Ann. Inst. Statist. Math.
Vol. 54, No. 2, 391–403 (2002)

ASYMPTOTIC BOUND ON THE CHARACTERISTIC FUNCTION OF SIGNED LINEAR SERIAL RANK STATISTICS

M'HAMMED KADRI[1] AND KHALID RIFI[2]

[1]*Département de Mathématiques et Informatique, Faculté des Sciences Dhar Mahraz, Fès, Morocco*
[2]*Ecole Normale Supérieure, B. P. 5206, Bensouda, Fès, Morocco*

(Received August 10, 1999; revised October 16, 2000)

Abstract. This work establishes an asymptotic bound on the characteristic function of signed linear serial rank statistics. The result is obtained under rather general conditions including the important case of van der Waerden scores. It generalizes the result of Seoh (1983, Ph.D. Thesis, Department of Mathematics, Indiana University) and constitutes an essential step to the elaboration of Berry-Esséen's bounds and the establishment of Edgeworth expansions. These statistics constitute a natural tool for testing the hypothesis of white noise with a symmetrical (unspecified) distribution in comparison to other alternative hypothesis of serial dependence.

Key words and phrases: Berry-Esséen bounds, characteristic function, Edgeworth expansions, signed serial rank statistics, time series.

1. Introduction

Let $\underline{X}_n = (X_{n1}, \dots, X_{nn})$ be a vector of independent random variables with probability density functions $f_{n1}, \dots, f_{nn}$ and distribution functions $F_{n1}, \dots, F_{nn}$, and let $\underline{R}_n^+ = (R_{n1}^+, \dots, R_{nn}^+)$ and $\underline{Z}_n = (Z_{n1}, \dots, Z_{nn})$ be respectively the vector of signed ranks and order statistics associated to absolute values $|X_{n1}|, \dots, |X_{nn}|$.

Linear signed rank statistics take the from

$$(1.1) \qquad T_{n,+} = (n-1)^{-1/2} \sum_{t=1}^{n} c_{nt} a_n(R_{nt}^+) \mathrm{sgn}(X_{nt}),$$

where $\underline{a}_n = (a_n(1), \dots, a_n(n))$ and $(c_{n1}, \dots, c_{nn})$, $n \in I\!N$, respectively denote a score vector of real numbers and a vector of regression constants and $\mathrm{sgn}(x) = 1$ if $x \geq 0$ and $\mathrm{sgn}(x) = -1$ elsewhere. These statistics are usually used to test the symmetry hypothesis (see, e.g. Hájek, (1962), Hájek and Šidák, (1967)) $H_1^{(n)} : F_{n1} = F_{n2} = \cdots = F_{nn} = F_n$, with $F_n(x) = 1 - F_n(-x)$ against some class of alternative hypothesis. The asymptotic normality of $T_{n,+}$ was established under $H_1^{(n)}$ and under suitable alternatives by several authors, namely by Hájek and Šidák (1967), Hájek (1968) and Hušková (1970). Under the symmetry hypothesis $H_1^{(n)}$ and for bounded score generating function, Puri and Wu (1986) have obtained the rate of convergence to the normality of the order $O(n^{-1/2+\delta})$, $\delta > 0$. An L^p bound for these statistics was obtained by Wu (1987). By adapting the van Zwet (1980) method, Puri and Seoh (1984*a*) have obtained, under the symmetry hypothesis $H_1^{(n)}$, Berry-Esséen bounds of order $O(n^{-1/2})$ for the statistic $T_{n,+}$, where the scores $(a_n(1), \dots, a_n(n))$ are derived from a score generating function not necessarily bounded. The Edgeworth expansions have been established by Puri and Seoh (1984*b*). The case of the unsigned linear rank statistics of the type

$$T_n = (n-1)^{-1/2} \sum_{t=1}^{n} c_{nt} a_n(R_{nt}),$$

where R_{nt} is the rank of X_{nt} among $(X_{ni} : 1 \le i \le n)$ were investigated by many authors. For the review, the reader is referred to von Bahr (1976), Hušková (1977, 1979) and Does (1982, 1983).

Asymptotic behaviour of the characteristic function and Berry-Esséen bounds for linear serial unsigned rank statistics have been established by Hallin and Rifi (1996, 1997). Under the symmetry hypothesis $H_1^{(n)}$ and the hypothesis $H^{(n)}$ where the variables $X_{n1}, \ldots, X_{nn}$ are independent (not necessarily having the same distribution), we establish the asymptotic behaviour of the characteristic function of serial linear signed rank statistics of the form

$$T_{n,+}^{(k)} = (n-k)^{-1/2} \sum_{t=k+1}^{n} a_n(R_{nt}^{+}) b_n(R_{nt-k}^{+}) \mathrm{sgn}(X_{nt}) \mathrm{sgn}(X_{nt-k}), \tag{1.2}$$

where $\underline{b}_n = (b_n(1), \ldots, b_n(n))$ is a score vector of real numbers, k is an integer $(1 \le k \le n-1)$.

The asymptotic normality of $T_{n,+}^{(k)}$ was established under the hypothesis $H_1^{(n)}$ by Hallin *et al.* (1987) and Hallin and Puri (1992).

2. Technical conditions and principal results

The first three conditions below are the same as those considered in Seoh (1983).

Assumption (A_1). There exist strictly positive real numbers a, A, b, B such that the scores $\underline{a}_n$ and $\underline{b}_n$ satisfy,

$$\sum_{i=1}^{n} |a_n(i)| \ge an, \qquad \sum_{i=1}^{n} a_n^2(i) \le An,$$
$$\sum_{i=1}^{n} |b_n(i)| \ge bn, \qquad \sum_{i=1}^{n} b_n^2(i) \le Bn.$$

Assumption (A_2). There exists $\delta > 0$ such that, for some $\zeta > n^{-3/2} \log n$, $\gamma(a_n(1), \ldots, a_n(n); \zeta) \ge n\delta\zeta$, where $\gamma(a_n(1), \ldots, a_n(n); \zeta) = \lambda\{x \in I\!R / \exists j \; 1 \le j \le n$ and $|x - a_n(j)| < \zeta\}$ with λ the Lebesgue measure.

