The Annals of Statistics
1976, Vol. 4, No. 4, 771–774

L_1 RATES OF CONVERGENCE FOR LINEAR RANK STATISTICS

BY R. V. ERICKSON AND H. L. KOUL

Michigan State University

This paper gives rates of convergence for the L_1 distances between the distributions of standardized linear rank statistics and the standard normal random variable. These rates are $O(N^{-\frac{1}{2}})$ under various conditions on the score function and the distributions of the underlying observations.

1. Introduction. This paper gives rates of convergence for the L_1 distances between the distributions of standardized linear rank statistics and the standard normal random variable. For score function φ with bounded derivative φ' we get the rate $O(N^{-\frac{1}{2}})$ under the null hypothesis (see Theorem 2.1), and if φ'' is bounded the rate is also $O(N^{-\frac{1}{2}})$ under the general alternative (see Theorem 2.2).

In Section 3 we mention a result which almost extends Theorem 2.1 to include φ equal inverse normal. We also mention implications of L_1 rates of convergence. Proofs are given in Section 4.

2. Notation and statement of results. Suppose given an array (X_{Nk}, F_{Nk}) where (in all that follows) $N = 1, 2, \cdots, k = 1, \cdots, N$. Say that this array depicts the *general alternative* if for each N

$$X_{N1}, \cdots, X_{NN} \quad \text{are independent,}$$

and

$$\text{all} \quad F_{Nk}(x) = P(X_{Nk} \leqq x) \quad \text{are continuous;}$$

the *null hypothesis* is satisfied if in addition for each N

$$X_{N1}, \cdots, X_{NN} \quad \text{are identically distributed.}$$

A simple linear rank statistic corresponding to a real valued score function φ on (0, 1) and regression constants $\{c_{Ni}, 1 \leqq i \leqq N\}$ is defined as

$$S \equiv S(N, \varphi, c) = \sum_1^N c_{Ni}\varphi(R_{Ni}/(N+1))$$

where R_{Ni} is the rank of X_{Ni} among $\{X_{Nj}, 1 \leqq j \leqq N\}$. Because everything depends on the row index N, there will be no harm if we suppress this subscript.

Introduce the notation

$$\bar{c} = \sum_1^N c_{Nk}/N = \sum_1^N c_k/N ,$$

$$\sigma_c^{\,2} = \sum_1^N (c_k - \bar{c})^2 , \qquad M_c^{\,2} = \max_{1 \leqq k \leqq N} (c_k - \bar{c})^2$$

$$\sigma_\varphi^{\,2} = \int_0^1 (\varphi(u) - \bar{\varphi})^2\, du , \qquad \bar{\varphi} = \int_0^1 \varphi(u)\, du ,$$

$$\sigma_S^{\,2} = \operatorname{Var} S , \qquad S^{\sim} = S - ES ,$$

$$G_N(x) = P(S^{\sim} \leqq x\sigma_S) , \qquad \underline{n}(x) = (2\pi e^{x^2})^{-\frac{1}{2}} , \qquad \mathscr{N}(x) = \int_{-\infty}^x \underline{n}(y)\, dy .$$

Received October 1974; revised January 1976.
AMS 1970 *subject classifications*. Primary 60F05; Secondary 62G99.
Key words and phrases. Linear rank statistics, L_1 rates for asymptotic normality.

Always assume $\sigma_c^2\sigma_\varphi^2 > 0$. Also let $\mathscr{N}$ be the random variable with distribution $\mathscr{N}$.

We often make the assumption

$$\mathbf{c}^*: \text{there is an absolute constant } K \text{ such that } N^{\frac{1}{2}}M_c \leqq K\sigma_c, \quad \text{all } N \geqq 1.$$

From now on K denotes a constant, which may depend on φ and c's but never on F's or N. The value of K varies with usage.

Under various hypotheses on φ we get L_1 rates of convergence to normality.

