doi: 10.1111/j.1467-9469.2006.00516.x
 Vol 33: 861–873, 2006

Two-stage *U*-statistics for Hypothesis Testing

HWAI-CHUNG HO and GRACE S. SHIEH

Institute of Statistical Science, Academia Sinica

ABSTRACT. A *U*-statistic is not easy to apply or cannot be applied in hypothesis testing when it is degenerate or has an indeterminate degeneracy under the null hypothesis. A class of two-stage *U*-statistics (TU-statistics) is proposed to remedy these drawbacks. Both the asymptotic distributions under the null and the alternative of TU-statistics are shown to have simple forms. When the degeneracy is indeterminate, the Pitman asymptotic relative efficiency of a TU-statistic dominates that of the incomplete *U*-statistics. If the kernel is degenerate under the null hypothesis but non-degenerate under the alternative, a TU-statistic is proved to be more powerful than its corresponding *U*-statistic. Applications to testing independence of paired angles in ecology and marine biology are given. Finally, a simulation study shows that a TU-statistic is more powerful than its corresponding incomplete *U*-statistic in almost all cases under two settings.

Key words: asymptotic normality, circular data, hypothesis testing, independence test, U-statistic

1. Introduction

Most estimators and test statistics can either be written as U-statistics or can be approximated by U-statistics (Halmos, 1946; Hoeffding, 1948). The theory of U-statistics has been applied in many tests of hypothesis, for instance, testing for independence, model misspecification, goodness-of-fit and others. When one applies U-statistics for testing hypothesis in the areas of geophysics (Stephens, 1979; Fisher & Lee, 1986), ecology (Fisher & Lee, 1982; example 1 in section 3), econometrics (Bierens & Ploberger, 1997; Fan & Li, 1999) and marine biology (example 2 in section 3), often one encounters some difficulties which are described in detail following the definition of degeneracy of U-statistics.

A U-statistic is an extension of the well-known sample mean. Let $X_1, \dots, X_n$ be independent and identically distributed (i.i.d.) random variables (r.v.) with distribution function F. Given a parameter of interest $\theta(F)$, to construct a U-statistic for θ, we first find a symmetric and real-valued function h such that

$$E\{h(X_1, \dots, X_k)\} = \theta,$$

where the function h is called a kernel of order k, k is a positive constant and $k \le n$. Let $C(n,k)$ denote the combinations of k distinct elements $\{i_1, \dots, i_k\}$ from $\{1, \dots, n\}$. A U-statistic is the average of all evaluations of h over the $C(n,k)$ distinct X_{i_j}s. Namely,

$$U_n = C(n,k)^{-1} \sum_{C(n,\,k)} h(X_{i_1}, \dots, X_{i_k}),$$

where $\sum_{C(n,\,k)}$ denotes summation over the combinations $C(n,k)$. An order-k kernel is *degenerate* of order d $(d \le k)$ if $\sigma_c^2 = 0$ for $c = 1, \dots, d$ and $\sigma_{d+1}^2 > 0$, where $\sigma_c^2 = \mathrm{var}\{h_c(X_1, \dots, X_c)\}$ and $h_c(x_1, \dots, x_c) = E\{h(x_1, \dots, x_c, X_{c+1}, \dots, X_k)\}$. If $\sigma_1^2 > 0$, h is called non-degenerate.

When using a degenerate U-statistic for hypothesis testing, we encounter two difficulties: (i) under the null hypothesis H_0, the limiting distribution of a degenerate order-d $(d \ge 1)$ U-statistic often does not have a closed form; or (ii) the null limiting distribution of a degenerate order-1 U-statistic suffers from an extensive calculation of weights (eigenvalues of the kernel). In some cases, identifying the eigenvalues of a kernel is prohibitive. One example is the multivariate case of the Cramér–von Mises statistic mentioned in Ahmad (1993).

Degenerate U-statistics were utilized in areas such as directional data, goodness-of-fit tests, econometrics and among others. In directional data, the test statistic for independence between two p-dimensional random vectors ($p \geq 3$) on the surfaces of spheres is a degenerate U-statistic of order $p-2$ (Fisher & Lee, 1986). Some degenerate of order-1 U-statistics arise in the context of testing for independence in paired circular data, e.g. the tests studied in Fisher & Lee (1982) and Shieh *et al.* (1994). Other degenerate U-statistics were proposed for testing goodness-of-fit, for instance, Watson's U^2 for two-sample goodness-of-fit on a circle (Watson, 1962; Persson, 1979), the Cramér–von Mises type statistics for one-sample goodness-of-fit in Anderson & Darling (1952), Gregory (1977), Shorack & Wellner (1986, ch. 5) and D'Agostino & Stephens (1986, ch. 4). Recently in econometrics, degenerate U-statistics were utilized in several tests for model mis-specification (Linton & Gozalo, 1995; Bierens & Ploberger, 1997; Fan & Li, 1999).

A U-statistic used in testing hypothesis may have an indeterminate degeneracy. Standard U-statistic theory cannot be applied as the order of degeneracy of the kernel is indeterminate. An incomplete U-statistic (U_I), which is an average of N randomly chosen evaluations of a kernel h with $N/n \to 0$ can be used. However, U_I is inefficient, since the rate of convergence of U_I is $\sqrt{N}$ (theorem 1 in p. 204 of Lee, 1990 & Janson, 1984). A typical example of an indeterminate U-statistic is the test proposed for independence of paired circular data (Fisher & Lee, 1983). This statistic can also test independence of a particular fish's spawning time and the low tide time in marine biology (Lund, 1999) or the paired dihedral angles of protein peptides (Singh *et al.*, 2002). Under the null hypothesis, the kernel is degenerate if at least one marginal assumes circular uniform distribution and the kernel is non-degenerate otherwise. See section 4 for details of this kernel. More generally, we encounter U-statistics with an indeterminate degeneracy when estimating a product type parameter $\theta = E\{g_1(X)g_2(Y)\}$, where g_1 and g_2 are real valued functions, and X and Y are i.i.d. r.v. with a positive variance. The kernel of the U-statistic for θ is

$$h(x, y) = \frac{\{g_1(x)g_2(y) + g_1(y)g_2(x)\}}{2}.$$

Given $\theta = 0$, h is degenerate if $E\{g_1(X)\} = E\{g_2(Y)\}$, and h is non-degenerate if $E\{g_1(X)\} \neq E\{g_2(Y)\}$.