The same holds true for $(b_n(1), \ldots, b_n(n))$, i.e., for some $\zeta \ge n^{-3/2} \log n$, $\gamma(b_n(1), \ldots, b_n(n); \zeta) \ge n\delta\zeta$.

Assumption (A_3). There exist a sequence of probability density functions $(\hat{f}_n)$ and a strictly positive sequence (ϵ_n) decreasing to zero, such that

$$\sum_{j=1}^{n} \int \frac{(f_{nj}(x) - \hat{f}_n(x))^2}{\hat{f}_n(x)} dx \le n\epsilon_n.$$

ASSUMPTION (A_4). If we put

$$\rho_n = \int \frac{(\hat{f}_n(x) - \hat{f}_n(-x))^2}{\hat{f}_n(x)} dx,$$

the sequence ρ_n decreases to zero.

Remark 2.1. The Assumption (A_4) is stronger than Assumption (A_5) of Theorem II.2.1 in Seoh (1983) since he supposes only that

$$\limsup_{n\to+\infty} \int |\hat{f}_n(x) - \hat{f}_n(-x)| dx < \infty.$$

Remark 2.2. If $(\hat{f}_n)$ satisfies Assumptions (A_3) and (A_4) and if $\hat{g}_n(x)$ is defined by

$$\hat{g}_n(x) = \frac{\hat{f}_n(x) + \hat{f}_n(-x)}{2},$$

then $\hat{g}_n$ is symmetrical. Furthermore, we can prove the existence of a sequence (α_n) of positive numbers converging to zero, such that

$$\sum_{j=1}^{n} \int \frac{(f_{nj}(x) - \hat{g}_n(x))^2}{\hat{g}_n(x)} dx \le n\alpha_n.$$

Denote $\mu_n^{(k)} = E(T_{n,+}^{(k)})$ and $(\sigma_n^{(k)})^2 = \sigma^2(T_{n,+}^{(k)})$ the mean and the variance of $T_{n,+}^{(k)}$, defined in (1.2), $T_n^* = (T_{n,+}^{(k)} - \mu_n^{(k)})/\sigma_n^{(k)}$ the standardized statistic and ψ_n^* its characteristic function.

Remark 2.3. Esséen's smoothing lemma (cf. Feller (1971), p. 538) reduces the proofs of Berry-Esséen bounds and Edgeworth expansion to the study of integrals containing the characteristic function $\psi_n^*(u)$ of T_n^* for large values of $|u|$. To bound these integrals, we will use respectively Theorem 2.2 and Theorem 2.1. For more details about this method we refer to van Zwet (1980).

In the sequel, we will use the following notations

$$\delta_0 = \frac{9b^2}{16B}, \quad \delta_1 = \frac{9a^2}{16A}, \quad \delta_2 = \min(\delta_0, \delta_1),$$

(2.1)
$$\delta_3 = \frac{\delta\delta_2}{32} \min\left(\frac{\delta b}{3b+8}, \delta_2\right), \quad \delta_4 = \frac{\delta_2^3}{2^8}, \quad \delta_5 = \frac{1}{6} \min\left(\frac{\delta_3}{2}, \delta_4\right).$$

The characteristic function of $T_{n,+}^{(k)} - \mu_n^{(k)}$ is given by $\varphi_{n,+}^{(k)}(u) = E\exp(iu(T_{n,+}^{(k)} - \mu_n^{(k)}))$, $u \in \mathbb{R}$.

THEOREM 2.1. *Under the independence hypothesis $H^{(n)}$, if the Assumptions (A_1), (A_2), (A_3) and (A_4) are satisfied, then there exist strictly positive numbers c, C, and κ depending only on a, A, b, B and the sequences (ρ_n) and (ϵ_n) such that, for $n > k$ and $\log n < |u| < cn^{3/2}$, $|\varphi_{n,+}^{(k)}(u)| < Cn^{-\kappa \log n}$.*

Note that in this theorem, constants c, C and κ are not depending on n, but only upon the whole sequences $(\rho_n, n \in I\!N)$ and $(\epsilon_n, n \in I\!N)$.

THEOREM 2.2. *Suppose that only Assumption* (A_1) *is satisfied. Then, under the hypothesis* $H_1^{(n)}$, *there exist strictly positive constants* c, C, *and* κ *such that, for* $n > k$ *and* $\log n < |u| < cn^{1/2}$, $|\varphi_{n,+}^{(k)}(u)| < Cn^{-\kappa \log n}$.

For convenience of notation, for fixed n, we put $a(i) = a_n(i)$, $b(i) = b_n(i)$, $X_i = X_{ni}$, $s_i = \mathrm{sgn}(X_{ni})$, $R_i^+ = R_{ni}^+$, $\underline{R}^+ = \underline{R}_n^+$ and $\underline{Z} = \underline{Z}_n$. $[x]$ denotes the integer part of x, $[x]^\star$ the smallest integer number greater or equal to x, and $|J|$ the cardinal of J.

3. Preliminary results and proofs

First, we consider the case of signed linear serial and simple rank statistics $T_{n,+}^{(1)}$ of order 1 (say T_n^+). Then, the results will be generalized in the case of any order $k \geq 1$. The characteristic function $\varphi_n^+(u)$ of the centered statistics corresponding to T_n^+ is given by $\varphi_n^+(u) = E\exp(iu(T_n^+ - ET_n^+))$. The proof of the theorems will be split to several steps.

