JOURNAL OF MULTIVARIATE ANALYSIS 3, 469–482 (1973)

Asymptotically Optimal Sequential Estimation of Regular Functionals of Several Distributions Based on Generalized U-Statistics*

GEORGE W. WILLIAMS

University of Michigan, Ann Arbor, Michigan 48104

AND

PRANAB KUMAR SEN

University of North Carolina, Chapel Hill, North Carolina 27514

For the several sample problem, a vector of estimable parameters is considered. For a fixed total sample size, a multistage (sequential) procedure based on generalized U-statistics is developed for choosing a partition of this sample size into individual sample sizes for which the generalized variance of the estimator of the parameter vector is asymptotically minimized.

1. Introduction

Let $\{\mathbf{X}_{ki}, i \geqslant 1\}$ be a sequence of independent and identically distributed random vectors (iidrv) with a $p(\geqslant 1)$-variate distribution function $(df) F_k(\mathbf{x})$, $\mathbf{x} \in R^p$, the p-dimensional Euclidean space, for $k = 1,..., c(\geqslant 2)$; all these c sequences are assumed to be mutually stochastically independent. Let us denote by

$$\mathbf{F} = (F_1,..., F_c)', \qquad \boldsymbol{\theta}(\mathbf{F}) = (\theta_1(\mathbf{F}),..., \theta_t(\mathbf{F}))', \qquad t \geqslant 1, \tag{1.1}$$

Received November 13, 1972; revised May 30, 1973.

AMS 1970 classification: 62G20, 62H99, 62L12.

Key words and phrases: Estimable parameters, generalized variance, generalized U-statistics, asymptotic optimality, asymptotic consistency and normality of the estimates.

* Work supported partially by National Institutes of Health, Training Grant No. T01 GM00038 and partially by the Aerospace Research Laboratories, U.S. Air Force Systems Command, Contract F33615-71-C-1927. Reproduction in whole or in part permitted for any purpose of the U.S. Government.

where the estimable parameter $\theta_i(\mathbf{F})$ corresponds to the *kernel*

$$\phi_i(\mathbf{X}_{kj}, 1 \leqslant j \leqslant m_{ki}, 1 \leqslant k \leqslant c) \text{ of } \textit{degree } \mathbf{m}_i = (m_{1i}, \dots, m_{ci})', \tag{1.2}$$

for $i = 1, \dots, t$, so that the m_{ki} are positive integers, and

$$\begin{aligned} \theta_i(\mathbf{F}) &= E\phi_i(\mathbf{X}_{kj}, 1 \leqslant j \leqslant m_{ki}, 1 \leqslant k \leqslant c) \\ &= \int \underset{R^{pm_i}}{\cdots} \int \phi_i(\mathbf{x}_{ki}, 1 \leqslant j \leqslant m_{kj}, 1 \leqslant k \leqslant c) \prod_{k=1}^{c} \prod_{j=1}^{m_{ki}} dF_k(\mathbf{x}_{kj}); \end{aligned} \tag{1.3}$$

$m_i = m_{1i} + \cdots + m_{ci}$, for $1 \leqslant i \leqslant t$. We may assume (without any loss of generality) that ϕ_i is symmetric in $\mathbf{X}_{kj}$, $1 \leqslant j \leqslant m_{ki}$, for each $k(=1, \dots, c)$, and $i = 1, \dots, t$. Also, by $\mathbf{a} \geqslant \mathbf{b}$ (or $\mathbf{a} \leqslant \mathbf{b}$), we mean the coordinatewise inequalities $a_i \geqslant b_i$ (or $a_i \leqslant b_i$), for all i. Then, for $\mathbf{n} = (n_1, \dots, n_c)' \geqslant \mathbf{m}_i$, the generalized U-statistics corresponding to $\theta_i(\mathbf{F})$ is

$$U_i(\mathbf{n}) = \binom{\mathbf{n}}{\mathbf{m}_i}^{-1} \sum\nolimits_{(\mathbf{n})}^{*} \phi_i(\mathbf{X}_{k\alpha_{kj}}, j = 1, \dots, m_{ki}, 1 \leqslant k \leqslant c) \tag{1.4}$$

where $\binom{\mathbf{a}}{\mathbf{b}} = \prod_{k=1}^{c} \binom{a_k}{b_k}$ and the summation $\sum_{(\mathbf{n})}^{*}$ extends over all possible $1 \leqslant \alpha_{k1} < \cdots < \alpha_{km_{ki}} \leqslant n_k$, $k = 1, \dots, c$, for $i = 1, \dots, t$. Let then

$$\mathbf{U}(\mathbf{n}) = (U_1(\mathbf{n}), \dots, U_t(\mathbf{n}))', \qquad \mathbf{n} \geqslant \mathbf{m}^*, \tag{1.5}$$

where $\mathbf{m}^* = (m_1^*, \dots, m_c^*)'$, $m_k^* = \max_{1 \leqslant i \leqslant t} m_{ki}$, $k = 1, \dots, c$.

Note that $\mathbf{U}(\mathbf{n})$, for fixed $\mathbf{n}$, is an unbiased estimator of $\boldsymbol{\theta}(\mathbf{F})$, and, in fact, for a general class of $\mathbf{F}$, it is the minimum concentration ellipsoid unbiased estimator. For other properties, we may refer to Section 3.2 of [6]. Our interest centers around the situation where the total sample size n is given, and we require to choose a partition $\mathbf{n} = (n_1, \dots, n_c)$ such that $\sum_{k=1}^{c} n_k = n$ and the corresponding $\mathbf{U}(\mathbf{n})$ is optimal in a certain sense. Let

$$\boldsymbol{\Gamma}(\mathbf{n}) = E([\mathbf{U}(\mathbf{n}) - \boldsymbol{\theta}(\mathbf{F})][\mathbf{U}(\mathbf{n}) - \boldsymbol{\theta}(\mathbf{F})]'), \qquad \mathbf{n} \geqslant \mathbf{m}^*. \tag{1.6}$$

For $t = 1$, $\Gamma(\mathbf{n})$ is the variance of $U_1(\mathbf{n})$, and interpreting optimality by the minimum variance criterion, our problem reduces to that of selecting $\mathbf{n}$ such that $\operatorname{var}[U_1(\mathbf{n})]$ is smallest among all $\mathbf{n}$ such that $\mathbf{n}'\mathbf{1} = n$. For $t \geqslant 2$, a natural extension of the minimum variance criterion is the ***minimum generalized variance criterion***, where

$$\text{generalized variance of } \mathbf{U}(\mathbf{n}) = \det \boldsymbol{\Gamma}(\mathbf{n}) = \| \boldsymbol{\Gamma}(\mathbf{n}) \|. \tag{1.7}$$

In fact, Wilks [9] has advocated the t-th root of the generalized variance as a measure of the efficacy of a t-vector of estimators, and we may refer to [6, Chap. 6] for details. Thus, our problem is to choose an $\mathbf{n}_0 = (n_{10}, \ldots, n_{c0})$ such that $\sum_{k=1}^{c} n_{k0} = n$ and

$$\| \boldsymbol{\Gamma}(\mathbf{n}_0) \| = \inf\{\| \boldsymbol{\Gamma}(\mathbf{n}) \| : \mathbf{n}'\mathbf{1} = n\}. \tag{1.8}$$

As will be seen in Section 2, $\boldsymbol{\Gamma}(\mathbf{n})$ involves, apart from $\mathbf{n}$, a set of unknown parameters which are all regular functionals of $\mathbf{F}$. Thus, in general, both $\mathbf{n}_0$ and $\boldsymbol{\Gamma}(\mathbf{n}_0)$ depend on $\mathbf{F}$, and we denote these by $\mathbf{n}_0(\mathbf{F})$ and $\boldsymbol{\Gamma}(\mathbf{n}_0, \mathbf{F})$, respectively. Hence, we require to find an estimator $\mathbf{N} = (N_1, \ldots, N_c)$ of $\mathbf{n}_0(\mathbf{F})$, such that for a broad class of df's, the corresponding $\mathbf{U}(\mathbf{N})$ has a generalized asymptotic variance "close to" $\| \boldsymbol{\Gamma}(\mathbf{n}_0, \mathbf{F}) \|$.

