

ON THE RATES OF CONVERGENCE IN DISTRIBUTION OF FIRST ORDER  
STATIONARY U-STATISTICS AND VON MISES STATISTICS

Robert Alan Smith  
Department of Statistics  
University of North Carolina at Chapel Hill  
Chapel Hill, NC 27514

---

Research supported by the National Science Foundation under Grant MCS78-01240.

1. Introduction and notation.

For a sequence  $X_1, X_2, X_3, \dots$  of independent, identically distributed random variables, and a symmetric Borel-measurable function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  satisfying

$$E|\Phi(X_{i_1}, \dots, X_{i_m})| < \infty$$

for all positive integers  $1 \leq i_1 \leq \dots \leq i_m \leq m$ , the corresponding sequences of U-statistics and von Mises statistics are defined by

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} \Phi(X_{i_1}, \dots, X_{i_m}) \quad n \geq m$$

and

$$V_n = n^{-m} \sum_{i_1=1}^n \dots \sum_{i_m=1}^n \Phi(X_{i_1}, \dots, X_{i_m}) \quad n \geq m,$$

respectively. Following Hoeffding (1948), let

$$\theta = E[\Phi(X_1, \dots, X_m)]$$

and for  $h = 1, \dots, m$ , let

$$\Phi_h(x_1, \dots, x_h) = E[\Phi(X_1, \dots, X_m) \mid X_1 = x_1, \dots, X_h = x_h],$$

$$\Psi_h(x_1, \dots, x_h) = \Phi_h(x_1, \dots, x_h) - \theta,$$

and

$$\zeta_h = E[\Psi_h(X_1, \dots, X_h)]^2.$$

Hoeffding showed that if  $\zeta_m < \infty$

$$\text{Var}[U_n] = \binom{n}{m}^{-1} \sum_{h=1}^m \binom{m}{h} \binom{n-m}{m-h} \zeta_h,$$

$$E[V_n - U_n]^2 = O(n^{-2}),$$

and if  $\zeta_1 > 0$ , then as  $n \rightarrow \infty$ ,

$$\sqrt{n}(U_n - \theta) \stackrel{L}{\rightarrow} m\sqrt{\zeta_1} Z \quad (1.1)$$

and

$$\sqrt{n}(V_n - \theta) \stackrel{L}{\rightarrow} m\sqrt{\zeta_1} Z$$

where  $Z$  is a standard normal random variable. Callaert and Janssen (1978) found the rate of convergence in (1.1) to be  $O(n^{-\frac{1}{2}})$ .

When  $\zeta_1 = 0$ , however, the asymptotic distributions of  $U_n$  and  $V_n$  are more cumbersome. Viewing

$$\theta(F) = \int \dots \int \Phi(x_1, \dots, x_m) dF(x_1) \dots dF(x_m)$$

as a functional on the space  $F$  of distributions on  $\mathbb{R}$  satisfying

$$\int \dots \int \Phi(x_1, \dots, x_m)^2 dF(x_1) \dots dF(x_m) < \infty,$$

and noting that  $\zeta_d = 0$  implies  $\zeta_{d-1} = 0$  for  $d = 2, \dots, m$ , Hoeffding called  $\theta(F)$  *stationary of order  $d$  at  $F_0 \in F$*  if

$$\zeta_d(F_0) = 0 \text{ and } \zeta_{d+1}(F_0) > 0.$$

Von Mises (1947), Philippova (1961), Gregory (1977), Neuhaus (1977), Eagleson (1979), Hall (1979) and Janson (1979) have all investigated the limiting behavior of  $U_n$  or  $V_n$  in this *degenerate* case, i.e., when the order of stationarity is greater than zero.

Following Gregory (1977) for the case  $m = 2$ , let  $\{\lambda_k: k = 0, 1, 2, \dots\}$  denote the sequence of eigenvalues and  $\{\psi_k: k = 0, 1, 2, \dots\}$  the corresponding sequence of orthonormal eigenfunctions of the kernel function

$$\Psi(x, y) = \Phi(x, y) - \theta.$$

That is,

$$E[\psi_k(X_1) \psi_k(X_1)] = \delta_{kk},$$

$$E[\Psi(X_1, X_2) \psi_k(X_2) | X_1] = \lambda_k \psi_k(X_1) \text{ a.s.},$$

and

$$\lim_{K \rightarrow \infty} E[\Psi(X_1, X_2) - \sum_{k=0}^K \lambda_k \psi_k(X_1) \psi_k(X_2)]^2 = 0. \quad (1.2)$$

When  $\zeta_1 = 0$ ,

$$E[\Psi(X_1, X_2) | X_1] = 0 \text{ a.s.}$$

so that  $\lambda_0$  may be taken to be zero with corresponding eigenfunction  $\psi_0 \equiv 1$ .

