

SEQUENTIAL SHRINKAGE U-STATISTICS:
GENERAL ASYMPTOTICS

by

P.K. Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1803

September 1986

SEQUENTIAL SHRINKAGE U-STATISTICS: GENERAL ASYMPTOTICS

Pranab Kumar Sen

Department of Biostatistics

University of North Carolina,

Chapel Hill, N.C. 27514, U.S.A.

For general estimable parameters in a nonparametric setup, shrinkage (Stein-rule) and preliminary test estimator versions of U-statistics are considered for the (multi-parameter) minimum risk sequential estimation problem. In the usual fashion, allowing the cost per unit sample to be small, an asymptotic model is framed, and in this setup, the asymptotic distributional risks of these versions of (sequential) U-statistics are studied. Related asymptotic risk-efficiency results are also considered.

Key words and phrases: Asymptotic (distributional) risk; local alternatives; preliminary test; preliminary test estimation; risk; risk-efficiency; sequential point estimation, Shrinkage estimation; stopping rule; U-statistics; von Mises' statistical functionals.

AMS Subject Classifications: 62G05, 62G20, 62L12.

Running Head: Sequential Shrinkage U-Statistics.

1. INTRODUCTION

In the nonparametric estimation theory, Hoeffding's (1948) U-statistics and the related von Mises (1947) statistical functionals play a central role. Optimality and other (asymptotic) properties of these estimators have been studied extensively in the literature; some account of these developments is given in Sen (1981; Ch. 3, 9 and 10). In the context of minimum risk estimation in a multi-parameter situation, Stein-rule (or shrinkage) estimation theory usually provides alternative estimators which dominate the classical ones. For an excellent account of the shrinkage estimation theory, (mostly) for multivariate normal distributional models, we may refer to Berger (1985). Usually, the Stein-rule estimators dominate the classical version, but it fails to dominate the preliminary test estimators (PTE). A PTE, on the other hand, may perform very well in a certain part of the parameter space, but fails to dominate the classical version. Both the shrinkage estimator and PTE have robust risk-efficiency pictures. This scenario for multi-normal distributional models is preserved in a general nonparametric setup, under suitable asymptotic considerations. Nonparametric shrinkage estimation of the multivariate location (vector) has been studied by Sen and Saleh (1985), and shrinkage as well as PTE versions of U-statistics were studied earlier by Sen (1984). For maximum likelihood estimators, parallel results have recently been studied by Sen (1986a). All these developments relate to the traditional fixed-sample size (i.e., non-sequential) case, where the sample size is non-stochastic.

In the minimum-risk estimation problem, even in the simplest uniparameter case, other nuisance parameter(s) may enter into the risk function, and hence, generally, no fixed sample size solution may exist. However, sequential estimation rules, generally, provide (at least, asymptotically)

efficient solutions. For the mean vector of a multivariate normal distribution (with unknown covariance matrix), sequential estimation procedures were considered by Ghosh, Sinha and Mukhopadhyay (1976) and others. For U-statistics, sequential estimation procedures are due to Sen and Ghosh (1981). For other nonparametric sequential estimation procedures, we may refer to Sen (1981, Ch. 10). There is a natural expectation on such sequential estimation procedures in a shrinkage estimation framework, and our current study centers around this theme.

For the multi-variate normal mean vector, two-stage shrinkage estimators are due to Ghosh and Sen (1983), and sequential shrinkage procedures are due to Takada (1984) and Ghosh, Nickerson and Sen (1986), among others. Using intricate properties of multi-normal distributions, the last paper exhibits the dominance of the sequential Stein-rule estimator over the classical sequential estimator (considered by Ghosh et al. (1976)) as well as the alternative one due to Takada (1984). In a meaningful asymptotic setup, the theory of Stein-rule and PTE for the maximum likelihood estimators has also been extended to the sequential case by Sen (1986b). In the current study, we concentrate on the Stein-rule and PTE versions of U-statistics and von Mises' statistical functionals in a sequential setup.

Our primary goal is to study the asymptotic risk and related risk-efficiency results for the Stein-rule and PTE versions of U-statistics and von Mises' functionals. This study is greatly facilitated by the incorporation of the notion of asymptotic distributional risk (ADR). This concept, in the non-sequential case, has been laid down systematically in Sen (1984) and Sen and Saleh (1985), and, in the sequential case, for the maximum likelihood estimators, in Sen (1986b). The ADR and related results for the sequential Stein-rule and PTE U-statistics, considered in later sections,

thus extend the earlier results of Sen (1984) to the sequential case.