LEMMA 3.1. (Bernstein's inequality) *Given* r *independent random variables* $y_1, \ldots, y_r$ *with Bernoulli distributions having the respective parameters* $\pi_1, \ldots, \pi_r$, *denote* $\pi = \sum_{i=1}^r \pi_i$, $B_r = \sum_{i=1}^r \pi_i(1-\pi_i)$ *and* $y = \sum_{i=1}^r y_i$. *Then, there exists a constant* $c > 0$ *such that, for any real* $\alpha \leq \pi$,

$$P(y \geq \alpha) \geq 1 - 2\exp\left\{\frac{-(\pi-\alpha)^2}{2B_r + (\pi-\alpha)c}\right\}.$$

PROOF. See Hallin and Rifi (1996).

The two following lemmas are the key to the proof of the main theorems.

LEMMA 3.2. (van Zwet (1980)) *Under the Assumption* (A_3) *and for any event* A *in the* σ*-algebra generated by the random variables* $X_1, \ldots, X_n$, *we have* $P(A) \leq 2(e^{n\epsilon_n} P_0(A))^{1/2}$, *where* P *and* P_0 *are respectively the probability measures calculated under the independence hypotheses* $H^{(n)}$ *(defined above) and* $H_0^{(n)}$, *a special case when* $X_1, \ldots, X_n$ *are independent and identically distributed.*

Consider real numbers $d_1, \ldots, d_m$ and $p_1, \ldots, p_m$ with $0 \leq p_j \leq 1$ for $j = 1, \ldots, m$. For $\zeta > 0$ and $0 < \epsilon < 1/2$, let $\gamma(d_1, \ldots, d_m; p_1, \ldots, p_m; \zeta, \epsilon)$ denote the Lebesgue measure λ of the ζ-neighorbood of the set of those d_j for which the corresponding p_j satisfy $\epsilon < p_j \leq 1 - \epsilon$; thus

$$\gamma(d_1, \ldots, d_m; p_1, \ldots, p_m; \zeta, \epsilon) = \lambda\{x : \exists j |x - d_j| < \zeta, \epsilon \leq p_j \leq 1 - \epsilon\}.$$

LEMMA 3.3. (van Zwet (1980)) *Suppose that positive number* d, D, δ' *and* ϵ *exist such that*

$$\sum_{j=1}^m p_j(1-p_j)d_j^2 \geq dm, \qquad \sum_{j=1}^m d_j^4 \leq Dm, \quad \gamma(d_1, \ldots, d_m; p_1, \ldots, p_m; \zeta', \epsilon) \geq \delta' m\zeta'$$

for some $\zeta' \geq m^{-3/2} \log m$. *Then, for every positive* b_1, *there exist positive numbers* b_2, B *and* β *depending only on* d, D, δ', ϵ *and* b_1 *and such that*

$$\prod_{j=1}^{m} (1 - 2p_j(1-p_j)(1-\cos(m^{-1/2} t d_j)))^{1/2} \leq B m^{-\beta \log m}$$

for $b_1 \log m \leq |t| \leq b_2 m^{3/2}$.

The statistic $(n-1)^{1/2} T_n^+$ can be decomposed as follows

$$(n-1)^{1/2} T_n^+ = T_{n1}^+ + T_{n2}^+, \tag{3.1}$$

where

$$T_{n1}^+ = \sum_{t=1}^{n_1} a(R_{2t}^+) b(R_{2t-1}^+) s_{2t} s_{2t-1},$$
$$T_{n2}^+ = \sum_{t=1}^{n_1 - 1} a(R_{2t+1}^+) b(R_{2t}^+) s_{2t+1} s_{2t} + a(R_n^+) b(R_{n-1}^+) s_n s_{n-1} I(n\,\text{odd}),$$

where $n_1 = [n/2]$ and $I(A)$ is the indicator function of a set A. For $2 \leq t \leq n_1$, we define

$$p_t = \frac{f_t(|X_t|)}{f_t(|X_t|) + f_t(-|X_t|)}, \qquad q_t = p_{2t-1}, \tag{3.2}$$

and

$$D_t = a(R_{2t}^+) b(R_{2t-1}^+) s_{2t} + a(R_{2t-1}^+) b(R_{2t-2}^+) s_{2t-2}. \tag{3.3}$$

LEMMA 3.4. *Under the independence hypothesis* $H^{(n)}$, *for all* $u \in I\!R$ *and* $n \geq 5$,

$$|\varphi_n^+(u)| \leq E\left\{ \prod_{t=2}^{n_1} [1 - 2q_t(1-q_t)(1-\cos((n-1)^{-1/2} 2uD_t))]^{1/2} \right\}. \tag{3.4}$$

PROOF. According to the decomposition (3.1), we can write

$$(n-1)^{1/2} T_n^+ = \sum_{t=2}^{n_1} D_t s_{2t-1} + a(R_2^+) b(R_1^+) s_2 s_1 + a(R_n^+) b(R_{n-1}^+) s_n s_{n-1} I(n\,\text{odd}).$$

Let $L = \{s_t \mid t \text{ an even number}, 1 \leq t \leq n\}$. Since conditionally giving $\underline{Z}$ and $\underline{R}^+$, $s_1, \ldots, s_n$ are independent with probabilities $p_t = P(s_t = 1 \mid \underline{Z}, \underline{R}^+) = 1 - P(s_t = -1 \mid \underline{Z}, \underline{R}^+)$, we have

$$\begin{aligned} E((n-1)^{1/2} T_n^+ \mid \underline{Z}, \underline{R}^+, L) = {} & \sum_{t=2}^{n_1} (2p_{2t-1} - 1) D_t + (2p_1 - 1) a(R_2^+) b(R_1^+) s_2 \\ & + (2p_n - 1) a(R_n^+) b(R_{n-1}^+) s_{n-1} I(n\,\text{odd}). \end{aligned}$$

