stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SW 71/81 APRIL
(DEPARTMENT OF MATHEMATICAL STATISTICS)

R.J.M.M. DOES

BERRY-ESSEEN THEOREMS FOR SIMPLE LINEAR RANK STATISTICS UNDER THE NULL-HYPOTHESIS

Preprint

kruislaan 413 1098 SJ amsterdam

Berry-Esseen theorems for simple linear rank statistics under the null-hypothesis *)

by

R.J.M.M. Does

ABSTRACT

Berry-Esseen bounds of order $\mathcal{O}(N^{-\frac{1}{2}})$ are established for simple linear rank statistics under the null-hypothesis. The theorems are proved for a wide class of scores generating functions which includes the normal quantile function. This improves earlier results under the null-hypothesis in HUŠKOVÁ (1977,1979).

KEY WORDS & PHRASES: *simple linear rank statistics, Berry-Esseen theorems, distributionfree tests*

---

*) This report will be submitted for publication elsewhere.

1. INTRODUCTION

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed random variables with a common continuous distribution function F. If $Z_1 < Z_2 < \ldots < Z_N$ denotes the sequence $X_1, X_2, \ldots, X_N$ arranged in increasing order, then the rank R_{jN} of X_j is defined by $X_j = Z_{R_{jN}}$ and the antirank D_{jN} is defined by $X_{D_{jN}} = Z_j$, $j = 1,2,\ldots,N$. For specified vectors of real numbers $c_N = (c_{1N}, c_{2N}, \ldots, c_{NN})$ (regression constants) and $a_N = (a_{1N}, a_{2N}, \ldots, a_{NN})$ (scores)

$$(1.1) \qquad T_N = \sum_{j=1}^{N} c_{jN}\, a_{R_{jN}}$$

is called a simple linear rank statistic. Well-known special cases are the two-sample linear rank statistics, which have $c_{jN} = 0$, for $j = 1,2,\ldots,n$, $c_{jN} = 1$, for $j = n+1,\ldots,N$ and Spearman's rank correlation coefficient ρ which, under the null-hypothesis, is distributed as T_N with $c_{jN} = j$ and $a_{R_{jN}} = R_{jN}$, $j = 1,2,\ldots,N$.

Throughout this paper we make the following assumption about the regression constants.

Assumption (A): The regression constants $c_{1N}, c_{2N}, \ldots, c_{NN}$ satisfy

$$\sum_{j=1}^{N} c_{jN} = 0 \quad , \sum_{j=1}^{N} c_{jN}^2 = 1 \quad \text{and} \quad \sum_{j=1}^{N} |c_{jN}|^3 = \mathcal{O}(N^{-\frac{1}{2}}).$$

The scores $a_{1N}, a_{2N}, \ldots, a_{NN}$ are generated by a function $J(t)$, $0 < t < 1$, in either one of the following two ways

$$(1.2) \qquad \text{(approximate scores)} \quad a_{jN} = J\left(\frac{j}{N+1}\right), \qquad j = 1,2,\ldots,N,$$

$$(1.3) \qquad \text{(exact scores)} \qquad a_{jN} = EJ(U_{j:N}), \qquad j = 1,2,\ldots,N.$$

Here $U_{j:N}$ denotes the j-th order statistic in a random sample of size N from the uniform distribution on (0,1). For almost all well-known linear rank tests the scores are of one of these two types.

Note that assumption (A) implies that $ET_N = 0$. Taking $\bar{a}_N = N^{-1} \Sigma_{j=1}^N a_{jN}$, we find that the variance σ_N^2 of T_N (cf.(1.1)) is given by

$$(1.4) \qquad \sigma_N^2 = \sigma^2(T_N) = \frac{1}{N-1} \sum_{j=1}^{N} (a_{jN} - \bar{a}_N)^2$$

(see e.g. Theorem II 3.1.c of HÁJEK & ŠIDÁK (1967)).