It is shown here that there exists a sequential (multistage) procedure which leads to a solution $\mathbf{N}$ (stochastic vector), such that $\sum_{k=1}^{c} N_k = n$ and for large n, $n^{1/2}[\mathbf{U}(\mathbf{N}) - \boldsymbol{\theta}(\mathbf{F})]$ has asymptotically a multinormal distribution with null mean vector and a dispersion matrix $\boldsymbol{\Gamma}^*$, where

$$\lim_{n\to\infty}\{\| \boldsymbol{\Gamma}^* \| / n^t \| \boldsymbol{\Gamma}(\mathbf{n}_0, \mathbf{F}) \|\} = 1, \tag{1.9}$$

for a class $\mathscr{F}$ of c-tuplets of df's $\{\mathbf{F}\}$. Since $\| \boldsymbol{\Gamma}^* \|$ is the asymptotic generalized variance of $n^{1/2}[\mathbf{U}(\mathbf{n}) - \boldsymbol{\theta}(\mathbf{F})]$, the asymptotic optimality of $\mathbf{U}(\mathbf{N})$ follows.

For $t = 1$, Yen [10] considered a two-stage procedure for finding the (asymptotically) minimum variance estimator of a regular functional of $(F_1, \ldots, F_c)$, where she estimated $\boldsymbol{\Gamma}(\mathbf{n})$ from an initial sample from each distribution. We consider here a multistage procedure which, besides including her procedure as a particular case, provides scope for updating the estimates of $\boldsymbol{\Gamma}(\mathbf{n})$ through the successive stages. Whereas Yen considered unbiased estimators of the variance of U-statistics, we adopt the structural convergence properties of U-statistics (cf. [7]), and thereby consider alternative estimators which are computationally simpler but asymptotically equally efficient. Moreover, through the use of some recently developed almost sure (a.s.) convergence results on generalized U-statistics (viz [8]), we are able to prove our results under regularity conditions weaker than those in [10].

In Section 2, we consider the proposed procedure along with the preliminary notions. Sections 3 deals with some basic results on U-statistics needed for proving the main theorem which is considered in Section 4. Throughout the paper, we consider the case of $c = 2$, while the last section includes a discussion of the general case of $c \geqslant 2$.

2. Preliminary Notions and the Proposed Procedure

For every $\mathbf{0} \leqslant \mathbf{d} = (d_1, d_2)' \leqslant \mathbf{m}_i \leqslant (m_{1i}, m_{2i})'$, we define the conditional expectation

$$\phi_{i,\mathbf{d}}(\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kd_k}, k = 1, 2) = E\{\phi_i(\mathbf{x}_{k1}, \ldots, \mathbf{x}_{kd_k}, \mathbf{X}_{kd_k+1}, \ldots, \mathbf{X}_{km_{ki}}, k = 1, 2)\}, \tag{2.1}$$

for $1 \leqslant i \leqslant t$, and let

$$\zeta_{ij}(\mathbf{d}, \mathbf{F}) = E\left\{\prod_{\ell=i,j} \phi_{\ell,\mathbf{d}}(\mathbf{X}_{k1}, \ldots, \mathbf{X}_{kd_k}, k = 1, 2)\right\} - \theta_i(\mathbf{F})\,\theta_j(\mathbf{F}), \tag{2.2}$$

for $\mathbf{0} \leqslant \mathbf{d} \leqslant ([\min(m_{1i}, m_{1j})],\ [\min(m_{2i}, m_{2j})])'$, $1 \leqslant i \leqslant j \leqslant t$. We term $\theta_i(\mathbf{F})$ as stationary of order zero if $\zeta_{ii}(1, 0; \mathbf{F})$ or $\zeta_{ii}(0, 1; \mathbf{F}) > 0$, $\forall 1 \leqslant i \leqslant t$. Hence, $\boldsymbol{\theta}(\mathbf{F})$ is stationary of order zero if $\theta_i(\mathbf{F})$ is stationary of order zero for every $i(=1, \ldots, t)$. Then, we have (cf. [6, p. 66])

$$\operatorname{cov}[U_i(\mathbf{n}), U_j(\mathbf{n})] = \binom{n_1}{m_{1j}}^{-1} \binom{n_2}{m_{2j}}^{-1} \sum_{d_1=0}^{m_{1j}} \sum_{d_2=0}^{m_{2j}} \binom{m_{1i}}{d_1}\binom{m_{2i}}{d_2}\binom{n_1 - m_{1i}}{m_{1j} - d_1}\binom{n_2 - m_{2i}}{m_{2j} - d_2} \zeta_{ij}(\mathbf{d}; \mathbf{F}) \tag{2.3}$$

for $1 \leqslant i \leqslant j \leqslant t$. Whenever, $\lambda_n = n^{-1}n_1$ is bounded away from 0 and 1, and n is large, the right-hand side (rhs) of (2.3) reduces to

$$n_1^{-1} m_{1i} m_{1j} \zeta_{ij}(1, 0; \mathbf{F}) + n_2^{-1} m_{2i} m_{2j} \zeta_{ij}(0, 1; \mathbf{F}) + O(n^{-2}). \tag{2.4}$$

Thus, under the assumption that there exists a λ_0: $0 < \lambda_0 \leqslant \frac{1}{2}$, such that

$$\lim_{n\to\infty} \lambda_n = \lambda \text{ exists and } \lambda \in [\lambda_0, 1 - \lambda_0], \tag{2.5}$$

$$n\boldsymbol{\Gamma}(\mathbf{n}) \to \boldsymbol{\Gamma}_\lambda = ((\lambda^{-1} m_{1i} m_{1j} \zeta_{ij}(1, 0; \mathbf{F}) + (1 - \lambda)^{-1} m_{2i} m_{2j} \zeta_{ij}(0, 1; \mathbf{F}))), \quad \text{as} \quad n \to \infty. \tag{2.6}$$

Moreover, it is well known (viz [6, p. 65]) that

$$\mathscr{L}(n^{1/2}[\mathbf{U}(\mathbf{n}) - \boldsymbol{\theta}(\mathbf{F})]) \to \mathscr{N}_t(\mathbf{0}, \boldsymbol{\Gamma}_\lambda), \quad \text{as} \quad n \to \infty. \tag{2.7}$$

