ASYMPTOTIC DISTRIBUTION THEORY FOR DEGENERATE U-STATISTICS FROM STATIONARY SEQUENCES

E. Carlstein
University of North Carolina
at Chapel Hill

SUMMARY

This paper initiates the study of degenerate U-statistics in the case of dependent data. Although the asymptotic distribution of non-degenerate U-statistics has been extensively investigated under various mixing assumptions, the study of degenerate U-statistics has until now been restricted to the iid case. Our asymptotic distribution theory covers *-mixing, Φ-mixing, and strong-mixing. The limiting distribution which we arrive at is a weighted sum of (dependent) χ^2 random variables; this is intuitively analogous to the well-known iid result. Moreover, our general treatment of degenerate U-statistics includes several important special cases: The asymptotic distributions of the χ^2 goodness-of-fit statistic (under strong-mixing) and of the generalized Cramer-von Mises statistic (under *-mixing) are obtained as applications of our theory.

Supported by NSF Grant DMS-8400602.
AMS 1980 subject classification: Primary 62E20, secondary 60G10.
Key words and phrases: dependence, strong-mixing, goodness-of-fit, Cramer-von Mises statistic, chi-squared.
Running heading: Degenerate U-statistics.

1. Introduction

Let $\{Z_i : -\infty < i < \infty\}$ be a strictly stationary sequence of R^p-valued random vectors $(1 \leq p < \infty)$; the marginal distribution of Z_0 is F. Let $\phi(z_1, z_2, \ldots z_r)$ be a real-valued function defined on $\overset{r}{\underset{i=1}{\chi}} R^p$ $(1 \leq r < \infty)$, and assume that ϕ is symmetric in its r vector arguments. Then the corresponding U-statistic with kernel ϕ, based on data $(Z_1, Z_2, \ldots, Z_n)$ $(n \geq r)$, is:

$$U_n = \sum_{1 \leq i_1 < i_2 < \cdots < i_r \leq n} \phi(Z_{i_1}, Z_{i_2}, \ldots, Z_{i_r}) / \binom{n}{r} .$$

For $0 \leq k \leq r$ denote $\phi_k(z_1, z_2, \ldots, z_k) = \int \cdots \int \phi(z_1, z_2, \ldots, z_r) \prod_{i=k+1}^{r} dF(z_i)$ and $\tilde{\phi}_k(z_1, z_2, \ldots, z_k) = \phi_k(z_1, z_2, \ldots, z_k) - \phi_0$. Observe that $\phi_r(\cdot) = \phi(\cdot)$ and $\phi_0 = \text{constant}$; write $\tilde{\phi}(\cdot) = \tilde{\phi}_r(\cdot)$. If $\int \tilde{\phi}_1^2(z) dF(z) = 0$, then we shall say that U_n is degenerate; otherwise U_n is nondegenerate.

Asymptotic distribution theory for U_n has been studied in great depth. Hoeffding (1948) introduced the concept of U-statistics and established asymptotic normality for nondegenerate U-statistics when $\{Z_i\}$ is iid. Asymptotic normality of nondegenerate U-statistics $(r \geq 2)$ has been extended to the dependent case in the following steps: Sen (1963) allowed for m-dependence in $\{Z_i\}$; Sen (1972) treated *-mixing sequences; Yoshihara (1976) considered absolutely regular sequences. Asymptotic normality for nondegenerate U-statistics has not been established for strong-mixing sequences. (Recall the hierarchy: iid ⇒ m-dependent ⇒ *-mixing ⇒ Φ-mixing ⇒ absolutely regular ⇒ strong-mixing (see Yoshihara (1976) and Sen (1972)).)

On the other hand, the asymptotic distribution of degenerate U-statistics has only been studied under the assumption that $\{Z_i\}$ is iid. Serfling (1980) and Gregory (1977) have shown (independently) that the asymptotic distribution in this case is equivalent to that of a weighted sum of independent χ^2 random variables. Our objective is to extend the asymptotic distribution theory for degenerate U-statistics to cases where $\{Z_i\}$ is dependent.

