JOURNAL OF MULTIVARIATE ANALYSIS 4, 453–468 (1974)

On Bounded Maximum Width Sequential Confidence Ellipsoids Based on Generalized U-Statistics*

GEORGE W. WILLIAMS AND PRANAB KUMAR SEN

University of Michigan, Ann Arbor, and University of North Carolina, Chapel Hill

For a vector of (estimable) functionals of several independent distributions, sequential confidence ellipsoids (of bounded maximum width) based on a class of generalized U-statistics are studied. A stopping rule along with a procedure for choosing the component sample sizes at each stage is developed, so that the proposed confidence ellipsoid has a confidence coefficient asymptotically (as the prescribed maximum width shrinks to zero) equal to a preassigned $1 - \alpha$ $(0 < \alpha < 1)$, and the expected total sample size is minimized for the procedure. Asymptotic efficiency of the procedure is also studied. The case of von Mises' functionals is treated briefly at the end.

1. INTRODUCTION

Let $\{\mathbf{X}_{ki}, i \geqslant 1\}$ be a sequence of independent and identically distributed random vectors (i.i.d.r.v.) 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 independent. Consider an estimable parameter (vector)

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

where

$$\theta_i(\mathbf{F}) = E\phi_i(X_{kj}, 1 \leqslant j \leqslant m_{ki}, 1 \leqslant k \leqslant c), \qquad i = 1,..., t, \tag{1.2}$$

the m_{ki} are all nonnegative integers, $\mathbf{m}_i = (m_{1i},..., m_{ci}) \neq \mathbf{0}$, $1 \leqslant i \leqslant t$, and, without any loss of generality, we assume that the kernel ϕ_i is symmetric in

Received June 8, 1973.

AMS 1970 classification: 62G15, 62H99, 62L12.

Key words and phrases: Asymptotic consistency and efficiency, average sample number, confidence ellipsoid, estimable parameters, generalized U-statistics, von Mises' differentiable statistical functions, and bounded maximum width.

* Work supported by the National Institutes of Health, Training Grant No. T01GM00038 and partially by 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.

$X_{k1}, \ldots, X_{km_{ki}}$ for every $1 \leqslant k \leqslant c$; $1 \leqslant i \leqslant t$. By (1.2), we have for every $i(=1, \ldots, t)$,

$$\theta_i(\mathbf{F}) = \int_{R^{pm_i}} \cdots \int \phi_i(x_{11}, \ldots, x_{cm_{ci}}) \prod_{k=1}^{c} \prod_{j=1}^{m_{ki}} dF_k(x_{kj}), \tag{1.3}$$

where $m_i = \mathbf{m}_i'\mathbf{1} = m_{1i} + \cdots + m_{ci}$, $1 \leqslant i \leqslant t$. On denoting the coordinate-wise inequalities $a_k <$ (or $\leqslant$) b_k, $1 \leqslant k \leqslant c$ by $\mathbf{a} <$ (or $\leqslant$) $\mathbf{b}$, and by $\mathbf{n} = (n_1, \ldots, n_c)'$, we note that for $\mathbf{n} \geqslant \mathbf{m}_i$, the generalized U-statistic corresponding to $\theta_i(\mathbf{F})$ is

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

where $\binom{\mathbf{a}}{\mathbf{b}} = \prod_{k=1}^{c} \binom{a_k}{b_k}$ and the summation $\sum_{(n)}^{*}$ extends over all possible $1 \leqslant \alpha_{k1} < \cdots < \alpha_{km_{ki}} \leqslant n_k$, $1 \leqslant k \leqslant c$, for $i = 1, \ldots, t$. Thus, if we let $\mathbf{m}^* = (m_1^*, \ldots, m_c^*)'$ where $m_k^* = \max_{1 \leqslant i \leqslant t} m_{ki}$, $k = 1, \ldots, c$, then for $\mathbf{n} \geqslant \mathbf{m}^*$,

$$\mathbf{U}(\mathbf{n}) = (U_1(\mathbf{n}), \ldots, U_t(\mathbf{n}))' \text{ unbiasedly estimates } \boldsymbol{\theta}(\mathbf{F}). \tag{1.5}$$

For a fixed $\mathbf{n}(\geqslant \mathbf{m}^*)$, $\mathbf{U}(\mathbf{n})$ is known to be an optimal unbiased estimator of $\boldsymbol{\theta}(\mathbf{F})$, for a general class of $\mathbf{F}$; for various properties of $\mathbf{U}(\mathbf{n})$, we may refer to Section 3.3 of Puri and Sen [7]. Our interest centers here in providing a confidence region for $\boldsymbol{\theta}(\mathbf{F})$ based on $\mathbf{U}(\mathbf{n})$. Specifically, we like to determine a closed convex region $I_{\mathbf{n}}(\in R^t)$ such that

$$P\{\boldsymbol{\theta}(\mathbf{F}) \in I_{\mathbf{n}}\} = 1 - \alpha : 0 < \alpha < 1, \tag{1.6}$$

and for some preassigned positive d,

$$\omega(\mathbf{U}(\mathbf{n}), I_{\mathbf{n}}) = \sup\{\rho(\mathbf{U}(\mathbf{n}), \boldsymbol{\theta}(\mathbf{F})) : \boldsymbol{\theta}(\mathbf{F}) \in I_{\mathbf{n}}\} \leqslant d, \tag{1.7}$$

where for two t-vectors $\mathbf{a}$ and $\mathbf{b}$,

$$\rho(\mathbf{a}, \mathbf{b}) = \sup\{|\boldsymbol{l}'(\mathbf{a} - \mathbf{b})| : \boldsymbol{l}'\boldsymbol{l} = 1\}. \tag{1.8}$$

In the above setup, α, d are given, and we require to determine $\mathbf{n}$ such that $\mathbf{n}'\mathbf{1} = n$ is minimized. More specifically, we desire to determine $\mathbf{n}_0^* = \mathbf{n}_0^*(d) = (n_{10}^*(d), \ldots, n_{c0}^*(d))'$ such that $\sum_{k=1}^{c} n_{k0}^*(d) = n_0^* = n_0^*(d)$ and

$$n_0^* = \inf\{\mathbf{n}'\mathbf{1} = n : P\{\boldsymbol{\theta}(\mathbf{F}) \in I_{\mathbf{n}} : \omega(\mathbf{U}(\mathbf{n}), I_{\mathbf{n}}) \leqslant d\} = 1 - \alpha\}. \tag{1.9}$$

Now, $\mathbf{F}$ is unknown, and the distribution of $\mathbf{U}(\mathbf{n})$, in general, depends on $\mathbf{F}$.

As a result, $\mathbf{n}_0^*(d)$ as well as $n_0^*(d)$ depends on $\mathbf{F}$ in addition to that of d and α. We therefore write

$$\mathbf{n}_0^* = \mathbf{n}_0^*(d, \alpha; \mathbf{F}), \qquad n_0^* = \mathbf{n}_0^{*\prime}\mathbf{1} = n_0^*(d, \alpha; \mathbf{F}). \tag{1.10}$$

We therefore require to find an estimator

$$\mathbf{N} = \mathbf{N}(d) = (N_1(d),..., N_c(d))', \tag{1.11}$$

of $\mathbf{n}_0^*(d, \alpha; \mathbf{F})$, such that for a broad class of $\{\mathbf{F}\}$, $N = N(d) = \mathbf{1}'\mathbf{N}(d)$ is "close" to $n_0^*(d, \alpha; F)$.