Under these assumptions, Gregory showed that if  $\zeta_2 > 0$ , then as  $n \rightarrow \infty$ ,

$$n(U_n - \theta) \stackrel{L}{\approx} \sum_{k \geq 1} \lambda_k (Z_k^2 - 1) \quad (1.3)$$

and if in addition  $\sum_{k \geq 1} |\lambda_k| < \infty$ , then

$$n(V_n - \theta) \stackrel{L}{\approx} \sum_{k \geq 1} \lambda_k Z_k^2 \quad (1.4)$$

where  $Z_1, Z_2, Z_3, \dots$  are independent standard normal random variables. Gregory cites several examples.

Gregory's results have been extended to two-sample U-statistics (Neuhaus (1977) and Eagleson (1979)), to functional limit theorems (Neuhaus (1977) and Hall (1979)), and to orders of stationarity  $d > 1$  (Eagleson (1977), and Janson (1979)).

This paper studies the rates of convergence in these results. Using the methods of Sazonov (1968b and 1969), Theorems 2.1 and 2.2 estimate the rate of convergence in (1.4) and (1.3), respectively. Section 3 considers first order stationary kernel functions of degree  $m > 2$ . Extensions to functional limit theorems and two-sample statistics are also possible.

2. Rates of convergence for first order stationary kernels of degree two.

For  $m = 2$ , let

$$W_n = n(V_n - \theta) = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \Psi(X_i, X_j),$$

$$H_n(x) = P[W_n \leq x]$$

and

$$H(x) = P\left[\sum_{k \geq 1} \lambda_k Z_k^2 \leq x\right]$$

where  $Z_1, Z_2, Z_3, \dots$  are independent standard normal random variables and  $\{\lambda_k: k = 0, 1, 2, \dots\}$  and  $\{\psi_k: k = 0, 1, 2, \dots\}$  are as in (1.2).

THEOREM 2.1 If  $\zeta_1 = 0$  and  $\zeta_2 > 0$ , if the eigenvalues  $\{\lambda_k\}$  satisfy

$$\lambda_1 \geq \lambda_2 \geq \dots > 0$$

and

$$\sum_{k \geq 1} \lambda_k^p < \infty \quad 0 < p \leq 1,$$

and if the eigenfunctions  $\{\psi_k\}$  satisfy

$$\sup_k E|\psi_k(X_1)|^3 < \infty$$

and

$$E\left[\Psi(X_1, X_1) - \sum_{k=1}^K \lambda_k \psi_k^2(X_1)\right]^2 = o(1) \quad \text{as } K \rightarrow \infty$$

then

$$\sup_x |H_n(x) - H(x)| = o(n^{-(p-2)/(22p+4)}).$$

PROOF For fixed positive integers  $K > 8$  and  $n > 2$ , let

$$\Psi^{(K)}(x, y) = \sum_{k=1}^K \lambda_k \psi_k(x) \psi_k(y)$$

$$\tilde{\Psi}^{(K)}(x, y) = \Psi(x, y) - \Psi^{(K)}(x, y),$$

$$W_n^{(K)} = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \psi^{(K)}(X_i, X_j)$$

$$\tilde{W}_n^{(K)} = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \tilde{\psi}^{(K)}(X_i, X_j),$$

and

$$H_n^{(K)}(x) = P[W_n^{(K)} \leq x]$$

$$H^{(K)}(x) = P\left[\sum_{k=1}^K \lambda_k Z_k^2 \leq x\right].$$

Then

$$\sup_x |H_n(x) - H(x)| \leq \Delta_1 + \Delta_2 + \Delta_3$$

where

$$\Delta_1 = \sup_x |H_n(x) - H_n^{(K)}(x - \sum_{k>K} \lambda_k)|,$$

$$\Delta_2 = \sup_x |H_n^{(K)}(x) - H^{(K)}(x)|$$

and

$$\Delta_3 = \sup_x |H^{(K)}(x - \sum_{k>K} \lambda_k) - H(x)|.$$

We proceed to estimate each of  $\Delta_1$ ,  $\Delta_2$ , and  $\Delta_3$  in turn.