To remedy the aforementioned drawbacks of U-statistics utilized in testing hypothesis, we propose a class of two-stage U-statistics (TU-statistics henceforth). The limiting null distribution of a TU-statistic has a simple closed form even when its kernel has a high-order degeneracy, and a TU-statistic is asymptotically normal when its kernel has indeterminate degeneracy. The proposed procedure for constructing a TU-statistic can be extended to multivariate U-statistics. Furthermore, when a kernel is degenerate under the null hypothesis but is non-degenerate under the alternative, the TU-statistic is shown to be more powerful than its associated U-statistic. For details, see section 2.2 and the examples in section 3 on testing independence for bivariate circular data.

In section 2, we first show how to construct TU-statistics in two cases where a U-statistic is either degenerate of order $d (d \geq 1)$ or indeterminate. Both the limiting distributions of TU-statistics under the null and the alternative have been derived. Under H_0, the limiting distribution is a Hermite polynomial function of the standard normal r.v. Criteria for choosing the tuning parameters of TU-statistics are also given. In section 3, two applications in testing independence of paired angles in ecology and marine biology are presented. The constructed TU-statistic has a limiting χ_1^2-type distribution. In section 4, we conduct a power study of the TU-statistic for testing bivariate circular independence with respect to its associated incomplete U-statistic under two settings. In section 5, the Pitman ARE of the TU-statistic is shown to dominate that of its associated incomplete U-statistic for the indeterminate degeneracy case. We close with some discussion in section 6.

2. Two-stage *U*-statistic

2.1. *The formulation*

Assume that $\sigma_k^2 \equiv \text{var}\{h(X_1, \ldots, X_k)\}$ is finite, where $k \geq 2$. We consider a general hypothesis testing problem, namely testing $H_0: \theta = c_1$ against the alternative $H_1: \theta > c_1$, where c_1 is a constant. Without loss of generality, we may assume that $c_1 = 0$. A test for H_0 can be based on a U-statistic of the form

$$\hat{\theta} = \sum_{C(n,k)} \frac{h(X_{i_1}, \ldots, X_{i_k})}{C(n,k)}.$$

Let $0 < \lambda < 1$. We first divide a sample of n observations into m subsamples of nearly equal size $l = [n^{\lambda}]$ or $l+1$ such that $m = [n/l]$. As $lm \leq n \leq (l+1)m$ so some subsamples have size l and the rest $l+1$, and we denote the jth subsample size by l_j. We note that m should not be < 2. For instance, if n can be evenly divided by m then we have the following subsamples:

$$L_i = \{(i-1)l+1, (i-1)l+2, \ldots, il\}, \quad i = 1, \ldots, m.$$

A TU-statistic ($T_{m,t}$) can be constructed by the following two stages.

Stage 1. We construct m independent normalized statistics as follows. Let l_j be the number of elements in the jth subsample from which a normalized statistic I_j is constructed under two different cases. (a) When the kernel is degenerate of known order $d \geq 1$, I_j is just the normalized complete U-statistic derived from the jth subsample,

$$I_j = l_j^{(d+1)/2} C(l_j, k)^{-1} \sum_{C(l_j, k)} h(X_{i_1}, \ldots, X_{i_k}). \tag{1}$$

(b) When the degeneracy of the kernel is indeterminate,

$$I_j = l_j^{-1/2} \sum_{l_j} h(X_{i_1}, \ldots, X_{i_k}),$$

where $\sum_{l_j}$ denotes the summation over l_j sets that are randomly chosen with replacement from $C(l_j, k)$ combinations of k distinct indices from the jth subsample. The chosen scheme with replacement implies a neat and closed form of variance (theorem 4 (i) of section 4.3.1, Lee, 1990).

Stage 2. We partition an information I_j into a signal $EI_j = l_j^{(d+1)/2}\theta$ and a noise part $I_j - l_j^{(d+1)/2}\theta$; then we estimate them by $l_j^{(d+1)/2}\hat{\theta}$ and $I_j - l_j^{(d+1)/2}\hat{\theta}$ respectively. Under the null hypothesis $\theta = 0$, the noise part dominates the limiting distribution. On the other hand, under the alternative hypothesis $\theta > 0$, the signal part dominates the limiting distribution. Thus the first and last term in the expansion of $\prod_{j=1}^{t}[(I_j - l_j^{(d+1)/2}\hat{\theta}) + l_j^{(d+1)/2}\hat{\theta}]$ dominate the limiting distribution under the null and alternative hypothesis, respectively, and we discard those in-between terms. A TU-statistic of order t is the sum of two multilinear U-type statistics with the kernel $g(x_1, \ldots, x_t) = x_1 x_2 \ldots x_t$ but using two sets of r.v., $\{I_j - l_j^{(d+1)/2}\hat{\theta}, 1 \leq j \leq m\}$ and $\{l_j^{(d+1)/2}\hat{\theta}, 1 \leq j \leq m\}$ respectively. Let $\tilde{I}_j$ denote $I_j - l_j^{(d+1)/2}\hat{\theta}$, the centred version of I_j, and put

$$S_t = C(m,t)^{-1} \sum_{C(m,\, t)} \prod_{i=1}^{t} l_{j_i}^{(d+1)/2}.$$

When the kernel is degenerate of order $d \geq 1$, a TU-statistic of order t has the form

$$T_{m,t} = C(m,t)^{-1} \sum_{C(m,\, t)} \tilde{I}_{j_1} \times \cdots \times \tilde{I}_{j_t} + S_t \hat{\theta}^t. \tag{2}$$

When the degeneracy of the kernel is indeterminate, we should use $T_{m,1}$ with $d = 0$ and $t = 1$, namely

$$T_{m,1} = m^{-1} \sum_{C(m,1)} I_j.$$

Note that, both I_j and $T_{m,t}$ depend on n which is suppressed for simplicity. From stage 1(b), it is clear that a TU-statistic is more efficient than its associated U_I since a TU is based on an average of l evaluations of the kernel while U_I is based on $\sqrt{N}$ evaluations and $N = o(l)$. There are two parameters λ and t involved when constructing a TU-statistic; some guidelines for choosing them are given in section 2.3.

2.2. The asymptotic properties of TU-statistics

In this section, the limiting distribution of $T_{m,t}$ under H_0 is derived. In the sequel, we shall need the recursive relation for the Hermite polynomials (Major, 1981)

$$H_k(x) = xH_{k-1}(x) - (k-1)H_{k-2}(x), \tag{3}$$

where

$$H_k(x) = (-1)^k e^{x^2/2} \left(\frac{d^k e^{-x^2/2}}{dx^k} \right)$$

is the kth Hermite polynomial.