Our procedure rests on suitable estimators of $\zeta_{ij}(1, 0; \mathbf{F})$ and $\zeta_{ij}(0, 1; \mathbf{F})$ for $i, j = 1, \ldots, t$. For this, as in [7], we define for every $\boldsymbol{\nu} = (\nu_1, \nu_2)' \geqslant \mathbf{m}^*$,

$$V_{\boldsymbol{\nu},r}^{(i)}(1, 0) = \binom{\nu_1 - 1}{m_{1i} - 1}^{-1} \binom{\nu_2}{m_{2i}}^{-1} \sum_{\boldsymbol{\nu} r}^{*} \phi_i(\mathbf{X}_{k\alpha_{kj}}, j = 1, \ldots, m_{ki}, k = 1, 2; \alpha_{11} = r), \tag{2.8}$$

where the summation $\sum_{\nu r}^{*}$ extends over all $1 \leqslant \alpha_{12} < \cdots < \alpha_{1m_{1i}} \leqslant \nu_1$, $1 \leqslant \alpha_{21} < \cdots < \alpha_{2m_{2i}} \leqslant \nu_2$ with $\alpha_{1j} \neq r$, $j = 2,..., m_{1i}$;

$$V_{\mathbf{v},r}^{(i)}(0, 1)$$
$$= \binom{\nu_1}{m_{1i}}^{-1} \binom{\nu_2 - 1}{m_{2i} - 1}^{-1} \sum_{\nu r}^{**} \phi_i(\mathbf{X}_{k\alpha_{kj}}, j = 1,..., m_{ki}, k = 1, 2; \alpha_{21} = r), \tag{2.9}$$

where the summation $\sum_{\nu r}^{**}$ extends over all $1 \leqslant \alpha_{11} < \cdots < \alpha_{1m_{1i}} \leqslant \nu_1$ and $1 \leqslant \alpha_{22} < \cdots < \alpha_{2m_{2i}} \leqslant \nu_2$ with $\alpha_{2j} \neq r$, $j = 2,..., m_{2i}$, $1 \leqslant i \leqslant t$. Let then

$$S_{ij,\mathbf{v}}(1, 0) = \frac{1}{\nu_1 - 1} \sum_{r=1}^{\nu_1} [V_{\mathbf{v},r}^{(i)}(1, 0) - U_i(\mathbf{v})][V_{\mathbf{v},r}^{(j)}(1, 0) - U_j(\mathbf{v})], \tag{2.10}$$

$$S_{ij,\mathbf{v}}(0, 1) = \frac{1}{\nu_2 - 1} \sum_{r=1}^{\nu_2} [V_{\mathbf{v},r}^{(i)}(0, 1) - U_i(\mathbf{v})][V_{\mathbf{v},r}^{(j)}(0, 1) - U_j(\mathbf{v})], \tag{2.11}$$

for $i, j = 1,..., t$, and let for $\mathbf{n} \geqslant \mathbf{v} \geqslant \mathbf{m}^*$,

$$\hat{\mathbf{\Gamma}}_{\mathbf{n}}(\mathbf{v}) = ((n_1^{-1} m_{1i} m_{1j} S_{ij,\mathbf{v}}(1, 0) + n_2^{-1} m_{2i} m_{2j} S_{ij,\mathbf{v}}(0, 1))). \tag{2.12}$$

For later use, we also define

$$\hat{\mathbf{\Gamma}}_{\mathbf{n}} = n \hat{\mathbf{\Gamma}}_{\mathbf{n}}(\mathbf{n}) = ((\lambda_n^{-1} m_{1i} m_{1j} S_{ij,\mathbf{n}}(1, 0) + (1 - \lambda_n)^{-1} m_{2i} m_{2j} S_{ij,\mathbf{n}}(0, 1))). \tag{2.13}$$

The proposed procedure. We conceive of a set of positive integers $\{m_0, m_1, ..., m_{n^*}\}$ such that $m_0 \geqslant \max_{1 \leqslant i \leqslant t} \max_{1 \leqslant k \leqslant 2} m_{ki}$, and

$$2m_0 + m_1 + m_2 + \cdots + m_{n^*} = n; \qquad n^* \geqslant 1. \tag{2.14}$$

When we conceive of a large n, there are two possible situations: (a) we regard n^* to be a fixed positive integer and allow $m_0, ..., m_{n^*}$ to be all large, and (b) n^* is allowed to increase with the increase in n. Case (a) represents the classical multistage sampling scheme (as will be explained later on), while case (b) is analogous to (group) sequential plan.

We start with m_0 observations drawn from each of the two distributions, and let

$$Z_1(\mathbf{n}) = \| \hat{\mathbf{\Gamma}}_{\mathbf{n}}(m_0, m_0) \| = n^{-t} Z_1^*(\lambda_n), \text{ say}, \tag{2.15}$$

where, by (2.12), $Z_1^*(\lambda_n)$ is a homogeneous polynomial of degree t in $(\lambda_n^{-1}, (1 - \lambda_n)^{-1})$ with the coefficients depending on $S_{ij,\mathbf{m}_0}(1, 0)$ and $S_{ij,\mathbf{m}_0}(0, 1)$, $1 \leqslant i \leqslant j \leqslant t$. Let then $\hat{\lambda}_n^{(1)} = \hat{\lambda}_n(\mathbf{m}_0)$ be the value of $\lambda_n (\in [0, 1])$ which minimizes $Z_1^*(\lambda_n)$; the existence of a unique solution (in probability) is proved in Theorem 3.4. In case there are multiple roots, a randomized procedure for selecting one of them should be employed. Define then a stochastic vector $\boldsymbol{\nu}_1 = (\nu_{11}, \nu_{12})$ by $\nu_{12} = 2m_0 + m_1 - \nu_{11}$ and

$$\nu_{11} = \begin{cases} m_0, & \hat{\lambda}_n^{(1)} < (m_0+1)/(2m_0+m_1), \\ [(2m_0+m_1)\hat{\lambda}_n^{(1)}], & (m_0+1) \leqslant (2m_0+m_1)\hat{\lambda}_n^{(1)} \leqslant (m_0+m_1), \\ m_0+m_1, & \hat{\lambda}_n^{(1)} > (m_0+m_1)/(2m_0+m_1), \end{cases} \tag{2.16}$$

where $[s]$ denotes the largest integer $\leqslant s$. Let then

$$Z_2(\mathbf{n}) = \| \hat{\boldsymbol{\Gamma}}_{\mathbf{n}}(\boldsymbol{\nu})\| = n^{-t} Z_2^*(\lambda_n), \text{ say}, \tag{2.17}$$

where again $Z_2^*(\lambda_n)$ is a homogeneous polynomial of degree t in $(\lambda_n^{-1}, (1-\lambda_n)^{-1})$. Let $\hat{\lambda}_n^{(2)} = \hat{\lambda}_n(\boldsymbol{\nu}_1)$ be the value of $\lambda_n(\in [0, 1])$ which minimizes $Z_2^*(\lambda_n)$. Define then a stochastic vector $\boldsymbol{\nu}_2 = (\nu_{21}, \nu_{22})$ by $\nu_{22} = 2m_0 + m_1 + m_2 - \nu_{21}$ and

$$\nu_{21} = \begin{cases} \nu_{11}, & \hat{\lambda}_n^{(2)} < (\nu_{11}+1)/(2m_0+m_1+m_2), \\ [(2m_0+m_1+m_2)\hat{\lambda}_n^{(2)}], & (\nu_{11}+1) \leqslant (2m_0+m_1+m_2)\hat{\lambda}_n^{(2)} \leqslant (\nu_{11}+m_2), \\ \nu_{11}+m_2, & \hat{\lambda}_n^{(2)} > (\nu_{11}+m_2)/(2m_0+m_1+m_2). \end{cases} \tag{2.18}$$