On the other hand,

$$\begin{aligned}|E\exp(iu(T_n^+ - E(T_n^+)))| &\le E|E(\exp(iu(T_n^+ - E(T_n^+ \mid \underline{Z}, \underline{R}^+, L))) \mid \underline{Z}, \underline{R}^+, L)| \\ &= E\prod_{t=2}^{n_1} |p_{2t-1}\exp(iu(n-1)^{-1/2}2(1-p_{2t-1})D_t) \\ &\qquad + (1-p_{2t-1})\exp(-iu(n-1)^{-1/2}2p_{2t-1}D_t)| \\ &= E\prod_{t=2}^{n_1}\{1 - 2q_t(1-q_t)(1-\cos((n-1)^{-1/2}2uD_t)\}^{1/2}.\end{aligned}$$

LEMMA 3.5. *Under the independence hypothesis $H^{(n)}$, if the Assumptions (A_3) and (A_4) are fulfilled, then for all $\epsilon \in [0, 1/4]$ and $\eta \in [0, \frac{1}{2} - \delta_5]$, there exist c_1 and κ_1, strictly positive numbers, such that*

$$P(\epsilon \le q_t \le 1-\epsilon \text{ for at least } [n\eta]^* \text{ indices } t) \ge 1 - c_1 e^{-\kappa_1 n}. \tag{3.5}$$

Remark 3.1. It is sufficient to prove relation (3.5) for large values of n. In fact, given a finite integer n_0, for $2 \le i \le n_0$, we can choose $c_1^{(i)}$ and $\kappa_1^{(i)}$ such that (3.5) holds. Take, then

$$c_1 = \sum_{i=2}^{n_0} c_1^{(i)} \quad \text{and} \quad \kappa_1 = \min_{2\le i\le n_0} \kappa_1^{(i)}.$$

This remark is available for all following results. In the sequel, we suppose n sufficiently large.

PROOF. According to Assumptions (A_3) and (A_4), we have

$$\begin{aligned}E(|2p_t - 1|) &= \int_{-\infty}^{+\infty} \frac{|f_t(x) - f_t(-x)|}{f_t(x) + f_t(-x)} f_t(x)dx \\ &= \int_{-\infty}^{+\infty} \frac{|f_t(x) - f_t(-x)|}{f_t(x) + f_t(-x)} f_t(-x)dx \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} |f_t(x) - f_t(-x)|dx \\ &\le \int_{-\infty}^{+\infty} |f_t(x) - \hat{f}_n(x)|dx + \frac{1}{2}\int_{-\infty}^{+\infty} |\hat{f}_n(x) - \hat{f}_n(-x)|dx.\end{aligned}$$

Now, from Markov's inequality,

$$P(|2p_t - 1| > 1 - 2\epsilon) \le \frac{1}{1-2\epsilon} E(|2p_t - 1|).$$

Put $\hat{f}_n = \hat{f}$. Then, applying both Cauchy-Schwarz and Hölder inequalities, we get

$$\frac{1}{n}\sum_{t=2}^{n_1} P(|2q_t - 1| > 1 - 2\epsilon) \le \frac{1}{n}\sum_{t=1}^{n} P(|2p_t - 1| > 1 - 2\epsilon)$$

$$\leq \frac{1}{1-2\epsilon}\left(\frac{1}{n}\sum_{t=1}^{n}\int\frac{(f_t-\hat{f})^2}{\hat{f}}\right)^{1/2}$$
$$+\frac{1}{2(1-2\epsilon)}\int_{-\infty}^{+\infty}|\hat{f}(x)-\hat{f}(-x)|dx$$
$$\leq 2\sqrt{\epsilon_n}+\int_{-\infty}^{+\infty}|\hat{f}(x)-\hat{f}(-x)|dx.$$

For $2 \leq t \leq n_1$, we define $y_t = 1$ if $q_t \in [\epsilon, 1-\epsilon]$ and $y_t = 0$ if not. Then, $y_2, \dots, y_{n_1}$ are Bernoulli's random variables with parameters $\pi_2, \dots, \pi_{n_1}$, respectively, such that $\pi_t = P(q_t \in [\epsilon, 1-\epsilon])$, and we have

$$P(\epsilon \leq q_t \leq 1-\epsilon \text{ for at least } [n\eta]^\star \text{ indices } t, 2 \leq t \leq n_1) = P\left(\sum_{t=2}^{n_1} y_t \geq [n\eta]^\star\right).$$

Since $\epsilon_n \downarrow 0$, for $\eta_0 = \frac{1}{2}(\frac{1}{2}-\delta_5-\eta)$ there exists d_0 such that for $n \geq d_0$, $2\sqrt{\epsilon_n}+\frac{3}{n} < \eta_0$. Assumption (A_4) assures the existence of an integer m_0 such that, for $n \geq m_0$, $\int|\hat{f}_n(x)-\hat{f}_n(-x)|dx < \delta_5$. There exists an integer n_0, $n_0 = \sup(d_0, m_0)$, such that, for $n \geq n_0$, $\sum_{t=2}^{n_1}\pi_t - [n\eta]^\star > n\eta_0$. According to Lemma 3.1, there exists a constant $c > 0$, such that

$$P\left(\sum_{t=2}^{n_1} y_t \geq [n\eta]^\star\right) \geq 1-2\exp\left\{\frac{-(\sum\pi_t-[n\eta]^\star)^2}{2\sum\pi_t(1-\pi_t)+c(\sum\pi_t-[n\eta]^\star)}\right\}$$
$$\geq 1-2\exp\left\{\frac{-n\eta_0^2}{4\sqrt{\epsilon_{n_0}}+2\delta_5+c\eta_0}\right\}$$
$$\geq 1-c_{1n_0}e^{-n\kappa_{1n_0}},$$

since

$$\frac{(\sum\pi_t-[n\eta]^\star)^2}{2\sum\pi_t(1-\pi_t)+c(\sum\pi_t-[n\eta]^\star)} \geq \frac{n\eta_0^2}{4\sqrt{\epsilon_{n_0}}+2\delta_5+c\eta_0},$$

where $c_{1n_0} = 2$ and $\kappa_{1n_0} = \eta_0^2/(4\sqrt{\epsilon_{n_0}}+2\delta_5+c\eta_0)$. The proof is complete.

