The Annals of Statistics
1986, Vol. 14, No. 2, 774–780

A LARGE DEVIATION RESULT FOR SIGNED LINEAR RANK STATISTICS UNDER THE SYMMETRY HYPOTHESIS

BY TIEE-JIAN WU

University of Houston

A Cramér type large deviation theorem for signed linear rank statistics under the symmetry hypothesis is obtained. The theorem is proved for a wide class of scores covering most of the commonly used ones (including the normal scores). Furthermore, the optimal range $0 < x \le o(n^{1/4})$ can be obtained for bounded scores, whereas the range $0 < x \le o(n^{\delta})$, $\delta \in (0, \frac{1}{4})$ is obtainable for many unbounded ones. This improves the earlier result under the symmetry hypothesis in Seoh, Ralescu, and Puri (1985).

1. Introduction and statement of the main theorem. For $n \ge 1$ let X_{ni}, $i = 1, \dots, n$, be independent and identically distributed random variables distributed according to the cumulative distribution function F. We assume that

$$F \text{ is continuous and symmetric about zero.} \tag{1.1}$$

Let R_{ni} be the rank of $|X_{ni}|$ among $|X_{n1}|, \dots, |X_{nn}|$. We shall consider the signed linear rank statistics of the forms

$$S_n = \sum_{i=1}^{n} c_{ni} a_n(R_{ni}) \operatorname{sgn}(X_{ni}), \qquad n = 1, 2, \dots, \tag{1.2}$$

where $c_{n1}, \dots, c_{nn}$ are known constants, $a_n(1), \dots, a_n(n)$ are known real numbers (called scores), and $\operatorname{sgn}(x) = 1$ or -1 according as $x \ge 0$ or < 0. Under suitable assumptions on the c_{ni}'s and $a_n(i)$'s, the asymptotic normality of S_n has been established [Hŭsková (1970) and Hájek and Šidák (1967)]. Recently, Seoh, Ralescu, and Puri (1985) obtained a Cramér type large deviation theorem with range $0 < x < o(n^{1/6})$ for the statistic S_n with bounded scores (in fact, they have considered the so-called generalized rank statistics which include S_n as a special case). The purpose of this paper is twofold. In the first place we extend their assertion on the large deviation probabilities for S_n under the symmetry hypothesis to a wide class of scores covering unbounded ones (including the normal scores). Secondly, we show that under the symmetry hypothesis the optimal range $0 < x \le o(n^{1/4})$ can be obtained for bounded scores. It should be remarked that in case of symmetry the optimal range equals $0 < x \le o(n^{1/4})$, while in general the optimal range is $0 < x \le o(n^{1/6})$ [cf. Feller (1971), page 553]. We also note that Kallenberg (1982) studied the same problem in the case of (unsigned) simple linear rank statistics with bounded scores.

Received January 1985; revised July 1985.

AMS 1980 *subject classifications*. Primary 60F10; secondary 62E20.

Key words and phrases. Signed linear rank statistics, score generating function, large deviation probabilities.

Throughout the paper we make the following assumptions (n_0 is some positive integer):

$$\bar{c}_n \neq 0, \qquad \max_{1 \le i \le n} |c_{ni} - \bar{c}_n| \le A_1 |\bar{c}_n| n^{-\delta_1}, \qquad n \ge n_0, \tag{1.3}$$

where $A_1 > 0$ is an absolute constant, $\delta_1 > 0$, and $\bar{c}_n = n^{-1}\sum_{i=1}^n c_{ni}$,

$$\max_{1 \le i \le n} |a_n(i)| > 0, \qquad n \ge n_0, \qquad \Bigl(\max_{1 \le i \le n} |a_n(i)|\Bigr)\Bigl(\sum_{i=1}^n a_n^2(i)\Bigr)^{-1/2} \to 0 \tag{1.4}$$

as $n \to \infty$.

Remark 1.1. We give two examples of constants c_{ni} satisfying (1.3):

1. $c_{n1} \neq 0$ and $c_{ni} = c_{n1}$ for all $i = 1, \dots, n$ and $n = 1, 2, \dots$;
2. $|\bar{c}_n| \ge n^{-\alpha}$, $\alpha \ge 0$, $\max_{1 \le i \le n} |c_{ni} - \bar{c}_n| = O(n^{-\delta_1 - \alpha})$.