As a first logical step (analogous to Sen's (1972) work for non-degenerate U_n), Theorem 1 (below) treats the case of $*$-mixing in $\{Z_i\}$. Next, by imposing an additional constraint on the kernel function ϕ, we obtain an asymptotic distributional result (Theorem 2) which allows $\{Z_i\}$ to be strong-mixing. This is particularly surprising since, as mentioned above, no strong-mixing result exists for nondegenerate U-statistics. These two new results are not mere mathematical curiosities, filling in obscure voids in our knowledge. Rather, they can be applied to several well-known statistical problems. For example, Moore (1982) and Chanda (1981) have studied the asymptotic distribution of the standard χ^2 goodness-of-fit statistic when the data come from a strong-mixing sequence. Their basic limiting distribution can actually be obtained as a special case of our Theorem 2. As another example, we consider the generalized Cramer-von Mises statistics. Their asymptotic distributions have not previously been studied in the presence of dependent data, since a result like Theorem 1 is needed in order to carry out the analysis. In further examples, we apply Theorem 2 to a measure of dependence, to the sample variance, and to the cross-product statistic.

In Section 2 we present and discuss our main theoretical results (Theorems 1 and 2). Applications of these results are found in Section 3. The proofs of Theorems 1 and 2 are deferred to Section 4.

2. Main Results

We are interested in the asymptotic distribution of the standardized U-statistic $\tilde{U}_n = n(U_n - \phi_0)$, where U_n is degenerate. To avoid trivialities, assume that $r \geq 2$ and that $\iint \tilde{\phi}_2^2(z_1,z_2)\, dF(z_1)\, dF(z_2) \in (0,\infty)$. Theorems 1 and 2 will actually assume $r = 2$, but this restriction is evidently not of much practical consequence: All of Gregory's (1977) distribution theory and applications for degenerate U-statistics in the iid case have $r = 2$; all of Serfling's (1980, Sec. 5.5.2) examples of degenerate U-statistics have $r = 2$; our examples in Section 3 (e.g. the χ^2 goodness-of-fit statistic and the generalized Cramer-von Mises statistic) only involve $r = 2$. Nevertheless, Theorem 4 (in Section 3) shows that Theorem 1 actually can be extended to the case $r \geq 3$ as well.

We shall make use of the following representation of $\tilde{\phi}_2$ (see Gregory (1977)). The equation

$$\int \tilde{\phi}_2(z_1,z_2)\, g(z_2)\, dF(z_2) = g(z_1)\lambda \quad \text{a.e. } [F]$$

has distinct nontrivial solutions $\{g_i\}$ (called eigenfunctions), with corresponding eigenvalues $\{\lambda_i\}$, satisfying the orthonormality condition $\int g_i(z) g_j(z)\, dF(z) = I\{i=j\}$. For fixed $K \geq 1$, denote $\vec{g}_K(z) = (g_1(z),\ldots,g_K(z))'$ and $\Lambda_K = \mathrm{diag}\{\lambda_1,\ldots,\lambda_K\}$. Then $\tilde{\phi}_2$ may be expressed (in the limit) in terms of $\{g_i\}$ and $\{\lambda_i\}$ by virtue of:

$$(*) \qquad \lim_{K\to\infty} \iint [\tilde{\phi}_2(z_1,z_2) - \vec{g}_K'(z_1)\Lambda_K \vec{g}_K(z_2)]^2\, dF(z_1)\, dF(z_2) = 0 .$$

The joint distribution of (Z_0, Z_i) $(i \geq 1)$ will be denoted by $F_i(\cdot,\cdot)$. Let $\sigma_n^+ = \sigma\{Z_i : i \geq n\}$ and $\sigma_n^- = \sigma\{Z_i : i \leq n\}$. We say that $\{Z_i\}$ is *-mixing if $\lim_{n\to\infty} \psi(n) = 0$, where $\psi(n) = \sup\{|P\{B|A\}/P\{B\} - 1| : A \in \sigma_0^-,\ B \in \sigma_n^+,\ P\{A\} > 0,\ P\{B\} > 0\}$. We say that $\{Z_i\}$ is Φ-mixing if $\lim_{n\to\infty} \Phi(n) = 0$, where $\Phi(n) = \sup\{|P\{B|A\} - P\{B\}| : A \in \sigma_0^-,\ B \in \sigma_n^+,\ P\{A\} > 0\}$. We say that $\{Z_i\}$ is

<u>strong-mixing</u> if $\lim_{n\to\infty} \alpha(n) = 0$, where

$\alpha(n) = \sup\{|P\{A \cap B\} - P\{A\}P\{B\}| : A \in \sigma_0^-,\ B \in \sigma_n^+\}$. We assume that $E\{\tilde{\phi}_2^2(Z_0, Z_i)\} < \infty \ \forall\, i \geq 1$.