LEMMA 3.6. *Under Assumption* (A_1), *there exist two subsets of* $\{1, 2, \dots, n\}$, I_a *and* I_b, *satisfying* $card(I_a) \geq [\delta_1 n]^\star$ *and* $card(I_b) \geq [\delta_0 n]^\star$ *and such that*

$$\forall t \in I_a, \quad |a(t)| \geq a/4 \quad \textit{and} \quad \forall t \in I_b, \ |b(t)| \geq b/4.$$

PROOF. Given $|a|^{(1)}, \dots, |a|^{(n)}$, the numbers $|a(t)|$ ranked in decreasing order, we put $\alpha_t = |a|^{(t)}$. Then for $t \leq [\delta_1 n]^\star$, $\alpha_t \geq a/4$. If not, there will exist $t_0 \leq [\delta_1 n]^\star$ satisfying $\alpha_{t_0} < \frac{a}{4}$. This implies that

$$\sum_{t=1}^{n}\alpha_t = \sum_{t=1}^{t_0-1}\alpha_t + \sum_{t_0}^{n}\alpha_t$$
$$< (t_0-1)\left(\frac{1}{t_0-1}\sum_{t=1}^{t_0-1}\alpha_t^2\right)^{1/2} + \frac{na}{4}$$
$$< na.$$

This is contradictory to (A_1). Let now π be the permutation that permits to rank the $|a(t)|$'s in decreasing order. We have $I_a = \pi^{-1}\{1, 2, \ldots, [\delta_1 n]^\star\}$. The same reasoning holds for the existence of the set I_b.

LEMMA 3.7. *Under the independence hypothesis $H^{(n)}$, Assumptions (A_1), (A_2) and (A_3), there exist strictly positive constants c_2 and κ_2 depending only on a, A, b, B and the sequence (ϵ_n), such that*

$$P(\gamma(D_2, \ldots, D_{n_1}, \zeta) > \zeta\delta_3 n) \geq 1 - c_2 e^{-\kappa_2 n}. \tag{3.6}$$

PROOF. Note that Lemma 3.2 reduces the proof of (3.6) to the case of independence and equidistribution hypothesis $H_0^{(n)}$. Putting $r = [\frac{n}{4}\min(\frac{\delta b}{3b+8}, \delta_2)]^\star$, for $k \leq r$, we have $\gamma(D_2, \ldots, D_k, \zeta) = \gamma(D_2, \ldots, D_{k-1}, \zeta) + 2\zeta$. This is the case unless $|D_k - D_t| < 2\zeta$ for some $t \leq k-1$, i.e. except that $D_k \in \cup_{t=2}^{k-1}]D_t - 2\zeta, D_t + 2\zeta[$. Now, if $|b(R_{2k-1}^+)| \geq b/4$, the relation above restricts $a(R_{2k}^+)$ to a set A_k which is union of $(k-2)$ intervals of length smaller or equal to $16\zeta/b$. Consequently, the set of $a(t)$ in A_k has a ζ-neighborhood of Lebesgue measure at most equal to $(k-2)(\frac{16\zeta}{b} + 2\zeta)$. According to Assumption (A_2), we have

$$\begin{aligned}\#(t, a(t) \notin A_k) &\geq \frac{1}{2\zeta}\left(\delta n\zeta - (k-2)\left(\frac{16\zeta}{b} + 2\zeta\right)\right)\\ &\geq \frac{\delta n}{2} - (k-2)(1 + 8/b).\end{aligned}$$

On the other hand,

$$\begin{aligned}&P_0(a(R_{2k}^+) \notin A_k \mid R_2^+, R_3^+, \ldots, R_{2(k-1)}^+)\\ &\geq P_0\left(a(R_{2k}^+) \notin A_k \mid |b(R_{2k-1}^+)| \geq \frac{b}{4}, R_2^+, R_3^+, \ldots, R_{2(k-1)}^+\right)\\ &\quad\times P_0\left(|b(R_{2k-1}^+)| \geq \frac{b}{4} \mid R_2^+, R_3^+, \ldots, R_{2(k-1)}^+\right)\\ &\geq \frac{\frac{\delta n}{2} - (k-2)\left(1 + \frac{8}{b}\right) - 2(k-1) + 1}{n - 2(k-1) + 1} \times \frac{\delta_o n - 2(k-1) + 1}{n - 2(k-1) + 1}\\ &\geq \left(\frac{\delta}{2} - \frac{(r-1)}{n}\left(3 + \frac{8}{b}\right)\right)\left(\delta_0 - \frac{2}{n}(r-1)\right)\\ &\geq \delta\frac{\delta_2}{8}.\end{aligned}$$

As $a(R_{2k}^+) \notin A_k$, 2ζ is added at the k-th step. Then $\frac{1}{2\zeta}\gamma(D_2, \ldots, D_r, \zeta)$ is stochastically larger than a binomial random variable, say $B(r, \frac{\delta\delta_2}{8})$. Since $r\frac{\delta\delta_2}{8} \geq n\delta_3$, $r < n$ and $\gamma(D_2, \ldots, D_{n_1}; \zeta) \geq \gamma(D_2, \ldots, D_r; \zeta)$, the proof is completed by applying Lemma 3.1.