It is shown here that there exists a sequential (multistage) procedure which leads to a solution $\mathbf{N}(d)$ (a stochastic c-vector), such that for a class $\mathscr{F}$ of c-tuplets of df's $\{\mathbf{F}\}$,

$$\lim_{d\to 0} P\{\boldsymbol{\theta}(\mathbf{F}) \in I_{\mathbf{N}(d)}\colon \omega(\mathbf{U}(\mathbf{N}(d)), I_{\mathbf{N}(d)}) \leqslant d\} = 1 - \alpha, \tag{1.12}$$

and

$$\lim_{d\to 0} E(\mathbf{1}'\mathbf{N}(d))/(\mathbf{1}'\mathbf{n}_0^*(d)) = 1. \tag{1.13}$$

In the sense of Chow and Robbins [2], (1.12) and (1.13) specify the asymptotic consistency and efficiency of the procedure.

For $t = 1$ and $\theta(\mathbf{F}) = \theta(F_1) - \theta(F_2)$ where $\theta(F_i) = \int_{-\infty}^{\infty} x\, dF_i(x)$, $i = 1, 2$ and the F_i, $i = 1, 2$, are univariate normal distributions with respective (unknown) variances σ_i^2, $i = 1, 2$ (not necessarily identical), Robbins *et al.* [8] considered the problem of obtaining a confidence interval of fixed width and coverage probability. Although the asymptotic consistency and efficiency results of their procedure were proven under the usual normality assumption they pointed out that these results remain true whenever F_1, F_2 have finite moments up to the fourth order, and even, that can be relaxed a little. A similar procedure (for $t = 1$) for the difference of locations in the two-sample case based on two-sample rank order statistics with exponentially integrable score functions was considered by Ghosh [4]. In the current paper, under general multivariate setup, the general case of estimable parameters (functionals of $\mathbf{F}$) is considered, and under conditions comparable to Robbins *et al.* [8], their solutions (based on generalized U-statistics) are studied.

In Section 2, the proposed procedure along with the preliminary notions is considered. Sections 3 and 4 are devoted to the study of the asymptotic consistency and efficiency of the procedure. In the last section, solutions based on von Mises' [6] differentiable statistical functions are briefly presented. Throughout the paper, for simplicity of presentation, we consider the case of $c = 2$, while the general case of $c \geqslant 2$ is treated briefly at the end.

2. Preliminary Notions and the Proposed Procedure

For every $\mathbf{0} \leqslant \mathbf{h} = (h_1, h_2)' \leqslant \mathbf{m}_i = (m_{1i}, m_{2i})'$, we let

$$\phi_{i,\mathbf{h}}(x_{k1},..., x_{kh_k}, k = 1, 2) = E\{\phi_i(x_{k1},..., x_{kh_k}, X_{kh_k+1},..., X_{km_{ki}}, k = 1, 2)\}, \tag{2.1}$$

for $i = 1,..., t$, and let

$$\zeta_{ij}(\mathbf{h}; \mathbf{F}) = E\{\phi_{i,h}(x_{k1},..., x_{kh_k}, k = 1, 2)$$
$$\cdot \phi_{j,\mathbf{h}}(x_{k1},..., x_{kh_k}, k = 1, 2)\} - \theta_i(\mathbf{F})\, \theta_j(\mathbf{F}),$$
$$\mathbf{0} \leqslant \mathbf{h} \leqslant (\min(m_{1i}, m_{1j}), \min(m_{2i}, m_{2j})), \qquad 1 \leqslant i \leqslant j \leqslant t. \tag{2.2}$$

We term $\theta_i(\mathbf{F})$ as stationary of order zero, if

$$\zeta_{ii}(1, 0; \mathbf{F}) \quad \text{or} \quad \zeta_{ii}(0, 1; \mathbf{F}) > 0, \qquad 1 \leqslant i \leqslant t. \tag{2.3}$$

Throughout the paper, it will be assumed that $\boldsymbol{\theta}(\mathbf{F})$ is stationary of order zero, so that (2.3) holds for every $i(=1,..., t)$. Then, we have (cf. [7, p. 66]) that

$$\operatorname{Cov}[U_i(\mathbf{n}), U_j(\mathbf{n})] = \binom{\mathbf{n}}{\mathbf{m}_i}^{-1} \sum_{\mathbf{h}=\mathbf{0}}^{\mathbf{m}_j} \binom{\mathbf{m}_i}{\mathbf{h}} \binom{\mathbf{n} - \mathbf{m}_i}{\mathbf{m}_j - \mathbf{h}} \zeta_{ij}(\mathbf{h}; \mathbf{F}), \tag{2.4}$$

for $1 \leqslant i \leqslant j \leqslant t$ where $\sum_{\mathbf{h}=\mathbf{0}}^{\mathbf{m}_j} = \sum_{h_1=0}^{m_{1j}} \sum_{h_2=0}^{m_{2j}}$. Let then

$$\boldsymbol{\Gamma}(\mathbf{n}) = ((\operatorname{Cov}[U_i(\mathbf{n}), U_j(\mathbf{n})]))_{i,j=1,...,t} \tag{2.5}$$

be the dispersion matrix of $[\mathbf{U}(\mathbf{n}) - \boldsymbol{\theta}(\mathbf{F})]$. We also let

$$\lambda_n = n^{-1}n_1 = n_1/(n_1 + n_2), \qquad 0 < \lambda < 1. \tag{2.6}$$

If there exists a $\lambda_0 : 0 < \lambda_0 < 1/2$ such that

$$\lim_{n\to\infty} \lambda_n = \lambda \text{ exists and } \lambda \in (\lambda_0, 1 - \lambda_0), \tag{2.7}$$

then by (2.4) and (2.5), we have

$$\begin{aligned} \lim_{n\to\infty} n\boldsymbol{\Gamma}(\mathbf{n}) &= ((\lambda^{-1} m_{1i} m_{1j} \zeta_{ij}(1, 0; F) + (1 - \lambda)^{-1} m_{2i} m_{2j} \zeta_{ij}(0, 1; F))) \\ &= \boldsymbol{\Gamma}(\lambda; \mathbf{F}), \quad \text{say.} \end{aligned} \tag{2.8}$$

Moreover, it is well-known (viz. [7, p. 65]) that as $n \to \infty$,

$$\mathscr{L}(n^{1/2}[\mathbf{U}(\mathbf{n}) - \boldsymbol{\theta}(\mathbf{F})]) \to \mathscr{N}_t(\mathbf{0}, \boldsymbol{\Gamma}(\lambda; \mathbf{F})). \tag{2.9}$$