Remark 1.2. Note that (1.4) is the only assumption we are making on the scores. However, they usually are generated by a real-valued Borel measurable function $\phi(u)$, $0 < u < 1$, in either one of the following two ways:

$$a_n(i) = \phi(i/(n+1)), \qquad i = 1, \dots, n, \tag{1.5}$$

$$a_n(i) = E\phi(U_n^{(i)}), \qquad i = 1, \dots, n, \tag{1.6}$$

where $U_n^{(i)}$ denotes the ith order statistic in a random sample of size n from a uniform distribution over $(0, 1)$. Now, suppose the score generating function ϕ satisfies:

(1.7) the set $\{u\colon \phi(u) \neq 0\}$ has positive Lebesgue measure and ϕ can be expressed as a finite linear combination of monotone functions $\{\phi_1, \dots, \phi_k\}$ with $|\phi_j(u)| \le M[u(1-u)]^{-1/2+\delta_2}$ for every $j = 1, \dots, k$ and $u \in (0, 1)$, where M is a positive constant and $0 < \delta_2 \le \frac{1}{2}$.

By Theorem V.1.4a and Lemma V.1.6a of Hájek and Šidák (1967), we obtain

$$\begin{gathered} 0 < \lim_{n \to \infty} n^{-1} \sum_{i=1}^n a_n^2(i) = \|\phi\|_2^2 < \infty, \\ \max_{1 \le i \le n} |a_n(i)| = O(n^{1/2 - \delta_2}) \end{gathered} \tag{1.8}$$

for both the cases (1.5) and (1.6). Since (1.8) clearly implies (1.4), thus any score generating function satisfying (1.7) generates scores satisfying (1.4). It may be noted that unlike in earlier papers referred to above where bounded derivatives of different orders on ϕ are assumed, here (1.7) is the only assumption on ϕ we need.

Let Φ denote the standard normal distribution function. We denote for $n = 1, 2, \dots$

$$\lambda_n = \Big(\max_{1\le i\le n}|a_n(i)|\Big)\Big(\sum_{i=1}^{n} a_n^2(i)\Big)^{-1/2}, \tag{1.9}$$

$$\tau_n^2 = \bar{c}_n^2 \sum_{i=1}^{n} a_n^2(i), \tag{1.10}$$

$$\sigma_n^2 = \operatorname{var} S_n = n^{-1}\sum_{i=1}^{n} c_{ni}^2 \sum_{j=1}^{n} a_n^2(j), \tag{1.11}$$

$$b_n = \min\big(\lambda_n^{-1/2}, n^{\delta_1/2}\big). \tag{1.12}$$

Note that b_n depends only on the scores and δ_1, the magnitude of the latter depending on that of $\max_{1\le i\le n}|c_{ni}\bar{c}_n^{-1} - 1|$ [see (1.3)]. Obviously, (1.4) and (1.9) imply $\lambda_n^{-1/2} \to \infty$ and $\lambda_n^{-1/2} \le n^{1/4}$. Thus

$$b_n \to \infty \quad \text{and} \quad b_n \le \min(n^{1/4}, n^{\delta_1/2}). \tag{1.13}$$

The main result of the paper is the following:

THEOREM 1.1. *Under assumptions* (1.1) *and* (1.3)–(1.4), *we have as* $n \to \infty$ *that*

$$\sup_{x\in I_n}\big|(1 - F_n(x))(1 - \Phi(x))^{-1} - 1\big| \to 0, \tag{1.14}$$

$$\sup_{x\in I_n}\big|(1 - G_n(x))(1 - \Phi(x))^{-1} - 1\big| \to 0, \tag{1.15}$$

where F_n *and* G_n *are the cdf's of* $S_n\tau_n^{-1}$ *and* $S_n\sigma_n^{-1}$, *respectively,* I_n *denotes the interval* $(0, \rho_n b_n]$, *and* ρ_n, $n \ge 1$, *is an arbitrary sequence of positive real numbers with* $\lim_{n\to\infty}\rho_n = 0$.

