*The Annals of Statistics*
1973, Vol. 1, No. 1, 153–160

# CONVERGENCE RATES FOR *U*-STATISTICS AND RELATED STATISTICS

BY WILLIAM F. GRAMS[1] AND R. J. SERFLING

*The Florida State University*

Bounds are provided for the rates of convergence in the central limit theorem and the strong law of large numbers for *U*-statistics. The results are obtained by establishing suitable bounds upon the moments of the difference between a *U*-statistic and its projection. Analogous conclusions for the associated von Mises statistical functions are indicated. Statistics considered for exemplification are the sample variance and the Wilcoxon two-sample statistic.

**1. Introduction.** The data consists of $c$ independent collections of independent observations $\{X_1^{(1)}, \cdots, X_{n_1}^{(1)}\}, \cdots, \{X_1^{(c)}, \cdots, X_{n_c}^{(c)}\}$ taken from distributions $F_1, \cdots, F_c$, respectively. Consider a parametric function $\theta = \theta(F_1, \cdots, F_c)$ for which there is an unbiased estimator. That is,

$$\theta = Eh(X_1^{(1)}, \cdots, X_{m_1}^{(1)}; \cdots; X_1^{(c)}, \cdots, X_{m_c}^{(c)}) \tag{1.0}$$

for some function $h$ which will be assumed, without loss of generality, to be *symmetric* within each of its $c$ blocks of arguments. Corresponding to the "kernel" $h$, and assuming $n_1 \geqq m_1, \cdots, n_c \geqq m_c$, the *U-statistic* for estimation of $\theta$ is obtained by averaging $h$ symmetrically over the data:

$$U = \prod_{j=1}^c \binom{n_j}{m_j}^{-1} \sum_c h(X_{i_{11}}^{(1)}, \cdots, X_{i_{1m_1}}^{(1)}; \cdots; X_{i_{c1}}^{(c)}, \cdots, X_{i_{cm_c}}^{(c)}) . \tag{1.1}$$

Here $\{i_{j1}, \cdots, i_{jm_j}\}$ denotes a set of $m_j$ distinct elements of the set $\{1, 2, \cdots, n_j\}$, $1 \leqq j \leqq c$, and $\sum_c$ denotes summation over all such combinations.

The one-sample ($c = 1$) $U$-statistics were introduced by Hoeffding [10] and a central limit theorem (CLT) covering a wide class of such statistics was proved. The treatment was generalized for $c \geqq 1$ by Lehmann [13] and Dwass [4]. In [11] Hoeffding proved the strong law of large numbers (SLLN) for $U$-statistics ($c = 1$). Later Berk [1] gave a different argument, exploiting the reverse martingale character of a sequence of one-sample $U$-statistics.

It is the purpose of the present paper to exhibit rates of convergence apropos to these limit theorems. Our method is to approximate $U$ by its *projection*,

$$\hat{U} = \sum_{j=1}^c \sum_{i=1}^{n_j} E(U \mid X_i^{(j)}) - (N-1)\theta , \tag{1.2}$$

where $N = n_1 + \cdots + n_c$. (See Hájek [8] for exposition of the notion of projection of a statistic upon the basic observations.) Since the summands of $\hat{U}$ are *independent*, it may be dealt with by standard theory. We then infer conclusions

Received January 31, 1972; revised June 13, 1972.

[1] Research of first author supported by the National Institute of General Medical Sciences under Training Grant 5-T01-GM00913-10 and presented as a portion of a dissertation in partial fulfillment of the requirements for the Ph. D. degree in Statistics at Florida State University

about $U$ by showing that $U - \hat{U}$ is negligible. This is accomplished by establishing, in Section 2, suitable bounds on the moments $E|U - \hat{U}|^a$. Application is made to the CLT in Section 3 and to the SLLN in Section 4. The main results of the paper are Theorems 2.1, 3.1 and 4.1.