Recall that $\sigma_c^2 = \text{var}\{h_c(X_1, \ldots, X_c)\}$, where $1 \leq c \leq k$. The limiting results in theorem 1 below are formulated for general $t \geq 1$, not limited to $t = 1$ or $t = 2$ that are suggested by some criteria in section 2.3.

Theorem 1

Assume $\theta = E\{h(X_1, \ldots, X_k)\}$ *and* $\sigma_k^2 = \text{var}\{h(X_1, \ldots, X_k)\} < \infty$. *Then*

(a) if h is degenerate of order $d \geq 1$, *under* $H_0 : \theta = 0$, $m^{t/2} T_{m,t}$ *converges in distribution to* $v_{d+1}^t H_t(Z)$ *as* $m \to \infty$, *where Z is the standard normal random variable and* $v_{d+1} = \sigma_{d+1} \sqrt{(d+1)!C(k, d+1)}$,

(b) if degeneracy of h is indeterminate, taking $d = 0$ *and* $t = 1$, *result (a) for* $m^{1/2} T_{m,t}$ *still holds and* $v_1^2 = \sigma_k^2 + \lim_{n\to\infty} n\,\text{var}(U_n)$,

(c) under the alternative $H_1 : \theta > 0$, *the limiting distribution of* $m^{t/2}(T_{m,t} - S_t\tilde{\theta})$ *is* $v_{d+1}^t H_t(Z)$, *where* $\tilde{\theta} = \hat{\theta}^t$ *if h is degenerate of order* $d \geq 1$ *and* $\tilde{\theta} = \theta$ *if h is indeterminate, thus* $d = 0$ *and* $t = 1$. *It follows that the test statistic* $m^{t/2} T_{m,t}$ *is consistent against the alternative* $\theta > 0$.

The variance of U_n in (b) can be computed by Monte Carlo, the jackknife or the bootstrap method. The proof of theorem 1 is an adaptation of the procedure in Rubin & Vitale (1980), and the details are in the appendix. Theorem 1(b) and the indeterminate case of part (c) are direct applications of the central limit theorem. For theorem 1 (a) and the degenerate case of part (c), a quick check to see why they hold is as follows. Suppose that $l_j = l$ for $1 \le j \le m$, then $T_{m,t}$ can be expressed as

$$T_{m,t} = C(m,t)^{-1} \sum_{s=0}^{t} C(m-s, t-s)[l^{(d+1)/2}(\theta - \hat{\theta})]^{t-s}$$
$$\times \sum_{C(m,\, s)} (I_{j_1} - l^{(d+1)/2}\theta) \cdots (I_{j_s} - l^{(d+1)/2}\theta) + l^{t(d+1)/2}\hat{\theta}^t,$$

where $\sum_{C(m,\, s)}$ is defined in section 1 and $\sum_{C(m,\, 0)} \equiv 1$. The above equation is equivalent to

$$m^{t/2}(T_{m,t} - l^{t(d+1)/2}\hat{\theta}^t)$$
$$= m^{t/2}[C(m,t)^{-1} \sum_{C(m,\, t)} (I_{j_1} - l^{(d+1)/2}\theta) \cdots (I_{j_t} - l^{(d+1)/2}\theta)] + o_p(1)$$

provided $d \ge 1$ [see also equation (A5) in the appendix]. Then the results follow since $(m^{1/2} l^{(d+1)/2}\hat{\theta})^t = o_p(1)$ under H_0 and $m^{t/2}[C(m,t)^{-1} \sum_{C(m,\, t)} (I_{j_1} - l^{(d+1)/2}\theta) \ldots (I_{j_t} - l^{(d+1)/2}\theta)]$ has the same limiting distribution as $v_{d+1}^t H_t(Z)$. We note that: (i) if $d = 0$ and $t \ge 2$ then the limiting distribution of $m^{t/2} T_{m,t}$ is different from that of theorem 1(a) and is quite complicated, and (ii) for the indeterminate case of theorem 1(b), the limiting distribution cannot be determined if $t > 1$. Also note that the extension of theorem 1 to the case of several samples is straightforward.

A TU-statistic has some advantages. First, under H_0, its limiting distribution is simple and its quantiles can be computed with required accuracy by a numerical or Monte Carlo method. On the other hand, the corresponding U-statistic's rejection region is approximate as its limiting distribution does not have a closed form. Secondly, under H_1, we can compute powers of TU-statistics according to a rejection region. Thirdly, a TU-statistic can be asymptotically more powerful than its corresponding U-statistic when a kernel is degenerate under the null hypothesis but is non-degenerate under the alternative. Some examples are the circular Kendall's tau (illustrated later in section 3) and other tests of independence for directional data using kernels of order $k \ge 3$ (section 6.2.1 of Lee, 1990). To demonstrate this advantage, we choose $t = 2$ (discussions on how t is chosen are given in the next subsection) for the TU-statistic. Let $C_U(\alpha), C_{\mathrm{TU}}(\alpha), \Phi(\cdot)$ and $G(\cdot)$ denote the size-α critical values of the null limiting distribution of the U- and TU-statistic, the distribution functions of the standard normal and $\chi_1^2 - 1$ respectively. The power of the TU-statistic with $t = 2$ is asymptotically $1 - G(\{C_{\mathrm{TU}}(\alpha) - n\theta^2 l^d\}/v_{d+1}^2)$ (by theorem 1(a,c)). While the power of the corresponding U-statistic is, by normal approximation, $1 - \Phi(\{C_U(\alpha)n^{-d/2} - \sqrt{n}\theta\}/\sigma)$ for large n. Because the latter converges to 1 slower than the former as $n \to \infty$, the TU-statistic is more powerful than its corresponding U-statistic.

When the degeneracy of a U-statistic is indeterminate, $T_{m,t}$ still works. Without loss of generality, this is shown by a $T_{m,t}$ with a kernel of degree 2 as below. We first take the number of subsamples equal to $m = [\sqrt{n}]$. In stage 1, the indeterminate degeneracy of h leads to $d = 0$ (to avoid false power due to a larger d) and compute

$$I_j = l_j^{-1/2} \sum_{l_j} h(X_{i_1}, X_{i_2}),$$

where $\sum_{l_j}$ is defined in stage 1(b) of section 2.1. In stage 2, we take $t = 1$ because the limit of $m^{t/2} T_{m,t}$ under H_0 is indeterminate for $t \ge 2$. Thus

$$T_{m,1} = m^{-1} \sum_{i=1}^{m} I_i.$$

By theorem 1(c), the limiting distribution of $\sqrt{m}(T_{m,1} - \sqrt{\bar{l}}\theta)$ is $N(0, v_1^2)$, namely $T_{m,1}/\sqrt{\bar{l}} \rightarrow_P \theta$. On the other hand, $\sqrt{N}(U_I - \theta) \rightarrow_D N(0, \sigma_k^2)$ and $N = o(\bar{l})$. Thus in an indeterminate case, a TU-statistic can be used for estimation and is more efficient than its corresponding U_I-statistic. Treating I_is as r.v., we can estimate v_1^2 by the jackknife or the bootstrap method. Furthermore, $T_{m,1}$ is more powerful than the associated incomplete U-statistic as $\sqrt{N/n} = o(1)$. We suggest that a TU-statistic be used when the degeneracy of a kernel is indeterminate.