REMARK 1.3. From (1.13), we get $I_n \subseteq (0, \rho_n n^{1/4}] \cap (0, \rho_n n^{\delta_1/2}]$. Let us consider the case that the scores are generated by ϕ according to either (1.5) or (1.6) with ϕ satisfying (1.7). We obtain from (1.8)–(1.9) that $A_2 n^{\delta_2/2} \le \lambda_n^{-1/2}$ for all sufficiently large n, where $A_2 > 0$ is a constant (independent of n). Thus the range I_n covers the range $0 < x \le o(n^{\delta})$, $\delta = \min(\delta_1/2, \delta_2/2)$. [Note that for unbounded ϕ we have $\delta_2 < \frac{1}{2}$ by (1.7), hence $\delta < \frac{1}{4}$ in this case.] For example, the range $0 < x \le o(n^{1/6+\delta'})$, $0 \le \delta' < \frac{1}{12}$, is obtained when $\delta_1 = \frac{1}{3} + 2\delta'$ and $\delta_2 \ge \frac{1}{3} + 2\delta'$ [in this case, $\max_{1\le i\le n}|c_{ni}\bar{c}_n^{-1} - 1| = O(n^{-1/3-2\delta'})$ and ϕ can be unbounded]. The widest range $0 < x \le o(n^{1/4})$ is obtained if and only if $\delta_1 = \delta_2 = \frac{1}{2}$ [in this case, $\max_{1\le i\le n}|c_{ni}\bar{c}_n^{-1} - 1| = O(n^{-1/2})$ and ϕ is bounded]. For the one sample normal scores test or van der Waerden's test [with $\phi(u) = \Phi^{-1}((u+1)/2)$ in (1.6) or (1.5), respectively] it holds that $(\log n)^{1/2} < \max_{1\le i\le n}|a_n(i)| = a_n(n) < (2\log n)^{1/2}$ for all sufficiently large n, which can be seen from Lemma VII.1.2 of Feller (1968) and from Section 4.4 of David (1981). It follows from (1.8), (1.9), and (1.12) that for both tests the widest possible range is $0 < x \le o(n^{1/4}(\log n)^{-1/4})$ (when $\delta_1 = \frac{1}{2}$).

To prove the theorem, we shall approximate S_n by the statistic

$$T_n = \bar{c}_n \sum_{i=1}^{n} a_n(R_{ni})\text{sgn}(X_{ni}). \tag{1.16}$$

Let $\mathbf{D} = (D_{n1}, \ldots, D_{nn})$ be the vector of antiranks associated with $\mathbf{R} = (R_{n1}, \ldots, R_{nn})$. Then T_n is equivalently expressible [in its dual form to (1.16)] as

$$T_n = \bar{c}_n \sum_{i=1}^{n} a_n(i)\text{sgn}(X_{nD_{ni}}). \tag{1.17}$$

But $\text{sgn}(X_{nD_{n1}}), \ldots, \text{sgn}(X_{nD_{nn}})$ are independent and identically distributed r.v.'s under assumption (1.1) with common symmetric Bernoulli distribution [see Theorem 19C of Hâjek (1969)]. Therefore T_n is actually expressible as a sum of discrete independent r.v.'s. A Cramér type large deviation theorem is applied to T_n, whereas a multinomial expansion is made use of to estimate the distance $E|S_n - T_n|^{2p}$ for any $p \in [1, n]$. In the sequel we suppress the index n whenever it is possible.

2. Some lemmas and the proof of the main theorem.

The following lemma deals with the large deviation probabilities of T_n.

LEMMA 2.1. *Under the assumption of Theorem* 1.1, *it holds true as* $n \to \infty$ *that*

$$\sup_{x \in I_n} \left|(1 - H_n(x_n))(1 - \Phi(x))^{-1} - 1\right| \to 0,$$

where $|x_n - x| = b_n^{-1}$, *and* H_n *denotes the cdf of* $T_n\tau_n^{-1}$.

PROOF. From (1.1), (1.3), (1.9)–(1.10), and (1.17), we get $\text{Var}\, T_n = \tau_n^2$ and for all n

$$\left|\bar{c}_n a_n(i)\text{sgn}(X_{D_i})\tau_n^{-1}\right| \le \lambda_n, \qquad i = 1, \ldots, n. \tag{2.1}$$