It is shown in Section 5 that analogous results hold for the associated von Mises statistic, which is given by replacing $F_1, \cdots, F_c$ by the respective sample df's in the formulation of $\theta$.

Many familiar statistics are of the $U$-type. See [10] and [13] for examples. We examine the sample variance and the Wilcoxon two-sample statistic in Section 6.

Finally, certain generalizations and further notions are discussed in Section 7.

**2. The moments of $U - \hat{U}$.** From (1.1) and (1.2) we readily obtain

$$\hat{U} - \theta = \sum_{j=1}^{c} \sum_{i=1}^{n_j} \frac{m_j}{n_j} h_j^*(X_i^{(j)}) , \tag{2.1}$$

where

$$h_j^*(x) = E[h(X_1^{(1)}, \cdots, X_{m_1}^{(1)}; \cdots; X_1^{(c)}, \cdots, X_{m_c}^{(c)}) \mid X_1^{(j)} = x] - \theta . \tag{2.2}$$

It follows that $\hat{U} - \theta$ may be expressed in the form (1.1) with the role of kernel played by

$$g(X_{i_{11}}^{(1)}, \cdots, X_{i_{1m_1}}^{(1)}; \cdots; X_{i_{c1}}^{(c)}, \cdots, X_{i_{cm_c}}^{(c)}) = \sum_{j=1}^{c} \sum_{k=1}^{m_j} h_j^*(X_{i_{jk}}^{(j)}) . \tag{2.3}$$

Define now a further "kernel" $H$ by

$$H = h - g - \theta , \tag{2.4}$$

and we have

$$U - \hat{U} = \prod_{j=1}^{c} \binom{n_j}{m_j}^{-1} \sum_c H(X_{i_{11}}^{(1)}, \cdots; \cdots; \cdots, X_{i_{cm_c}}^{(c)}) . \tag{2.5}$$

Thus $U - \hat{U}$ is of the form (1.1) with kernel $H$. Note that $E[H] = 0$ and $E[H \mid X_{i_{jk}}^{(j)}] = 0$.

Let $n = \min_{1 \leq i \leq c} n_i$.

THEOREM 2.1. *Let $r$ be a positive integer. If $E[H^{2r}] < \infty$ (implied by $E[h^{2r}] < \infty$), then*

$$E(U - \hat{U})^{2r} = O(n^{-2r}) , \qquad n \to \infty . \tag{2.6}$$

PROOF. By (2.5), the quantity in (2.6) may be written

$$\prod_{j=1}^{c} \binom{n_j}{m_j}^{-2r} \sum E[\prod_{a=1}^{2r} H(X_{i_{a11}}^{(1)}, \cdots; \cdots; \cdots, X_{i_{acm_c}}^{(c)})] , \tag{2.7}$$

where the indices are as in (1.1), with the additional suffix "$a$" identifying the factor within the product, and $\sum$ denotes summation over all $\prod_{j=1}^{c} \binom{n_j}{m_j}^{2r}$ of the indicated terms. Clearly, the hypothesis of the theorem implies that $E(U - \hat{U})^{2r} < \infty$.

Let $M = m_1 + \cdots + m_c$ and consider a typical term in the sum in (2.7). If all $M$ indices occurring in one of the factors inside the expectation occur only in that factor, then the independence of that factor from the other factors implies that the product of factors has expectation zero. If $(M-1)$ of the indices in one of the factors occur only in that factor, then again the product of factors has expectation zero, since the conditional expectation, given all variables but the $(M-1)$ designated, is zero. Hence a term in (2.7) may have nonzero expectation only if each factor in the product contains at least two indices which appear in other factors of the product.