Define

$$(1.5) \qquad T_N^* = \sigma_N^{-1} T_N$$

and

$$(1.6) \qquad F_N^*(x) = P(T_N^* \leq x) \quad \text{for } -\infty < x < \infty .$$

The asymptotic normality of T_N^* has been established under very general conditions (cf. HÁJEK & ŠIDÁK (1967). Chapter V). Recently a Berry-Esseen bound of order $\mathcal{O}(\Sigma_{j=1}^N |c_{jN}|^3)$ for the distribution function F_N^* of T_N^* (cf.(1.5) and (1.6)) has been obtained for bounded scores generating functions (cf. HUŠKOVÁ (1977,1979)), i.e.

$$(1.7) \qquad \sup_{x \in \mathbb{R}} |F_N^*(x) - \Phi(x)| = \mathcal{O}(\sum_{j=1}^{N} |c_{jN}|^3),$$

where Φ is the standard normal distribution function. The purpose of this paper is to extend the assertion (1.7) to a large class of scores generating functions including the normal quantile function. The related problem of establishing Edgeworth expansions for simple linear rank statistics will be discussed in the author's forthcoming Ph.D. thesis.

In Section 2 we state our results in the form of two theorems. Section 3 contains a number of preliminaries. The proofs of the theorems are contained in Section 4. In the sequel we suppress the index N whenever it is possible.

2. BERRY-ESSEEN THEOREMS

We start this section by introducing a condition on the derivative of a function which ensures that this derivative does not oscillate too wildly near 0 and 1 (see also Appendix 2 of ALBERS, BICKEL & VAN ZWET (1976)).

Condition R_r: For real $r>0$, a function h on (0,1) is said to satisfy condition R_r if h is twice continuously differentiable on (0,1) and

$$\limsup_{t\to 0,1} t(1-t)\left|\frac{h''(t)}{h'(t)}\right| < 1 + \frac{1}{r}.$$

To formulate our theorems we need some smoothness assumptions for the scores generating function J.

Assumption (B): The scores generating function J satisfies

$$\int_0^1 J(t)dt = 0, \quad \int_0^1 J^2(t)dt = 1 \quad \text{and} \quad \int_0^1 |J(t)|^3 dt < \infty .$$

Assumption (C): The scores generating function J is continuously differentiable on (0,1). There exist positive numbers $\Gamma > 0$ and $\alpha < 5/4$ such that its first derivative J' satisfies

$$|J'(t)| \le \Gamma \{t(1-t)\}^{-\alpha} \qquad \text{for } t \in (0,1).$$

For exact scores Theorem 2.1 provides a Berry-Esseen theorem for the distribution function F_N^* (cf.(1.6)) of T_N^* (cf.(1.5)). Theorem 2.2 deals with the case of approximate scores.

<u>THEOREM 2.1</u>. *Take* $a_j = EJ(U_{j:N})$ for $j = 1,2,\ldots,N$. *Assume that assumptions* (A) *and* (B) *are satisfied and that*

$$\sum_{j=1}^{N} \sigma^2(J(U_{j:N})) = O(N^{\frac{1}{2}}(\log N)^{-2}). \tag{2.1}$$

Then

$$\sup_{x\in\mathbb{R}} |F_N^*(x) - \Phi(x)| = O(N^{-\frac{1}{2}}). \tag{2.2}$$

If J *also satisfies condition* R_1, *then uniformly in* k *and* ℓ,

$$(3.3)\qquad \sum_{j=k}^{\ell} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}^2 = O\left(N^{-1}\{\tfrac{k}{N+1}\}^{-3/2+2\delta} + N^{-1}\{\tfrac{N+1-\ell}{N+1}\}^{-3/2+2\delta}\right).$$

Finally, if in addition J *satisfies assumption* (B), *then*

$$(3.4)\qquad \sum_{j=1}^{N} \{J(\tfrac{j}{N+1}) - \tfrac{1}{N}\sum_{i=1}^{N} J(\tfrac{i}{N+1})\}^2 = N + O(N^{\frac{1}{2}-2\delta}).$$

PROOF. Without loss of generality we suppose that assumption (C) holds for $\alpha \in (1,5/4)$ and we take $\delta = 5/4 - \alpha$. Let h be a function on (0,1) with $h'(t) \equiv \Gamma\{t(1-t)\}^{-5/4+\delta}$ and write $\lambda_j = j/(N+1)$. Since h satisfies condition R_2, Lemma A.2.3 of ALBERS, BICKEL & VAN ZWET (1976) yields

$$E\{h(U_{j:N}) - h(\lambda_j)\}^2 = O\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-3/2+2\delta}}{N}\right)$$

uniformly in j. Because $|J'(t)| \le h'(t)$ we have $|J(s)-J(t)| \le |h(s)-h(t)|$ for every $s,t \in (0,1)$ and hence

$$E\{J(U_{j:N}) - J(\lambda_j)\}^2 = O\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-3/2+2\delta}}{N}\right)$$

uniformly in j. Now (3.1) follows by summation and (3.2) is implied by (3.1) as $\sigma^2(J(U_{j:N})) \le E\{J(U_{j:N}) - J(\lambda_j)\}^2$.