2.3. *Criteria for choosing t and λ*

We shall take both power and size of $T_{m,t}$ into consideration when choosing parameters λ and t. For ease of illustration, we let each $l_j, j = 1, \ldots, m$, equal $\bar{l}$, where $\bar{l} = [n^{\lambda}]$ and $0 < \lambda < 1$.

The second term of a TU-statistic as defined in the right-hand side of (2), $((m\bar{l}^{d+1})^{1/2}\hat{\theta})^t$, becomes large under the alternative hypothesis, and this leads to rejection of H_0 and large power. Suppose that θ is positive under H_1. To have a TU-statistic as powerful as its associated U-statistic, it is required that $((m\bar{l}^{d+1})^{1/2}\hat{\theta})^t$ dominate $n^{(d+1)/2}\hat{\theta}$ or equivalently $m^{(d+1)/2} = o(m^{t/2}\bar{l}^{(d+1)(t-1)/2})$. A simple sufficient condition for this to hold is $t \geq 2$ and $\lambda > 1/2$. Thus for a degenerate case, we should take $t \geq 2$ and $\lambda > 1/2$ in order to have large power. However, for a fixed sample size, as θ tends to zero, a large t will reduce power of a TU-statistic. For any given $\lambda > 1/2$, the optimal choice of t under H_1 will depend on the parameter θ and how fast the centred and normalized TU-statistic converges to its limiting distribution, and is in general a very difficult problem to solve. For simplicity, we suggest using $t = 2$ for a degenerate kernel case. Clearly, with this chosen t and λ, a TU-statistic will also outperform the associated incomplete U-statistic. However, letting λ tend to 1 may induce serious size distortion due to the finite sample distribution of a TU-statistic deviating greatly from its limiting distribution. As in the extreme case where $\lambda = 1 (m = 1)$, $T_{m,t}$ reduces to a U-statistic which we avoid using in high-order degenerate cases. Similar size biases also occur in the indeterminate case where asymptotic normality holds. This is evident in Tables 1–3 in section 4. Thus among those $\lambda > 1/2$, we suggest choosing one that is close to 1/2.

We now demonstrate how one applies these criteria by an example. In case I of section 4, as the degeneracy of the kernel is indeterminate, we take $d = 0$ and $t = 1$. When $n = 200$, for instance, taking $\lambda = 0.7$ results in five subsamples makes the size bias greater than those of other designs. The large λ value results in a poor approximation of $T_{m,t}$ to its null limiting distribution. Having both power and type I error considered, we suggest using $m = 14$ and thus $\lambda = 0.5$ in this case.

3. Applications

In this section, TU-statistics are applied to test independence of paired angles that are of interest in ecology and in marine biology. In the former, the dependence of bird nest direction on creek bed direction is of interest, whereas in the latter the association of the spawning time of a particular fish on the low tide time.

Example 1. The orientations (θ_i) of the nests of 51 noisy scrub birds along the creek bed, together with the corresponding directions (ϕ_i) of creek flow at the nearest point to the nest were from Fisher (1993, pp. 252), as supplied by Dr Graham Smith. Here both θ_is and ϕ_is are assumed uniformly distributed on [0, 1]. We applied a TU-statistic to test the association (Δ) of nest direction on creek bed direction, $H_0: \Delta = 0$ versus $H_1: \Delta > 0$. Among the 51 pairs of orientations, the first 50 were utilized. Let $p_i^T = (\theta_i, \phi_i)$, $1 \leq i \leq n$ and $sgn(x) = 1, 0$ or -1 for $x >, =$ or < 0. The kernel, studied in Fisher & Lee (1982),

$$h(p_1, p_2, p_3) = sgn(\theta_1 - \theta_2)sgn(\theta_2 - \theta_3)sgn(\theta_3 - \theta_1)sgn(\phi_1 - \phi_2)sgn(\phi_2 - \phi_3)sgn(\phi_3 - \phi_1)$$

was utilized. The U-statistic generated by this kernel, called circular Kendall's tau, was used to test independence for bivariate circular data. In this case, the kernel $h(p_1, p_2, p_3)$ is degenerate of order 1 under H_0, and is non-degenerate under H_1 (section 6.2.1 in Lee, 1990). Therefore, as previously discussed in section 2.2, the TU-statistic based on this kernel is more powerful than its corresponding U-statistic in testing the hypothesis. Three designs were studied: $(m, \bar{l}) = (5, 10)$, $(6, 8)$ and $(7, 7)$. In stage 1, taking $d = 1$ (since h is degenerate order-1), we computed m normalized U-statistics $I_j, j = 1, \ldots, m$, and $m = 5, 6$ and 7; then in stage 2, $T_{m,t}$ with $t = 2$. For all three designs, values of $mT_{m,t}$ (–0.15, 0.30 and 0.40) are smaller than the 99% critical value ($v_2^2(\chi_1^2 - 1) = 10.68$), where $v_2 = \sigma_2\sqrt{2}_{C(3,2)}$ and $\sigma_2 = 1/3$. Thus we do not reject H_0. This agrees with the results from the approaches of a U-statistic and an incomplete U-statistic in Fisher & Lee (1982) and in Fisher (1993) respectively.

Example 2. The spawning times of a particular fish and the time of the low tide were from a marine biology study by Robert R. Warner at University of California, Santa Barbara (Lund, 1999). The data were gathered in St Croix, the US Virgin Islands. Before one provides a model for the spawning time of the fish (T_S) on the corresponding low tide time (T_{LT}), one should test the dependence of T_S on T_{LT}. Let $\mathbf{p}_i^T = (\theta_i, \phi_i)$, for $i = 1, 2$.