For the $a$th factor and the $j$th sample, let $q_a^{(j)}$ be the number of indices not repeated in other factors, and let $p_a^{(j)} = m_j - q_a^{(j)}$ be the number of indices repeated elsewhere. Among the repeated indices within the $j$th sample, let $q_0^{(j)}$ be the number of distinct elements. Then clearly

$$2q_0^{(j)} \leqq \sum_{a=1}^{2r} p_a^{(j)} . \tag{2.8}$$

The number of ways of selecting these $\sum_{a=0}^{2r} q_a^{(j)}$ indices for the $j$th sample is, making use of (2.8), of order

$$O(n_j^{\sum_{a=0}^{2r} q_a^{(j)}}) = O(n_j^{2rm_j - \frac{1}{2}\sum_{a=1}^{2r} p_a^{(j)}}) = O(n_j^{2rm_j} n^{-\frac{1}{2}\sum_{a=1}^{2r} p_a^{(j)}}) . \tag{2.9}$$

Here the implicit constants depend upon $r$ and $m_1, \cdots, m_c$, but not upon $n_1, \cdots, n_c$. This is true also of the number of ways of selecting the values $q_0^{(j)}, \cdots, q_{2r}^{(j)}$. Therefore the *overall* number of ways of selecting the indices for the $j$th sample is of order given by the right-most term in (2.9). It follows that the number of terms in the sum in (2.7) for which the expectation is possibly nonzero is of order

$$O(\prod_{j=1}^{c} n_j^{2rm_j} n^{-\frac{1}{2}\sum_{j=1}^{c}\sum_{a=1}^{2r} p_a^{(j)}}) = O(\prod_{j=1}^{c} n_j^{2rm_j} n^{-2r}) , \tag{2.10}$$

since $\sum_{j=1}^{c} p_a^{(j)} \geqq 2$. Thus (2.6) follows. □

The case $r = 1$ of Theorem 2.1 was proved by Hoeffding [10] and suffices for applications such as the CLT and SLLN. For information on the *rates* of these convergences, however, the generalization for $r > 1$ is relevant.

**3. Rate of convergence in the** CLT. The variance of the projection $\hat{U}$ is found from (2.1) to be

$$\sigma^2(\hat{U}) = \sum_{i=1}^{c} m_i^2 \zeta_i / n_i , \tag{3.1}$$

where $\zeta_i = \mathrm{Var}\,[h_i^*(X_1^{(i)})]$. Asymptotic normality theorems [10], [4], [13] for $U$-statistics state that

$$P[(U-\theta)/\sigma(\hat{U}) \leqq t] \to \Phi(t) , \qquad n \to \infty , \tag{3.2}$$

where $\Phi(t) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{t} \exp(-u^2/2)\, du$, $n = \min\{n_1, \cdots, n_c\}$ as previously, and it is assumed that $Eh^2 < \infty$ and $n\sigma^2(\hat{U}) \geqq B > 0$ as $n \to \infty$.

The rate of convergence in (3.2) is seen in the theorem below to satisfy a bound which improves with the order of the moments that may be assumed on $h$ (or $H$). If moments of all orders may be assumed, the bound may be brought "close" to the order $O(n^{-\frac{1}{2}})$, which in view of the Berry–Esséen theorem [3] is the best possible without specific assumptions on the underlying distributions $F_1, \cdots, F_c$ of $X_1^{(1)}, \cdots, X_1^{(c)}$. Thus, e.g., regarding the two-sample Wilcoxon statistic, which has a bounded kernel, we are able to corroborate remarks of Stoker [17] on improvement of the order $O(n^{-\frac{2}{5}})$ which he obtained.

THEOREM 3.1. *Assume that $h$ satisfies*

$$E|h_i^*(X_1^{(i)})|^3 < \infty\,, \qquad 1 \leqq i \leqq c \ (\textit{implied by } E|h|^3 < \infty) \tag{3.3}$$

*and*

$$n \textstyle\sum_{i=1}^{c} m_i^2\zeta_i/n_i \geqq B > 0\,, \qquad n \to \infty\,. \tag{3.4}$$

*If, further, $EH^{2r} < \infty$ (implied by $Eh^{2r} < \infty$) for a positive integer $r$, then*

$$\sup_t |P[(U - \theta)/\sigma(\hat{U}) \leqq t] - \Phi(t)| = O(n^{-r/(2r+1)})\,, \qquad n \to \infty\,. \tag{3.5}$$