THEOREM 2.1. *Assume* $\mathbf{c}^*$, *the null hypothesis and that* φ *has a bounded derivative* φ'. *Then*

$$||G_N - \mathscr{N}||_1 \leqq KN^{-\frac{1}{2}}.$$

THEOREM 2.2. *Assume the general alternative and that* φ *has a bounded second derivative* φ''. *Then*

$$||G_N - \mathscr{N}||_1 \leqq KN^{-\frac{1}{2}},$$

under appropriate variance conditions (see (4.4)).

From the proofs of these theorems it will be clear that similar rates for asymptotic normality can be given for signed rank statistics. For Theorem 2.2 one uses Hušková (1970) in place of Hájek (1968).

3. Remarks. We state, without proof, two additional results, and we mention some implications of L_1 rates.

Assume $\mathbf{c}^*$, the null hypothesis, and that φ is nondecreasing and has two derivatives, $\varphi^{(1)}$ and $\varphi^{(2)}$, in (0, 1). By arguing as in Section 7 of Chernoff and Savage (1958) we can prove

(a) if $|\varphi^{(k)}(u)| \leqq K|u(1-u)|^{-\alpha-k}$, for all u in (0, 1), $k = 1, 2$ and some $\alpha < 0$, then $||G_N - \mathscr{N}||_1 \leqq KN^{-\frac{1}{2}}$, and

(b) if $\varphi = \mathscr{N}^{-1}$ then $||G_N - \mathscr{N}||_1 \leqq KN^{-\frac{1}{2}} \log (N+1)$.

The proofs of (a) and (b) are omitted since result (b) indicates a lack of strength in the technique.

Let us now point out that L_1 rates imply L_∞ rates. More precisely:

$$||G_N - \mathscr{N}||_\infty \leqq (4||G_N - \mathscr{N}||_1/5)^{\frac{1}{2}}.$$

To see this, notice that if $||G_N - \mathscr{N}||_\infty = h$, then it is possible to draw a right triangle with height h and base $5h/2$ between the graphs of G_N and $\mathscr{N}$. This is because $\mathscr{N}$ has maximum slope $(2\pi)^{-\frac{1}{2}} < \frac{2}{5}$. It is unfortunate that the square root spoils this L_∞ bound.

One way to use an L_1 bound is to contaminate both the standardized statistic S, write S^*, and the standardized normal $\mathscr{N}$ with an independent disturbance D which has bounded density function f. Then

$$\sup_x |P(S + D \leqq x) - P(\mathscr{N} + D \leqq x)| \leqq ||f||_\infty ||G_N - \mathscr{N}||_1.$$

Finally, it follows from (4.1 c) below that

$$|E|S^*| - (2/\pi)^{\frac{1}{2}}| \leqq ||G_N - \mathscr{N}||_1 .$$

4. Proofs. Our basic technique is to replace S by a sum of independent variables and to bound the error made by so doing.

LEMMA 4.1. *Let V and W have respective distributions G and H. Then*

(a) $||G - \mathscr{N}||_1 \leqq ||H - \mathscr{N}||_1 + (E|V - W|^2)^{\frac{1}{2}}$,
(b) $|EV^+ - E\mathscr{N}^+| \leqq ||G - \mathscr{N}||_1$,
(c) $|E|V| - E|\mathscr{N}|| \leqq ||G - \mathscr{N}||_1$.

PROOF. For (a) use the L_1 triangle inequality and the bound

$$\begin{aligned} ||G - H||_1 &= \textstyle\int_{-\infty}^{\infty} |EI(V > x) - EI(W > x)|\, dx \\ &\leqq E \textstyle\int_{-\infty}^{\infty} |I(V > x) - I(W > x)|\, dx = E|V - W| \leqq (E|V - W|^2)^{\frac{1}{2}} . \end{aligned}$$

Similarly,

$$EV^+ = \textstyle\int_0^{\infty} |1 - P(V \leqq x)|\, dx \leqq E\mathscr{N}^+ + \textstyle\int_0^{\infty} |P(V \leqq x) - P(\mathscr{N} \leqq x)|\, dx ,$$

giving (b), (c). □

(4.2) PROOF OF THEOREM 2.1. Let F denote the common cdf of $\{X_i, 1 \leqq i \leqq n\}$ and let $U_i = F(X_i)$, $1 \leqq i \leqq N$. Introduce

$$T_i = (c_i - \bar{c})[\varphi(U_i) - \bar{\varphi}] , \qquad T = \textstyle\sum_1^N T_i , \qquad \sigma_T^2 = \operatorname{Var}(T) = \sigma_c^2\sigma_\varphi^2 .$$