To this end, we used a TU-statistic based on the kernel

$$h(\mathbf{p}_1, \mathbf{p}_2) = \frac{\sin(\theta_1 - \theta_2)\sin(\phi_1 - \phi_2)}{\{E\sin^2(\theta_1 - \theta_2)E\sin^2(\phi_1 - \phi_2)\}^{1/2}}. \quad (4)$$

This kernel is equal to that of the test for circular association used in Fisher & Lee (1983) as $n \to \infty$. To see the pattern of the paired data, we coverted the period 0–20 h of T_{LT} to $[0, 2\pi)$ (called θ) and the period 11.76–14.98 h of T_S to $[0, 2\pi)$ (called ϕ). One hundred pair angles that have no missing values were utilized to compute the TU-statistic.

The p.d.f. of a von-Mises distribution with mean μ and concentration parameter κ (VMS(μ, κ) henceforth) is:

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa\cos(\theta - \mu)\}, \quad 0 < \theta < 2\pi, \ 0 < \kappa < \infty, \quad (5)$$

where $I_0(\kappa) = (2\pi)^{-1}\int_0^{2\pi} \exp\{\kappa\cos(\phi - \mu)\}\,d\phi$, is the modified Bessel function of order zero. We note that the concentration of the data around the mean is greater with larger κ values. Under the null hypothesis, we applied the algorithm in Shieh & Johnson (2005) to obtain that the marginal distribution of θ_is and ϕs follow VMS(π, 0.5) and VMS(π, 1.1) respectively. As VMS(π, 0.5) is close to circular uniform $[0, 2\pi)$, the degeneracy of this kernel is indeterminate.

Under three designs: $(m, \bar{l}) = (3, 33)$, $(5, 20)$, $(9, 16)$, we computed the TU-statistic as follows. In stage 1, taking $d = 0$, we computed m normalized U-statistics I_j, where $j = 1, \ldots, m$, $m = 3, 5$ and 9; in stage 2, $T_{m,t}$ with $t = 1$. For all three designs, values of $\sqrt{m}T_{m,t}$ (0.205, 0.082 and −0.180) are smaller than the 95% critical value as $\sigma_k^2 = 1.011$ and n var$(U_n) = 0.250$. Thus, we do not reject H_0 and conclude that there is no evidence to claim dependence of the spawning time of the fish on the corresponding low tide time.

4. Simulations

In this section, we compare the power of TU-statistic to its associated incomplete U-statistic under two cases with some finite samples. The test statistic for association between angular r.v. (Θ and Φ) in Fisher & Lee (1983) assumes the form

$$\sum_{1 \le i < j \le n} \frac{\sin(\theta_i - \theta_j)\sin(\phi_i - \phi_j)}{\{\sum \sin^2(\theta_i - \theta_j) \sum \sin^2(\phi_i - \phi_j)\}^{1/2}} = \frac{U_n^{(1)}}{(U_n^{(2)} U_n^{(3)})^{1/2}},$$

where $\mathbf{p}_i^T = (\theta_i, \phi_i)$, for $i = 1, 2, \ldots, n$; $U_n^{(j)}$, $j = 1, 2$ and 3 are U-statistics based on kernels $g_1(\mathbf{p}_1, \mathbf{p}_2) = \sin(\theta_1 - \theta_2)\sin(\phi_1 - \phi_2)$, $g_2(\mathbf{p}_1, \mathbf{p}_2) = \sin^2(\theta_1 - \theta_2)$ and $g_3(\mathbf{p}_1, \mathbf{p}_2) = \sin^2(\phi_1 - \phi_2)$ respectively. As $n \to \infty$, the kernel of this test is equal to that in (4). For simplicity, we study the power of $T_{m,t}$ and its associated incomplete U-statistic U_I with the kernel in (4). As mentioned in section 1, U_I, an average of N randomly chosen evaluations of h with $N/n \to 0$, may be used in the indeterminate degeneracy case. Thus it is of interest to compare $T_{m,t}$ with the associated U_I with $N = [n^{0.9}]$ randomly chosen number of evaluations.

The alternative model is constructed by taking

$$\Theta = \Phi \cdot I[U_1 \le p] + \Upsilon \cdot I[U_1 > p], \tag{6}$$

where Φ and Υ assume VMS(π, κ) with $\kappa = 0.5$ and 0 in cases I and II studied respectively. U_1 is uniform (0, 1) and p is some fixed constant in the interval [0, 1]. We control the association of Θ and Φ by p, $p = \sqrt{\rho_T}$, where $\rho_T = E[h(\mathbf{p}_1, \mathbf{p}_2)]$ with $h(\mathbf{p}_1, \mathbf{p}_2)$ defined in (4) and $0 \le \rho_T \le 1.0$. From (5), it is clear that a VMS with $\kappa = 0$ reduces to a circular uniform $[0, 2\pi)$. Under the null hypothesis, the kernel in (4) is degenerate of order 1 when either θ_i or ϕ_i assumes VMS(π, κ) $\kappa = 0$ (section 4 in Fisher & Lee, 1983). The kernel is non-degenerate when κ deviates from 0 enough that the data concentrate on their non-zero mean (i.e. the mean resultant length of the data is non-zero).

We study two interesting cases. Case I, both Θ and Υ, in (6) follow VMS(π, 0.5) and Φ is generated by the alternative. In this case, κ is close to but not equal to 0. Hence from data, it is hard to tell which distributions Θ and Φ follow. We further study case II, in which both Θ and Υ, in (6) follow VMS(π, 0). After some computation, we obtained that the kernel is degenerate of order 1 under H_0 and non-degenerate under H_1.

Under the alternative model, we compute both $T_{m,t}$ and U_I with sample sizes 50, 100 and 200. Their powers are obtained from comparing against the 0.05 level critical values from their limiting distributions. For a given sample size, five designs for $T_{m,t}$ are studied in both cases. The number of subsamples (m) and the average subsample size ($\bar{l}$) are listed in Tables 1–6. When computing $T_{m,t}$, we take $t = 1$ and $d = 0$ in case I and $t = d + 1 = 2$ in case II according to theorem 1. The results of cases I and II are summarized in Tables 1–3 and Tables 4–6 respectively.