Along with the preliminary notions on U-statistics (and von Mises' functionals), the proposed shrinkage and PTE (sequential) U-statistics are considered in Section 2. The notion of ADR is then introduced in Section 3; the general case of sequential estimators is considered in this setup. Section 4 deals with the ADR of the sequential Stein-rule and PTE versions of U-statistics. Related asymptotic risk-efficiency (ARE) results are presented in Section 5. The concluding section is devoted to some general remarks.

2. SEQUENTIAL SHRINKAGE AND PTE U-STATISTICS

Consider a sequence $\{X_i; i \geq 1\}$ of independent and identically distributed (i.i.d.) random vectors (r.v.) with a distribution function (d.f.) F , defined on the q -dimensional Euclidean space E^q , for some $q \geq 1$. Let \mathcal{F} be the space of all d.f.'s belonging to a class, and for every $F \in \mathcal{F}$, consider a vector of functionals

$$\theta = \theta(F) = (\theta_1(F), \dots, \theta_p(F))^T, \text{ for some } p \geq 1, \quad (2.1)$$

whose domain is \mathcal{F} . If there exist kernels $\phi_j(x_1, \dots, x_{m_j})$, symmetric in its arguments, of degree $m_j (\geq 1)$, such that

$$\theta_j(F) = E_F \phi_j(X_1, \dots, X_{m_j}), \quad \forall F \in \mathcal{F}, \quad 1 \leq j \leq p, \quad (2.2)$$

then θ is an estimable parameter (vector) of degree $m^* = \max\{m_1, \dots, m_p\}$.

For $n \geq m^*$, we may define $U_n = (U_{n1}, \dots, U_{np})'$, the vector of U-statistics, by letting

$$U_{nj} = \binom{n}{m_j}^{-1} \sum_{1 \leq i_1 < \dots < i_{m_j} \leq n} \phi_j(X_{i_1}, \dots, X_{i_{m_j}}), \quad 1 \leq j \leq p. \quad (2.3)$$

U_n is a symmetric, unbiased estimator of θ having some optimal properties [viz. Hoeffding(1948)]. In particular, if T_n is an unbiased estimator of θ , then the corresponding U_n has a risk smaller than or equal to that of T_n .

For every $n (\geq 1)$, define the sample (empirical) d.f. F_n , by

$$F_n(x) = n^{-1} \sum_{i=1}^n I(X_i \leq x), \quad x \in E^q. \quad (2.4)$$

Then the von Mises' (1947) statistical functional is defined by

$$V_n = \theta(F_n) = (V_{n1}, \dots, V_{np}) = (\theta_1(F_n), \dots, \theta_p(F_n)), \quad (2.5)$$

where

$$\theta_j(F_n) = n^{-m_j} \sum_{i_1=1}^n \cdots \sum_{i_{m_j}=1}^n \phi_j(X_{i_1}, \dots, X_{i_{m_j}}), \quad 1 \leq j \leq p. \quad (2.6)$$

In general, V_n is not unbiased for θ , although U_n and V_n are generally very close to each other. Actually,

$$\|U_n - V_n\| = O(n^{-1}) \text{ almost surely (a.s.)}, \text{ as } n \rightarrow \infty. \quad (2.7)$$

[viz., Ch. 3 of Sen (1981)]. For this reason, we shall mainly study the case of U_n , and, in the concluding section, append a general discussion on V_n .

We assume that the kernels (ϕ_j) are all square integrable. Let then

$$\zeta_{jj',c} = \text{Cov}\{\phi_j(X_1, \dots, X_{m_j}), \phi_{j'}(X_{m_j-c+1}, \dots, X_{m_j+m_{j'}-c})\}, \quad (2.8)$$

for $c = 0, 1, \dots, \min\{m_j, m_{j'}\}$, $j, j' = 1, \dots, p$. Then, by an appeal to Hoeffding (1948), we have

$$\begin{aligned} nE_F[(U_n - \theta)(U_n - \theta)^T] &= \Gamma_n = ((\gamma_{njj'})) \\ &= n \left(\binom{n}{m_j}^{-1} \sum_{c=1}^{m_j} \binom{m_{j'}}{c} \binom{n-m_{j'}}{m_j-c} \zeta_{jj',c} \right) \\ &= ((m_j m_{j'} \zeta_{jj',1}(F) + O(n^{-1}))) \\ &= \Gamma + O(n^{-1}); \quad \Gamma = ((\zeta_{jj'})) = ((m_j m_{j'} \zeta_{jj',1})). \end{aligned} \quad (2.9)$$