<u>THEOREM 1.</u> Let $\{Z_i\}$ be $*$-mixing and let $r=2$. If

$$\text{(1.a)} \quad \sum_{n=0}^{\infty} (n+1)\psi^{\frac{1}{2}}(n) < \infty$$

and

$$\text{(1.b)} \quad \sum_{i=1}^{\infty} |\lambda_i| < \infty$$

then $\tilde{U}_n \xrightarrow{D} \sum_{i=1}^{\infty} \lambda_i (W_i^2 - 1)$ as $n \to \infty$. The joint distribution of any finite set of W_i's (with $\lambda_i \neq 0$) is multivariate normal with means 0 and $E\{W_I W_J\} = \sum_{i=-\infty}^{\infty} E\{g_I(Z_0) g_J(Z_i)\}$. (Put $W_i \equiv 0$ if $\lambda_i = 0$).

This result is analogous to Gregory's (1977) Theorem 2.1 for iid data (with $Q_{n1} \equiv P_0$ in his notation). The W_i's in our conclusion are not necessarily iid, because our Z_i's are not necessarily iid. (Note, however, that if $\{Z_i\}$ are iid, our asymptotic distribution is identical to his.)

When the set of eigenfunctions is finite, we can relax the conditions on dependence. The χ^2 goodness-of-fit statistic is an important example where $\{g_i\}$ is finite and Theorem 2 applies (see Section 3).

<u>THEOREM 2</u>. Let $\{Z_i\}$ be strong-mixing, let $r=2$, and assume that $\#\{g_i\} =: K < \infty$.

If $\exists \delta > 0$ s.t. $E\{|g_i(Z_0)|^{2+\delta}\} < \infty \ \ \forall\, i \in \{1,\ldots,K\}$ and s.t.

$\sum_{n=1}^{\infty} (\alpha(n))^{\delta/(2+\delta)} < \infty$, and if $F_i \ll [F \times F]$ $\forall i \geq 1$, then

$\tilde{U}_n \xrightarrow{D} \sum_{i=1}^{K} \lambda_i (W_i^2 - 1)$ as $n \to \infty$. The joint distribution of the W_i's is as in Theorem 1 .

Note that, for nondegenerate U_n (with $r \geq 2$) , there are no asymptotic distributional results allowing strong-mixing. If we make the more restrictive Φ-mixing assumption, we can eliminate the moment conditions on g_i.

THEOREM 3. Let $\{Z_i\}$ be Φ-mixing, let $r = 2$, and assume that $\#\{g_i\} =: K < \infty$. If $\sum_{n=1}^{\infty} \Phi^{\frac{1}{2}}(n) < \infty$ and $F_i \ll [F \times F]$ $\forall i \geq 1$, then the conclusion of Theorem 2 holds.

The proofs of Theorems 1, 2, and 3 are deferred to Section 4.

3. Applications

Example 1: χ^2 Goodness-of-Fit Statistic.

Suppose that the range of Z_0 is partitioned into $I \geq 2$ mutually exclusive and exhaustive sets $\{A_i : 1 \leq i \leq I\}$, with $p_i := P\{Z_0 \in A_i\} > 0$ $\forall i$. For evaluating the agreement between the model (characterized by $\{p_i : 1 \leq i \leq I\}$) and the data $(Z_1, \ldots, Z_n)$, we consider the standard χ^2 goodness-of-fit statistic:

$$X_n = \sum_{i=1}^{I} \Big(\sum_{j=1}^{n} I\{Z_j \in A_i\} - np_i \Big)^2 / np_i$$

$$= (n-1) \sum_{1 \leq j < k \leq n} \tilde{\phi}(Z_j, Z_k) / \binom{n}{2} + \sum_{j=1}^{n} V_j / n - 1 ,$$

where $\phi(z_1,z_2) = \sum_{i=1}^{I} I\{z_1 \in A_i\} I\{z_2 \in A_i\}/p_i$, $\phi_0 = 1$, and $V_j = \sum_{i=1}^{I} I\{Z_j \in A_i\}/p_i$. Observe that the implicit U-statistic with kernel ϕ is degenerate. The Ergodic Theorem implies that $\sum_{j=1}^{n} V_j/n \xrightarrow{a.s.} I$, provided $\{Z_i\}$ is strong-mixing (say). Thus it suffices to determine the asymptotic distribution of $\tilde{U}_n + I-1$.