Tables 1–6 show that the power of $T_{m,t}$ increases as n and ρ_T increase. For all sample sizes studied, as λ tends to 1/2, the empirical size of $T_{m,t}$ tends to 0.05. There is a positive size

Table 1. *Empirical powers of $T_{m,t}$ and U_I observed in 5000 simulations of sample 50*

ρT	$(m, \bar{l}) = (13,3)$	(10,5)	(7,7)	(5,10)	(3,16)	U_I
0.0	0.073	0.059	0.049	0.045	0.040	0.047
0.1	0.217	0.224	0.219	0.209	0.205	0.200
0.2	0.375	0.433	0.411	0.429	0.414	0.381
0.3	0.569	0.640	0.641	0.640	0.630	0.584
0.4	0.734	0.805	0.805	0.813	0.799	0.762
0.5	0.859	0.916	0.916	0.918	0.923	0.873
0.6	0.932	0.969	0.967	0.973	0.968	0.948
0.7	0.975	0.993	0.992	0.991	0.993	0.985
0.8	0.994	0.999	0.998	0.999	0.999	0.997
0.9	0.999	1.000	1.000	1.000	1.000	1.000
1.0	1.000	1.000	1.000	1.000	1.000	1.000

Table 2. *Empirical powers of $T_{m,t}$ and U_I observed in 5000 simulations of sample 100*

ρT	$(m,\bar{l})=(13,7)$	(9,11)	(6,16)	(5,20)	(3,33)	U_I
0.0	0.074	0.040	0.041	0.038	0.034	0.056
0.1	0.354	0.301	0.289	0.302	0.299	0.275
0.2	0.659	0.632	0.617	0.632	0.633	0.546
0.3	0.860	0.859	0.860	0.867	0.857	0.786
0.4	0.962	0.961	0.961	0.965	0.965	0.922
0.5	0.995	0.994	0.995	0.994	0.999	0.977
0.6	1.000	0.999	1.000	1.000	1.000	0.998
0.7	1.000	1.000	1.000	1.000	1.000	1.000

Table 3. *Empirical powers of $T_{m,t}$ and U_I observed in 5000 simulations of sample 200*

ρT	$(m,\bar{l})=(25,8)$	(14,14)	(8,25)	(6,33)	(5,40)	U_I
0.0	0.050	0.043	0.042	0.035	0.038	0.050
0.1	0.469	0.458	0.464	0.462	0.473	0.395
0.2	0.852	0.850	0.874	0.861	0.862	0.757
0.3	0.985	0.979	0.985	0.986	0.980	0.941
0.4	0.999	1.000	0.999	0.999	0.999	0.994
0.5	1.000	1.000	1.000	1.000	1.000	0.999
0.6	1.000	1.000	1.000	1.000	1.000	1.000

Table 4. *Empirical powers of $T_{m,t}$ and U_I observed in 5000 simulations of sample 50*

ρT	$(m,\bar{l})=(13,3)$	(10,5)	(7,7)	(5,10)	(3,16)	U_I
0.0	0.083	0.062	0.046	0.039	0.034	0.056
0.1	0.130	0.201	0.253	0.329	0.410	0.165
0.2	0.336	0.553	0.638	0.719	0.778	0.325
0.3	0.650	0.856	0.900	0.930	0.949	0.536
0.4	0.879	0.966	0.979	0.987	0.988	0.705
0.5	0.977	0.995	0.997	0.997	0.997	0.849
0.6	0.997	0.999	1.000	1.000	0.999	0.934
0.7	1.000	1.000	1.000	1.000	1.000	0.977

Table 5. *Empirical powers of $T_{m,t}$ and U_I observed in 5000 simulations of sample 100*

ρT	$(m,\bar{l})=(13,7)$	(9,11)	(6,16)	(5,20)	(3,33)	U_I
0.0	0.058	0.049	0.041	0.036	0.035	0.051
0.1	0.432	0.585	0.664	0.720	0.779	0.218
0.2	0.911	0.957	0.969	0.976	0.980	0.465
0.3	0.994	0.998	0.999	0.999	0.998	0.728
0.4	1.000	1.000	1.000	1.000	1.000	0.889
0.5	1.000	1.000	1.000	1.000	1.000	0.971

Table 6. *Empirical powers of $T_{m,t}$ and U_I observed in 5000 simulations of sample 200*

ρT	$(m,\bar{l})=(25,8)$	(14,14)	(8,25)	(6,33)	(5,40)	U_I
0.0	0.051	0.048	0.042	0.041	0.036	0.049
0.1	0.816	0.912	0.961	0.973	0.977	0.308
0.2	1.000	1.000	1.000	1.000	1.000	0.671
0.3	1.000	1.000	1.000	1.000	1.000	0.910

bias when m is large ($m=13$) for $n=50$ and 100 in both cases. The pattern that the power of $T_{m,t}$ increases as values of m decreases (as λ increases) is clearly shown in case II. For all sample sizes studied, except the design $(m,\bar{l})=(13,3)$ under $n=50$ and $\rho_T=0.1$, $T_{m,t}$ is more powerful than U_I when $\rho_T\geq 0.1$ (when Θ and Φ are correlated). This shows that TU is more efficient than U_I in almost all cases studied.

5. Asymptotic relative efficiency of TU-statistics versus U_I

For a fixed size α, the Pitman asymptotic relative efficiency (ARE) (Pitman, 1949) is an index of the relative sample sizes required for two sequences of tests to achieve the same power against the same alternative hypothesis. This ARE exists only when both sequences of tests converges to the same limiting distribution. Thus we compare the ARE of TU versus U_I based on an indeterminate kernel as in this case both sequences of tests have asymptotically a normal distribution. To be concrete, we can use the kernel for testing independence of paired bivariate circular data in (4) of section 4 as an example.

For testing $H_0:\theta=0$ versus $H_1:\theta>0$, the sequences of tests of TU and U_I converge to normal distributions. Let $\sigma_k^2=\text{var}\{h(X_1,\dots,X_k)\}$. By theorem 1(b) in section 2 and theorem 1 of p. 204 in Lee (1990),

$$\sqrt{m}T_{m,1}\rightarrow_D N(0,v_{d+1}^2)$$

and let $N/n\rightarrow 0$ as $n\rightarrow\infty$,

$$\sqrt{N}U_I\rightarrow_D N(0,\sigma_k^2),$$

respectively. $T_{m,1}$ rejects H_0 if $T_{m,1}\geq a_n v_{d+1}/\sqrt{m}$ and U_I rejects H_0 if $U_I\geq a_n\sigma_k/\sqrt{N}$, where $a_n\rightarrow Z_{1-\alpha}$ as $n\rightarrow\infty$. Both tests will have asymptotically a size α. Under $H_1:\theta>0$, by theorem 1(c) the asymptotic power function of TU is

$$\psi_n^{(1)}(\theta)\approx\Phi\left(\frac{\sqrt{n}\theta}{v_{d+1}}-Z_{1-\alpha}\right)$$

since

$$S_1=\sum_{j=1}^{m}\frac{\sqrt{l_j}}{m}\sim\sqrt{l}.$$

Similarly, the asymptotic power function of U_I is

$$\psi_n^{(2)}(\theta)\approx\Phi\left(\frac{\sqrt{N}\theta}{\sigma_k}-Z_{1-\alpha}\right).$$