LEMMA 3.8. *Under the independence hypothesis $H^{(n)}$, Assumptions (A_1), (A_3) and (A_4), there exist strictly positive numbers c_3 and κ_3, depending only on a, A, b, B and the sequences (ϵ_n) and (ρ_n) such that,*

$$P\left(|D_t| \geq \frac{ab}{16} \text{ for at least } [\delta_4 n]^\star \text{ indices } t\right) \geq 1 - c_3 e^{-\kappa_3 n}. \tag{3.7}$$

PROOF. Let $r = [n\delta_2/4]^*$, and suppose that $X_1, \ldots, X_n$ are independent and identically distributed random variables having common probability density function $\hat{f}_n$. Without loss of generality, we can suppose that $\hat{f}_n$ is symmetrical. If not, by virtue of Assumption (A_4), $\alpha_n = 3\epsilon_n + 2\rho_n$ and $\hat{g}_n$ may be used instead of ϵ_n and $\hat{f}_n$ to apply Lemma 3.2. For $t \leq r$, put $W_{1t} = a(R^+_{2t})b(R^+_{2t-1})$ and $W_{2t} = a(R^+_{2t-1})b(R^+_{2t-2})$. Denote $(r^+_2, \ldots, r^+_{2t-3})$ an observation of $(R^+_2, \ldots, R^+_{2t-3})$, $\omega_{2t-3} = \{r^+_2, \ldots, r^+_{2t-3}\}$ and P_c the conditional probability given $R^+_2, \ldots, R^+_{2t-3}$. Observe that under $H_1^{(n)}$, the vectors $s = (s_1, \ldots, s_n)$, $\underline{Z}$ and $\underline{R}^+$ are mutually independent (see, e.g., Hájek and Šidák (1967)). We get

$$\begin{aligned} P_c\left(|D_t| \geq \frac{ab}{16}\right) &= P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} \geq 0, W_{2t} \geq 0\right) \\ &+P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} \geq 0, W_{2t} < 0\right) \\ &+P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} < 0, W_{2t} < 0\right) \\ &+P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} < 0, W_{2t} \geq 0\right). \end{aligned}$$

On the other hand, we have

$$\begin{aligned} &P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} \geq 0, W_{2t} \geq 0\right) \\ &\geq P_c(s_{2t} = 1, s_{2t-2} = 1) \times P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} \geq 0, W_{2t} \geq 0 \mid s_{2t} = 1, s_{2t-2} = 1\right) \\ &= \frac{1}{2^2} P_c\left(|W_{1t} + W_{2t}| \geq \frac{ab}{16}, W_{1t} \geq 0, W_{2t} \geq 0\right) \\ &= \frac{1}{2^2} P_c\left(|W_{1t}| + |W_{2t}| \geq \frac{ab}{16}, W_{1t} \geq 0, W_{2t} \geq 0\right). \end{aligned}$$

By the same way, we obtain

$$\begin{aligned} P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} \geq 0, W_{2t} < 0\right) &\geq \frac{1}{2^2} P_c\left(|W_{1t}| + |W_{2t}| \geq \frac{ab}{16}, W_{1t} \geq 0, W_{2t} < 0\right), \\ P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} < 0, W_{2t} < 0\right) &\geq \frac{1}{2^2} P_c\left(|W_{1t}| + |W_{2t}| \geq \frac{ab}{16}, W_{1t} < 0, W_{2t} < 0\right), \\ P_c\left(|D_t| \geq \frac{ab}{16}, W_{1t} < 0, W_{2t} \geq 0\right) &\geq \frac{1}{2^2} P_c\left(|W_{1t}| + |W_{2t}| \geq \frac{ab}{16}, W_{1t} < 0, W_{2t} \geq 0\right). \end{aligned}$$

Using the above inequalities, we have for $t \leq r$,

$$\begin{aligned} P_c\left(|D_t| \geq \frac{ab}{16}\right) &\geq \frac{1}{2^2} P_c\left(|W_{1t}| + |W_{2t}| \geq \frac{ab}{16}\right) \\ &\geq \frac{1}{2^2} P_c\left(|W_{1t}| \geq \frac{ab}{16}\right) \geq \frac{1}{2^2} P_c(R^+_{2t} \in I_a, R^+_{2t-1} \in I_b) \\ &\geq \frac{1}{2^2} \sum_{(i_a, i_b) \in (I_a \backslash \omega_{2t-3}) \times (I_b \backslash \omega_{2t-3})} P_c(R^+_{2t} = i_a, R^+_{2t-1} = i_b) \end{aligned}$$

$$\geq \frac{1}{2^2(n-2t+4)(n-2t+3)} \sum_{(i_a,i_b)\in(I_a\backslash\omega_{2t-3})\times(I_b\backslash\omega_{2t-3})} I[i_a \neq i_b]$$
$$\geq \frac{(\delta_2 n - 2t+3)^2 - n}{2^2 n^2} \geq \frac{1}{2^2}\left(\left(\delta_2 - \frac{2}{n}(r-1)\right)^2 - \frac{1}{n}\right)$$
$$\geq \frac{1}{2^2}\left(\frac{\delta_2^2}{4} - \frac{1}{n}\right) \geq \frac{\delta_2^2}{2^5}.$$

Then, under $H_1^{(n)}$, the number of $t \leq r$ for which $|D_t| \geq ab/16$ is stochastically larger than a binomial random variable $B(r, \delta_2^2/2^5)$. Note that $2\delta_4 n \leq r\delta_2^2/2^5$, and by using Lemma 3.1, the relation (3.7) is thus proved under $H_1^{(n)}$. By using Lemma 3.2, the proof is complete.