If J also satisfies R_1 then, in view of (A.2.11) in ALBERS, BICKEL & VAN ZWET(1976), we have

$$(3.5)\qquad |EJ(U_{j:N}) - J(\lambda_j)| = O\left(\frac{\lambda_j(1-\lambda_j) + |J'(\lambda_j|)}{N}\right) = O\left(\frac{\{\lambda_j(1-\lambda_j)\}^{-5/4+\delta}}{N}\right)$$

uniformly in j; (3.3) follows by summation.

If J also satisfies assumption (B), then

$$\left|\frac{1}{N}\sum_{j=1}^{N} J(\tfrac{j}{N+1})\right| = \left|\frac{1}{N}\sum_{j=1}^{N}\{J(\tfrac{j}{N+1}) - EJ(U_{j:N})\}\right| = O(N^{-3/4-\delta})$$

because of (3.5). Furthermore, in view of (3.1) and (3.5),

$$\left| \sum_{j=1}^{N} J^2\left(\frac{j}{N+1}\right) - N \right| = \left| \sum_{j=1}^{N} \{J^2\left(\frac{j}{N+1}\right) - EJ^2(U_{j:N})\} \right| \leq$$

$$\leq \sum_{j=1}^{N} E\{J(U_{j:N}) - J\left(\frac{j}{N+1}\right)\}^2 + 2\sum_{j=1}^{N} \left|J\left(\frac{j}{N+1}\right)\right| \left| EJ(U_{j:N}) - J\left(\frac{j}{N+1}\right) \right| =$$

$$= \mathcal{O}(N^{\frac{1}{2}-2\delta})$$

which proves (3.4) and the lemma. □

We now consider the behavior of the characteristic function of T_N^* for large values of the argument. Let

$$\psi_N(t) = Ee^{itT_N^*} . \tag{3.6}$$

LEMMA 3.2. *Suppose that the conditions of either Theorem* 2.1 *or Theorem* 2.2 *are satisfied. Then there exist positive numbers* B, β *and* γ *such that*

$$|\psi_N(t)| \leq BN^{-\beta \log N} \tag{3.7}$$

for $\log N \leq |t| \leq \gamma N^{\frac{1}{2}}$ *and* $N = 2,3,\ldots$.

PROOF. The present lemma is essentially a special case of Theorem 2.1 of VAN ZWET (1980) where (3.7) is proved for $\log N \leq |t| \leq \gamma N^{3/2}$. Since we are concerned with independent and identically distributed random variables $X_1, X_2, \ldots, X_N$ - which we may assume to be uniformly distributed without loss of generality - condition (2.7) of this theorem is clearly satisfied. Moreover, it is easy to see that condition (2.6) is superfluous in our case since we are only concerned with values of $|t| \leq \gamma N^{\frac{1}{2}}$. Finally, it follows from Section 3 in VAN ZWET (1980) that the existence of positive numbers c and C such that

$$\sum_{j=1}^{N} c_j^2 \geq c \quad , \quad \sum_{j=1}^{N} |c_j|^3 \leq CN^{-\frac{1}{2}} , \tag{3.8}$$

$$(3.9) \qquad \sum_{j=1}^{N} (a_j-\bar{a})^2 \geq cN \quad , \quad \sum_{j=1}^{N} |a_j-\bar{a}|^3 \leq CN$$

suffices to prove the present lemma. Since assumption (A) guarantees the validity of (3.8) it remains to check (3.9).