In an analogous manner, for each $j(=2, 3,..., n^*)$, define the stochastic variables $Z_j(\mathbf{n}) = \| \hat{\boldsymbol{\Gamma}}_{\mathbf{n}}(\boldsymbol{\nu}_{j-1})\| = n^{-t} Z_j^*(\lambda_n)$ and the vectors $\boldsymbol{\nu}_j$, and terminating this procedure at the n^*-th stage, we define our desired partition of the total sample size as

$$\mathbf{N} = \boldsymbol{\nu}_{n^*} = (N_1, N_2), \qquad N_k = \nu_{n^*k}, \qquad k = 1, 2, \tag{2.19}$$

and the corresponding estimator of $\boldsymbol{\theta}(\mathbf{F})$ by

$$\mathbf{U}(\mathbf{N}) = \mathbf{U}(\boldsymbol{\nu}_{n^*}). \tag{2.20}$$

In the asymptotic set-up, we conceive of the cases (a) and (b) as stated earlier, and assume that $m_0(= m_0(n))$ increases with n, in such a way that

$$\lim_{n\to\infty} m_0(n) = \infty \qquad \text{and} \qquad \lim_{n\to\infty} n^{-1} m_0(n) = 0 \tag{2.21}$$

[If $n^{-1}m_0(n) \to b(>0)$, then for optimal $\lambda < b$, we may encounter difficulties in deriving the properties of our proposed procedure.] The main theorem of the paper is the following.

THEOREM 1. *If* $\boldsymbol{\theta}(\mathbf{F})$ *is stationary of order zero,* $\phi_1, \ldots, \phi_t$ *all have finite fourth order moments and* (2.21) *holds, then as* $n \to \infty$,

$$\text{(i)} \quad n^{-1} | \mathbf{N} - \mathbf{n}_0 | \to 0 \text{ a.s.}, \tag{2.22}$$

$$\text{(ii)} \quad \mathscr{L}(n^{1/2}[\mathbf{U}(\mathbf{N}) - \boldsymbol{\theta}(\mathbf{F})]) \to \mathscr{N}_t(\mathbf{0}, \boldsymbol{\Gamma}^*(\mathbf{F})), \tag{2.23}$$

where $| \mathbf{x} | = \max_{1 \leqslant j \leqslant c} | x_j |$, $\boldsymbol{\Gamma}^*(\mathbf{F}) = \lim_{n \to \infty} \{ n \boldsymbol{\Gamma}(\mathbf{n}_0, \mathbf{F}) \}$, *and* $\mathbf{n}_0$ *and* $\boldsymbol{\Gamma}(\mathbf{n}_0, \mathbf{F})$ *are defined in* (1.8) *and thereafter.*

The proof is postponed to Section 4.

3. Strong Convergence of U-Statistics and Derived Estimates

Consider a (double) sequence $\{Y(\mathbf{n}) = Y(n_1, n_2), \mathbf{n} \geqslant \mathbf{1}\}$ of nonnegative random variables, and a (double) sequence $\{\mathscr{C}(\mathbf{n}), \mathbf{n} \geqslant \mathbf{1}\}$ of product σ-fields where $\mathscr{C}(\mathbf{n}) = \mathscr{C}_1(n_1) \times \mathscr{C}_2(n_2)$, $\mathbf{n} \geqslant \mathbf{1}$, $\mathscr{C}_i(n)$ is $\downarrow$ in n, $i = 1, 2$, and $\mathscr{C}_1(m)$ and $\mathscr{C}_2(n)$ are σ-fields. Assume that

$$\text{(i)} \quad E[Y(m, n) \mid \mathscr{C}_2(n')] \geqslant Y(m, n'), \text{ for all } n' \geqslant n,\ m \geqslant 1, \tag{3.1}$$

$$\text{(ii)} \quad E[Y(m, n) \mid \mathscr{C}_1(m')] \geqslant Y(m', n), \text{ for all } m' \geqslant m,\ n \geqslant 1, \tag{3.2}$$

$$\text{(iii)} \quad E[Y^2(\mathbf{n})] < \infty, \text{ for every } \mathbf{n} \geqslant \mathbf{1}. \tag{3.3}$$

LEMMA 3.1. *For nonnegative* $\{Y(\mathbf{n}), \mathbf{n} \geqslant \mathbf{1}\}$ *satisfying* (3.1), (3.2), *and* (3.3),

$$E\{[\sup_{\mathbf{m} \geqslant \mathbf{n}} Y(\mathbf{m})]^2\} \leqslant 16 E Y^2(\mathbf{n}). \tag{3.4}$$

Proof. For some arbitrary $\mathbf{N}(\geqslant \mathbf{n})$, consider the array

$$\begin{pmatrix} Y(n_1, n_2) & Y(n_1, n_2+1) & Y(n_1, N_2) \\ \vdots & \vdots & \vdots \\ Y(N_1, n_2) & Y(N_1, n_2+1) & Y(N_1, N_2) \end{pmatrix} = (\mathbf{Y}(n_2), \ldots, \mathbf{Y}(N_2)), \tag{3.5}$$

where, by (3.1),

$$E[\mathbf{Y}(j) \mid \mathscr{C}_2(k)] \geqslant \mathbf{Y}(k) \qquad \text{for} \quad n_2 \leqslant j \leqslant k \leqslant N_2, \tag{3.6}$$

i.e., $\{\mathbf{Y}(j), \mathscr{C}_2(j), j \geqslant n_2\}$ is a nonnegative reverse semimartingale (in $R^{N_1 - n_1}$), and $\max_{n_1 \leqslant k \leqslant N_1} | Y(k, j) | = | \mathbf{Y}(j) |$ is a convex function of $\mathbf{Y}(j)$, so that by the Doob [3, p. 317] inequality

$$\begin{aligned} E[(\max_{n_2 \leqslant j \leqslant N_2} | \mathbf{Y}(j) |)^2 \mid \mathscr{C}_1(n_1)] &\leqslant 4E[| \mathbf{Y}(n_2) |^2 \mid \mathscr{C}_1(n_1)] \\ &= 4E[(\max_{n_1 \leqslant k \leqslant N_1} | Y(k, n_2) |)^2 \mid \mathscr{C}_1(n_1)]. \end{aligned} \tag{3.7}$$

Therefore,

$$
\begin{aligned}
E[(\max_{n_1 \leqslant k \leqslant N_1} \max_{n_2 \leqslant j \leqslant N_2} | Y(k,j)|)^2] &= E[(\max_{n_2 \leqslant j \leqslant N_2} | Y(j)|)^2] \\
&= E\{E[(\max_{n_2 \leqslant j \leqslant N_2} | \mathbf{Y}(j)|)^2 \mid \mathscr{C}_1(n_1)]\} \\
&\leqslant E\{4E[(\max_{n_1 \leqslant k \leqslant N_1} | Y(k, n_2)|)^2 \mid \mathscr{C}_1(n_1)]\} \\
&= 4E[(\max_{n_1 \leqslant k \leqslant N_1} | Y(k, n_2)|)^2] \\
&\leqslant 16E[Y^2(n_1, n_2)], \qquad (3.8)
\end{aligned}
$$

where the last inequality follows by (3.2) and the Doob inequality. Since (3.8) holds for every $\mathbf{N} \geqslant \mathbf{n}$, the proof of the lemma follows by letting $N_i \to \infty$, $i = 1, 2$. Q.E.D.