Without loss of generality, the equation defining the eigenfunctions and eigenvalues of $\tilde{\phi}_2$ may be expressed in matrix form as $\vec{\gamma} - \vec{1}\,\vec{p}'\vec{\gamma} = \lambda\,\vec{\gamma}$. Here $\vec{1}$ is an I-dimensional column vector of 1's, $\vec{p}' = (p_1,\ldots,p_I)$, $\lambda \in R^1$, and $\vec{\gamma} = (\gamma_1,\ldots,\gamma_I)' \in R^I$ characterizes the eigenfunction $g(\cdot)$ by the relation $g(z) = \sum_{i=1}^{I} \gamma_i \, I\{z \in A_i\}$. Corresponding to $\lambda_1 = 0$ is the solution $\vec{\gamma}_1 = \vec{1}$. The remaining eigenvalues $\{\lambda_2,\lambda_3,\ldots,\lambda_I\}$ are each equal to 1, and their corresponding eigenfunctions are simply a set of I-1 vectors $\{\vec{\gamma}_2,\vec{\gamma}_3,\ldots,\vec{\gamma}_I\}$ in R^I satisfying $\vec{p}'\vec{\gamma}_i = 0$ and $\vec{\gamma}_i' \operatorname{diag}\{p_1,\ldots,p_I\}\vec{\gamma}_j = I\{i=j\}$ (i.e. the $\vec{\gamma}_i$'s are mutually orthonormal w.r.t. the weights $\{p_i\}$).

We now wish to apply Theorem 2. Assume that $\sum_{n=1}^{\infty} (\alpha(n))^{\nu} < \infty$ for some $\nu \in (0,1)$; the moment condition on g_i is trivial here. Since ϕ may be rewritten in terms of the random vectors $\{\tilde{Z}_i\}$, where $\tilde{Z}_i' = (I\{Z_i \in A_1\},\ldots, I\{Z_i \in A_I\})$, the relevant absolute continuity condition is $\tilde{F}_i \ll [\tilde{F} \times \tilde{F}]$ $\forall\, i \geq 1$, where $\tilde{F}_i$ is the joint distribution of $(\tilde{Z}_0,\tilde{Z}_i)$ and $\tilde{F}$ is the marginal. Since $\{\tilde{Z}_i\}$ are discrete r.v.s , this condition is satisfied. Hence we may conclude that $X_n \xrightarrow{D} \sum_{i=2}^{I} W_i^2$, where $(W_2,\ldots,W_I) \sim N_{I-1}(\vec{0}_{I-1}, \sum_{t=-\infty}^{\infty} \Gamma_t)$, Γ_t is a square matrix of dimension I-1 with $(u,v)^{\underline{th}}$ entry

$\vec{\gamma}_u' P_t \vec{\gamma}_v$ $(2 \leq u \leq I,\ 2 \leq v \leq I)$, and P_t is an $I \times I$ matrix with $(i,j)\underline{th}$ entry $P\{Z_0 \in A_i,\ Z_t \in A_j\}$. Note that when $\{Z_i\}$ is iid the asymptotic distribution reduces to the familiar $\chi^2_{(I-1)}$. Also note that the asymptotic distribution is in general equivalent to that of $\sum_{i=2}^{I} \xi_i^2 \omega_i$, where $\{\xi_i\}$ are iid $N(0,1)$ and $\{\omega_2,\ldots,\omega_I\}$ are the eigenvalues of the matrix $\Gamma = \sum_{t=-\infty}^{\infty} \Gamma_t$.

To see that our asymptotic distribution is the same as that in Chanda's (1981) Theorem 2.1, let $T = \mathrm{diag}\{0,\tau_2,\tau_3,\ldots,\tau_I\}$ be the diagonal matrix containing the eigenvalues of his matrix Λ, and let $\Pi = (\vec{\pi}_1,\vec{\pi}_2,\ldots,\vec{\pi}_I)$ be the corresponding matrix of eigenvectors. Then $\Lambda = \Pi T \Pi' = \tilde{\Pi}\tilde{T}\tilde{\Pi}'$, where $\tilde{\Pi} = (\vec{\pi}_2,\ldots,\vec{\pi}_I)$ and $\tilde{T} = \mathrm{diag}\{\tau_2,\ldots,\tau_I\}$. Observe that $\Gamma = D\Lambda D'$, where $D = (\vec{\gamma}_2,\ldots,\vec{\gamma}_I)'\ \mathrm{diag}\{p_1^{\frac{1}{2}},\ldots,p_I^{\frac{1}{2}}\}$. Hence $\Gamma = B\tilde{T}B'$, where $B = D\tilde{\Pi}$ is a square orthonormal matrix of dimension $I-1$. This shows that the eigenvalues of Γ (i.e. the ω_i's) are equal to $(\tau_2,\ldots,\tau_I)$, and that Chanda's (1981) Theorem 2.1 is just a special case of our Theorem 2. Chanda (1981) provides no proof of his result.