Estimation of  $\Delta_1$ : For any  $\epsilon > 0$ ,

$$\begin{aligned} & |H_n(x) - H_n^{(K)}(x - \sum_{k>K} \lambda_k)| \leq |P[W_n^{(K)} \leq x - \tilde{W}_n^{(K)}, |\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k| \leq \epsilon] \\ & - P[W_n^{(K)} \leq x - \sum_{k>K} \lambda_k, |\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k| \leq \epsilon]| + |P[W_n^{(K)} \leq x - \tilde{W}_n^{(K)}, |\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k| > \epsilon] \\ & - P[W_n^{(K)} \leq x - \sum_{k>K} \lambda_k, |\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k| > \epsilon]| \leq |P[W_n^{(K)} \leq x - \tilde{W}_n^{(K)}, |\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k| \leq \epsilon] \\ & - P[W_n^{(K)} \leq x - \sum_{k>K} \lambda_k, |\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k| > \epsilon]| + P[|\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k| > \epsilon] \\ & \leq \sup_x P[x - \epsilon < W_n^{(K)} \leq x + \epsilon] + P[|\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k| > \epsilon]. \end{aligned}$$

The first term is

$$\begin{aligned} \sup_x P[x - \varepsilon < W_n^{(K)} \leq x + \varepsilon] &= \sup_x (P[W_n^{(K)} \leq x + \varepsilon] - P[W^{(K)} \leq x + \varepsilon]) \\ &+ P[W^{(K)} \leq x + \varepsilon] - P[W^{(K)} \leq x - \varepsilon] + P[W^{(K)} \leq x - \varepsilon] + P[W_n^{(K)} \leq x - \varepsilon]) \\ &\leq \Delta_2 + \sup_x \int_{x-\varepsilon}^{x+\varepsilon} h^{(K)}(y) dy + \Delta_2 \leq 2\Delta_2 + 2\varepsilon \sup_x h^{(2)}(x) \end{aligned}$$

where  $h^{(K)}$  denotes the probability density function of  $\sum_{k=1}^K \lambda_k Z_k^2$ .

This and an application of Chebyshev's inequality imply

$$\Delta_1 \leq 2\Delta_2 + C_1 \varepsilon + \varepsilon^{-2} E[\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k]^2$$

where  $C_1$  is a constant which may be taken to be  $(\lambda_1 \lambda_2)^{-1/2}$ . Minimizing the RHS with respect to  $\varepsilon$  yields

$$\Delta_1 \leq 2\Delta_2 + C_2 (E[\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k]^2)^{1/3}$$

for some constant  $C_2$ . Noting that

$$E[\tilde{\Psi}^{(K)}(X_1, X_2)]^2 = \sum_{k>K} \lambda_k^2$$

and

$$E[\tilde{\Psi}^{(K)}(X_1, X_2) \tilde{\Psi}^{(K)}(X_1, X_3)] = 0,$$

one can show that

$$\begin{aligned} E[\tilde{W}_n^{(K)} - \sum_{k>K} \lambda_k]^2 &= 2(n-1)^{-1} n \sum_{k>K} \lambda_k^2 + n^{-1} E[\tilde{\Psi}^{(K)}(X_1, X_1) - \sum_{k>K} \lambda_k]^2 \\ &\leq 2\lambda_K^{2-p} \sum_{k>K} \lambda_k^p + n^{-1} E[\tilde{\Psi}^{(K)}(X_1, X_1) - \sum_{k>K} \lambda_k]^2 \leq C_3 K^{1-2/p} + C_4 n^{-1} \end{aligned}$$

where  $C_3$  and  $C_4$  are constants such that

$$k \lambda_k^p \leq C_3^{p/(2-p)}$$

and

$$E[\tilde{\Psi}^{(k)}(X_1, X_1) - \sum_{j>k} \lambda_j]^2 \leq C_4$$

for sufficiently large  $k$ . Thus,

$$\Delta_1 \leq 2\Delta_2 + C_2(C_3 k^{1-2/p} + C_4 n^{-1})^{1/3}. \quad (2.1)$$

Estimation of  $\Delta_2$ : Applying Theorem 1 of Sazonov (1968a) with equation (10) of Sazonov (1968b) to the sequence of  $K$ -dimensional sample means