Our procedure rests on suitable estimators of $\zeta_{ij}(1, 0; \mathbf{F})$ and $\zeta_{ij}(0, 1; \mathbf{F})$ for $i, j = 1,..., t$, based on the structural convergence properties of U-statistics studied by Sen [9]. For every $\mathbf{v} = (\nu_1, \nu_2)' \geqslant \mathbf{m}^*$, we let

$$V_{\mathbf{v},r}^{(i)}(1, 0) = \binom{\nu_1 - 1}{m_{1i} - 1}^{-1} \binom{\nu_2}{m_{2i}}^{-1} \sum_{\mathbf{v},r}^{*} \cdot \phi_i(\mathbf{X}_{k\alpha_{kj}}, 1 \leqslant j \leqslant m_{ki}, k = 1, 2; \alpha_{11} = r), \tag{2.10}$$

where the summation $\sum_{\mathbf{v},r}^{*}$ extends over all possible $1 \leqslant \alpha_{12} < \cdots < \alpha_{1m_{1i}} \leqslant \nu_1$, and $1 \leqslant \alpha_{21} < \cdots < \alpha_{2m_{i2}} \leqslant \nu_2$, with $\alpha_{1j} \neq r$, $2 \leqslant j \leqslant m_{1i}$;

$$V_{\mathbf{v},r}^{(i)}(0, 1) = \binom{\nu_1}{m_{1i}}^{-1} \binom{\nu_2 - 1}{m_{2i} - 1}^{-1} \sum_{\mathbf{v},r}^{**} \cdot \phi_i(\mathbf{X}_{k\alpha_{kj}}, 1 \leqslant j \leqslant m_{ki}, k = 1, 2; \alpha_{21} = r), \tag{2.11}$$

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

$$S_{ij,\mathbf{v}}(1, 0) = \frac{1}{\nu_1 - 1} \sum_{r=1}^{\nu_1} [V_{\mathbf{v},r}^{(i)}(1, 0) - U_i(\mathbf{v})][V_{\mathbf{v},r}^{(j)}(1, 0) - U_j(\mathbf{v})], \tag{2.12}$$

$$S_{ij,\mathbf{v}}(0, 1) = \frac{1}{\nu_2 - 1} \sum_{r=1}^{\nu_2} [V_{\mathbf{v},r}^{(i)}(0, 1) - U_i(\mathbf{v})][V_{\mathbf{v},r}^{(j)}(0, 1) - U_j(\mathbf{v})], \tag{2.13}$$

for $i, j = 1,..., t$, and for every $0 < \lambda < 1$, let

$$\hat{\mathbf{\Gamma}}(\lambda, \mathbf{v}) = \left(\left(\frac{1}{\lambda} m_{1i} m_{1j} S_{ij,\mathbf{v}}(1, 0) + \frac{1}{1 - \lambda} m_{2i} m_{2j} S_{ij,\mathbf{v}}(0, 1)\right)\right). \tag{2.14}$$

We shall use $\hat{\mathbf{\Gamma}}(\lambda, \mathbf{v})$ as an estimator of $\mathbf{\Gamma}(\lambda, \mathbf{F})$, defined by (2.8).

Let now $\mathbf{I}_t$ be the identity matrix of order $t \times t$ and let the roots of the equation

$$| \hat{\mathbf{\Gamma}}(\lambda, \mathbf{v}) - \gamma \mathbf{I}_t | = 0, \tag{2.15}$$

i.e., the characteristic roots of $\hat{\mathbf{\Gamma}}(\lambda, \mathbf{v})$, be denoted by $g_i(\lambda, \mathbf{v})$, $i = 1,..., t$. Finally, let

$$g^*(\lambda, \mathbf{v}) = \max_{1 \leqslant i \leqslant t} g_i(\lambda, \mathbf{v}). \tag{2.16}$$

The Proposed Procedure. We conceive of a set of positive integers $\{n_\alpha, \alpha \geqslant 0\}$, such that

$$n_0 \geqslant \max_{1 \leqslant i \leqslant t} \max_{1 \leqslant k \leqslant 2} m_{ki}, \tag{2.17}$$

while the subsequent entries n_α, $\alpha \geqslant 1$, need not satisfy (2.17). We start with n_0 observations from each distribution, and for $\mathbf{v}_0 = \mathbf{n}_0 = n_0(1, 1)'$, compute $\hat{\mathbf{\Gamma}}(\frac{1}{2}, \mathbf{n}_0)$ as well as $g^*(\frac{1}{2}, n_0)$. Let $\chi^2_{t,\alpha}$ be the upper $100\alpha\%$ point of the χ^2 distribution with t degrees of freedom. If then

$$2n_0 \geqslant d^{-2}\chi^2_{t,\alpha} g^*(\tfrac{1}{2}, \mathbf{n}_0), \tag{2.18}$$

we construct the confidence region

$$I_0 = \{\boldsymbol{\theta}(\mathbf{F})\colon 2n_0[\mathbf{U}(\mathbf{n}_0) - \theta(\mathbf{F})]' [\hat{\mathbf{\Gamma}}(\tfrac{1}{2}, \mathbf{n}_0)]^{-1} [\mathbf{U}(\mathbf{n}_0) - \boldsymbol{\theta}(\mathbf{F})] \leqslant \chi^2_{t,\alpha}\}. \tag{2.19}$$

Note that by definition in (1.7)–(1.8) and by (2.18) and (2.19),

$$\begin{aligned}
\omega(\mathbf{U}(\mathbf{n}_0), I_0) &= \sup\{\rho(\mathbf{U}(\mathbf{n}_0), \boldsymbol{\theta}(\mathbf{F}))\colon \boldsymbol{\theta}(\mathbf{F}) \in I_0\} \\
&= \sup\{| \boldsymbol{l}'[\mathbf{U}(\mathbf{n}_0) - \boldsymbol{\theta}(\mathbf{F})]|\colon \boldsymbol{l}'\boldsymbol{l} = 1 \text{ and } \boldsymbol{\theta}(\mathbf{F}) \in I_0\} \\
&= \sup\{([\mathbf{U}(\mathbf{n}_0) - \boldsymbol{\theta}(\mathbf{F})]' [\hat{\mathbf{\Gamma}}(\tfrac{1}{2}, \mathbf{n}_0)]^{-1} [\mathbf{U}(\mathbf{n}_0) - \boldsymbol{\theta}(\mathbf{F})] \\
&\quad \cdot [\boldsymbol{l}'\hat{\mathbf{\Gamma}}(\tfrac{1}{2}, \mathbf{n}_0)\boldsymbol{l}])^{1/2}\colon \boldsymbol{l}'\boldsymbol{l} = 1,\ \boldsymbol{\theta}(\mathbf{F}) \in I_0\} \\
&= \sup\{([\mathbf{U}(\mathbf{n}_0) - \boldsymbol{\theta}(\mathbf{F})]' [\hat{\mathbf{\Gamma}}(\tfrac{1}{2}, \mathbf{n}_0)]^{-1} [\mathbf{U}(\mathbf{n}_0) - \boldsymbol{\theta}(\mathbf{F})] \\
&\quad \cdot g^*(\tfrac{1}{2}, \mathbf{n}_0))^{1/2}\colon \boldsymbol{\theta}(\mathbf{F}) \in I_0\} \\
&= [\chi^2_{t,\alpha} g^*(\tfrac{1}{2}, \mathbf{n}_0)/2n_0]^{1/2} \\
&\leqslant d.
\end{aligned} \tag{2.20}$$