Moore's (1982) matrix C, in Case 1 of his Theorem 2.1, is identical to Chanda's (1981) matrix Λ. Thus Moore's limiting distribution also is obtained from our Theorem 2. Note that we do not assume $\{Z_i\}$ to be Gaussian. □

<u>Example 2: Generalized Cramer-von Mises Statistic.</u>

Suppose that $\{Z_i\}$ are *-mixing real-valued r.v.s with absolutely continuous strictly increasing distribution function F. The generalized Cramer-von Mises goodness-of-fit statistic is $n\int_{-\infty}^{\infty} w(F(z))[F(z)-\hat{F}_n(z)]^2 dF(z)$, where $\hat{F}_n$ is the empirical c.d.f. of $(Z_1,\ldots,Z_n)$ and w is a non-negative weight function. Without loss of generality, we consider $Y_n =$

$n \sum_{1\le i<j\le n} \phi(U_i, U_j)/\binom{n}{2} + \sum_{i=1}^{n} \phi(U_i, U_i)/n$, where $\{U_i : -\infty < i < +\infty\}$ is a strictly stationary sequence of $*$-mixing Uniform [0,1] r.v.s, and

$\phi(u_1, u_2) = \int_0^1 w(u)(u - I\{u_1 \le u\})(u - I\{u_2 \le u\})du$. Observe that $\int_0^1 \phi(u_1, u_2)du_2 = 0$, so $\phi_1 \equiv \phi_0 = 0$ and the corresponding U-statistic is degenerate.

Two commonly used weight functions are $w_1(u) \equiv 1$ and $w_2(u) = 1/u(1-u)$ $(0 < u < 1)$. It is straightforward to verify that all the integrability conditions on ϕ are satisfied for these weight functions. Hence we can apply Theorem 1 and the Ergodic Theorem to the first and second terms (respectively) in Y_n. (The limiting value of the second term is 1/6 for w_1 and 1 for w_2.) De Wet and Venter (1973) have studied the eigenvalues and eigenfunctions of ϕ in great generality. In particular, their work shows that $\lambda_i = 1/\pi^2 i^2$ (not $1/i^2$) for w_1 and that $\lambda_i = 1/i(i+1)$ for w_2. Thus condition (1.b) is satisfied. Note that $\{g_i\}$ are based on the infinite set of Jacobi polynomials, so Theorems 2 and 3 cannot be applied to the Cramer-von Mises statistic. Moreover, our treatment of the $*$-mixing case is the first investigation of this statistic allowing for any nontrivial dependence. □

<u>Example 3: A Measure of Dependence.</u>

Suppose Z_0 has discrete distribution F concentrating on the set $\{z_i : 1 \le i \le I\}$; let $p_i = P\{Z_0 = z_i\} > 0 \ \forall i$. The sequence $\{Z_i\}$ is supposed to be independent, but, upon observing $(Z_1, \ldots, Z_n)$, the practitioner suspects the presence of nontrivial dependence within ranges of m consecutive observations ($2 \le m \le n$). If in fact no dependence is present, then $P\{Z_{i+1} = z_{i_1}, Z_{i+2} = z_{i_2}, \ldots, Z_{i+m} = z_{i_m}\} = \prod_{j=1}^{m} p_{i_j}$ for any i and any $(i_1, i_2, \ldots, i_m)$. We can measure the degree of departure from this null hypothesis by comparing how many subseries $\vec{Z}_m^i := (Z_{i+1}, Z_{i+2}, \ldots, Z_{i+m})$ $(0 \le i \le n-m)$ in our sample equal $(z_{i_1}, z_{i_2}, \ldots, z_{i_m})$ vs how many were expected to equal $(z_{i_1}, z_{i_2}, \ldots, z_{i_m})$ assuming independence in $\{Z_i\}$. This is the standard observed vs expected concept, but with the following important distinction: Even if we assume independence in $\{Z_i\}$, the random vectors $\{\vec{Z}_m^i : 0 \le i \le n-m\}$ are still nontrivially (m-1)-dependent. (Of course, one could look only at nonoverlapping subseries $\{\vec{Z}_m^{im} : 0 \le i \le [n/m]-1\}$, which are independent under H_0. But in doing so, the majority of the available information about serial dependence would be thrown away.) Calculation of the asymptotic null distribution of the statistic