$$n^{-1} \sum_{i=1}^n (\psi_1(X_i), \dots, \psi_K(X_i))^T$$

yields

$$\Delta_2 \leq C_5 K^3 n^{-\frac{1}{2}} \sum_{k=1}^K E|\psi_k(X_1)|^3 \leq C_6 K^4 n^{-\frac{1}{2}} \quad (2.2)$$

where  $C_5$  is an absolute constant and  $C_6 = C_5 \cdot \sup_k E|\psi_k(X_1)|^3$ .

Estimation of  $\Delta_3$ : By Lemma XVI,3.2 of Feller (1971)

$$\Delta_3 \leq \frac{1}{\pi} \int_{-\infty}^{\infty} \left| \frac{\exp(-it \sum_{k=1}^K \lambda_k) \hat{h}^{(K)}(t) - \exp(-it \sum_{k \geq 1} \lambda_k) \hat{h}(t)}{t} \right| dt$$

where

$$\hat{h}(t) = \int e^{itx} dH(x)$$

and

$$\hat{h}^{(K)}(t) = \int e^{itx} dH^{(K)}(x).$$

The numerator of the integrand on the right can be written

$$\left| \hat{h}^{(K)}(t) \left| \exp(it \sum_{k>K} \lambda_k) - \hat{h}^{(K)}(t) \right| \right|$$



where

$$\begin{aligned}\hat{h}^{(K)}(t) &= \frac{\hat{h}(t)}{\hat{h}^{(K)}(t)} = E[e^{it\tilde{W}^{(K)}}] \\ &= 1 + it \sum_{k>K} \lambda_k - \frac{t^2}{2} E[\tilde{W}^{(K)}]^2 + o(|t|^2) \\ &= 1 + it \sum_{k>K} \lambda_k - \frac{t^2}{2} \left( \left( \sum_{k>K} \lambda_k \right)^2 + 2 \sum_{k>K} \lambda_k^2 \right) + o(|t|^2)\end{aligned}$$

Writing

$$\exp(it \sum_{k>K} \lambda_k) = 1 + it \sum_{k>K} \lambda_k - \frac{t^2}{2} \left( \sum_{k>K} \lambda_k \right)^2 + o(|t|^2)$$

yields

$$\begin{aligned}\Delta_3 &\leq \frac{1}{\pi} \int_{-\infty}^{\infty} |\hat{h}^{(K)}(t)| \left| \frac{t^2 \sum_{k>K} \lambda_k^2 + o(|t|^2)}{t} \right| dt \\ &= \frac{\sum_{k>K} \lambda_k^2}{\pi} \int_{-\infty}^{\infty} |\hat{h}^{(K)}(t)| |t + o(|t|)| dt\end{aligned}$$

where the integral is finite for  $K \geq 8$ . Since  $\sum_{k>K} \lambda_k^2 \leq C_3 K^{1-2/p}$ ,

$$\Delta_3 \leq C_7 K^{1-2/p}. \quad (2.3)$$

Combining (2.1), (2.2), and (2.3), one obtains

$$\begin{aligned}\sup_x |H_n(x) - H(x)| &\leq 3C_6 K^4 n^{-\frac{1}{2}} \\ &+ C_2 (C_3 K^{1-2/p} + C_4 n^{-1})^{1/3} + C_7 K^{1-2/p}.\end{aligned}$$

Letting  $K = n^{3p/(22p+4)}$  yields the desired result.

**REMARKS** (1) In the above theorem, the kernel function  $\Psi$  is assumed to have infinitely many positive eigenvalues. If, however, only finitely many eigenvalues,  $\lambda_1, \dots, \lambda_K$ , say, are non-zero, then the estimation of  $\Delta_2$  above shows

that

$$\sup_x \left| H_n(x) - H(x) \right| \leq C K^3 n^{-\frac{1}{2}} \sum_{k=1}^K E |\psi_k(X_1)|^3$$

where  $\psi_1, \dots, \psi_K$  are the orthonormal eigenfunctions of  $\psi$  corresponding to  $\lambda_1, \dots, \lambda_K$  and  $C$  is an absolute constant. This is the case for the familiar chi-square goodness-of-fit test.