It then follows from (1.4), (1.9), (1.12), (2.1), and from Theorem 1 of Feller (1943) that there exist $n_1 > 0$ such that for all $n \ge n_1$, and $x \in (b_n^{-1}, \rho_n b_n]$

$$1 - H_n(x_n) = \exp[-2^{-1}x_n^2 Q_n(x_n)]\{1 - \Phi(x_n) + \theta_n\lambda_n\exp(-2^{-1}x_n^2)\}, \tag{2.2}$$

where

$$\begin{aligned} &|\theta_n| < 9, \qquad Q_n(x_n) = \sum_{i=1}^{\infty} q_{ni}x_n^i, \\ &q_{n1} = 0, \qquad |q_{ni}| < 7^{-1}(12\lambda_n)^i, \qquad i = 2, 3, \ldots . \end{aligned} \tag{2.3}$$

Note that $q_{n1} = 0$ because the third moment of $\bar{c}_n a_n(i)\text{sgn}(X_{D_i})$ vanishes for all $i = 1, \ldots, n$ under assumption (1.1). Clearly (2.3) implies for all sufficiently large n that

$$\left|x_n^2 Q_n(x_n)\right| \le 7^{-1}(144)(\lambda_n x_n^2)^2(1 - 12\lambda_n x_n)^{-1}, \tag{2.4}$$

which converges to zero as $n \to \infty$ uniformly in $x \in (b_n^{-1}, \rho_n b_n]$. Now, by Lemma VII.1.2 of Feller (1968), it can be readily seen that uniformly in $x \in (b_n^{-1}, \rho_n b_n]$

$$\lambda_n \exp(-2^{-1}x_n^2)(1 - \Phi(x))^{-1} \to 0, \tag{2.5}$$

$$(1 - \Phi(x_n))(1 - \Phi(x))^{-1} \to 1 \tag{2.6}$$

as $n \to \infty$. Combining (2.2)–(2.6) yields

$$\left|(1 - H_n(x_n))(1 - \Phi(x))^{-1} - 1\right| \to 0 \tag{2.7}$$

uniformly in $x \in (b_n^{-1}, \rho_n b_n]$. Next, by (1.1), (1.4), (1.17), and by Theorem V.1.2 of Hájek and Šidák (1967), we have $\|H_n - \Phi\|_\infty \to 0$, which implies uniformly in $x \in (0, b_n^{-1}]$

$$\left|(1 - H_n(x_n))(1 - \Phi(x))^{-1} - 1\right| \to 0 \tag{2.8}$$

as $n \to \infty$. The proof follows from (2.7) and (2.8) immediately. □

The following lemma gives us an upper bound for the distance $E|S_n - T_n|^{2p}$.

LEMMA 2.2. *Under the assumptions of Theorem* 1.1, *for all $n \geq n_0$ and real $p \in [1, n]$,*

$$E|(S_n - T_n)|^{2p} \leq A_3^p p^p n^{-2p\delta_1} \tau_n^{2p}, \tag{2.9}$$

where $A_3 > 0$ is an absolute constant.

PROOF. By Hölder's inequality, it is sufficient to prove (2.9) only for $p = 1, 2, \ldots, n$. The dual form of S_n is $S_n = \sum_{i=1}^n c_{D_i} a_n(i)\operatorname{sgn}(x_{D_i})$. Thus, in view of (1.17), we get $S_n - T_n = \sum_{i=1}^n a_n(i)(c_{D_i} - \bar{c}_n)\operatorname{sgn}(X_{D_i})$. Furthermore, let $\{p_1, \ldots, p_n\}$ be an arbitrary collection of nonnegative integers containing at least one odd number, then (1.1) implies $E(\prod_{i=1}^n W_i^{p_i}) = 0$, where $W_i = (c_{D_i} - \bar{c}_n)\operatorname{sgn}(X_{D_i})$. It follows, by using the multinomial expansion, that for any $p = 1, \ldots, n$

$$E(S_n - T_n)^{2p} = \sum_{m=1}^{p} \sum_{\mathbf{i}_m \in A(m)} \sum_{\mathbf{p}_m \in B(m)} C_{2\mathbf{p}_m}^{2p} \prod_{j=1}^{m} \left(a_n(i_j)\right)^{2p_j} E\left(\prod_{j=1}^{m} W_{i_j}^{2p_j}\right), \tag{2.10}$$

where

$$\mathbf{i}_m = (i_1, \ldots, i_m), \qquad \mathbf{p}_m = (p_1, \ldots, p_m),$$

$$A(m) = \{\mathbf{i}_m : 1 \leq i_1 < \cdots < i_m \leq n\},$$

$$B(m) = \left\{\mathbf{p}_m : \sum_{j=1}^{m} p_j = p;\ p_j = 1, \ldots, p \text{ for each } j\right\},$$