For exact scores $a_j = EJ(U_{j:N})$, assumption (B) and (2.1) imply that $\bar{a} = \int J = 0$ and

$$\sum_{j=1}^{N} a_j^2 = \sum_{j=1}^{N} EJ^2(U_{j:N}) - \sum_{j=1}^{N} \sigma^2(J(U_{j:N})) = N - \mathcal{O}(N^{\frac{1}{2}}),$$

$$\sum_{j=1}^{N} |a_j|^3 \leq \sum_{j=1}^{N} E|J(U_{j:N})|^3 = N \int_0^1 |J(t)|^3 dt$$

and (3.9) follows. For approximate scores $a_j = J(j/(N+1))$, (3.9) is an immediate consequence of assumption (B) and the continuity of J (cf. also (3.4)). □

Let [x] denote the largest integer not exceeding x. Define $m = [N^{1/3}]$ and $I = \{1,2,\ldots,m,N-m+1,\ldots,N-1,N\}$. Take $\delta \in (0,\frac{1}{4})$ as in Lemma 3.1.

LEMMA 3.3. *If assumptions* (A) *and* (C) *are satisfied, then*

$$(3.10) \qquad E\Big(\sum_{j\in I} c_{D_j} \, J(\tfrac{j}{N+1})\Big)^4 = \mathcal{O}(N^{-2/3-8\delta/3}).$$

PROOF. According to assumption (A), $\Sigma c_j = 0$, $\Sigma c_j^2 = 1$, $\Sigma |c_j|^3 = \mathcal{O}(N^{-\frac{1}{2}})$ and $\Sigma c_j^4 \leq \max |c_j| . \Sigma |c_j|^3 = \mathcal{O}(N^{-2/3})$. Hence, straightforward computation shows that for distinct $i,j,h,k \in I$,

$$Ec_{D_i}^4 = \mathcal{O}(N^{-5/3}), \quad Ec_{D_i}^3 c_{D_j} = \mathcal{O}(N^{-8/3}), \quad E\, c_{D_i}^2 c_{D_j}^2 = \mathcal{O}(N^{-2}),$$

$$Ec_{D_i}^2 c_{D_j} c_{D_h} = \mathcal{O}(N^{-3}), \quad Ec_{D_i} c_{D_j} c_{D_h} c_{D_k} = \mathcal{O}(N^{-4}).$$

Assumption (C) ensures that for $\ell = 1,2,3,4$.

$$(3.11) \qquad \frac{1}{N} \sum_{j\in I} |J(\tfrac{j}{N+1})|^{\ell} \sim \int_0^{N^{-2/3}} \{|J(t)|^{\ell} + |J(1-t)|^{\ell}\}dt = \\ = \mathcal{O}(N^{-2/3 + \ell/6 - 2\ell\delta/3})$$

Direct computation of the left-hand side of (3.10) now produces the result of the lemma. □

4. PROOFS OF THE THEOREMS

To establish a Berry-Esseen theorem one usually invokes Esseen's smoothing lemma (see e.g. FELLER (1971) page 538), which implies that for all $\gamma > 0$

$$\sup_{x\in\mathbb{R}} |F_N^*(x) - \Phi(x)| \le \frac{1}{\pi}\int_{-\gamma N^{\frac12}}^{\gamma N^{\frac12}} |t|^{-1}|\psi_N(t) - e^{-\frac12 t^2}|dt + O(N^{-\frac12}) \tag{4.1}$$

where $\psi_N(t)$ denotes the characteristic function of T_N^* (cf.(3.6)).

It follows from Lemma 3.2 that in order to prove Theorems 2.1 and 2.2 it is sufficient to show that

$$\int_{|t|\le \log N} |t|^{-1}|\psi_N(t) - e^{-\frac12 t^2}|dt = O(N^{-\frac12}). \tag{4.2}$$

We first prove Theorem 2.1. Let $R = (R_1,R_2,\ldots,R_N)$ and $D = (D_1,D_2,\ldots,D_N)$ denote the vectors of ranks and antiranks respectively and define

$$S_N = \sum_{j=1}^{N} c_j\ J(U_j) = \sum_{j=1}^{N} c_{D_j}\ J(U_{j:N}), \tag{4.3}$$

where $U_1,U_2,\ldots,U_N$ are independent and uniformly distributed random variables on (0,1). Since the vector of order statistics is independent of R, we have

$$E(S_N|R) = \sum_{j=1}^{N} c_{D_j}\ EJ(U_{j:N}) = T_N \tag{4.4}$$

and it follows that

$$E(e^{itT_N}(S_N-T_N)) = E(E(e^{itT_N}(S_N-T_N)|R)) =$$

$$= E(e^{itT_N}\ E(S_N-T_N|R)) = 0.$$

Hence

$$(4.5)\qquad Ee^{itS_N} = Ee^{itT_N} + \mathcal{O}(t^2\, E(S_N-T_N)^2)$$

and because of (4.4), assumption (B) and (2.1)