(2) The assumed positivity of the  $\lambda_k$ 's is used to ensure the convexity of certain sets in  $\mathbb{R}^k$  to which Theorem 1 of Sazonov (1968a) is applied to obtain (2.2). The assumption can be circumvented.

(3) For the Cramer-von Mises goodness-of-fit statistic

$$\int [F(x) - F_n(x)]^2 dF(x)$$

where  $F_n$  denotes the corresponding empirical distribution function, the eigenvalues have the form

$$\lambda_k = (\pi k)^{-2} \quad k \geq 1$$

so that  $p$  could be taken to be slightly larger than  $\frac{1}{2}$ . The resulting rate of  $O(n^{\epsilon-1/10})$  for arbitrarily small positive  $\epsilon$  matches the first estimate by Sazonov (1968b) but falls short of his later (1969) result in which advantage was taken of the particular kernel function.

(4) The rate may be improved by assuming  $L_{2r}$ -convergence ( $r \geq 2$ ) in place of  $L_2$ -convergence in (1.2). There does not seem to be any practical justification for doing so, however.

Now let

$$T_n = n(U_n - \theta)$$

$$G_n(x) = P[T_n \leq x]$$

and

$$G(x) = P\left[\sum_{k \geq 1} \lambda_k (Z_k^2 - 1) \leq x\right]$$

where the  $Z_k$ 's,  $\lambda_k$ 's, and  $\psi_k$ 's are as before.

THEOREM 2.2 If the hypotheses of Theorem 2.1 hold, then

$$\sup_x \left| G_n(x) - G(x) \right| = O(n^{(p-2)/(22p+4)}).$$

PROOF

$$\begin{aligned} G_n(x) &= P[T_n \leq x, \left| n^{-1} \sum_{i=1}^n \Psi(X_i, X_i) - \sum_{k \geq 1} \lambda_k \right| \leq \epsilon] \\ &+ P[T_n \leq x, \left| n^{-1} \sum_{i=1}^n \Psi(X_i, X_i) - \sum_{k \geq 1} \lambda_k \right| > \epsilon] \\ &\leq P[T_n + n^{-1} \sum_{i=1}^n \Psi(X_i, X_i) \leq x + \sum_{k \geq 1} \lambda_k + \epsilon] \\ &+ P\left[ \left| n^{-1} \sum_{i=1}^n \Psi(X_i, X_i) - \sum_{k \geq 1} \lambda_k \right| > \epsilon \right] \\ &= H_n(x + \sum_{k \geq 1} \lambda_k + \epsilon) + P\left[ \left| n^{-1} \sum_{i=1}^n \Psi(X_i, X_i) - \sum_{k \geq 1} \lambda_k \right| > \epsilon \right]. \end{aligned}$$

Similarly

$$H_n(x + \sum_{k \geq 1} \lambda_k - \epsilon) \leq G_n(x) + P\left[ \left| n^{-1} \sum_{i=1}^n \Psi(X_i, X_i) - \sum_{k \geq 1} \lambda_k \right| > \epsilon \right].$$

Thus,

$$\begin{aligned} \sup_x |G_n(x) - G(x)| &\leq \sup_x |H_n(x) - H(x)| \\ &+ \sup_x |H(x+\epsilon) - H(x)| + P\left[ \left| n^{-1} \sum_{i=1}^n \Psi(X_i, X_i) - \sum_{k \geq 1} \lambda_k \right| > \epsilon \right] \end{aligned}$$

where the first term on the right is estimated by Theorem 2.1, the second term

is

$$\begin{aligned} \sup_x (H(x+\epsilon) - H(x)) &= \sup_x \int_x^{x+\epsilon} dH(t) \\ &\leq \epsilon \sup_x h^{(2)}(x) = \epsilon / \sqrt{\lambda_1 \lambda_2}, \end{aligned}$$

and the third term is

$$\begin{aligned} P\left[\left|n^{-1} \sum_{i=1}^n \Psi(X_i, X_i) - \sum_{k \geq 1} \lambda_k\right| > \epsilon\right] \\ \leq \epsilon^{-2} n^{-1} E\left[\Psi(X_1, X_1) - \sum_{k \geq 1} \lambda_k\right]^2, \end{aligned}$$

so that

$$\begin{aligned} |G_n(x) - G(x)| &\leq C_9 n^{(p-2)/(22p+4)} + \epsilon / \sqrt{\lambda_1 \lambda_2} \\ &+ \epsilon^{-2} n^{-1} E\left[\Psi(X_1, X_1) - \sum_{k \geq 1} \lambda_k\right]^2. \end{aligned}$$