Let now $\mathbf{A}$ be a $t \times t$ positive definite (*pd*) matrix, and define $\mathbf{U}(\mathbf{n})$, $\theta(\mathbf{F})$ and $\mathbf{\Gamma}(\mathbf{n})$ as in (1.1)–(1.3), (1.5), and (1.6).

Lemma 3.2. *If* $\mathbf{U}(\mathbf{n})$ *possesses second order moments, then for every* $\mathbf{n} \geqslant \mathbf{m}$, $h > 0$,

$$
P\{\sup_{\mathbf{N} \geqslant \mathbf{n}} [\sup_{\ell \neq 0} (\ell' \mathbf{A} \ell)^{-1/2} \mid \ell' [\mathbf{U}(\mathbf{N}) - \boldsymbol{\theta}(\mathbf{F})]|] > h\} \leqslant 16 h^{-2} \operatorname{Trace}\{\mathbf{A}^{-1} \mathbf{\Gamma}(\mathbf{n})\}. \quad (3.9)
$$

Proof. By the Schwarz inequality,

$$
\begin{aligned}
\sup_{\ell \neq 0} (\ell' \mathbf{A} \ell)^{-1/2} \mid \ell' [\mathbf{U}(\mathbf{N}) - \boldsymbol{\theta}(\mathbf{F})]| &= ([\mathbf{U}(\mathbf{N}) - \boldsymbol{\theta}(\mathbf{F})]' \mathbf{A}^{-1} [\mathbf{U}(\mathbf{N}) - \boldsymbol{\theta}(\mathbf{F})])^{1/2} \\
&= Y(\mathbf{N}), \quad \text{say.} \qquad (3.10)
\end{aligned}
$$

Then, by (1.6) and (3.10)

$$
EY^2(\mathbf{n}) = \operatorname{Trace}(\mathbf{A}^{-1} \mathbf{\Gamma}(\mathbf{n})). \qquad (3.11)
$$

Also, if $\mathscr{C}_k(n)$ be the σ-field generated by the unordered collection $\{\mathbf{X}_{k1}, \ldots, \mathbf{X}_{kn}\}$ and by $\mathbf{X}_{kn+1}, \mathbf{X}_{kn+2}, \ldots$, $k = 1, 2$, $n \geqslant 1$, then by a direct extension of the reverse martingale property of one-sample U-statistics (cf. [2]), it follows that

$$
\begin{aligned}
E[\mathbf{U}(n_1, N) \mid \mathscr{C}_2(n_2')] &= U(n_1, n_2'), \qquad \text{for all} \quad m_2 \leqslant N \leqslant n_2', \\
E[\mathbf{U}(N, n_2) \mid \mathscr{C}_1(n_1')] &= U(n_1', n_2), \qquad \text{for all} \quad m_1 \leqslant N \leqslant n_1',
\end{aligned}
$$

and as $Y(\mathbf{n})$ is convex in $[\mathbf{U}(\mathbf{n}) - \boldsymbol{\theta}(\mathbf{F})]$, it follows that (3.1) and (3.2) both hold. Hence, the lemma follows from (3.10), (3.11), and Lemma 3.1.

Let us denote by

$$
c(\mathbf{n}) = [16 \operatorname{Trace}(\mathbf{A}^{-1} \mathbf{\Gamma}(\mathbf{n}))]^{-1/2}, \qquad (3.12)
$$

so that by (1.6) and (2.4), $c(\mathbf{n})$ is $\uparrow$ in each argument of $\mathbf{n}$, and

$$c(\mathbf{n}) \to \infty \quad \text{as} \quad \min_{1 \leqslant k \leqslant 2} n_k \to \infty. \tag{3.13}$$

Hence, from (3.9) and (3.12), we have for every $h > 0$, $\mathbf{n} \geqslant \mathbf{m}$,

$$P\{\sup_{\mathbf{N} \geqslant \mathbf{n}} \sup_{\ell \neq 0} (\ell' \mathbf{A} \ell)^{-1/2} \mid \ell'[\mathbf{U}(\mathbf{N}) - \boldsymbol{\theta}(\mathbf{F})]| > h/c(\mathbf{n})\} \leqslant h^{-2}, \tag{3.14}$$

so that by (3.13) and (3.14)

$$\sup_{\ell \neq 0} (\ell' \mathbf{A} \ell)^{-1/2} \mid \ell'[\mathbf{U}(\mathbf{N}) - \boldsymbol{\theta}(\mathbf{F})]| \to 0 \quad \text{a.s. as } \mathbf{n} \to \infty. \tag{3.15}$$

THEOREM 3.3. *For $\mathbf{n} \geqslant \boldsymbol{\nu}$ and λ_n satisfying* (2.5), *when* $E[\phi_i^4(...)] < \infty$, $1 \leqslant i \leqslant t$,

$$n | \hat{\mathbf{\Gamma}}_{\mathbf{n}}(\boldsymbol{\nu}) - \mathbf{\Gamma}(\mathbf{n})| \to 0 \text{ a.s.} \quad \textit{as} \quad \boldsymbol{\nu} \to \infty. \tag{3.16}$$

Proof. By virtue of (2.5), (2.6), (2.10), (2.11), and (2.12), it suffices to show that for every $i, j = 1,..., t$, as $\nu_1 \to \infty$, $\nu_2 \to \infty$,

$$S_{ij,\boldsymbol{\nu}}(1, 0) \xrightarrow{\text{a.s.}} \zeta_{ij}(1, 0; \mathbf{F}) \quad \text{and} \quad S_{ij,\boldsymbol{\nu}}(0, 1) \xrightarrow{\text{a.s.}} \zeta_{ij}(0, 1; \mathbf{F}). \tag{3 17}$$