Consider the sequence of power functions $\{\psi_n^{(1)}(\theta_n),n\geq 1\}$ evaluated at $\theta_n=\delta/\sqrt{n}$ with $\delta>0$,

$$\lim_{n\rightarrow\infty}\psi_n^{(1)}\left(\frac{\delta}{\sqrt{n}}\right)=\Phi\left(\frac{\delta}{v_{d+1}}-Z_{1-\alpha}\right)=\psi^*,$$

where $0<\psi^*<1$. Similarly, consider the sequence of power functions of U_I, $\{\psi_{n'}^{(2)}(\theta_n),n\geq 1\}$. The sample size n' required for U_I to achieve the same asymptotic power as $T_{m,t}$, namely $\psi_{n'}^{(2)}(\theta_n)\rightarrow\psi^*$ as $n\rightarrow\infty$, satisfies $n=o(n')$. Thus the Pitman ARE of TU dominates that of its associated U_I.

6. Discussion

We have shown that for testing hypothesis when the kernel of a U-statistic has an indeterminate degeneracy or is high-order degenerate under H_0, a TU-statistic can be applied while its associate U-statistic fails. In the case of indeterminate degeneracy, an incomplete U-statistic would work for testing hypothesis, but it has much smaller Pitman ARE than its associated TU-statistic. For estimation, a U_I is also less efficient than its corresponding TU-statistic as shown in section 2.2. When the kernel is high-order degenerate, the limiting distribution of a TU-statistic has a chi-square-type closed form, while the associated U-statistic does not.

For a class of tests based on kernels that are degenerate under the null hypothesis but non-degenerate under the alternative, TU-statistics are shown to be more powerful than their corresponding U-statistics. When a U-statistic with a kernel of order k is non-degenerate under H_0, the corresponding TU-statistic may be reduced to an incomplete U-statistic as follows. In stage 1, taking every subsample size l equal to k and discarding the rest of the $n-mk$ observations, we have the number of subsamples $m=[n/k]$. Further, we take $I_j = h(X_{i_1},\ldots,X_{i_k})$ for $j=1,\ldots,m$. With $t=1$, we have that $T_{m,t}=m^{-1}\sum_m h(X_{i_1},\ldots,X_{i_k})$, where $\sum_m$ denotes a summation over m randomly chosen evaluations from those $C(m,t)$ combinations of k distinct indices. The limiting distribution of $T_{m,t}$ is normal which agrees with the incomplete U-statistic theory (Janson, 1984). However, the ARE of $T_{m,t}$ is lower than that of the corresponding U-statistic. Hence for a non-degenerate kernel, we suggest that a regular U-statistic be used.

We are aware of a competitive class of statistics, subsampling U-statistics (U_s) (Politis *et al.*, 1999), which can also resolve the problems we aim at. However, the procedure of a subsampling U-statistic is complicated to implement in the indeterminate degeneracy case. First, one has to estimate the convergence rate with a few indeterminate parameters in the formula. The estimated convergence rate may differ significantly if various values are chosen for the parameters. For a large sample size n, a subsampling U-statistic may be difficult to compute, because $C(n,b)$ can be very large. Randomly choosing N evaluations from $C(n,b)$ can approximate its associated subsampling U-statistic (section 2.4, Politis *et al.*, 1999). However, there is no theoretical guidelines on how large N should be, and assessing the performance of the approximation empirically would not be possible if a complete evaluation of U_s is not available.

Acknowledgements

We are grateful to three referees, an associate editor and the editor for constructive comments which improved the manuscript much. We wish to thank Dr Min-Te Chao for carefully reading our manuscript and Y. C. Jiang and Chi-Ming Chuang for computational assistance while the manuscript was re-written. This research was supported in part by Academia Sinica Thematic grant and a grant from National Science Council, ROC.

References

Ahmad, I. A. (1993). Modification of some goodness-of-fit statistics to yield asymptotically normal null distributions. *Biometrika* **80**, 466–472.

Anderson, T. W. & Darling, D. A. (1952). Asymptotic theory of certain ‘goodness of fit’ criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193–212.

Bierens, H. J. & Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65**, 1129–1151.

D’Agostino, R. B. & Stephens, M. A. (1986). *Goodness of fit tests*. Marcel Dekker, New York, NY.

Fan, Y. & Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *J. Nonparametr. Statist.* **10**, 245–271.

Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge University Press, Cambridge.

Fisher, N. I. & Lee, A. J. (1982). Nonparametric measures of angular-angular association. *Biometrika* **69**, 315–321.

Fisher, N. I. & Lee, A. J. (1983). A correlation coefficient for circular data. *Biometrika* **70**, 327–332.
Fisher, N. I. & Lee, A. J. (1986). Correlation coefficients for random variables on a unit sphere or hypersphere. *Biometrika* **73**, 327–332.
Gregory, G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5**, 110–123.
Halmos, H. C. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17**, 34–43.
Hoeffding, W. (1948). A class of statistics with asymptotically normal distributions. *Ann. Math. Statist.* **19**, 293–325.
Janson, S. (1984). The asymptotic distributions of incomplete U-statistics. *Z. Wahrsch. Verw. Gebiete* **66**, 495–505.
Lee, A. J. (1990). *U-statistics – theory and practice*. Marcel Dekker, New York, NY.
Linton, O. & Gozalo, P. (1995). *A nonparametric test of conditional independence*. Yale University, New Haven, CT.
Lund, U. (1999). Least circular distance regression for directional data. *J. Appl. Statist.* **26**, 723–733.
Major, P. (1981). *Multiple Wiener-Ito integrals: with applications to limit theorems.* Lecture Notes in Mathematics no. 849. Springer-Verlag, New York, NY.
Persson, T. (1979). A new way to obtain Watson's U^2. Scand. J. Statist. **6**, 119–122.
Pitman, E. J. G. (1949). *Lecture notes on nonparametric statistical inference*. Columbia University, New York, NY.
Politis, D. N., Romano, J. P. & Wolf, M. (1999). *Subsampling*. Springer, New York, NY.
Rubin, H. & Vitale, R. A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8**, 165–170.
Shieh, G. S. & Johnson, R. A. (2005). Inferences based on a bivariate distribution with von Mises marginals. *Ann. Inst. Statist. Math.* **57**, 789–802.
Shieh, G. S., Johnson, R. A. & Frees, E. W. (1994). Weighted degenerate U-statistics and testing independence of bivariate circular data. *Statist. Sinica* **4**, 729–747.
Shorack, G. R. & Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley, New York, NY.
Singh, H., Hnizdo, V. & Demchuk, E. (2002). Probabilistic model for two dependent circular variables. *Biometrika* **89**, 719–723.
Stephens, M. A. (1979). Vector correlation. *Biometrika* **66**, 41–48.
Watson, G. S. (1962). Goodness-of-fit tests on a circle. II. *Biometrika* **49**, 57–63.