$$(4.6)\qquad E(S_N-T_N)^2 = ES_N^2 - ET_N^2 = 1 - \frac{1}{N-1}\sum_{j=1}^{N} \{EJ(U_{j:N})\}^2 =$$

$$= \frac{1}{N-1}\sum_{j=1}^{N} \sigma^2(J(U_{j:N})) - \frac{1}{N-1} = \mathcal{O}(N^{-\frac{1}{2}}(\log N)^{-2}).$$

As S_N is a sum of independent random variables with $ES_N = 0$, $\sigma^2(S_N) = 1$ and $\Sigma|c_j|^3 E|J(U_j)|^3 = \mathcal{O}(N^{-\frac{1}{2}})$ (cf. assumptions (A) and (B)), we may apply Lemma V 2.1 of PETROV (1972) to obtain that for $|t| \leq \log N$,

$$(4.7)\qquad |Ee^{itS_N} - e^{-\frac{1}{2}t^2}| = \mathcal{O}(N^{-\frac{1}{2}}|t|^3 e^{-t^2/3}).$$

Finally we note that (4.6) implies that

$$(4.8)\qquad \sigma_N^2 = \sigma^2(T_N) = 1 + \mathcal{O}(N^{-\frac{1}{2}}(\log N)^{-2}).$$

Combining (4.5) through (4.8) we arrive at (4.2) and the proof of Theorem 2.1 is complete.

We now turn to the proof of Theorem 2.2. To distinguish simple linear rank statistics with exact scores and with approximate scores we define

$$(4.9)\qquad T_N' = \sum_{j=1}^{N} c_j\, J\left(\frac{R_j}{N+1}\right) = \sum_{j=1}^{N} c_{D_j}\, J\left(\frac{j}{N+1}\right)$$

and

$$(4.10)\qquad T_N = \sum_{j=1}^{N} c_{D_j}\, EJ(U_{j:N}).$$

Because of Lemma 3.1, the conditions of Theorem 2.2 imply those of Theorem 2.1 and we may therefore conclude from the proof of Theorem 2.1 that

$$(4.11)\qquad \int_{|t|\le \log N} |t|^{-1}\left|Ee^{itT_N} - e^{-\frac{1}{2}t^2}\right|dt = O(N^{-\frac{1}{2}})$$

A Taylor expansion yields

$$(4.12)\qquad Ee^{itT_N} = Ee^{itT'_N} + it\, Ee^{itT'_N}(T_N-T'_N) + O(t^2E(T_N-T'_N)^2)\,.$$

In the situation of Theorem 2.2 the scores generating function satisfies both assumption (C) and condition R_1, so that we may apply Lemma 3.1 to find that for some $\delta\in(0,1/4)$,

$$E(T_N-T'_N)^2 = E\Big(\sum_{j=1}^{N} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}\Big)^2 =$$

$$= \frac{1}{N}\sum_{j=1}^{N}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}^2 - \frac{1}{N(N-1)}\sum_{(i,j)\neq}\sum\{EJ(U_{i:N}) - J(\tfrac{i}{N+1})\}.$$

$$(4.13)\qquad \cdot\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\} =$$

$$= \frac{1}{N-1}\sum_{j=1}^{N}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}^2 - \frac{1}{N(N-1)}\Big(\sum_{j=1}^{N}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}\Big)^2 =$$

$$= O(N^{-\frac{1}{2}-2\delta}).$$

Define $m = [N^{1/3}]$ and $I = \{1,2\ldots,m,N-m+1,\ldots,N-1,N\}$ as in Section 3. Repeating the argument of (4.13) for restricted sums we find

$$(4.14)\qquad E\Big|\sum_{j=m+1}^{N-m} c_{D_j}(EJ(U_{j:N}) - J(\tfrac{j}{N+1}))\Big| \le$$

$$\le \Big\{E\Big(\sum_{j=m+1}^{N-m} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}\Big)^2\Big\}^{\frac{1}{2}} = O(N^{-\frac{1}{2}-2\delta/3}).$$