The idea of the proof is to use (4.1 a) with $V = S^\sim/\sigma_S$, $W = T/\sigma_T$, to use known L_1 rates for $T/\sigma_T \to \mathscr{N}$ and to find the right bounds for $||S^\sim/\sigma_S - T/\sigma_T||_2$.

If H_N is the distribution of T/σ_T then Erickson ((1974), Theorem A) has shown that

$$||H_N - \mathscr{N}||_1 \leqq 13 \textstyle\sum_1^N (E|T_k/\sigma_T|^3 + E|T_k/\sigma_T|^4) .$$

Since $\sum E|T_k|^3 \leqq \sum |c_k - \bar{c}|^3 \int_0^1 |\varphi(u) - \bar{\varphi}|^3\, du \leqq 2||\varphi||_\infty M_c \sigma_T^2$ and $\sum E|T_k|^4 \leqq (2||\varphi||_\infty M_c \sigma_T)^2$, **c*** implies

$$||H_N - \mathscr{N}||_1 \leqq KN^{-\frac{1}{2}} .$$

It remains to bound $d^2 = E(S^\sim/\sigma_S - T/\sigma_T)^2$. From the Schwarz inequality it follows that $d^2 \leqq 4E(S^\sim - T)^2/\sigma_T^2$. Next recall (Hájek–Šidák, 1967, page 160) $E(S^\sim - T)^2 \leqq 2\sigma_c^2E(\varphi(R) - \varphi(U_1))^2 \leqq 2\sigma_c^2||\varphi'||_\infty^2E(R - U_1)^2$, $R = R_1/(N + 1)$. Finally, given U_1, $R_1 - 1$ is binomial $(N - 1, U_1)$, which implies $E(R - U_1)^2 = (6N + 6)^{-1}$ and $d \leqq 2||\varphi'||_\infty/\sigma_\varphi N^{\frac{1}{2}} < K/N^{\frac{1}{2}}$. □

(4.3) PROOF OF THEOREM 2.2. For this proof argue exactly as above but with T replaced by $Z = \sum_1^N Z_k$, where $Z_k = N^{-1} \sum_j (c_j - \bar{c}) \int_{-\infty}^{\infty} [I(X_k \leqq x) - F_k(x)]\varphi'(H(x))F_j(dx)$, and $H = N^{-1} \sum_1^N F_j$. Note that $|Z_k| \leqq M_c||\varphi'||_\infty$. From Theorem 4.2 of Hájek (1968) we have $E(S^\sim - Z)^2 \leqq K\sigma_c^2/N$. As above

$$||G_N - \mathscr{N}||_1 \leqq 13[2||\varphi'||_\infty M_c \sigma_Z^{-1} + 4||\varphi'||_\infty^2 M_c^2 \sigma_Z^{-2}] + 2KN^{-\frac{1}{2}}\sigma_c/\sigma_Z ,$$

and the conclusion of the theorem follows provided

$$NM_c^2 \leqq K\sigma_Z^2, \qquad \text{all } N, \tag{4.4}$$

since then $\sigma_c^2/\sigma_Z^2 \leqq NM_c^2/\sigma_Z^2 \leqq K$. □

REFERENCES

CHERNOFF, H. and SAVAGE, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics. *Ann. Math. Statist.* **29** 972–994.

ERICKSON, R. V. (1974). L_1 bounds for asymptotic normality of m-dependent sums using Stein's technique. *Ann. Probability* **2** 522–529.

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.

HUŠKOVÁ, M. (1970). Asymptotic distribution of simple linear rank statistics for testing symmetry. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **14** 308–322.

DEPARTMENT OF STATISTICS AND PROBABILITY
WELLS HALL
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824