PROOF OF THEOREM 2.1. (i) The case $k = 1$. Suppose that n is sufficiently large and let $\epsilon \in [0, 1/4]$. Put $\delta' = \delta_3/100$, $D = [(A+B)/\delta_5]^4$ and $d = \epsilon(1-\epsilon)(ab)^2 2^{-9}\delta_4$, where δ_2, δ_4 and δ_5 are given by the relation (2.1). Let $J = \{2 \leq t \leq n_1, |D_t| \leq D^{1/4}\}$ and $m = |J|$ be the cardinal of J. It is easy to see that $n_1 - 1 - \delta_5 n \leq m \leq n_1 - 1$. We have $\delta_5 < 1/4$ and hence $\frac{n}{20} \leq m \leq n/2$. Let $\zeta_0 = n^{-3/2}\log n$ and define the set F by:

$$F = \left\{\epsilon \leq q_t \leq 1-\epsilon \text{ for at least } \left[\left(\frac{1}{2} - 2\delta_5\right)n\right]^\star \text{ indices } t\right\}$$
$$\cap \ \{\gamma(D_2, \ldots, D_{n_1}, \zeta_0) \geq \delta_3 n \zeta_0\}$$
$$\cap \left\{|D_t| \geq \frac{ab}{16} \text{ for at least } [\delta_4 n]^\star \text{ indices } t\right\}.$$

By Lemma 3.5 (with $\eta = \frac{1}{2} - 2\delta_5$) and Lemmas 3.7 and 3.8, we obtain by Bonferroni's inequality

$$P(F) \geq 1 - c_4 e^{-\kappa_4 n}, \tag{3.8}$$

where $c_4 = c_1 + c_2 + c_3$ and $\kappa_4 = \min(\kappa_1, \kappa_2, \kappa_3)$ are strictly positive numbers depending only on a, A, b, B, δ and the sequences (ϵ_n) and (ρ_n). Moreover, in F, the indices t for which $\epsilon \leq q_t \leq 1-\epsilon$ and $|D_t| \geq \frac{ab}{16}$ is at least equal to $(\delta_4 - 2\delta_5)n$ since the number of indices t not satisfying at least one of these assumptions is at most equal to

$$n_1 - 1 - [\delta_4 n]^\star + n_1 - 1 - [(1/2 - 2\delta_5)n]^\star \leq n_1 - 1 - (\delta_4 - 2\delta_5)n.$$

Thus at least $(\delta_4 - 3\delta_5)n$ of these indices are in J. For any sample in F, we obtain the following inequalities

$$\sum_{t\in J} q_t(1-q_t)D_t^2 \geq \epsilon(1-\epsilon)\frac{(ab)^2}{(16)^2}(\delta_4 - 3\delta_5 n) \geq \epsilon(1-\epsilon)2^{-9}\delta_4 n \geq dm.$$

Similarly, in F the number of indices t satisfying $t \notin J$ or $q_t \notin [\epsilon, 1-\epsilon]$ is at most $3\delta_5 n$. Thus Lebesgue measure of the ζ_0-neighborhood of the set $\{D_t, t \in J\}$ for which the corresponding q_t satisfies $\epsilon \leq q_t \leq 1-\epsilon$, fulfils

$$\gamma(D_t, q_t, t \in J, \zeta, \epsilon) \geq (\delta_3 - 6\delta_5)n\zeta \geq n\frac{\delta_3}{2}\zeta_0 \geq \delta_3 m \zeta_0.$$

By putting $\zeta' = m^{-3/2}\log m$, for $m \geq 2$ we get $\frac{1}{100} \leq \zeta_0/\zeta' \leq 1$. Since $\gamma(\ldots;\zeta)$ is increasing in ζ,

$$\gamma(D_t, q_t, t \in J, \zeta', \epsilon) \geq \gamma(D_t, q_t, t \in J, \zeta_0, \epsilon) \geq m\delta'\zeta'.$$

And according to the definition of J, $\sum_{t\in J} D_t^4 \leq Dm$. Consequently in F, the D_t, q_t, $t \in J$ satisfy the assumptions of Lemma 3.3, with constants d, D, δ' and ϵ. Lemma 3.4 and the relation (3.8) complete this proof.

(ii) For the serial statistic $T_{n,+}^{(k)}$ with any order k, put $n_1^{(k)} = [\frac{n}{2k}]$ and for $2 \leq t \leq n_1^{(k)}$,

$$D_t^{(k)} = a(R_{2tk}^+)b(R_{(2t-1)k}^+)s_{2tk} + a(R_{(2t-1)k}^+)b(R_{(2t-2)k}^+)s_{(2t-2)k},$$

$$S_{n,+}^k = \sum_{t=2}^{n_1^{(k)}} D_t^{(k)} s_{(2t-1)k}, \qquad Q_{n,+}^{(k)} = (n-k)^{1/2}T_{n,+}^{(k)} - S_{n,+}^{(k)}.$$

Let $S = \{s_{2tk}\colon 2 \leq t \leq n_1^{(k)}\}$ and Π be the set of all s_t which appear in $Q_{n,+}^{(k)}$. Given S, Π, $\underline{Z}$ and $\underline{R}^+$, $Q_{n,+}^{(k)}$ is constant and consequently,

$$\begin{aligned}|\varphi_{n,+}^{(k)}(u)| &\leq E|E[\exp\{iu(T_{n,+}^{(k)} - E(T_{n,+}^{(k)} \mid \underline{Z}, \underline{R}^+, \Pi, S))\} \mid \underline{Z}, \underline{R}^+, S, \Pi]| \\ &= E\prod_{t=2}^{n_1^{(k)}} \left\{1 - 2q_t^{(k)}(1-q_t^{(k)})\left(1 - \cos\left(\frac{2u}{(n-k)^{1/2}}D_t^{(k)}\right)\right)\right\}^{1/2},\end{aligned}$$

with $q_t^{(k)} = p_{(2t-1)k}$. Thus we obtain an analogous result to Lemma 3.4. The rest of the proof is similar to the former, since we consider the ranks $(R_{tk}^+, t = 2, \ldots, n_1^{(k)})$ instead of $(R_2, \ldots, R_{n_1})$ for the proof of Lemmas 3.7 and 3.8.