For $\mathbf{d} = (d_1, d_2)'$ and $\mathbf{n} = (n_1, n_2)'$, we define

$$U_{ij}(\mathbf{d}; \mathbf{n}) = \prod_{k=1}^{2} \left\{ \binom{n_k}{m_{ki}} \binom{m_{ki}}{d_k} \binom{n_k - m_{ki}}{m_{kj} - d_k} \right\}^{-1} \sum \phi_i(\mathbf{X}_{k\alpha_{kj}}, 1 \leqslant j \leqslant m_{ki}, k = 1, 2)$$
$$\times \phi_j(\mathbf{X}_{k\beta_{kj}}, 1 \leqslant j \leqslant m_{kj}, k = 1, 2), \tag{3.18}$$

where the summation extends over all possible $1 \leqslant \alpha_{k1} < \cdots < \alpha_{km_{ki}} \leqslant n_k$, $1 \leqslant \beta_{k1} < \cdots < \beta_{km_{kj}} \leqslant n_k$, $k = 1, 2$ with the restriction that $\alpha_{kj} = \beta_{kj}$ for $j = 1,..., d_k$, $k = 1, 2$, while the remaining $m_{ki} - d_k$ of the α_{kj}'s are different from the remaining $m_{kj} - d_k$ of the β_{kj}'s, for $k = 1, 2$ and $\mathbf{d} \geqslant \mathbf{0}$. Then, by (2.8), (2.10) and a few routine steps, it follows that

$$S_{ij,\boldsymbol{\nu}}(1, 0) = U_{ij}(1, 0; \boldsymbol{\nu}) - U_{ij}(0, 0; \boldsymbol{\nu}) + \sum{}^{*} \alpha(\mathbf{d}, \boldsymbol{\nu}) \, U_{ij}(\mathbf{d}; \boldsymbol{\nu}), \tag{3.19}$$

where the summation $\sum^*$ extends over permissible values of $\mathbf{d} \geqslant \mathbf{0}$, and where the $\alpha(\mathbf{d}, \boldsymbol{\nu})$ are all $0(n^{-1})$. By (3.15), as $\boldsymbol{\nu} \to \infty$,

$$U_{ij}(1, 0; \boldsymbol{\nu}) \xrightarrow{\text{a.s.}} \zeta_{ij}(1, 0; \mathbf{F}) + \theta_i(\mathbf{F}) \, \theta_j(\mathbf{F}), \; U_{ij}(\mathbf{0}, \boldsymbol{\nu}) \xrightarrow{\text{a.s.}} \theta_i(\mathbf{F}) \, \theta_j(\mathbf{F}), \tag{3.20}$$

and for each $\mathbf{d} \geqslant \mathbf{0}$, $U_{ij}(\mathbf{d}, \nu) \to^{\text{a.s.}} EU_{ij}(\mathbf{d}, \nu)$ as $\nu \to \infty$, where $E[\phi_i^4(\cdots)] < \infty$ $(i = 1,\ldots, t)$ imply that $[EU_{ij}^2(\mathbf{d}, \nu)]^2 \leqslant E[\phi_i^4]\, E[\phi_j^4] < \infty$. Consequently, $| EU_{ij}(\mathbf{d}, \nu)| < \infty$ for all $\mathbf{d} \geqslant \mathbf{0}$, and hence, by (3.19) and (3.20), as $\nu \to \infty$, $S_{ij,\nu}(1, 0; \nu) \to \zeta_{ij}(1, 0; \mathbf{F})$ a.s. A similar case follows with $S_{ij,\nu}(0, 1; \nu)$, and hence, the proof of the theorem is complete.

Let us now assume that $\| \mathbf{\Gamma}_\lambda \|$ (defined by (2.6)) assumes a unique minimum at the point λ^*. By (2.6) and the fact that λ^{-1} (and $(1 - \lambda)^{-1}$) go to ∞ as $\lambda \to 0$ (and $\to 1$), we may assume without any loss of generality that

$$\lambda^* \in [\lambda_0, 1 - \lambda_0] \qquad \text{where} \quad 0 < \lambda_0 \leqslant \tfrac{1}{2}. \tag{3.21}$$

THEOREM 3.4. *If $E\phi_i^4 < \infty$, $i = 1,\ldots, t$ and λ^* is unique, then under* (2.21), *$Z_s^*(\lambda_n)$ attains a global minimum either at a unique point $\hat{\lambda}_{ns}$ or at multiple points all converging to a common $\hat{\lambda}_{ns}$, a.s., as $n \to \infty$. Moreover, in either case, for every $\epsilon' > 0$,*

$$P\{U_{s=1}^{n^*}[| \hat{\lambda}_{ns} - \lambda^* | > \epsilon']\} \to 0 \qquad as \quad n \to \infty. \tag{3.22}$$

Proof. By (2.12) and the definition of $Z_s^*(\lambda_n)$, $s \geqslant 1$, it suffices to show that as $\nu \to \infty$, $\| n\hat{\mathbf{\Gamma}}_n(\nu)\|$ attains (a.s.) a unique infimum (over $0 < \lambda < 1$) at $\lambda_n = \hat{\lambda}_{n\nu}$ where $\hat{\lambda}_{n\nu} \to \lambda^*$ a.s., as $\nu \to \infty$. Let $\mathbf{r} = (r_1, \ldots, r_t)$ (where $r_j = 0, 1$, for $j = 1,\ldots, t$) and let $A(\nu, \mathbf{r})$ be the determinant of the matrix (of order $t \times t$) whose ℓ-th column has the elements $m_{1i}m_{1\ell}S_{ij,\nu}(1, 0)$, $1 \leqslant i \leqslant t$, if $r_\ell = 0$, and $m_{2i}m_{2\ell}S_{ij,\nu}(0, 1)$, $1 \leqslant i \leqslant t$, if $r_\ell = 1$, for $\ell = 1,\ldots, t$. Then

$$g(\lambda_n ; \nu) = \| n\hat{\mathbf{\Gamma}}_n(\nu)\| = \sum_{(\mathbf{r})} \lambda_n^{-t+\Sigma_{\ell=1}^t r_\ell}(1 - \lambda_n)^{-\Sigma_{\ell=1}^t r_\ell} A(\nu, \mathbf{r}) \tag{3.23}$$

On replacing $S_{ij,\nu}(1, 0)$ and $S_{ij,\nu}(0, 1)$ by $\zeta_{ij}(1, 0; \mathbf{F})$ and $\zeta_{ij}(0, 1; \mathbf{F})$, respectively, we define analogously the determinants $A^*(\mathbf{r})$, so that

$$g(\lambda_n) = \| \mathbf{\Gamma}_{\lambda_n} \| = \sum_{(\mathbf{r})} \lambda_n^{-t+\Sigma_{\ell=1}^t r_\ell}(1 - \lambda_n)^{-\Sigma_{\ell=1}^t r_\ell} A^*(\mathbf{r}) \tag{3.24}$$

Now, by Theorem 3.3, as $\nu \to \infty$,

$$A(\nu, \mathbf{r}) \xrightarrow{\text{a.s.}} A^*(\mathbf{r}) \qquad \text{for every} \quad \mathbf{r}: r_\ell = 0, 1,\ 1 \leqslant \ell \leqslant t. \tag{3.25}$$

Moreover, for $\lambda_0 \leqslant \lambda_n \leqslant 1 - \lambda_0$, both λ_n^{-1} and $(1 - \lambda_n)^{-1}$ are $\leqslant \lambda_0^{-1} < \infty$, so that by (3.23), (3.24), and (3.25),

$$\sup_{\lambda_0 \leqslant \lambda_n \leqslant 1-\lambda_0} \left| \| n\hat{\mathbf{\Gamma}}_n(\nu)\| - \| \mathbf{\Gamma}_{\lambda_n} \| \right| \to 0 \quad \text{a.s. as } \nu \to \infty \tag{3.26}$$

