\varepsilon)a^2c^2\delta_4 M = dM, \end{aligned}$$

for every sample point in F. Similarly we see that on F the number of indices j for which $j \notin J$ or $P_j \notin [\varepsilon, 1-\varepsilon]$, equals at most $2\delta_5 N$ and hence

$$\theta(D_j, \ j\in J : P_j, \ j \in J ; \zeta, \varepsilon) \ge (\delta_2-4\delta_5)N\zeta \ge \delta_2 M\zeta/2.$$

Take $\zeta' = 3M^{-3/2}\log M$. If $M \ge 2$, then (3.14) ensures that $1/24 \le \zeta/\zeta' < 1$ and as θ is obviously nondecreasing in ζ

$$\text{(3.17)} \qquad \theta(D_j, \ j\in J ; P_j, \ j\in J ; \zeta', \varepsilon) \ge (\delta_2 M/2)(\zeta'/24) = \delta_2 M\zeta'/48.$$

Since this is trivially true for $M=1$ also, (3.17) holds for every sample point in F. Finally the definition of J implies that

$$\text{(3.18)} \qquad \sum_{j\in J} D_j^4 \le DM.$$

We have shown that on the set F the sequences D_j and P_j, $j \in J$, satisfy the conditions (2.17) and (2.18) of Lemma 2.3 of Van Zwet (1982) for values d, D, ε and $\delta' = \delta_2/48$ which depend only on c, C, a, A, δ and the sequence ε_N. Application of this lemma with $b_1 = 1$ yields the existence of positive numbers b_2, B_5 and β_5 depending only on c, C, a, A, δ and ε_N and such that for every sample point in F,

$$\begin{aligned} &\prod_{j=1}^{N} [1-2P_j(1-P_j)\{1-\cos(N^{-\frac{1}{2}}2tD_j)\}]^{\frac{1}{2}} \\ &\le \prod_{j\in J} [1-2P_j(1-P_j)\{1-\cos(M^{-\frac{1}{2}}\left(\frac{M}{N}\right)^{\frac{1}{2}}2tD_j)\}]^{\frac{1}{2}} \le B_5 M^{-\beta_5 \log M} \end{aligned}$$

for $\log M \le \left(\frac{M}{N}\right)^{\frac{1}{2}} 2|t| \le b_2 M^{\frac{3}{2}}$. An easy calculation based on (3.14) shows that this implies that there exist positive B_6 and β_6 depending only on B_5 and β_5 such that on the set F

$$\text{(3.19)} \qquad \prod_{j=1}^{N} [1-2P_j(1-P_j)\ \{1-\cos(N^{-\frac{1}{2}}2tD_j)\}]^{\frac{1}{2}} \le B_6 N^{-\beta_6 \log N}$$

for $\log N \leq |t| \leq \gamma N^{3/2}$, where $\gamma = b_2/16$. Combining (3.15), (3.19) and Lemma 3.1, we find that for every t, $\log N \leq |t| \leq \gamma N^{3/2}$,

$$|\phi_N(t)| \leq B_6 N^{-\beta_6 \log N} + B_4 e^{-\beta_4 N} \leq BN^{-\beta \log N}$$

with $B = B_4 + B_6$ and $\beta = \min(\beta_4, \beta_6)$. The proof is complete.

ACKNOWLEDGEMENT. The author wishes to express deep and sincere gratitude to Professor Madan L. Puri for suggesting this problem (a part of the author's thesis), and for his kind help and encouragement during the preparation of the author's dissertation.

Indiana University
Department of Mathematics
Bloomington, IN 47405
U.S.A.

REFERENCES

[1] Albers, W., Bickel, P.J. and Van Zwet, W.R. *Asymptotic expansions for the power of distribution free tests in the one sample problem*, Ann. Stat. 4, (1976) 108—156.

[2] Bickel, P.J. *Edgeworth expansions in nonparametric statistics*, Ann. Stat. 2, (1974) 1—20.

[3] Bickel, P.J. and Van Zwet, W.R. *Asymptotic expansions for the power of distribution free tests in the two-sample problem*, Ann. Stat. 6, (1978) 937—1004.

[4] Feller, W. *An introduction to probability theory and its applications*, Vol.2, 2nd edition, Wiley, N.Y. (1971).

[5] Hájek, J. *Asymptotically most powerful rank order tests*, Ann. Math. Stat. 33, (1962) 1129—1147.

[6] Hájek, J. and Šidak, Z. *Theory of rank tests*, Academic Press, N.Y. (1967).

[7] Hoeffding, W. *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. 58, (1963) 13—30.

[8] Hušková, M. *Asymptotic distribution of simple linear rank statistics for testing symmetry*, Z. Wahrscheinlichkeits theorie verw. Geb. 14, (1970) 308—322.

[9] Puri, M.L. and Ralescu, S.S. *The asymptotic distribution theory of one sample signed rank statistic*, Statistical decision theory and related topics Ⅲ, Vol.2., Academic Press, N.Y. (1982), 213—232.

[10] Puri, M.L. and Seoh, M. *Berry-Esséen theorems for signed linear rank statistics*

with regression constants, To appear in Colloquium on Limit Theorems in Probability and Statistics. Veszprem (Hungary) (1983a).

[11] Puri, M.L. and Seoh, M. *Edgeworth expansions for signed linear rank statistics with regression constants*, Submitted (1983b).

[12] Seoh, M. and Puri, M.L. *Berry-Esseen theorem for signed linear rank statistics under near location alternatives*, Submitted (1983c).

[13] Seoh, M. and Puri, M.L. *Asymptotic expansion of signed linear rank statistics under near location alternatives*, Submitted (1983d).

[14] Van Zwet, W.R. *On the Edgeworth expansion for the simple linear rank statistic*, Nonparametric Statistical Inference, Vol. II, (B.V. Gnedenko, M.L. Puri and I. Vincze, Eds), North Holland Publishing Company, 889–909(1982).