On the other hand, if (2.18) does not hold, we consider

$$\hat{\lambda}_0 = \inf\{\lambda_0\colon g^*(\lambda_0, \mathbf{n}_0) = \inf_{0<\lambda<1} g^*(\lambda, \mathbf{n}_0)\}. \tag{2.21}$$

Note that in order to avoid the possibility of multiplicity of roots, we take the minimum $\hat{\lambda}_0$ among the possible values. Define then a stochastic vector

$$\mathbf{v}_1 = (\nu_{11}, \nu_{12})', \qquad \nu_{12} = 2n_0 + n_1 - \nu_{11}, \tag{2.22}$$

where

$$\nu_{11} = \begin{cases} n_0, & \hat{\lambda}_0 < (n_0 + 1)/(2n_0 + n_1), \\ [(2n_0 + n_1)\hat{\lambda}_0], & (n_0 + 1) \leqslant (2n_0 + n_1)\hat{\lambda}_0 \leqslant n_0 + n_1, \\ n_0 + n_1, & \hat{\lambda}_0 > (n_0 + n_1)/(2n_0 + n_1), \end{cases} \tag{2.23}$$

where $[s]$ denotes the greatest integer contained in $s(\geqslant 1)$. Let then

$$\hat{\lambda}_0^* = \nu_{11}/(\nu_{11} + \nu_{12}) = \nu_{11}/(2n_0 + n_1), \tag{2.24}$$

compute $g^*(\hat{\lambda}_0^*, \mathbf{v}_1)$, and see if

$$2n_0 + n_1 = \nu_{11} + \nu_{12} \geqslant d^{-2}\chi^2_{t,\alpha} g^*(\hat{\lambda}_0^*, \mathbf{v}_1). \tag{2.25}$$

If (2.25) holds, we construct the confidence region

$$I_1 = \{\boldsymbol{\theta}(\mathbf{F}): (2n_0 + n_1)[\mathbf{U}(\boldsymbol{\nu}_1) - \boldsymbol{\theta}(\mathbf{F})]' [\hat{\boldsymbol{\Gamma}}(\hat{\lambda}_0{}^*, \boldsymbol{\nu}_1)]^{-1} [\mathbf{U}(\boldsymbol{\nu}_1) - \boldsymbol{\theta}(\mathbf{F})] \leqslant \chi^2_{t,\alpha}\}, \tag{2.26}$$

where as in (2.20), we have

$$\omega(\mathbf{U}(\boldsymbol{\nu}_1), I_1) \leqslant d. \tag{2.27}$$

If (2.25) does not hold, we consider

$$\hat{\lambda}_1 = \inf\{\lambda_0: g^*(\lambda_0, \boldsymbol{\nu}_1) = \inf_{0<\lambda<1} g^*(\lambda, \boldsymbol{\nu}_1)\}, \tag{2.28}$$

and define a stochastic vector

$$\boldsymbol{\nu}_2 = (\nu_{21}, \nu_{22})', \qquad \nu_{22} = 2n_0 + n_1 + n_2 - \nu_{21},$$

where

$$\nu_{21} = \begin{cases} \nu_{11}, & \hat{\lambda}_1 < (\nu_{11} + 1)/(2n_0 + n_1 + n_2), \\ [(2n_0 + n_1 + n_2)\hat{\lambda}_1], & (\nu_{11} + 1) \leqslant (2n_0 + n_1 + n_2)\hat{\lambda}_1 \leqslant (\nu_{11} + n_2), \\ \nu_{11} + n_2, & \hat{\lambda}_1 > (\nu_{11} + n_2)/(2n_0 + n_1 + n_2). \end{cases}$$

Let then

$$\hat{\lambda}_1{}^* = \nu_{21}/(\nu_{21} + \nu_{22}) = \nu_{21}/(2n_0 + n_1 + n_2), \tag{2.30}$$

and compute $g^*(\hat{\lambda}_1{}^*, \boldsymbol{\nu}_2)$ and see if

$$\nu_{21} + \nu_{22} \geqslant d^{-2}\chi^2_{t,\alpha} g^*(\hat{\lambda}_1{}^*, \boldsymbol{\nu}_2). \tag{2.31}$$

If (2.31) holds, construct the confidence region I_2, defined as in (2.26) with $2n_0 + n_1(=\nu_{11} + \nu_{12})$, $\mathbf{U}(\boldsymbol{\nu}_1)$ and $\hat{\boldsymbol{\Gamma}}(\hat{\lambda}_0{}^*, \boldsymbol{\nu}_1)$ being replaced by $\nu_{21} + \nu_{22}$, $\mathbf{U}(\boldsymbol{\nu}_2)$ and $\hat{\boldsymbol{\Gamma}}(\hat{\lambda}_1{}^*, \boldsymbol{\nu}_2)$, respectively; then (2.27) holds for $\mathbf{U}(\boldsymbol{\nu}_1)$ and I_1 being replaced by $\mathbf{U}(\boldsymbol{\nu}_2)$ and I_2, respectively. If (2.31) does not hold, we proceed to the next stage. The process continues until for some $k(\geqslant 0)$,

$$\nu_{k1} + \nu_{k2} \geqslant d^{-2}\chi^2_{t,\alpha} g^*(\hat{\lambda}^*_{k-1}, \boldsymbol{\nu}_k), \qquad \text{where} \quad \hat{\lambda}^*_{-1} = \tfrac{1}{2}, \tag{2.32}$$

and then our proposed confidence region for $\boldsymbol{\theta}(\mathbf{F})$ is

$$I_k = \{\boldsymbol{\theta}(\mathbf{F}): [\mathbf{U}(\boldsymbol{\nu}_k) - \boldsymbol{\theta}(\mathbf{F})]' [\hat{\boldsymbol{\Gamma}}(\hat{\lambda}^*_{k-1}, \boldsymbol{\nu}_k)]^{-1} \cdot [\mathbf{U}(\boldsymbol{\nu}_k) - \boldsymbol{\theta}(\mathbf{F})] \leqslant (\chi^2_{t,\alpha}/(\nu_{k1} + \nu_{k2}))\}. \tag{2.33}$$

Thus, our stopping number is

$$N = \nu_{k1} + \nu_{k2} = 2n_0 + n_1 + \cdots + n_k\,, \tag{2.34}$$

where, of course, both k (and hence), N are integer valued random variables.

Remark. One should note that in the above treatment, the obvious dependence of N (and k) on d and α have been suppressed in our notation, which, hopefully, will lead to no confusion. We may also note that in (2.18), (2.19), (2.25), (2.26), (2.31), (2.32), (2.33), instead of using $\chi^2_{t,\alpha}$, one could have used a sequence of positive numbers $\{a_k\,, k \geqslant 0\}$, such that

$$\lim_{k\to\infty} a_k = a = \chi^2_{t,\alpha}\,. \tag{2.35}$$

The motivation of using a_k instead of $\chi^2_{t,\alpha}$ (at the kth stage with sample size $\nu_{k1} + \nu_{k2}$) lies in providing a better approximation to the coverage probabilities of (2.19), (2.26), (2.33), etc., (to the desired $1 - \alpha$). However, precise choice of $\{a_k\}$ depends on the behavior of $\mathbf{U}(\boldsymbol{\nu}_k)$, and hence, on the original $\{\mathbf{F}\}$.