Note that $\mathscr{g}(\lambda_n)$ [or $\mathscr{g}(\lambda_n; \boldsymbol{\nu})$] is a homogeneous polynomial of degree t in $[\lambda_n^{-1}, (1-\lambda_n)^{-1}]$, so that it has at most $(t+1)$ extrema within the interval $[\lambda_0, 1-\lambda_0]$. Since $A^*(\mathbf{1})$ and $A^*(\mathbf{0})$ are both positive, $\mathscr{g}'(\lambda_n)$ is negative as $\lambda_n \to 0$ and positive as $\lambda_n \to 1$, so that the existence of at least one minimum is insured; the number of minima (t^*) is bounded by $[(t+2)/2]$. By the assumed uniqueness of λ^*, we have the following: if λ_d^*, $d = 1,..., t^*$, be the points at which $\mathscr{g}(\lambda_n)$ attains local minima, then there exists a $\epsilon > 0$, such that $\| \boldsymbol{\Gamma}_{\lambda_d^*} \| > \| \boldsymbol{\Gamma}_{\lambda^*} \| + \epsilon$, for all $\lambda_d^* \neq \lambda^*$, and

$$|\lambda_d^* - \lambda^*| > \epsilon > 0 \text{ for every } d\colon \lambda_d^* \neq \lambda^*. \tag{3.27}$$

Moreover, $\| \boldsymbol{\Gamma}_\lambda \|$ is a continuous function of $\lambda \in (0, 1)$, so that for $\lambda \in [\lambda^* - \epsilon, \lambda^* + \epsilon]$, $\mathscr{g}(\lambda)$ is monotonically decreasing in $[\lambda^* - \epsilon, \lambda^*)$ and monotonically increasing in $(\lambda^*, \lambda^* + \epsilon]$. Consequently, within $[\lambda^* - \epsilon, \lambda^*]$, $\mathscr{g}^{-1}$ exists and is continuous, and similarly for $[\lambda^*, \lambda^* + \epsilon]$. Also, by (3.23) and (3.25), we conclude that $\mathscr{g}(\lambda_n, \boldsymbol{\nu})$ has also at most $[(t+2)/2]$ minima within $[\lambda_0, 1-\lambda_0]$, there being at least one (a.s.). Since the left-hand side of (3.26) can be bounded (a.s.) by $\epsilon/2$, $\epsilon > 0$, for $\boldsymbol{\nu} \geqslant [n_0(\epsilon)]\mathbf{1}$, we have as $\boldsymbol{\nu} \to \boldsymbol{\infty}$,

$$\inf_{\lambda_0 \leqslant \lambda_n \leqslant 1-\lambda_0} \mathscr{g}(\lambda_n; \boldsymbol{\nu}) \leqslant \mathscr{g}(\lambda^*, \boldsymbol{\nu}) \leqslant \mathscr{g}(\lambda^*) + \epsilon/2 \text{ a.s.}, \tag{3.28}$$

and if $\lambda^*(\boldsymbol{\nu})$ be any point where $\mathscr{g}(\lambda_n; \boldsymbol{\nu})$ attains a global minimum,

$$\inf_{\lambda_0 \leqslant \lambda_n \leqslant 1-\lambda_0} \mathscr{g}(\lambda_n, \boldsymbol{\nu}) = \mathscr{g}(\lambda^*(\boldsymbol{\nu}), \boldsymbol{\nu}) \geqslant \mathscr{g}(\lambda^*(\boldsymbol{\nu})) - \epsilon/2 \geqslant \mathscr{g}(\lambda^*) - \epsilon/2 \text{ a.s.}, \tag{3.29}$$

as $\mathscr{g}(\lambda) \geqslant g(\lambda^*)$, for all λ. Consequently, by (3.26), (3.28), and (3.29), as $\boldsymbol{\nu} \to \boldsymbol{\infty}$,

$$\mathscr{g}(\lambda^*) - \epsilon \leqslant \mathscr{g}(\lambda^*(\boldsymbol{\nu})) \leqslant \mathscr{g}(\lambda^*) + \epsilon \text{ a.s.}, \tag{3.30}$$

for every root $\lambda^*(\boldsymbol{\nu})$ for which $\mathscr{g}(\lambda^*(\boldsymbol{\nu}))$ is a global minimum. As such, by (3.27) and the discussions following it, we have (i) if $\lambda^*(\boldsymbol{\nu})$ is unique then $\lambda^*(\boldsymbol{\nu}) \to \lambda^*$ a.s., as $\boldsymbol{\nu} \to \boldsymbol{\infty}$, and (ii) if there are multiple roots, all of these are (a.s.) close to a $\lambda^*(\boldsymbol{\nu})$ and (3.22) holds. Q.E.D.

LEMMA 3.5. *If $E\phi_i^2(\cdots) < \infty$, $1 \leqslant i \leqslant t$, then for every $\epsilon > 0$ and $\eta > 0$, there exist a $\delta > 0$ and an integer $n_0(=n_0(\epsilon, \eta))$, such that for $\boldsymbol{\nu} \geqslant \mathbf{1}n_0$,*

$$P\{\max_{|\boldsymbol{\nu}-\mathbf{n}|<\delta|\mathbf{n}|} \sup_{\boldsymbol{\ell}\neq\mathbf{0}} (\boldsymbol{\ell}'\mathbf{A}\boldsymbol{\ell})^{-1/2} \,|\, \boldsymbol{\ell}'[\mathbf{U}(\mathbf{n}) - \mathbf{U}(\boldsymbol{\nu})]| > \epsilon[(n_1+n_2)/n_1n_2]^{1/2}\} < \eta, \tag{3.31}$$

where $|\mathbf{x}| = \max\{|x_1|, |x_2|\}$ and $\boldsymbol{\nu} = (\nu_1, \nu_2)$ satisfies (2.5) with n_i being replaced by ν_i, $i = 1, 2$.

Proof. The proof follows along the lines of the proof of Lemma 3.2. Here we note that

$$E\{[\mathbf{U}(\nu_1, \nu_2) - \mathbf{U}(\nu_1, \nu_2')] \mid \mathscr{C}_2(\nu_2')\} = 0, \qquad \text{for all} \quad \nu_2' \geqslant \nu_2,$$
$$E\{[\mathbf{U}(\nu_1, \nu_2) - \mathbf{U}(\nu_1', \nu_2)] \mid \mathscr{C}_1(\nu_1')\} = 0, \qquad \text{for all} \quad \nu_1' \geqslant \nu_1.$$

Also, by (1.6) and (2.6), for any *pd* $\mathbf{A}$, we have under (2.5),

$$| [\mathrm{Tr}(\mathbf{A}^{-1}\mathbf{\Gamma}(\boldsymbol{\nu})) - \mathrm{Tr}(\mathbf{A}^{-1}\mathbf{\Gamma}(\boldsymbol{\nu}'))]/\mathrm{Tr}(\mathbf{A}^{-1}\mathbf{\Gamma}(\boldsymbol{\nu}))| \leqslant \epsilon^2/16,$$
$$\text{whenever} \quad \| \boldsymbol{\nu} - \boldsymbol{\nu}' \| < \delta(>0),$$

where for every $\epsilon > 0$, we can choose $\delta(> 0)$ adequately small. The rest of the proof follows by using the Schwarz inequality [as in (3.10)] and Lemma 3.1. For brevity, the details are omitted.