$$\sum_{\forall(i_1, i_2, \ldots, i_m)} \left[\sum_{i=0}^{n-m} I\{\vec{Z}_m^i = (z_{i_1}, z_{i_2}, \ldots, z_{i_m})\} - (n-m+1) \prod_{j=1}^{m} p_{i_j} \right]^2 \Big/ (n-m+1) \prod_{j=1}^{m} p_{i_j}$$

now fits easily into the framework of Example 1. Since $\{\vec{Z}_m^i\}$ are (m-1)-dependent, summability of the mixing coefficients is immediate. For $|t| \ge m$ we have $\Gamma_t = 0$. For $|t| < m$ the matrix P_t can be calculated explicitly in terms of the original p_i's, because the precise dependence structure in $\{\vec{Z}_m^i\}$ is known. □

Example 4: The Sample Variance.

The standard sample variance $s_n^2 = \sum_{i=1}^{n} (Z_i-\bar{Z}_n)^2/(n-1)$ may be expressed as the U-statistic with $\phi(z_1,z_2) = (z_1-z_2)^2/2$ and $\phi_0 = \mathrm{Var}\{Z_0\}$. This statistic is a useful estimator even under dependence, because $E\{s_n^2\} \to \mathrm{Var}\{Z_0\}$ whenever $\sum_{i=1}^{\infty} \mathrm{Cov}\{Z_0,Z_i\}$ converges. Since $\tilde{\phi}_1(z) = [(z-E\{Z_0\})^2 - \mathrm{Var}\{Z_0\}]/2$, U_n is degenerate if $|Z_0 - E\{Z_0\}| = \mathrm{s.d.}\{Z_0\}$ a.s. Note that this condition does not necessarily imply that Z_0 is a degenerate r.v. In particular, consider $Z_0 \sim$ Binomial $(1,\frac{1}{2})$. Serfling (1980, p. 194) shows that, in this case, $K=1$, $g_1(z) = 2z-1$, and $\lambda_1 = -\frac{1}{4}$. We now extend his results to the case of a strong-mixing sequence. Since the moment condition and the absolute continuity condition of Theorem 2 are trivial here, we may conclude that $n(s_n^2 - \frac{1}{4}) \xrightarrow{D} \frac{1}{4}(1-W^2)$, provided $\sum_{n=1}^{\infty} (\alpha(n))^{\nu} < \infty$ for some $\nu \in (0,1)$. The r.v. W is normal with mean 0 and variance $\sum_{i=-\infty}^{\infty} (4E\{Z_0Z_i\}-1)$. If $\{Z_i\}$ is independent, this result reduces to that given by Serfling (1980).

Example 5: Cross-Product Statistic (Serfling (1980), Ex. 5.5.2B).

Let $\phi(z_1,z_2) = z_1z_2$, $E\{Z_0\} = 0$, and $E\{Z_0^2\} > 0$. Then $\phi_1(z) \equiv \phi_0 = 0$, so $U_n = \sum_{1\le i<j\le n} Z_iZ_j/\binom{n}{2}$ is degenerate. In this situation, $K=1$, $g_1(z) = z/E^{\frac{1}{2}}\{Z_0^2\}$, and $\lambda_1 = E\{Z_0^2\}$. To apply Theorem 2 or 3, we assume $E\{Z_0^4\} < \infty$; we also must assume that the absolute continuity condition and the summability condition (on the mixing coefficients) are satisfied. Note that, in Theorem 2, the moment condition reduces to simply

$E\{|Z_0|^{2+\delta}\} < \infty$. The limiting distribution of $\tilde{U}_n$ is $E\{Z_0^2\}(W^2-1)$, where W is normal with mean 0 and variance $\sum_{i=-\infty}^{\infty} E\{Z_0 Z_i\}/E\{Z_0^2\}$. If $\{Z_i\}$ is independent, this reduces to Serfling's result. □