3. Asymptotic Consistency of the Proposed Procedure

In the asymptotic setup, we let $d \to 0$, and desire to study the properties of the proposed confidence procedure. Here, we set $n_0 = n_0(d)$ and assume that

$$\lim_{d\to 0} n_0(d) = \infty \qquad \text{but} \quad \lim_{d\to 0} d^2 n_0(d) = 0, \tag{3.1}$$

that is $n_0 = n_0(d)$ increases at a slower rate than d^{-2}, as $d \to 0$. Compared to (2.15) and (2.16), we consider the roots of

$$|\, \boldsymbol{\Gamma}(\lambda; \mathbf{F}) - \gamma \mathbf{I}_t \,| = 0, \tag{3.2}$$

denote these by $\gamma_1(\lambda, \mathbf{F}),\ldots, \gamma_t(\lambda, \mathbf{F})$, and define

$$\gamma^*(\lambda, \mathbf{F}) = \max_{1\leqslant j\leqslant t} \gamma_j(\lambda, \mathbf{F}), \qquad 0 < \lambda < 1. \tag{3.3}$$

Our second assumption is that $\gamma^*(\lambda, \mathbf{F})$ assumes a unique minimum $\gamma_0{}^*(\mathbf{F})$ at $\lambda = \lambda^*$, where $\lambda^* \in [\lambda_0\,, 1 - \lambda_0]$, $0 < \lambda_0 \leqslant 1/2$, i.e.,

$$\gamma_0{}^*(\mathbf{F}) = \gamma^*(\lambda^*, \mathbf{F}) = \inf_{0<\lambda<1} \gamma^*(\lambda, F) \qquad \text{and} \qquad \gamma^* \in [\lambda_0\,, 1 - \lambda_0]. \tag{3.4}$$

Our main theorem of the section is the following.

THEOREM 3.1. *If* $E\phi_i^4(\mathbf{X}_{kj}, 1 \leqslant j \leqslant m_{ki}, k = 1, 2) < \infty, \forall 1 \leqslant i \leqslant t$, *and* (3.1), (3.4) *hold, then for the stopping variable* $N = N(d)$ *and the corresponding confidence region* $I_k = I(N(d))$, *say, defined by* (2.33) *and* (2.34), *we have the following.*

$$\text{(i)} \quad N(d) \text{ is a well-defined, nonincreasing function of } d(>0), \tag{3.5}$$

$$\text{(ii)} \quad \lim_{d\to 0} N(d) = \infty \quad a.s., \qquad \lim_{d\to 0} E[N(d)] = \infty, \tag{3.6}$$

$$\text{(iii)} \quad \lim_{d\to 0}\{[d^2N(d)]/[\gamma_0^*(\mathbf{F})\,\chi^2_{t,\alpha}]\} = 1 \quad a.s., \tag{3.7}$$

and

$$\text{(iv)} \quad \lim_{d\to 0} P(\boldsymbol{\theta}(\mathbf{F}) \in I(N(d))\} = 1 - \alpha, \tag{3.8}$$

where $I(N(d))$ *is given by*

$$\{\boldsymbol{\theta}(\mathbf{F})\colon [\mathbf{U}(\mathbf{N}(d)) - \boldsymbol{\theta}(\mathbf{F})]' [\hat{\mathbf{\Gamma}}(\hat{\lambda}_N^*, \mathbf{N}(d))]^{-1} [\mathbf{U}(\mathbf{N}(d)) - \boldsymbol{\theta}(\mathbf{F})] \leqslant \chi^2_{t,\alpha}/N(d)\}, \tag{3.9}$$

$$\mathbf{N}(d) = (N_1(d), N_2(d))', \qquad N_1(d) + N_2(d) = N(d); \tag{3.10}$$

$$\hat{\lambda}_N^* = \hat{\lambda}^*_{N(d)} = N_1(d)/N(d). \tag{3.11}$$

First, we consider several basic lemmas which will subsequently be needed in the proof of the main theorem.

LEMMA 3.2. *If* $E\phi_i^4(X_{11}, ..., X_{1m_{1i}}, X_{21}, ..., X_{2m_{2i}}) < \infty$, $1 \leqslant i \leqslant t$, *then for every* $\eta(0 < \eta < \frac{1}{2})$, *as* $\mathbf{n} \to \infty$,

$$\sup_{\eta<\lambda<1-\eta} \{\sup_{\mathbf{a}\neq\mathbf{0}}[\mathbf{a}'\{\hat{\mathbf{\Gamma}}(\lambda, \mathbf{n}) - \mathbf{\Gamma}(\lambda, \mathbf{F})\}\mathbf{a}]\} \to 0 \quad a.s., \tag{3.12}$$

where $\mathbf{\Gamma}(\lambda, \mathbf{F})$ *and* $\hat{\mathbf{\Gamma}}(\lambda, \mathbf{n})$ *are defined by* (2.8) *and* (2.14).

The proof follows directly from Theorem 3.3 of Williams and Sen [12], and hence, for intended brevity, is omitted.

Now, the largest characteristic root of a square matrix is a continuous function of its elements (viz. [11]). Consequently, by (2.8), (3.2), and (3.3), for every $0 < \lambda < 1$,

$$\gamma^*(\lambda, \mathbf{F}) = g(\zeta_{ij}(1, 0; \mathbf{F}), \zeta_{ij}(0, 1; \mathbf{F}), 1 \leqslant i < j \leqslant t, \lambda), \tag{3.13}$$

where g is a continuous function of its arguments. Similarly, for $0 < \lambda < 1$, $\boldsymbol{\nu} \geqslant \mathbf{m}^*$,

$$g^*(\lambda; \boldsymbol{\nu}) = g(S_{ij,\boldsymbol{\nu}}(1, 0), S_{ij,\boldsymbol{\nu}}(0, 1), 1 \leqslant i \leqslant j \leqslant t, \lambda) \tag{3.14}$$

is a continuous function of its arguments.

LEMMA 3.3 *If* $E\phi_i^4(\cdots) < \infty$, $1 \leqslant i \leqslant t$ *and* (3.1), (3.4) *hold, then*

$$\hat{\lambda}_n \to \lambda^* \ a.s., \qquad as \quad n \to \infty, \tag{3.15}$$

where $\hat{\lambda}_n = \inf\{\lambda_0 : g^*(\lambda_0, \mathbf{n}) = \inf_{0<\lambda<1} g^*(\lambda, \mathbf{n})\}$ *and* $g^*(\lambda, \mathbf{n})$ *is defined by* (2.16).