4. The Proof of Theorem 1

By (1.8) and (2.6), we have

$$\lim_{n\to\infty} n\mathbf{\Gamma}(\mathbf{n}_0, \mathbf{F}) = \mathbf{\Gamma}^*(\mathbf{F})$$
$$= (((m_{1i}m_{1j}/\lambda^*)\, \zeta_{ij}(1, 0; \mathbf{F}) + (m_{2i}m_{2j}/1 - \lambda^*)\, \zeta_{ij}(0, 1; \mathbf{F}))), \tag{4.1}$$

where λ^*, defined before (3.21), is the (assumed) unique point where $\| \mathbf{\Gamma}_\lambda \|$, defined by (2.6), assumes a unique minimum. Thus,

$$\lim_{n\to\infty} n^{-1}\mathbf{n}_0 = (\lambda^*, 1 - \lambda^*). \tag{4.2}$$

Also, by Theorem 3.4, on defining $\boldsymbol{\nu}_j, j \geqslant 1$, as in (2.16), (2.18), etc., and noting that (2.21) holds, we have

$$n^{-1}\boldsymbol{\nu}_{n^*} \to (\lambda^*, 1 - \lambda^*) \text{ a.s.} \qquad \text{as} \quad n \to \infty. \tag{4.3}$$

Then, (2.22) follows directly from (4.2) and (4.3).

Since $\lambda^* \in [\lambda_0, 1 - \lambda_0]$ and the kernels are all square integrable by the asymptotic normality of generalized U-statistics (cf. [6, p. 66]),

$$\mathscr{L}(n^{1/2}[\mathbf{U}(\mathbf{n}_0) - \boldsymbol{\theta}(\mathbf{F})]) \to \mathscr{N}_t(\mathbf{0}, \mathbf{\Gamma}^*(\mathbf{F})), \qquad \text{as} \quad n \to \infty, \tag{4.4}$$

where $\mathbf{\Gamma}^*(\mathbf{F})$ is defined by (4.1) and is *pd* as $\boldsymbol{\theta}(\mathbf{F})$ is assumed to be stationary of order zero. Since, by (2.22), for every $\delta > 0$,

$$P\{n^{-1}|\,\mathbf{N} - \mathbf{n}_0\,| > \delta\} \to 0, \qquad \text{as} \quad n \to \infty, \tag{4.5}$$

the proof of (2.23) follows directly from (4.4), (4.5), Lemma 3.5 and the well-known Anscombe (1952) theorem.

5. Some Concluding Remarks

We note that our estimators of $\zeta_{ij}(1, 0; \mathbf{F})$ and $\zeta_{ij}(0, 1; \mathbf{F})$ are based on the structural components $V_{\mathbf{v},r}^{(i)}(1, 0)$ and $V_{\mathbf{v},r}^{(i)}(0, 1)$, defined in (2.8) and (2.9). On the other hand, Yen [10] considered unbiased estimators of these parameters which involve averages over all possible choice of $m_{1i} + m_{1j} - 1$ of the $\mathbf{X}_{1\alpha}$, $\alpha = 1,\ldots, \nu_1$ and $m_{2i} + m_{2j}$ of the $\mathbf{X}_{2\alpha}$, $\alpha = 1,\ldots, \nu_2$ or $m_{1i} + m_{1j}$ from the ν_1 and $m_{2i} + m_{2j} - 1$ from the ν_2 observations. As a result, her computation involves for large (ν_1, ν_2) a number of terms of the order $\nu^{m_{1i}+m_{1j}+m_{2i}+m_{2j}-1}$, whereas ours involve a number of terms of the order $[\nu^{m_{1i}+m_{2i}} + \nu^{m_{1j}+m_{2j}}]$, so that considerable gain in computation time is expected by using our procedure.

In the general case of $c \geqslant 2$, we have analogous to (2.6), the covariance matrix $\mathbf{\Gamma}_{\boldsymbol{\lambda}}$ given by

$$\left(\left(\sum_{k=1}^{c} \lambda_k^{-1} m_{ki} m_{kj} \zeta_{ij}(\boldsymbol{\delta}_k\,; \mathbf{F})\right)\right) \tag{5 1}$$

where $\boldsymbol{\delta}_k$ has 1 in the k-th place and 0 elsewhere, $1 \leqslant k \leqslant c$, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_c)$. Here (3.23) or (3.24) will be a homogeneous polynomial of degree t in $\lambda_1, \ldots, \lambda_c$ where $\sum_{k=1}^{c} \lambda_k = 1$. Theorem 3.4 and Lemma 3.5 extend readily for λ being replaced by $\boldsymbol{\lambda}$, so that the proof of Theorem 1 follows on parallel lines. However, from computational aspects, this will be naturally more laborious for $c > 2$.

If we define the empirical *df* $F_{n_k,k}(\mathbf{x})$ as

$$F_{n_k,k}(\mathbf{x}) = n_k^{-1}\,[\text{number of } \mathbf{X}_{ki} \leqslant \mathbf{x},\ i = 1,\ldots, n_k]$$

for $k = 1,\ldots, c$, then another statistic that can be used to estimate $\theta_i(\mathbf{F})$ is $\theta_i(F_{n_1,1}, \ldots, F_{n_c,c})$ where

$$\theta_i(F_{n_1,1}, \ldots, F_{n_c,c}) = \prod_{k=1}^{c} n_k^{-m_{ki}} \sum{}' \phi_i(\mathbf{X}_{k\alpha_{kj}}, j = 1,\ldots, m_{ki}, 1 \leqslant k \leqslant c)$$

and the summation $\sum'$ extends over all $1 \leqslant \alpha_{kj} \leqslant n_k$, $k = 1,\ldots, c$; $j = 1,\ldots, m_{ki}$

for $i = 1,\ldots, t$. A study of the asymptotic distribution theory of such functionals of the empirical *df*'s has been made by von Mises [5]. Since, whenever

$$E \mid \theta_i(F_{n_1,1},\ldots, F_{n_c,c})| < \infty,$$

$$n^{\alpha}[\theta_i(F_{n_1,1},\ldots, F_{n_c,c}) - U_i(n_1,\ldots, n_c)] \to 0 \quad \text{a.s.}$$

as $n \to \infty$ for $\alpha < 1$, then it is possible to prove a result analogous to Theorem 1 for this alternative type of estimator.

Acknowledgment

Thanks are due to the referee for his useful comments on the paper.

References

[1] Anscombe, F. J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Phil. Soc.* **48** 600–607.

[2] Berk, R. H. (1966). Limiting behavior of posterior distribution when the model is incorrect. *Ann. Math. Statist.* **37** 51–58.

[3] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[4] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

[5] von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.

[6] Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York.

[7] Sen, P. K. (1960). On some convergence properties of *U*-statistics. *Calcutta Statist. Assoc. Bull.* **10** 1–18.

[8] Sen, P. K. (1972). Weak convergence of generalized *U*-statistics. *Univ. North Carolina, Inst. Statist. Mimeo report no.* 837.

[9] Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika* **24** 471–494.

[10] Yen, E. H. (1964). On two-stage nonparametric estimation. *Ann. Math. Statist.* **35** 1099–1114.