PROOF. We will apply a standard device (see, e.g., [2]). Namely, if

$$\sup_t |P[(\hat{U} - \theta)/\sigma(\hat{U}) \leqq t] - \Phi(t)| = O(a_n)\,, \qquad n \to \infty\,, \tag{3.6}$$

for a sequence of constants $\{a_n\}$, then

$$\begin{aligned} \sup_t |P[(U - \theta)/\sigma(\hat{U}) \leqq t] - \Phi(t)| \\ = O(a_n) + P[|U - \hat{U}|/\sigma(\hat{U}) > a_n]\,, \qquad n \to \infty\,. \end{aligned} \tag{3.7}$$

The proof is elementary.

Now, by the classical Berry–Esséen theorem, as stated in Loève ([15], page 288), it follows directly from (2.1) that

$$\sup_t |P[(\hat{U} - \theta)/\sigma(\hat{U}) \leqq t] - \Phi(t)| \leqq C\, \frac{\sum_{j=1}^{c} \sum_{i=1}^{n_j} (m_j/n_j)^3 E|h_j^*(X_i^{(j)})|^3}{\sigma^3(\hat{U})} \tag{3.8}$$

where $C$ is a universal constant. It is checked easily that the RHS of (3.8) is $O(n^{-\frac{1}{2}})$ as $n \to \infty$ subject to (3.4).

Consequently, for any sequence of constants $a_n$ satisfying $n^{-\frac{1}{2}} = O(a_n)$ as $n \to \infty$, we have (3.6) and thus (3.7). We now utilize Theorem 2.1 in selecting the best sequence $\{a_n\}$. By (3.1), (3.4) and Markov's inequality, we have

$$P[|U - \hat{U}|/\sigma(\hat{U}) > a_n] \leqq n^r a_n^{-2r} B^{-r} E(U - \hat{U})^{2r}\,. \tag{3.9}$$

By Theorem 2.1, the RHS of (3.9) is $O(n^{-r}a_n^{-2r})$. Setting $a_n = O(n^{-r}a_n^{-2r})$, we obtain $a_n = O(n^{-r/(2r+1)})$. □

COROLLARY 3.1. *Assume that $h$ has finite moments of all orders and that* (3.4) *holds. Then, for every $\varepsilon > 0$,*

$$\sup_t |P[(U - \theta)/\sigma(\hat{U}) \leqq t] - \Phi(t)| = O(n^{-\frac{1}{2}+\varepsilon})\,, \qquad n \to \infty\,. \tag{3.10}$$

**4. Rate of convergence in the SLLN.** In this section we consider one-sample $U$-statistics ($c = 1$). The following lemma will be required.

LEMMA 4.1 (Katz–Baum [9]). *Let $\xi_1, \xi_2, \cdots$ be* i.i.d. *random variables. If*

$r > 1$, *the following are equivalent*:

$$P[|\xi_1| > n] = O(n^{-r}) \quad \text{and} \quad E\xi_1 = \mu\,; \tag{4.1}$$

$$P[|n^{-1} \textstyle\sum_1^n \xi_i - \mu| > \varepsilon] = O(n^{1-r})\,, \qquad \textit{for each } \varepsilon > 0\,; \tag{4.2}$$

$$P[\sup_{k \geqq n} |k^{-1} \textstyle\sum_1^k \xi_i - \mu| > \varepsilon] = O(n^{1-r})\,, \qquad \textit{for each } \varepsilon > 0\,. \tag{4.3}$$

A corollary of this lemma is that if $E|\xi_1|^r < \infty$, then (4.3) holds. This corollary is generalized to $U$-statistics in the following result.