Proof. By virtue of (2.15), (2.16), (3.2), (3.3), Lemma 3.2, and (3.13)–(3.14), it follows that as $\mathbf{n} \to \boldsymbol{\infty}$, for every $0 < \lambda_0 \leqslant \frac{1}{2}$,

$$g^*(\lambda, \mathbf{n}) \xrightarrow{\text{a.s.}} \gamma^*(\lambda, \mathbf{F}), \qquad \text{for every} \quad \lambda \in [\lambda_0, 1 - \lambda_0]. \tag{3.16}$$

By carefully examining the characteristic Eqs. in (3.2)–(3.3), we have

$$\begin{aligned}(0 \leqslant) \min_{1 \leqslant i \leqslant t} &\left\{\frac{m_{1i}^2}{\lambda} \zeta_{ii}(1, 0; \mathbf{F}) + \frac{m_{2i}^2}{1-\lambda} \zeta_{ii}(0, 1; \mathbf{F})\right\} \\ &\leqslant \gamma^*(\lambda; \mathbf{F}) \leqslant \sum_{i=1}^{t} \left\{\frac{m_{1i}^2}{\lambda} \zeta_{ii}(1, 0; \mathbf{F}) + \frac{m_{2i}^2}{1-\lambda} \zeta_{ii}(0, 1; \mathbf{F})\right\} (<\infty),\end{aligned} \tag{3.17}$$

for every $0 < \lambda < 1$. Since both the left and right hand side terms in (3.17) go to ∞ as $\lambda \to 0$ or to 1, and on the other hand, for every $0 < \lambda < 1$, these terms are finite, it readily follows that $\gamma^*(\lambda, \mathbf{F})$ attains a minimum at $\lambda = \lambda^*$ where $0 < \lambda^* < 1$. Our assumption (3.4) insures the uniqueness of λ^*. The remainder of the proof follows from (3.12), (3.16), the continuity (in λ) of $g^*(\lambda, \mathbf{n})$ and $\gamma^*(\lambda, \mathbf{F})$, and some standard reasonings. Because of its essential similarity with the technique of the proof of Theorem 3.4 of Williams and Sen [12], the details are omitted.

COROLLARY 3.3. *Under the conditions of Lemma* 3.3, *as* $\mathbf{n} \to \boldsymbol{\infty}$,

$$g^*(\hat{\lambda}_n, \mathbf{n}) \to \gamma^*(\lambda^*, \mathbf{F}) \ a.s. \tag{3.18}$$

The proof follows from (3.12), (3.15), (3.16), and a few routine steps, and is therefore omitted.

Let us now return to the proof of Theorem 3.1. Note that by (3.12), (3.17), and an analogous inequality for $g^*(\lambda, \mathbf{n})$,

$$g^*(\hat{\lambda}_n, \mathbf{n})/\gamma^*(\lambda^*, \mathbf{F}) > 0 \text{ a.s.}, \tag{3.19}$$

while by Corollary 3.3, $g^*(\hat{\lambda}_n, \mathbf{n})/\gamma^*(\lambda^*, \mathbf{F}) \to 1$ a.s., as $\mathbf{n} \to \infty$. As such the proof of (3.5), (3.6), and (3.7) follows from Lemma 1 of Chow and Robbins [2] and our (3.1).

To prove (3.8), we note that by (3.1), (3.4), Lemma 3.2, and (3.6)

$$\hat{\lambda}^*_{N(d)} \to \lambda^* \quad \text{a.s.,} \qquad \text{as} \quad d \to 0. \tag{3.20}$$

The uniform continuity, in probability, of generalized U-statistics, already established in Lemma 3.5 of Williams and Sen [12], implies that for every $\epsilon > 0$ and $\eta > 0$, there exist a $\delta > 0$ and an n_0, such that for $\mathbf{n} \geqslant n_0 \mathbf{1}$,

$$P\left\{\max_{\|\mathbf{v}-\mathbf{n}\|<\delta\|\mathbf{n}\|} \sup_{\mathbf{l}\neq\alpha} (\mathbf{l}'\mathbf{A}\mathbf{l})^{-1/2} \,|\, \mathbf{l}'[\mathbf{U}(\mathbf{n}) - \mathbf{U}(\mathbf{v})]| > \epsilon \left(\frac{n_1 + n_2}{n_1 n_2}\right)^{1/2}\right\} < \eta, \tag{3.21}$$

where $\mathbf{A}$ is a positive-definite matrix, $\| \mathbf{n} \| = \max(n_1, n_2)$ and $\mathbf{n}$ satisfies the condition that $n_1/(n_1 + n_2) \in [\lambda_0, 1 - \lambda_0]$. By virtue of (2.9), (3.7), (3.20), and (3.21), we obtain on using the Anscombe [1] theorem that as $d \to 0$,

$$\mathscr{L}([N(d)]^{1/2} [\mathbf{U}(\mathbf{N}(d)) - \boldsymbol{\theta}(\mathbf{F})]) \to \mathscr{N}_t(0, \boldsymbol{\Gamma}(\lambda^*, \mathbf{F})), \tag{3.22}$$

where $\boldsymbol{\Gamma}(\lambda^*, \mathbf{F}) = \boldsymbol{\Gamma}(\lambda, \mathbf{F})|_{\lambda=\lambda^*}$. Moreover, by (3.12) and (3.20) as $d \to 0$

$$\hat{\boldsymbol{\Gamma}}(\hat{\lambda}^*_{N(d)}\,;\, \mathbf{N}(d)) \to \boldsymbol{\Gamma}(\lambda^*, \mathbf{F}) \quad \text{a.s.} \tag{3.23}$$

Consequently, by (3.22) and (3.23), as $d \to 0$,

$$\mathscr{L}(N(d)[\mathbf{U}(\mathbf{N}(d)) - \boldsymbol{\theta}(\mathbf{F})]' \,[\hat{\boldsymbol{\Gamma}}(\hat{\lambda}^*_{N(d)}\,;\, N(d))]^{-1} \,[\mathbf{U}(\mathbf{N}(d)) - \boldsymbol{\theta}(\mathbf{F})]) \to \chi_t^2, \tag{3.24}$$

where χ_t^2 has the central chi-square distribution with t d.f., and the proof of (3.8) is complete.

4. Asymptotic Efficiency

If $\mathbf{F}$ were specified, for positive α and d, one could have obtained the desired sample size $n_0^*(d, \alpha, \mathbf{F})(=n_0^*(d)$, say), as in (1.9)–(1.10). In such a case, one can construct a confidence region $I(n_0^*(d))$, defined by (2.33) with $\mathbf{v}_k$ and $\hat{\boldsymbol{\Gamma}}(\hat{\lambda}^*_{1-k}, \mathbf{v}_k)$ being replaced by $\mathbf{n}_0^*(d)$ and $\boldsymbol{\Gamma}(\lambda^*, \mathbf{F})$, respectively. Then, by some standard steps, we have

$$\lim_{d\to 0}\{d^2 n_0^*(d)/\gamma_0^*(\mathbf{F})\, \chi^2_{t,\alpha}\} = 1. \tag{4.1}$$

In our problem, $\mathbf{F}$ is unknown, and for our proposed sequential procedure, $N(\mathbf{d})$ as well as $\mathbf{N}(\mathbf{d})$ is stochastic. Our contention is to strengthen (3.7) to

$$\lim_{d\to 0}\{(d^2E[N(d)])/(\gamma_0^*(\mathbf{F})\,\chi^2_{t,\alpha})\} = 1, \tag{4.2}$$

so that by comparison with (4.1), we can term that our proposed procedure is asymptotically efficient.