Choosing

$$\epsilon = (2\sqrt{\lambda_1 \lambda_2} E\left[\Psi(X_1, X_1) - \sum_{k \geq 1} \lambda_k\right]^2 n^{-1})^{1/3}$$

to minimize the right hand side yields the desired result.

### 3. Rates of convergence for first order stationary kernels of arbitrary degree m.

Again let

$$G_n(x) = P[T_n \leq x] \text{ and } H_n(x) = P[W_n \leq x]$$

where

$$T_n = n(U_n - \theta) = n \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} \Psi(X_{i_1}, \dots, X_{i_m})$$

and

$$W_n = n(V_n - \theta) = n^{-m+1} \sum_{i_1=1}^n \dots \sum_{i_m=1}^n \Psi(X_{i_1}, \dots, X_{i_m}).$$

Hoeffding (1961) showed that

$$U_n - \theta = \sum_{h=1}^m \binom{m}{h} U_n^{(h)} \quad (3.1)$$

where

$$U_n^{(h)} = \binom{n}{h}^{-1} \sum_{1 \leq i_1 < \dots < i_h \leq n} \dots \sum \psi^{(h)}(x_{i_1}, \dots, x_{i_h}),$$

$$\psi^{(1)}(x) = \psi_1(x),$$

and for  $h = 2, \dots, m$ ,

$$\psi^{(h)}(x_1, \dots, x_h) = \psi_h(x_1, \dots, x_h) - \sum_{j=1}^{h-1} \sum_{1 \leq i_1 < \dots < i_j \leq h} \psi^{(j)}(x_{i_1}, \dots, x_{i_j}).$$

Similarly,

$$V_n - \theta = \sum_{h=1}^m \binom{m}{h} V_n^{(h)} \quad (3.2)$$

where

$$V_n^{(h)} = n^{-h} \sum_{i_1=1}^n \dots \sum_{i_h=1}^n \psi^{(h)}(x_{i_1}, \dots, x_{i_h}).$$

When  $\zeta_1 = 0$ ,

$$U_n^{(1)} = V_n^{(1)} = n^{-1} \sum_{i=1}^n \psi_1(x_i)$$

is almost surely zero so that (3.1) and (3.2) become

$$U_n - \theta = \binom{m}{2} U_n^{(2)} + R_n^{(2)}$$

and

$$V_n - \theta = \binom{m}{2} V_n^{(2)} + Q_n^{(2)}$$

where

$$R_n^{(2)} = \sum_{h=3}^m \binom{m}{h} U_n^{(h)}$$

and

$$Q_n^{(2)} = \sum_{h=3}^m \binom{m}{h} V_n^{(h)}.$$

Since  $E[R_n^{(2)}]^2$  and  $E[Q_n^{(2)}]^2$  are both  $O(n^{-3})$ , (1.3) and (1.4) applied to  $U_n^{(2)}$  and  $V_n^{(2)}$ , respectively, yield the following results:

THEOREM 3.1 If  $\zeta_1 = 0$  and  $\zeta_2 > 0$ , then as  $n \rightarrow \infty$ ,

$$n(U_n - \theta) \stackrel{L}{\rightarrow} \binom{m}{2} \sum_{k \geq 1} \lambda_k (Z_k^2 - 1)$$

where  $\{\lambda_k\}$  is the sequence of eigenvalues of the kernel function  $\psi_2$  and  $\{Z_k\}$  is a sequence of independent standard normal random variables.

THEOREM 3.2 If in addition to the hypotheses of Theorem 3.1,  $\sum_{k \geq 1} |\lambda_k| < \infty$ , then as  $n \rightarrow \infty$

$$n(V_n - \theta) \stackrel{L}{\rightarrow} \binom{m}{2} \sum_{k \geq 1} \lambda_k Z_k^2.$$

Estimates of the rates of convergence in Theorems 3.1 and 3.2 can be obtained from applications of Theorems 2.2 and 2.1 to  $U^{(2)}$  and  $V^{(2)}$ , respectively.