THEOREM 4.1. *Let $\{U_n\}$ be the sequence of U-statistics generated by a kernel h applied to a sequence of observations $\{X_i\}$. Assume $Eh^{2r} < \infty$ for a positive integer r. Then, for any $\varepsilon > 0$,*

$$P[\sup_{k \geqq n} |U_k - \theta| > \varepsilon] = O(n^{1-2r})\,, \qquad n \to \infty\,. \tag{4.4}$$

PROOF. Let $\varepsilon > 0$ be given. Then

$$\begin{aligned} P[\sup_{k \geqq n} |U_k - \theta| > \varepsilon] \\ \leqq P\left[\sup_{k \geqq n} |U_k - \hat{U}_k| > \frac{\varepsilon}{2}\right] + P\left[\sup_{k \geqq n} |\hat{U}_k - \theta| > \frac{\varepsilon}{2}\right]. \end{aligned} \tag{4.5}$$

Since $\hat{U}_k - \theta = k^{-1} \sum_{j=1}^k m_1 h_1^*(X_j)$, the right-most term in (4.5) is $O(n^{1-2r})$ by Lemma 4.1. Now the sequence $\{U_n - \hat{U}_n\}$ is a reverse martingale (see Geertsema [7] for discussion), so that by Loève ([15], page 391),

$$P[\sup_{k \geqq n} |U_k - \hat{U}_k| > C] \leqq C^{-2r} E(U_n - \hat{U}_n)^{2r} \tag{4.6}$$

for any constant $C$. Therefore, by Theorem 2.1, the first term of the RHS of (4.5) is $O(n^{-2r})$. Thus (4.4) follows. □

**5. Analogous results for related von Mises statistics.** We shall deal with the one-sample case. Let $\{X_1\}$ be an i.i.d. sequence with df $F$. Let $\theta = \theta(F) = Eh(X_1, \cdots, X_m)$. For a sample of size $n$, the $U$-statistic for estimation of $\theta$ is, recalling (1.1),

$$U_n = \binom{n}{m}^{-1} \textstyle\sum_c h(X_{i_1}, \cdots, X_{i_m})\,, \tag{5.1}$$

where $\sum_c$ denotes summation over all combinations $\{i_1, \cdots, i_m\}$ from $\{1, \cdots, n\}$. The *associated von Mises statistic* (see Hoeffding [10] for discussion) is

$$V_n = n^{-m} \textstyle\sum_{i_1=1}^n \cdots \sum_{i_m=1}^n h(X_{i_1}, \cdots, X_{i_m})\,. \tag{5.2}$$

The following result parallels Theorem 2.1 and shows that properties of $V_n$ may be inferred from those of $U_n$, just as in Section 3 and Section 4 properties of $U_n$ were inferred from those of $\hat{U}_n$.

THEOREM 5.1. *Assume $E|h(X_{i_1}, \cdots, X_{i_m})|^r \leqq A < \infty$ for all $1 \leqq i_1, \cdots, i_m \leqq m$ and r a positive integer. Then*

$$E|U_n - V_n|^r = O(n^{-r})\,. \tag{5.3}$$

PROOF. Let $n_{(m)} = n(n-1)\cdots(n-m+1)$. Clearly,

$$n^m(U_n - V_n) = [n^m - n_{(m)}]U_n - \sum_* h(X_{i_1}, \cdots, X_{i_m}), \tag{5.4}$$

where $\sum_*$ denotes summation over choices $\{i_1, \cdots, i_m\}$ from $\{1, 2, \cdots, n\}$ where at least one equality $i_a = i_b$, $a \neq b$, holds. As the number of terms in $\sum_*$ is $n^m - n_{(m)}$, we have by Minkowski's inequality that

$$E|\sum_* h(X_{i_1}, \cdots, X_{i_m})|^r \leqq A[n^m - n_{(m)}]^r. \tag{5.5}$$

Likewise, $E|U_n|^r \leqq A$. Thus

$$n^{rm}E|U_n - V_n|^r \leqq 2^r A[n^m - n_{(m)}]^r. \tag{5.6}$$

But $n^m - n_{(m)} = O(n^{m-1})$, which yields (5.3). □

With the use of Theorem 5.1 where earlier Theorem 2.1 was needed, Theorem 3.1 may be extended to apply to $V_n$ in place of $U_n$. Theorem 4.1 can also be extended, since by (5.3), Markov's inequality, and a crude summation, we have $P[\sup_{k \geqq n} |V_k - U_k| > \varepsilon] = O(n^{1-2r})$.