As in (3.3), (3.4), we assume that $\gamma^*(\lambda, \mathbf{F})$ has a unique minimum $\gamma_0^*(\mathbf{F}) = \gamma^*(\lambda^*, \mathbf{F})$, where $\lambda^* \in [\lambda_0, 1 - \lambda_0]$. Therefore, for every $\epsilon > 0$, there exists an $\eta > 0$, such that

$$\gamma^*(\lambda, \mathbf{F}) \geqslant \gamma_0^*(\mathbf{F}) + 2\epsilon, \qquad \text{for every} \quad |\lambda - \lambda^*| \geqslant \eta. \tag{4.3}$$

Also, by (3.13) and (3.14), for every $0 < \lambda < 1$, the continuity of g implies that for every $\epsilon > 0$, there exists a $\epsilon' > 0$, such that

$$[|S_{ij,\mathbf{n}}(\mathbf{u}) - \zeta_{ij}(\mathbf{u}; \mathbf{F})| \leqslant \epsilon', \mathbf{u} = (1, 0), (0, 1), 1 \leqslant i \leqslant j \leqslant t] \tag{4.4}$$

insures that for $0 < \lambda_0 < \frac{1}{2}$,

$$|g^*(\lambda, \mathbf{n}) - \gamma^*(\lambda, \mathbf{F})| \leqslant \epsilon \;\forall \lambda_0 \leqslant \lambda \leqslant 1 - \lambda_0\,. \tag{4.5}$$

Since $g^*(\lambda, \mathbf{n})$ and $\gamma^*(\lambda, \mathbf{F})$ attain minima at $\lambda = \hat{\lambda}_n^*$ and λ^*, respectively' it follows from (4.3) and (4.5) that

$$(4.4) \Rightarrow |g^*(\hat{\lambda}_n^*, \mathbf{n}) - \gamma^*(\lambda^*, \mathbf{F})| \leqslant \epsilon, \qquad |\hat{\lambda}_n^* - \lambda^*| \leqslant \eta. \tag{4.6}$$

Let us now assume that for some $\delta > 0$,

$$E\,|\phi_i(X_{11}, \ldots, X_{2m_{2i}})|^{4(1+\delta)} < \infty, \qquad \text{for} \quad i = 1, \ldots, t. \tag{4.7}$$

Then, we have the following lemma.

LEMMA 4.1. *Under* (4.7), *for every* $\epsilon' > 0$, *there exist a positive* $c_{\epsilon'}(<\infty)$ *and an integer* $n_0(\epsilon')$, *such that for* $\mathbf{n} \geqslant n_0(\epsilon')\mathbf{1}$,

$$P\{|S_{ij,\mathbf{n}}(\mathbf{u}) - \zeta_{ij}(\mathbf{u}; \mathbf{F})| > \epsilon', \textit{for some } \mathbf{u} = (1, 0), (0, 1) \textit{ and } 1 \leqslant i \leqslant j \leqslant t\} \leqslant c_{\epsilon'} n^{-1-\delta}. \tag{4.8}$$

Proof. Let us write

$$S^*_{ij,\mathbf{n}}(1, 0) = \frac{1}{n_1}\sum_{\alpha=1}^{n_1} \psi_{i,10}(X_{1\alpha})\,\psi_{j,10}(X_{1\alpha}), \qquad \psi_{i,10}(x) = \phi_{i,10}(x) - \theta_i(\mathbf{F}), \tag{4.9}$$

$$S^*_{ij,\mathbf{n}}(0, 1) = \frac{1}{n_2}\sum_{\alpha=1}^{n_2} \psi_{i,01}(X_{2\alpha})\,\psi_{j,01}(X_{2\alpha}), \qquad \psi_{i,01}(x) = \phi_{i,01}(x) - \theta_i(F), \tag{4.10}$$

for $1 \leqslant i \leqslant j \leqslant t$. Then (4.7) implies that

$$E \mid S_{ij,\mathbf{n}}^{*}(\mathbf{u})|^{2(1+\delta)} < \infty, \quad \text{for} \quad \mathbf{u} = (1,0), (0,1), \quad 1 \leqslant i \leqslant j \leqslant t. \tag{4.11}$$

Since the $S_{ij,\mathbf{n}}^{*}(\mathbf{u})$ involve an average over independent and identically distributed random variables, by a well-known result on the sample mean (viz. [3]), we have under (4.11),

$$P\{| S_{ij,\mathbf{n}}^{*}(\mathbf{u}) - \zeta_{ij}(\mathbf{u}; \mathbf{F})| > \tfrac{1}{2}\epsilon_1\} \leqslant c_{\epsilon'}^{(1)} n^{-(1+\delta)}; \qquad c_{\epsilon'}^{(1)} < \infty, \tag{4.12}$$

for every $\mathbf{u} = (1, 0)$, $(0, 1)$ and $1 \leqslant i \leqslant j \leqslant t$. Thus, by the Bonferroni inequality,

$$P\{| S_{ij,\mathbf{n}}^{*}(\mathbf{u}) - \zeta_{ij}(\mathbf{u}; \mathbf{F})| > \tfrac{1}{2}\epsilon_1, \quad \text{for some} \quad 1 \leqslant i \leqslant j \leqslant t$$

and

$$\mathbf{u} = (1,0), (0,1)\} \leqslant t(t+1)\, c_{\epsilon'}^{(1)} n^{-1-\delta}. \tag{4.13}$$

It has been shown by Williams and Sen [12] that each $S_{ij,\mathbf{n}}(\mathbf{u})$ can be expressed as a linear combination of several generalized U-statistics whose kth moments exist when $E \mid \phi_i \mid^{2k} < \infty$, $1 \leqslant i \leqslant t$. As such, by Theorem 2.1 of Grams and Serfling [5], it follows by a few standard steps that

$$E \mid S_{ij,\mathbf{n}}(\mathbf{u}) - S_{ij,\mathbf{n}}^{*}(\mathbf{u})|^2 = O(n^{-2}), \tag{4.14}$$

for every $1 \leqslant i \leqslant j \leqslant t$ and $\mathbf{u} = (0, 1)$, $(1, 0)$. Hence, for $n \geqslant n_0(\epsilon')$,

$$P\{| S_{ij,\mathbf{n}}(\mathbf{u}) - S_{ij,\mathbf{n}}^{*}(\mathbf{u})| > \tfrac{1}{2}\epsilon_1, \quad \text{for some} \quad 1 \leqslant i \leqslant j \leqslant t$$

and

$$\mathbf{u} = (1,0), (0,1)\} \leqslant c_{\epsilon'}^{(2)} n^{-2}, \qquad c_{\epsilon'}^{(2)} < \infty. \tag{4.15}$$

The lemma directly follows from (4.13) and (4.15) when $0 < \delta < 1$. For $\delta \geqslant 1$, let k be the largest even integer contained in $2(1 + \delta)$. Then, we show by using Theorem 2.1 of Grams and Serfling [5] that for the kth order moment, we have n^{-k} in (4.14), so that the proof again follows on parallel lines.