Lastly, we show that Theorem 1 can in fact be extended to degenerate U-statistics with r-argument kernels ($r \geq 3$). We use the usual projection technique (see Sen (1972)) to write

$$\tilde{U}_n = \hat{U}_n + n \sum_{h=3}^{r} \binom{r}{h} U_n^{(h)} ,$$

where $\hat{U}_n := \binom{r}{2} n \sum_{1 \leq i < j \leq n} \tilde{\phi}_2(Z_i, Z_j)/\binom{n}{2}$ and $U_n^{(h)}$ is as defined in Sen (1972, eq. 3.25). Note that $\hat{U}_n = n\left[\binom{r}{1} U_n^{(1)} + \binom{r}{2} U_n^{(2)}\right]$ because of U_n's degeneracy. The assumptions and notation of Sections 1 and 2 are retained.

<u>THEOREM 4</u>. Let $\{Z_i\}$ be *-mixing and let $r \geq 3$. Assume that

$$\int \cdots \int \phi^2(z_1, \ldots, z_r) \prod_{i=1}^{r} dF(z_i) < \infty \text{ and } \sum_{n=0}^{\infty} (n+1)^{r-1} \psi^{\frac{1}{2}}(n) < \infty .$$

Then $(\tilde{U}_n - \hat{U}_n) \xrightarrow{P} 0$ as $n \to \infty$. Consequently, if (1.b) also holds, then $\tilde{U}_n \xrightarrow{D} \binom{r}{2} \sum_{i=1}^{\infty} \lambda_i (W_i^2 - 1)$ as $n \to \infty$. (The $\{W_i\}$ have the same distribution as in Theorem 1.)

<u>PROOF</u>. The required convergence in distribution of $\tilde{U}_n$ will follow immediately by applying Theorem 1 to $\hat{U}_n$. To see that $\hat{U}_n$ approximates $\tilde{U}_n$ (in probability), consider $E\{(\hat{U}_n - \tilde{U}_n)^2\} \leq n^2 (r-2) \sum_{h=3}^{r} \binom{r}{h}^2 E\{[U_n^{(h)}]^2\}$. By equation (3.27) of Sen (1972), the r.h.s. of the above inequality is $O(1/n)$. □

4. Proofs

Proof of Theorem 1. We will show that $T_n := 2 \sum_{1 \le i<j \le n} \tilde{\phi}(Z_i, Z_j)/n \approx \tilde{U}_n$ converges in distribution to the required r.v. For $K \ge 1$ define $T_{nK} = 2 \sum_{1 \le i<j \le n} \vec{g}_K'(Z_i)\Lambda_K \vec{g}_K(Z_j)/n$, and let τ_n and τ_{nK} be the characteristic functions of T_n and T_{nK} respectively. Let $s \in R^1$ be arbitrary but fixed.

As in Serfling (1980, p. 197) we have $|\tau_n(s)-\tau_{nK}(s)| \le |s| E^{\frac{1}{2}}\{(T_n-T_{nK})^2\}$. Furthermore, $(T_n-T_{nK})/(n-1)$ is itself a 2-argument U-statistic with kernel $G_K(z_1,z_2) = \tilde{\phi}(z_1,z_2) - \vec{g}_K'(z_1)\Lambda_K \vec{g}_K(z_2)$. Observe that $\tilde{\phi}_1(Z_0) = 0$ a.s., and $E\{g_i(Z_0)\} = 0$ whenever $\lambda_i \ne 0$, so that $\int G_K(z_1,z_2)\, dF(z_1) = 0$ a.e. [F] (see Serfling (1980, pp. 196-197)). Hence, $(T_n-T_{nK})/(n-1)$ is a degenerate U-statistic and therefore is precisely of the form $U_n^{(2)}$ (as defined by Sen (1972, eq. 3.25)). Now applying equation (3.27) of Sen (1972), we have $E\{(T_n-T_{nK})^2\} \le c \iint G_K^2(z_1,z_2)\, dF(z_1)\, dF(z_2)$, $\forall n$, where $c < \infty$ depends only on $\{\psi(n)\}$. Therefore, by (*), $E\{(T_n-T_{nK})^2\} < \varepsilon$ for $K \ge K(\varepsilon)$, uniformly in n.