THEOREM 3.3 If  $\zeta_1 = 0$  and  $\zeta_2 > 0$ , if the eigenvalues  $\{\lambda_k\}$  of  $\psi^{(2)}$  satisfy

$$\lambda_1 \geq \lambda_2 \geq \dots > 0$$

and

$$\sum_{k \geq 1} \lambda_k^p < \infty,$$

and if the corresponding eigenfunctions  $\{\psi_k\}$  satisfy

$$\sup_k E |\psi_k(X_1)|^3 < \infty$$

and

$$E[\psi^{(2)}(X_1, X_1) - \sum_{k=1}^k \lambda_k \psi_k^2(X_1)]^2 = o(1) \text{ as } k \rightarrow \infty,$$

then

$$\sup_x |G_n(x) - G(x)| = o(n^{-(p-2)/(22p+4)}).$$

PROOF For any  $\varepsilon > 0$ ,

$$\begin{aligned} G_n(x) &\leq P[T_n \leq x, |nR_n^{(2)}| \leq \varepsilon] + P[|nR_n^{(2)}| > \varepsilon] \\ &\leq P[T_n - nR_n^{(2)} \leq x + \varepsilon] + P[|nR_n^{(2)}| > \varepsilon] \\ &= P[n \binom{m}{2} U_n^{(2)} \leq x + \varepsilon] + P[|nR_n^{(2)}| > \varepsilon]. \end{aligned}$$

Similarly,

$$P[n \binom{m}{2} U_n^{(2)} \leq x + \varepsilon] \leq G_n(x) + P[|nR_n^{(2)}| > \varepsilon].$$

Thus,

$$\begin{aligned} \sup_x |G_n(x) - G(x)| &\leq \sup_x \left| P[n \binom{m}{2} U_n^{(2)} \leq x + \varepsilon] - G(x + \varepsilon) \right| \\ &+ \sup_x |G(x + \varepsilon) - G(x)| + P[|nR_n^{(2)}| > \varepsilon]. \end{aligned} \quad (3.3)$$

The first term on the right hand side of (3.3) is  $O(n^{(p-2)/(22p+4)})$  by Theorem 2.2. The second term is  $O(\varepsilon)$  as  $\varepsilon \rightarrow 0$ . Since  $E[R_n]^2 = O(n^{-3})$ , the third term is  $O(n^{-1}\varepsilon^{-2})$ . Choosing  $\varepsilon = n^{(p-2)/(22p+4)}$  completes the proof.

The proof of the next result is exactly parallel to that above.

THEOREM 3.4 Under the assumptions of Theorem 3.3,

$$\sup_x |H_n(x) - H(x)| = O(n^{(p-2)/(22p+4)}).$$

#### References

- Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics. *Ann. Statist.* 6, 417-421.
- Eagleson, G. K. (1979). Orthogonal expansions and U-statistics. *Austral. J. Statist.* 21, 221-237.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II, Second Ed., Wiley, New York.

- Filippova, A. A. (1961). Von Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theor. Prob. Appl.* 7, 24-57.
- Gregory, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* 5, 110-123.
- Hall, P. (1979). On the invariance principle for U-statistics. *Stochastic Processes and their Applications* 9, 163-174.
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* 19, 293-325.
- Hoeffding, W. (1961). The strong law of large numbers for U-statistics. Univ. of North Carolina, Institute of Statistics, Mimeo Series, No. 302.
- Janson, S. (1979). The asymptotic distribution of degenerate U-statistics. Uppsala University, Department of Mathematics Report No. 1979:5, Uppsala University, Uppsala, Sweden.
- Mises, R. v. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* 18, 309-348.
- Neuhaus, G. (1977). Functional limit theorems for U-statistics in the degenerate case. *J. Multivar. Anal.* 7, 424-439.
- Sazonov, V. V. (1968a). On the multidimensional central limit theorem. *Sankhyā, Ser. A* 30, 181-204.
- Sazonov, V. V. (1968b). On  $\omega^2$  criterion. *Sankhyā, Ser. A* 30, 205-210.
- Sazonov, V. V. (1969). An improvement of a convergence-rate estimate. *Theory of Probability and Its Applications* 14, 640-651.