Combining (4.11) through (4.14) we see that, in order to prove (4.2), we have to show that

$$(4.15)\qquad \int_{|t|\le \log N}\Big|E\Big(e^{itT'_N}\sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}\Big)\Big|dt = O(N^{-\frac{1}{2}}).$$

We note that (4.13) and (4.14) imply that

$$(4.16)\qquad E\Big(\sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}\Big)^2 = O(N^{-\frac{1}{2}-2\delta}).$$

Let $\Omega = \{D_j : j \in I\}$ be the set of antiranks D_j with indices in I and let $\omega = \{d_j : j \in I\}$ be a possible realization of Ω. We have

$$E\Big(e^{itT_N'} \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}\Big) =$$

$$(4.17)\qquad = E\{E(\exp\{it \sum_{j=m+1}^{N-m} c_{D_j} J(\tfrac{j}{N+1})\}|\Omega)E(\exp\{it \sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\}.$$

$$\cdot \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}|\Omega)\}.$$

Conditionally on $\Omega = \omega$, $\Sigma_{j=m+1}^{N-m} c_{D_j} J(j/(N+1))$ is distributed as a simple linear rank statistic for sample size N-2m based on a set of regression constants $\{c_1,c_2,\ldots,c_N\} \setminus \{c_{d_j} : j \in I\}$ and having a scores generating function

$$J_N(x) = J\Big(\frac{m + (N-2m+1)x}{N+1}\Big) \quad \text{for } x \in (0,1).$$

We write this simple linear rank statistic as

$$(4.18)\qquad T_{\omega N} = \sum_{j=1}^{M} b_j J_N\Big(\frac{Q_j}{M+1}\Big)$$

where $M = N-2m$, $\{b_1,b_2,\ldots,b_M\} = \{c_1,c_2,\ldots,c_N\} \setminus \{c_{d_j} : j \in I\}$, $Q_1,Q_2,\ldots,Q_M$, are the ranks of $V_1,V_2,\ldots,V_M$, which are independent and uniformly distributed random variables on (0,1). Define

$$(4.19)\qquad S_{\omega N} = \sum_{j=1}^{M} b_j J_N(V_j).$$

LEMMA 4.1. *Under the assumptions of Theorem* 2.2 *we have*

$$(4.20)\qquad E(T_{\omega N} - S_{\omega N})^2 = \Big(1 + \big(\sum_{j\in I} c_{d_j}\big)^2\Big)\, O(N^{-2/3-4\delta/3}).$$

PROOF.

$$E(T_{\omega N} - S_{\omega_N})^2 = \sum_{j=1}^{M} b_j^2 \, E(J_N\left(\frac{Q_1}{M+1}\right) - J_N(V_1))^2 +$$

$$+ \sum_{(i,j)\neq}\sum b_i b_j \, E(J_N\left(\frac{Q_1}{M+1}\right) - J_N(V_1))(J_N\left(\frac{Q_2}{M+1}\right) - J_N(V_2)).$$

Because $\sum_{j=1}^{M} b_j^2 \le 1$ and

$$\left|\sum_{(i,j)\neq}\sum b_i b_j\right| = \left|(\sum_{j=1}^{M} b_j)^2 - \sum_{j=1}^{M} b_j^2\right| = \left|(\sum_{j\in I} c_{d_j})^2 - \sum_{j=1}^{M} b_j^2\right| \le$$

$$\le 1 + (\sum_{j\in I} c_{d_j})^2$$

the Cauchy-Schwarz inequality yields

$$E(T_{\omega N} - S_{\omega N})^2 \le (2 + (\sum_{j\in I} c_{d_j})^2) \, E(J_N\left(\frac{Q_1}{M+1}\right) - J_N(V_1))^2.$$