Proof of Theorem 2.2. For the sake of simplicity, we restrict ourselves to the case $k = 1$, the general case follows along the same lines as explained in part (ii) of the proof of Theorem 2.1. Under $H_1^{(n)}$, the Assumptions (A_3) and (A_4) are satisfied. Consequently, Lemmas 3.1, 3.2, 3.4, 3.5, 3.6 and 3.8 remain valid. For sufficiently large n, put $\delta_6 = \delta_4/2$, $D = ((A+B)/\delta_6)^4$, $d = (ab)^2 2^{-9}\delta_4$, $J_1 = \{2 \leq t \leq n_1, |D_t| \leq D^{1/4}\}$ with D_t defined in (3.3) and $m_1 = |J_1|$. It is easy to see that $\frac{n}{20} \leq m_1 \leq \frac{n}{2}$. Let the set F_1 defined by

$$F_1 = \left\{|D_t| \geq \frac{ab}{16} \text{ for at least } [\delta_4 n]^\star \text{ indices } t\right\}.$$

Then, according to Lemma 3.8, there exist positive numbers c_4 and κ_4, such that

$$P(F_1) \geq 1 - c_4 e^{-k_4 n}. \tag{3.9}$$

On the other hand, in the space F_1, the number of indices t not belonging to J_1 is at most equal to $\delta_6 n$. Consequently

$$\sum_{t\in J_1} D_t^2 \geq (ab)^2 2^{-9}\delta_4 n = nd.$$

From the definition of J_1, we have $\sum_{t\in J_1} D_t^4 \leq m_1 D \leq n\frac{D}{2}$. Now, since $X_1, \ldots, X_n$ are independent and identically distributed random variables having a symmetrical probability density, the q_t defined in (3.2) are constants with $q_t = 1/2$. According to Lemma 3.4, if $|u| \leq cn^{1/2}$, with $c = (d/D)^{1/2}$, we have

$$\begin{aligned}
|\varphi_n^+(u)| &\leq E\left(\prod_{t=2}^{n_1}\left\{1+\cos\frac{2uD_t}{(n-1)^{1/2}}\right\}^{1/2} 2^{-1/2}\right)\\
&\leq E\left(\prod_{t\in J_1}\left\{1+\cos\frac{2uD_t}{(n-1)^{1/2}}\right\}^{1/2} 2^{-1/2}\right)\\
&\leq E\left(\exp\left\{-\frac{1}{16}\frac{4u^2}{(n-1)}\sum_{t\in J_1} D_t^2 + \frac{1}{96}\frac{2^4u^4}{(n-1)^2}\sum_{t\in J_1} D_t^4\right\}\right)\\
&\leq e^{-u^2d/12},
\end{aligned}$$

since for all real x, we have $(\frac{1+\cos x}{2})^{1/2} \leq \exp\{-\frac{x^2}{16} + \frac{x^4}{96}\}$. According to Lemma 3.4 and the relation (3.9), there exist positive numbers C and κ such that, for $\log n < |u| < cn^{1/2}$, $|\varphi_n^+(u)| < Cn^{-\kappa \log n}$. This completes the proof of Theorem 2.2.

References

Does, R. J. M. M. (1982). Berry-Esséen theorems for simple linear rank statistics under the null-hypothesis, *Ann. Probab.*, **10**, 982–991.

Does, R. J. M. M. (1983). An Edgeworth expansion for simple linear rank statistics under the null hypothesis, *Ann. Statist.*, **11**, 607–642.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Wiley, New York.

Hájek, J. (1962). Asymptotically most powerful rank order test, *Ann. Math. Statist.*, **33**, 1124–1147.

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives, *Ann. Math. Statist.*, **39**, 325–346.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

Hallin, M. and Puri, M. L. (1992). Rank tests for time series analysis: A survey, *New Directions in Time Series Analysis* (eds. D. Brillinger, E. Parzen and Rosenblatt), 111–154, Springer, New York.

Hallin, M. and Rifi, Kh. (1996). The asymptotic behavior of the characteristic function of simple serial rank statistics, *Math. Methods Statist.*, **5**, 199–213.

Hallin, M. and Rifi, Kh. (1997). A Berry Esséen theorem for a serial rank statistics, *Ann. Inst. Statist. Math.*, **49** (4), 777–799.

Hallin, M., Ingenbleek, J. Fr. and Puri, M. L. (1987). Linear and quadratic serial rank tests for randomness against serial dependence, *J. Time Ser. Anal.*, **8**, 409–424.

Hušková, M. (1970). Asymptotic distribution of simple linear rank statistics for testing symetry, *Z. Wahrsch. Verw. Gebiete*, **14**, 308–322.

Hušková, M. (1977). The rate of convergence of simple linear rank statistic under hypothesis and alternatives, *Ann. Statist.*, **5**, 658–670.

Hušková, M. (1979). The Berry Esséen theorem for rank statistics, *Comment. Math. Univ. Carolin.*, **20**, 399–415.

Puri, M. L. and Seoh, M. (1984*a*). Berry-Esséen theorems for signed rank statistics with regression constants, *Limit Theorems in Probability and Statistics* (ed. P. Revéz), 875–905, Colloq. Math. Sci. János. Bolyai, **36**, North-Holland, Amsterdam.

Puri, M. L. and Seoh, M. (1984*b*). Edgeworth expansions for signed rank statistics with regression constants, *J. Statist. Plann. Inference*, **10**, 137–149; 289–309.

Puri, M. L. and Wu, T. J. (1986). The order of normal approximation for signed linear rank statistics, *Theory Probab. Appl.*, **31**, 145–151.

Seoh, M. (1983). Rate of convergence to normality and Edgeworth expansions for signed linear rank statistics with regression constants, Ph. D. Thesis, Department of Mathematics, Indiana University.

van Zwet, W. R. (1980). On the Edgeworth expansion for simple rank statistics, *Nonparametric Statistical Inference*, II (eds. B. V. Gnedenko, M. L. Puri and I. Vincze), 889–909, North Holland, Amsterdam.

von Bahr, B. (1976). Remainder estimate in a combinatorial limit theorem, *Z. Wahrsch. Verw. Gebiete*, **35**, 131–139.

Wu, T. J. (1987). An L^p error bound in normal approximation for signed linear rank statistics, *Sankhyā Ser. A*, **49**, 122–127.