THEOREM 4.2. *If* $\gamma^*(\lambda, \mathbf{F})$ *has a unique minimum* $\gamma_0^*(\mathbf{F})$ *at* $\lambda = \lambda^*$ *and* (4.7) *holds, then* (4.2) *holds, that is, the proposed procedure is asymptotically efficient.*

Proof. Let us define $n_0^*(d)$ as in (1.9), (1.10), and let

$$n_{0,1}^{*}(d) = [n_0^*(d)(1-\epsilon)], \qquad n_{0,2}^{*}(d) = [n_0^*(d)(1+\epsilon)] + 1, \tag{4.16}$$

where $\epsilon > 0$ is arbitrarily small. Then, by virtue of Lemma 4.1, it readily follows that

$$\lim_{d \to 0} \sum_{n=n_{0,2}^{*}(d)}^{\infty} P\{N(d) > n\} = 0. \tag{4.17}$$

Noting that

$$d^2E[N(d)] = d^2\left[\left(\sum_{n<n_{0,1}^*(d)} + \sum_{n=n_{0,1}^*(d)+1}^{n_{0,2}^*(d)-1} + \sum_{n\geqslant n_{0,2}^*(d)}\right) nP\{N(d) = n\}\right], \tag{4.18}$$

where as $d \to 0$,

$$\begin{aligned} d^2 \sum_{n<n_{0,1}^*(d)} nP\{N(d) = n\} &\leqslant [d^2n_{0,1}^*(d)]\, P\{N(d) \leqslant n_{0,1}^*(d)\} \\ &\leqslant (1-\epsilon)\, \gamma_0^*(\mathbf{F})\, \chi_{t,\alpha}^2 P\{N(d)/n_0^*(d) \leqslant 1-\epsilon\} \\ &\leqslant \eta, \quad \text{by (4.1) and (3.7)}, \end{aligned} \tag{4.19}$$

$\eta(>0)$ being some arbitrarily small positive number,

$$\begin{aligned} &d^2 \sum_{n>n_{0,2}^*(d)} nP\{N(d) = n\} \\ &\quad = d^2n_{0,2}^*(d)\, P\{N(d) \geqslant n_{0,2}^*(d)\} + \sum_{n>n_{0,2}^*(d)} P\{N(d) \geqslant n\} \\ &\quad = [\gamma_0^*(\mathbf{F})\, \chi_{t,\alpha}^2 + O(d^2)](1+\epsilon)\, P\{N(d) \geqslant n_{0,2}^*(d)\} + \sum_{n>n_{0,2}^*(d)} P\{N(d) \geqslant n\} \\ &\quad \leqslant \eta, \quad \text{by (4.17), (3.7), and (4.1)}, \end{aligned} \tag{4.20}$$

and finally, by (4.16) and (3.7)

$$\begin{aligned} &\left| d^2 \sum_{n=n_{1,0}^*(d)+1}^{n_{0,2}^*(d)} nP\{N(d) = n\} - \gamma_0^*(\mathbf{F})\, \chi_{t,\alpha}^2 \right| \\ &\quad \leqslant (d^2n_0^*(d)\epsilon)\, P\{n_{0,1}^*(d) < N(d) < n_{0,2}^*(d)\} + o(1) \\ &\quad \leqslant \eta, \qquad \text{as} \quad d \to 0, \quad \text{by (3.7) and (4.1)}. \end{aligned} \tag{4.21}$$

The proof of (4.2) follows readily from (4.19), (4.20), and (4.21). Q.E.D.

5. Some Concluding Remarks

First, we sketch the case of $c \geqslant 2$. Here, analogous to (2.8), the dispersion matrix $\mathbf{\Gamma}(\boldsymbol{\lambda}, \mathbf{F})$ is 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 kth place and 0 elsewhere, $1 \leqslant k \leqslant c$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_c)$, $\lambda_k > 0$, $\sum_{k=1}^{c} \lambda_k = 1$. Here, Theorems 3.1 and 4.2 can be extended on parallel lines to the general case of $c \geqslant 2$. From computational aspects, however, the procedure becomes more laborious for $c > 2$, as one has to consider simultaneously the variation of $\lambda_1, \ldots, \lambda_c$, subject to $\sum_{k=1}^{c} \lambda_k = 1$. The characteristic roots in (2.16) or elsewhere, will be a function of $\boldsymbol{\lambda}$, so that minimizing these with respect to $\boldsymbol{\lambda}$ (subject to $\boldsymbol{\lambda}'\mathbf{1} = 1$) becomes computationally tedious.

If we define for $k = 1, \ldots, c$ the empirical *df* $F_{n_k,k}(\mathbf{x})$ as

$$F_{n_k,k}(\mathbf{x}) = n_k^{-1} \sum_{i=1}^{n_k} c(\mathbf{x} - \mathbf{X}_{ki}), \qquad \mathbf{x} \in R^p,$$

where $c(u)$ is 1 if all its p arguments are nonnegative, and otherwise $c(u)$ is equal to 0, then the von Mises statistic is

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

where the summation $\sum$ extends over all $1 \leqslant \alpha_{kj} \leqslant n_k$, $k = 1, \ldots, c$; $j = 1, \ldots, m_{ki}$. A study of the asymptotic distribution theory of such functionals of the empirical *df*'s has been made by von Mises [6]. Since, whenever the variance of $\theta_i(F_{n_1,1}, \ldots, F_{n_c,c})$ exists,

$$n^{1/2}[\theta_i(F_{n_1,1}, \ldots, F_{n_c,c}) - U_i(n_1, \ldots, n_c)] \to 0 \quad \text{a.s.}$$

as $n \to \infty$, Theorems 3.1 and 4.2 may be extended to apply for von Mises' statistics as opposed to U-statistics.

Some applications will be considered in a separate paper.

References

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

[2] Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence interval for the mean. *Ann. Math. Statist.* **36** 457–462.

[3] Chung, K. L. (1951). The strong law of large numbers. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 341–352, University of California Press, Berkeley, California.

[4] Ghosh, M. (1971). Robust sequential confidence intervals for the Behrens–Fisher problem. *Calcutta Statist. Assoc. Bull.* **20** 77–82.

[5] Grams, W. F. and Serfling, R. J. (1973). Convergence rates for U-statistics and related statistics. *Ann. Statist.* **1** 153–160.

[6] VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18** 309–348.
[7] PURI, M. L. AND SEN, P. K. (1971). *Nonparametric Methods in Multivariate Analysis.* John Wiley, New York.
[8] ROBBINS, H., SIMONS, G., AND STARR, N. (1967). A sequential analogue of the Behrens–Fisher problem. *Ann. Math. Statist.* **38** 1384–1391.
[9] SEN, P. K. (1960). On some convergence properties of U-statistics. *Calcutta Statist. Assoc. Bull.* **10** 1–18.
[10] SEN, P. K. AND GHOSH, M. (1971). On bounded length sequential confidence intervals based on one-sample rank order statistics. *Ann. Math. Statist.* **42** 189–203.
[11] TAUSSKY, D. (1964). On the variation of the characteristic roots of a finite matrix under various changes of its elements. *In* "Recent Advances in Matrix Theory," (H. Schneider, Ed.), pp. 125–138. University Wisconsin Press, Wisconsin.
[12] WILLIAMS, G. W. AND SEN, P. K. (1973). Asymptotically optimal sequential estimation of regular functionals of several distributions based on generalized U-statistics. *J. Multivariate Anal.* **3** 469–482.