and

$$C_{2\mathbf{p}_m}^{2p} = (2p)!((2p_1)! \cdots (2p_m)!)^{-1}.$$

Using the multinomial expansion again, we have

$$\left(\sum_{i=1}^{n} a_n^2(i)\right)^p = \sum_{m=1}^{p} \sum_{\mathbf{i}_m \in A(m)} \sum_{\mathbf{p}_m \in B(m)} C_{\mathbf{p}_m}^p \prod_{j=1}^{m} \left(a_n(i_j)\right)^{2p_j}, \tag{2.11}$$

where $C_{\mathbf{p}_m}^p = p!((p_1!)\dots(p_m!))^{-1}$. Note also that

$$C_{2\mathbf{p}_m}^{2p} \le (2p)^p C_{\mathbf{p}_m}^p, \qquad p = 1,\dots,n, \qquad \mathbf{p}_m \in B(m). \tag{2.12}$$

The proof follows from (1.3), (1.10), and (2.10)–(2.12) quickly. □

Proof of Theorem 1.1. We get by standard arguments that

$$-Q_n + \left(1 - H_n(x + b_n^{-1})\right) \le 1 - F_n(x) \le \left(1 - H_n(x - b_n^{-1})\right) + Q_n, \tag{2.13}$$

where $Q_n = P[|S_n - T_n| > b_n^{-1}\tau_n]$. Put $p_n = A_3^{-1}e^{-1}b_n^2$. Then (1.13) implies that $p_n \in [1, n]$ for all sufficiently large n. It follows from Markov's inequality, Lemma 2.2, and (1.12) that

$$Q_n \le \left(A_3 p_n b_n^2 n^{-2\delta_1}\right)^{p_n} \le e^{-p_n}, \tag{2.14}$$

which, together with Lemma VII.1.2 of Feller (1968), imply uniformly in $x \in I_n$

$$Q_n(1 - \Phi(x))^{-1} \le e^{-p_n}(1 - \Phi(\rho_n b_n))^{-1} \to 0 \tag{2.15}$$

as $n \to \infty$. (1.14) may be concluded from (2.13), (2.15), and Lemma 2.1 immediately. Next, from (1.3)–(1.4) and (1.10)–(1.11) it follows for all sufficiently large n that

$$\left|\sigma_n\tau_n^{-1} - 1\right| \le \tau_n^{-2}|\sigma_n^2 - \tau_n^2| \le \max_{1 \le i \le n} |c_i\bar{c}_n^{-1} - 1|^2 \le A_1^2 n^{-2\delta_1}. \tag{2.16}$$

By (1.12), (1.14), (2.16), and by Lemma VII.1.2 of Feller (1968), we obtain uniformly in $x \in I_n$

$$1 - G_n(x) = 1 - F_n\left(\sigma_n\tau_n^{-1}x\right) = \left[1 - \Phi\left(\sigma_n\tau_n^{-1}x\right)\right](1 + o(1)), \tag{2.17}$$

$$1 - \Phi\left(\sigma_n\tau_n^{-1}x\right) = [1 - \Phi(x)](1 + o(1)) \tag{2.18}$$

as $n \to \infty$. Now (2.17)–(2.18) imply (1.15). This completes the proof. □

Acknowledgment. The author wishes to thank the referee and the associate editor for their helpful suggestions which improved the presentation of the paper.

REFERENCES

David. H. A. (1981). *Order Statistics.* 2nd ed. Wiley, New York.

Feller, W. (1943). Generalization of a probability limit theorem of Cramér. *Trans. Amer. Math. Soc.* **54** 361–372.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* **1.** Wiley, New York.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* **2.** Wiley, New York.

Hájek, J. (1969). *A Course in Nonparametric Statistics.* Holden-Day, San Francisco.

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

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

KALLENBERG, W. C. M. (1982). Cramér type large deviations for simple linear rank statistics. *Z. Wahrsch. verw. Gebiete* **60** 403–409.

SEOH, M., RALESCU, S. S. and PURI, M. L. (1985). Cramér type large deviations for generalized rank statistics. *Ann. Probab.* **13** 115–125.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON, UNIVERSITY PARK
HOUSTON, TEXAS 77004