Furthermore we have

$$E(J_N\left(\frac{Q_1}{M+1}\right) - J_N(V_1))^2 = \frac{1}{M}\sum_{j=1}^{M} E(J_N(V_{j:M}) - J_N(\frac{j}{M+1}))^2$$

where $V_{1:M} < V_{2:M} < \ldots < V_{M:M}$ denote the order statistics of $V_1, V_2, \ldots, V_M$. We note that $|J_N'(t)|$ is bounded above by

$$h_N'(t) = (\frac{N-2m+1}{N+1}) \, h'(\frac{m+(N-2m+1)t}{N+1})$$

where h is defined as in the proof of Lemma 3.1. Since h_N satisfies condition R_2, we can argue as in the proof of Lemma 3.1 to show that

$$\sum_{j=1}^{M} E\,\{J_N(V_{j:M}) - J_N(\frac{j}{M+1})\}^2 = \mathcal{O}(\frac{1}{M}\sum_{j=1}^{M} \frac{j}{M+1}(1-\frac{j}{M+1})\{h_N'(\frac{j}{M+1})\}^2) =$$

$$= \mathcal{O}(\int_0^1 t(1-t)\{h_N'(t)\}^2 dt) = \mathcal{O}(\int_{\frac{m}{N+1}}^{1-\frac{m}{N+1}} \{t(1-t)\}^{-3/2+2\delta} dt) =$$

$$= \mathcal{O}(N^{1/3-4\delta/3})$$

and the proof of the lemma is complete. □

It follows from Lemma 4.1 that

$$|Ee^{itT_{\omega N}} - Ee^{itS_{\omega N}}| = O(|t|N^{-1/3-2\delta/3}\{1+(\sum_{j\in I} c_{d_j})^2\}^{\frac{1}{2}}). \tag{4.21}$$

Since $S_{\omega N}$ is a sum of independent random variables, with variance (cf. assumption (A))

$$\tau_\omega^2 = \sigma^2(S_{\omega N}) = (1 - \sum_{j\in I} c_{d_j}^2)\sigma^2(J_N(V_1)), \tag{4.22}$$

Lemma V 2.1 of PETROV (1972) together with assumptions (A) and (B) yield

$$|Ee^{it(S_{\omega N}-ES_{\omega N})} - e^{-\frac{1}{2}\tau_\omega^2 t^2}| = O(N^{-\frac{1}{2}}|t|^3) \tag{4.23}$$

for all $|t| \leq \log N$. Furthermore, in view of assumptions (A), (B) and (C),

$$|Ee^{it(S_{\omega N} - ES_{\omega N})} - Ee^{itS_{\omega N}}| \leq |t|\ |ES_{\omega N}| =$$

$$= |t\sum_{j=1}^{M} b_j \int_0^1 J_N(t)dt| \leq |t|\ \frac{N}{M}\ |\sum_{j\in I} c_{d_j}|\ |\int_{\frac{m}{N+1}}^{1-\frac{m}{N+1}} J(t)dt| \leq \tag{4.24}$$

$$\leq 2|t|\ \frac{mN}{M} \max_{1\leq j\leq N} |c_j|\ \left|\int_0^{\frac{m}{N+1}} \{J(t) + J(1-t)\}dt\right| = O(|t|N^{-1/3-2\delta/3}).$$

Defining

$$\tau_N^2 = MN^{-1}\ \sigma^2(J_N(V_1)) \tag{4.25}$$

and noting that

$$|e^{-\frac{1}{2}\tau_\omega^2 t^2} - e^{-\frac{1}{2}\tau_n^2 t^2}| \leq \tfrac{1}{2}\ |\tau_\omega^2 - \tau_N^2|\, t^2,$$

we combine (4.21) through (4.24) to arrive at

$$(4.26)\qquad |Ee^{itT_{\omega N}} - e^{-\frac{1}{2}\tau_N^2 t^2}| = O(\ |t|N^{-1/3-2\delta/3}\{1+(\sum_{j\in I} c_{d_j})^2\}^{\frac{1}{2}} + t^2|\tau_\omega^2 - \tau_N^2|)$$

for all $|t| \le \log N$ and uniformly in ω. Substituting this result in (4.17) we obtain by repeated use of the Cauchy-Schwarz inequality,

$$Ee^{itT'_N} \sum_{j\in I} c_{D_j} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}) =$$

$$= e^{-\frac{1}{2}\tau_N^2 t^2} E(e^{it\sum_{j\in I} c_{D_j} J(\frac{j}{N+1})} \sum_{j\in I} c_{D_j} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}) +$$

$$+ O(E|\sum_{j\in I} c_{D_j} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}|(\ |t|N^{-1/3-2\delta/3}\{1+(\sum_{j\in I} c_{D_j})^2\}^{\frac{1}{2}} +$$