**6. Examples.** Typical examples of $U$-statistics are given by the sample variance, Fisher's $k$-statistics, Gini's mean difference, Kendall's $\tau$, the grade correlation coefficient, and Wilcoxon's one- and two-sample statistics. (See [10], [13].) Let us briefly examine two of these examples.

(i) *Sample variance.* Let $X_1, \cdots, X_n$ be independent observations from a distribution $F$. Assume $E(X_1) = 0$ and $\theta(F) = \sigma^2 = E(X_1^2) > 0$. The sample variance is given by

$$U = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \binom{n}{2}^{-1} \sum_{i<j} h(X_i, X_j), \tag{6.1}$$

where $\bar{X} = n^{-1} \sum_1^n X_i$ and $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$. Thus $h_1^*(x) = \frac{1}{2}(x^2 - \sigma^2)$, $H(x_1, x_2) = x_1 x_2$, and $\sigma^2(\hat{U}) = n^{-1}E(X_1^2 - \sigma^2)^2$. If $E(X_1^{2r}) < \infty$ for an integer $r \geqq 3$, then clearly the hypothesis of Theorem 3.1 is satisfied and the rate of convergence in (3.2) is $O(n^{-r/(2r+1)})$. See also Section 7, Remark (vi).

(ii) *Wilcoxon two-sample statistic.* Let $\{X_1^{(1)}, \cdots, X_{n_1}^{(n)}\}$ and $\{X_1^{(2)}, \cdots, X_{n_2}^{(2)}\}$ be independent observations from continuous distributions $F_1$ and $F_2$. Then, for $\theta(F_1, F_2) = \int F_2\, dF_1$, an unbiased estimator is the Wilcoxon two-sample statistic, which may be written

$$U = (n_1 n_2)^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I_{[0,\infty)}(X_i^{(1)} - X_j^{(2)}), \tag{6.2}$$

where $I_A(\cdot)$ is the indicator of the set $A$. In this case the kernel is bounded and (3.4) is satisfied. Hence, by Corollary 3.1, the rate of convergence in (3.2) is, for any $\varepsilon > 0$, of order $O(n^{-\frac{1}{2}+\varepsilon})$.

**7. Concluding remarks.** (i) Since the $X_i^{(j)}$'s enter into the definition of $U$ only through the kernel $h$, our results apply also to vector observations.

(ii) In similar fashion as we have dealt with the CLT and SLLN, rates of convergence could also be established for the law of the iterated logarithm for $U$-statistics, which has been given in [16].

(iii) Although our Theorem 4.1 was restricted to the 1-sample case, a generalization to the $c$-sample case proceeds as follows. Consider a sequence of $c$-vectors $\{(n_1, \cdots, n_c)\}$ in which all components increase strictly from one vector to the next. For the associated sequence of $c$-sample $U$-statistics, the sequence of differences $U - \hat{U}$ again forms a reverse martingale, as established by the smoothing properties of conditional expectations (see [15], page 350, #4). Hence, in view of (2.1), and by application of Lemma 4.1 $c$ times, the method of proof of Theorem 4.1 extends to the general case.

(iv) In Theorem 3.1, condition (3.3) may be weakened to

$$E|h_i^*(X_1^{(i)})|^{2+\delta} < \infty\,, \qquad 1 \leqq i \leqq c\,, \tag{3.3'}$$

for some positive constant $\delta \leqq \frac{2}{3}$. In this case we can conclude only

$$\sup_t |P[(U - \theta)/\sigma(\hat{U}) \leqq t] - \Phi(t)| = O(n^{-\delta/2})\,. \tag{3.5'}$$

The appropriate Berry–Esséen result needed here stems from Liapounov's work [14] and is given as Theorem 1 of Esséen [5].