For fixed $K \ge 1$ we have $T_{nK} = \sum_{k=1}^{K} \lambda_k \left[(n^{\frac{1}{2}}\bar{g}_{kn})^2 - \bar{\xi}_{kn}\right]$, where $\bar{g}_{kn} = \sum_{i=1}^{n} g_k(Z_i)/n$ and $\bar{\xi}_{kn} = \sum_{i=1}^{n} g_k^2(Z_i)/n$. By the Ergodic Theorem, $\bar{\xi}_{kn} \xrightarrow{a.s.} 1$ as $n \to \infty$ for each k. By the multivariate analog of Ibragimov's (1962) Theorem 1.5, the asymptotic joint distribution of those $n^{\frac{1}{2}}\bar{g}_{kn}$'s with $\lambda_k \ne 0$ $(1 \le k \le K)$ is precisely the joint distribution of the corresponding W_k's. Thus,

$$T_{nK} \xrightarrow{D} Y_K := \sum_{k=1}^{K} \lambda_k (W_k^2 - 1) \quad \text{as} \quad n \to \infty .$$

Now, for any $M > N$ we have $E\{(Y_M - Y_N)^2\} \leq \left(\sum_{i=N+1}^{M} |\lambda_i| E^{\frac{1}{2}}\{(W_i^2-1)^2\} \right)^2$, with $E\{(W_i^2-1)^2\} \leq 3E^2\{W_i^2\} + 1$. For each i s.t. $\lambda_i \neq 0$, we find $E\{W_i^2\} \leq 1 + 4 \sum_{n=1}^{\infty} \Phi^{\frac{1}{2}}(n) < \infty$ (the first inequality follows from Lemma 1.1 of Ibragimov (1962); the second follows from our condition (1.a)). Thus, condition (1.b) implies that $E\{(Y_M-Y_N)^2\} \to 0$ as $N \to \infty$, uniformly in $M > N$. By the Cauchy convergence criteria, we may conclude that an L_2 r.v. $Y := \sum_{i=1}^{\infty} \lambda_i (W_i^2-1)$ exists s.t. $\lim_{K\to\infty} E\{(Y_K-Y)^2\} = 0$. Let θ and θ_K be the characteristic functions of Y and Y_K respectively. Again we have $|\theta_K(s) - \theta(s)| \leq |s| E^{\frac{1}{2}}\{(Y_K-Y)^2\}$. Combining the results of this paragraph with those of the previous 2 paragraphs, we see that $\tau_n(s) \to \theta(s)$ as $n \to \infty$. □

Proof of Theorem 3. Since (*) implies that $\tilde{\phi}_2(z_1, z_2) = \vec{g}_K'(z_1) \Lambda_K \vec{g}_K(z_2)$ a.e. [F×F], and since each F_i is absolutely continuous w.r.t. F×F, it follows that $T_n = T_{nK}$ a.s. (T_n and T_{nK} are as defined in the proof of Theorem 1.) The 3rd paragraph in the proof of Theorem 1 now applies. □

Proof of Theorem 2. Exactly the same as the proof of Theorem 3, but replace Ibragimov's (1962) Theorem 1.5 with his Theorem 1.7. □

REFERENCES

Chanda, K.C. (1981) Chi-square goodness-of-fit tests based on dependent observations. In Statistical Distributions in Scientific Work 5 (C. Taillie, G. Patil and B. Baldessari, eds.) 35-49. D. Reidel, Dordrecht, Holland.

de Wet, T. and Venter, J.H. (1973) Asymptotic distributions for quadratic forms with applications to tests of fit. Ann. Statist. 1, 380-387.

Gregory, G.G. (1977) Large sample theory for U-statistics and tests of fit. Ann. Statist. 5, 110-123.

Hoeffding, W. (1948) A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19, 293-325.

Ibragimov, I.A. (1962) Some limit theorems for stationary processes. Theory Probab. Appl. 7, 349-382.

Moore, D.S. (1982) The effect of dependence on chi squared tests of fit. Ann. Statist. 10, 1163-1171.

Sen, P.K. (1963) On the properties of U-statistics when the observations are not independent. Part one: estimation of non-serial parameters in some stationary stochastic process. Cal. Statist. Assoc. Bull. 12, 69-92.

Sen, P.K. (1972) Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. Z. Wahrsch. verw. Gebiete 25, 71-82.

Serfling, R.J. (1980) Approximation Theorems of Mathematical Statistics. Wiley, New York.

Yoshihara, K. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. Z. Wahrsch. verw. Gebiete 35, 237-252.

Department of Statistics
321 Phillips Hall 039A
University of North Carolina
Chapel Hill, NC 27514