$$(4.27)\qquad + t^2|1 - \sum_{j\in I} c_{D_j}^2 - \frac{M}{N}|)) =$$

$$= e^{-\frac{1}{2}\tau_N^2 t^2} E(\{1+it\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\}\sum_{j\in I} c_{D_j} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}) +$$

$$+ O(\{E(\sum_{j\in I} c_{D_j} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\})^2\}^{\frac{1}{2}} [\{t^2(E\{\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\}^4)^{\frac{1}{2}}\} +$$

$$+ |t|N^{-1/3-2\delta/3}\{E(1+(\sum_{j\in I} c_{D_j})^2)\}^{\frac{1}{2}} + t^2\{E(\sum_{j\in I} c_{D_j}^2 - \frac{2m}{N})^2\}^{\frac{1}{2}}])$$

for all $|t| \le \log N$.

We note that the assumptions (A) and (C), (3.5) and (3.11) imply

$$|E\{1+it\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1})\}\sum_{j\in I} c_{D_j} \{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}.$$

$$(4.28)\qquad = |t|\left|\frac{1}{N-1}\sum_{j\in I} J(\tfrac{j}{N+1})\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\} - \frac{1}{N(N-1)}\{\sum_{j\in I} J(\tfrac{j}{N+1})\}\cdot\right.$$

$$\left.\cdot\sum_{j\in I}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}\right| = O(|t|N^{-\frac{1}{2}-2\delta}).$$

Furthermore, we obtain by applying Lemma 3.3

$$(4.29)\qquad \{E(\sum_{j\in I} c_{D_j} J(\tfrac{j}{N+1}))^4\}^{\frac{1}{2}} = O(N^{-1/3-4\delta/3}).$$

Finally,

$$(4.30)\qquad \{E(1+(\sum_{j\in I} c_{D_j})^2)\}^{\frac{1}{2}} = 1 + O(N^{-2/3})$$

and

$$(4.31)\qquad \{E(\sum_{j\in I} c_{D_j} - 2MN^{-1})^2\}^{\frac{1}{2}} = O(N^{-2/3})$$

according to assumption (A). Combining (4.16) and (4.28) through (4.31) and substituting these results in the right-hand side of (4.27) we find that

$$(4.32)\qquad E(e^{itT'_N} \sum_{j\in I} c_{D_j}\{EJ(U_{j:N}) - J(\tfrac{j}{N+1})\}) = O(|t|N^{-\frac{1}{2}-2\delta})$$

for all $|t| \leq \log N$. To conclude we note that it follows from (3.4) that

$$(4.33)\qquad \sigma^2(T'_N) = \frac{1}{N-1}\sum_{j=1}^{N}(J(\tfrac{j}{N+1}) - \frac{1}{N}\sum_{i=1}^{N} J(\tfrac{i}{N+1}))^2 = 1 + O(N^{-\frac{1}{2}-2\delta}).$$

We see that the proof of Theorem 2.2 is complete by combining (4.1), Lemma 3.2, (4.11) through (4.14), (4.32) and (4.33). □

ACKNOWLEDGEMENT

The author is very grateful to Professor W.R. van Zwet for suggesting the problem and for his essential help and stimulating discussions during the preparation of this paper.

REFERENCES

[1] ALBERS, W., P.J. BICKEL, & W.R. VAN ZWET, (1976), *Asymptotic expansions for the power of distributionfree tests in the one-sample problem*, Ann. Statist. 4, 108-156.

[2] FELLER, W., (1971), *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd Edition, Wiley, New York.

[3] HÁJEK, J. & Z. ŠIDÁK, (1967), *Theory of Rank Tests*, Academic Press, New York.

[4] HUŠKOVÁ, M., (1977), *The rate of convergence of simple linear rank statistics under hypothesis and alternatives*, Ann. Statist. 5, 658-670.

[5] HUŠKOVÁ, M., (1979), *The Berry-Esseen theorem for rank statistics*, Comment. Math. Univ. Carolin. 20, 399-415.

[6] PETROV, V.V., (1972), *Sums of Independent Random Variables*, Springer-Verlag, Berlin.

[7] VAN ZWET, W.R., (1980), *On the Edgeworth expansion for the simple linear rank statistic*, Mathematical Centre Report SW 70, Amsterdam, to appear in the Proceedings of the Colloquium on Nonparametric Statistical Inference, I. Vincze editor, Budapest.