(v) It is clear that Theorem 2.1 holds when the distributions $F_1, \cdots, F_c$ depend upon $n$, provided that $\sup E(H^{2r}) < \infty$, where the supremum is taken over the class of distributions $\mathscr{F} = \{(F_{1,n}, \cdots, F_{c,n}), n \geqq 1\}$. Similar generalizations hold for Theorems 3.1 and 4.1.

(vi) Our results are limited by the crudity of (3.7). By a direct analysis for any particular $U$-statistic, one might obtain a sharper rate of convergence. For example, for the sample variance, Hsu [12] has obtained the rate $O(n^{-\frac{1}{2}})$ under 6th moment assumptions whereas in 6 we can give only $O(n^{-\frac{3}{7}})$ under the same assumptions and require all moments to achieve $O(n^{-\frac{1}{2}+\varepsilon})$.

(vii) It has come to our attention since preparing this paper that Theorem 2.1 in the 1-sample case has been proved (by different methods) by Funk [6] and applied to establish a moderate deviation theorem for $U$-statistics. However, Funk deals with the $c$-sample case by a different approach that entails some cumbersome restrictions. It appears that the use of our Theorem 2.1 would provide a more direct approach leading to simpler conditions to check.

**Acknowledgment.** The authors are indebted to Professor J. Sethuraman for stimulating conversations in the genesis of this work and to the referee for several helpful suggestions.

REFERENCES

[1] Berk, Robert H. (1966). Limiting behavior of posterior distributions. *Ann. Math. Statist.* **37** 51–58.

[2] Cramér, Harald (1946). *Mathematical Methods of Statistics.* Princeton University Press.

[3] Cramér, Harald (1970). *Random Variables and Probability Distributions.* Cambridge University Press.

[4] DWASS, MEYER (1956). The large-sample power of rank tests in the two-sample problem. *Ann. Math. Statist.* **27** 352–374.

[5] ESSÉEN, C. G. (1945). Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. *Acta Math.* **77** 1–125.

[6] FUNK, GERALD M. (1970). The probabilities of moderate deviations of $U$-statistics and excessive deviations of Kolmogorov–Smirnov and Kuiper statistics. Ph. D. dissertation, Michigan State Univ.

[7] GEERTSEMA, J. C. (1970). Sequential confidence intervals based on rank tests. *Ann. Math. Statist.* **41** 1016–1026.

[8] HÁJEK, JAROSLAV (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

[9] HANSON, D. L. (1970). Some results on convergence rates for weighted sums. *Lecture Notes in Mathematics* **160** 53–63.

[10] HOEFFDING, WASSILY (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

[11] HOEFFDING, WASSILY (1961). The strong law of large numbers for $U$-statistics. Institute of Statistics Mimeo Series No. 302, University of North Carolina, Chapel Hill, N. C.

[12] HSU, P. L. (1945). The approximate distributions of the mean and the variance of a sample of independent variables. *Ann. Math. Statist.* **16** 1–29.

[13] LEHMANN, E. L. (1951). Consistency and unbiasedness of certain non-parametric tests. *Ann. Math. Statist.* **22** 165–179.

[14] LIAPOUNOV, A. (1901). Nouvelle forme du theorème sur la limite de la probabilité. Mem. Acad. Sci. St. Petersbourg **12**(5), 1–24.

[15] LOÈVE, M. (1963). *Probability Theory* (3rd ed.). Van Nostrand, Princeton.

[16] SERFLING, R. J. (1971). The law of the iterated logarithm for $U$-statistics and related von Mises statistics (abstract). *Ann. Math. Statist.* **42** 1794.

[17] STOKER, D. J. (1954). An upper bound for the deviation between the distribution of Wilcoxon's test statistic for the two-sample problem and its limiting normal distribution for finite samples. I and II. *Indag. Math.* **16** 599–606, 607–614.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306