Received March 2005, in final form February 2006

Grace S. Shieh, Institute of Statistical Science, Academia Sinica, Nan-Gang district, Taipei 115, Taiwan.
E-mail: gshieh@stat.sinica.edu.tw

Appendix

Proof of theorem 1

As part (b) and the indeterminate case of part (c) are straightforward, we focus on part (a) and the degenerate case of part (c). As in the text, the sample size n is suppressed for quantities that depend on it. We may assume without loss of generality that the sizes l_j, $1 \le j \le m$, of the m subsamples are all equal to an integer $l \equiv n/m$. For each given j, $1 \le j \le m$, the normalized U-statistic I_j constructed from the jth subsample in stage 1 has variance converging to

$$\lim_{l\to\infty} \mathrm{var}(I_j) \equiv v_{d+1}^2 = \begin{cases} \sigma_k^2 + \lim_{n\to\infty} n\ \mathrm{var}(U_n) & \text{if } d=0 \\ C(k,d+1)^2(d+1)!\sigma_{d+1}^2 & \text{if } d \ge 1 \end{cases}, \tag{A1}$$

uniformly for all $j=1,\ldots,m$ (theorem 4 of section 4.3.1 and theorem 3 of section 1.3, Lee, 1990). Let $\{Z_{m,i}, 1 \le m, 1 \le i \le m\}$ be an array of row-wise independent r.v. with $E(Z_{m,i})=0$ and $E(Z_{m,i}^2)=1$ for all m, i. Furthermore, assume that the sequence of r.v. $V_{m,1}=m^{-1/2}\sum_i Z_{m,i}$ converge to a standard normal in distribution as $m\to\infty$. Let t be a positive integer. Define $V_{m,t}=m^{-t/2}\sum_{\{m,t\}} Z_{m,i_1}\ldots Z_{m,i_t}$, where $\sum_{\{m,t\}}$ denotes summation over all t-tuple mutually distinct indices $(i_1,\ldots,i_t)$ with $1 \le i_1,\ldots,i_t \le m$. Obviously, $V_{m,t}=O_p(1)$ for all $t \ge 1$. As

$$m^{-t/2}\sum_{\{m,t\}} Z^2_{m,i_1} Z_{m,i_2}\dots Z_{m,i_t}$$
$$= m^{-t/2}\sum_{\{m,t\}}(Z^2_{m,i_1}-1)Z_{m,i_2}\dots Z_{m,i_t} + m^{-t/2}(m-t+1)\sum_{\{m,t-1\}} Z_{m,i_1}\dots Z_{m,i_{t-1}},$$

we have

$$V_{m,1}V_{m,t} = V_{m,t+1} + t[m^{-1/2}\{m^{-t/2}\sum_{\{m,t\}}(Z^2_{m,i_1}-1)Z_{m,i_2}\dots Z_{m,i_t}$$
$$+ m^{-t/2}(m-t+1)\sum_{\{m,t-1\}} Z_{m,i_1}\dots Z_{m,i_{t-1}}\}]$$
$$= V_{m,t+1} + tV_{m,t-1} + O_p(m^{-1/2}). \tag{A2}$$

If, for all $t \geq 2$,

$$H_t(V_{m,1}) = V_{m,t} + O_p(m^{-1/2}) \quad \text{as } m \to \infty, \tag{A3}$$

then immediately

$$V_{m,t} \to H_t(Z) \tag{A4}$$

in distribution, where Z is the standard normal r.v. In the following, we shall prove (A3) by an induction argument. It is clear that (A3) holds for $t=2$. Assume (A3) is true for all t and $2 \leq t \leq u$. By recursive relations (3) in the main text and (A2)

$$H_{u+1}(V_{m,1}) = V_{m,1}H_u(V_{m,1}) - uH_{u-1}(V_{m,1})$$
$$= V_{m,1}(V_{m,u} + O_p(m^{-1/2})) - u(V_{m,u-1} + O_p(m^{-1/2})) = V_{m,u+1} + O_p(m^{-1/2}),$$

which establishes (A3). To apply (A4) to prove (a) and (c), we first see that, for $d \geq 1$ and $0 \leq s \leq t$,

$$C(m-s, t-s)[l^{(d+1)/2}(\theta - \hat{\theta})]\sum_{C(m,\,s)}\prod_{i=1}^{s}(I_{ji} - l^{(d+1)/2}\theta) = O_p(m^{-(t-s)((d+1)/2-1)+s}).$$

Thus

$$m^{t/2}(T_{m,t} - l^{t(d+1)/2}\hat{\theta}^t)$$
$$= m^{t/2}[C(m,t)^{-1}\sum_{C(m,\,t)}(I_{j_1} - l^{(d+1)/2}\theta)\dots(I_{j_t} - l^{(d+1)/2}\theta)] + o_p(1) \tag{A5}$$

and, under H_0,

$$m^{t/2}T_{m,t} = m^{t/2}(T_{m,t} - l^{t(d+1)/2}\hat{\theta}^t) + o_p(1).$$

Note that $EI_j = l^{(d+1)/2}\theta$. Define

$$I^*_j = \{\text{var}(I_j)\}^{-1/2}(I_j - l^{(d+1)/2}\theta)$$

to form a collection of arrays $\{I^*_j, 1 \leq m, j = 1,\dots,m\}$ of row-wise independent r.v. with $E(I^*_j) = 0$ and $E(I^*_j)^2 = 1$. Set

$$W_{m,t} = m^{-t/2}\sum_{\{m,t\}} I^*_{j_1}\dots I^*_{j_t}.$$

Then, by (A1) and (A5),

$$m^{t/2}(T_{m,t} - l^{t(d+1)/2}\hat{\theta}^t) = v^t_{d+1}W_{m,t} + o_p(1).$$

Analogous to (A4), replacing $V_{m,1}$ and $V_{m,t}$ of (A3) by $m^{-1/2}\sum_j I^*_j$ and $W_{m,t}$, respectively, we obtain that $W_{m,t}$ converges in distribution to $H_t(Z)$ as $m \to \infty$. This completes the proof.