Generally, Γ_n (or Γ) is an unknown matrix, and we consider the following jackknifed estimator of [viz., Sen (1977, 1981)]. For every $i (= 1, \dots, n)$, we let

$$U_{n-1}^{(i)} = U(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n), \quad U_{n,i} = nU_n - (n-1)U_{n-1}^{(i)}. \quad \text{Then}$$

$$\hat{\Gamma} = (n-1)^{-1} \sum_{i=1}^n (U_{n,i} - U_n)(U_{n,i} - U_n)^T \quad (2.10)$$

is the jackknifed estimator of Γ .

First, we introduce the non-sequential versions. For an estimator δ_n of θ , we consider a quadratic loss function $L_n(\delta_n, \theta) = (\delta_n - \theta)^T Q_n (\delta_n - \theta)$,

where Q_n is a given positive definite (p.d.) matrix. Then the risk of δ_n is given by

$$\rho_n(\delta_n, \theta) = E_{\theta} L_n(\delta_n, \theta) = \text{Tr}(Q_n \Sigma_n), \quad (2.11)$$

where Σ_n stands for the dispersion matrix of δ_n . The following Stein-rule (shrinkage) U-statistic was considered in Sen (1984):

$$U_n^S = (I - s d_n (n U_n^T \hat{\Gamma}_n^{-1} U_n)^{-1} Q_n^{-1} \hat{\Gamma}_n^{-1}) U_n, \quad (2.12)$$

where $(0 < s < 2(p-2))$ is a positive constant (for $p > 2$), $d_n = \text{ch}(Q_n \hat{\Gamma}_n)$ is the smallest characteristic root of $Q_n \hat{\Gamma}_n$, and in (2.12), we use the null pivot. [Otherwise, for a given pivot θ_0 , replace U_n^S and U_n by $U_n^S - \theta_0$ and $U_n - \theta_0$, respectively. Without any loss of generality, we may take $\theta_0 = 0$ and consider the simpler form in (2.12)]

Note that $L_n = n U_n^T \hat{\Gamma}_n^{-1} U_n$ is a test statistic for testing the null hypothesis $H_0: \theta = 0$. Under the null hypothesis H_0 , L_n has asymptotically the central chi square distribution with p degrees of freedom (DF). Thus, $\ell_{n,\alpha}$, the α -level critical value of L_n can be approximated by $\chi_{p,\alpha}^2$, the upper $100\alpha\%$ point of the chi square d.f. with p DF. Then PTE version of U_n , considered in Sen (1984), is given by

$$U_n^{PT} = \begin{cases} U_n, & \text{if } L_n > \ell_{n,\alpha} \\ 0, & \text{otherwise.} \end{cases} \quad (2.13)$$

Thus, both U_n^S and U_n^{PT} are adaptive U-statistics, where the information on the validity of the pivot (0) based on the statistic L_n is incorporated in a suitable form. In this respect, U_n^S may also be regarded as a smoother version of U_n^{PT} .

In a sequential setup, the cost of sampling (i.e., drawing the observations X_i , $i \geq 1$) is also incorporated in the formulation of the loss function as well as the corresponding risk. If $c(>0)$ is the (known) cost per unit sample, then, we may consider the following loss function

$$L(\delta_n, \theta; c) = (\delta_n - \theta)^T Q_n (\delta_n - \theta) + cn, \quad (2.14)$$

where we take $Q_n = Q$ (given). In this setup, the risk function is

$$\begin{aligned} \rho(\delta_n, \theta; c) &= E_F L(\delta_n, \theta; c) \\ &= \text{Tr}(Q \Sigma_n) + cn. \end{aligned} \quad (2.15)$$

Note that for $\delta_n = U_n$, by virtue of (2.9) and the reversed martingale property of $\{U_n\}$, $\text{Tr}(Q \Sigma_n)$ is nonincreasing in $n (> m^*)$, while for every $c (> 0)$, cn is increasing in n . Thus, the risk in (2.15), for $\delta_n = U_n$, is given by $cn + n^{-1} \text{Tr}(Q \Gamma) + o(n^{-2})$. Hence, proceeding as in Sen and Ghosh (1981), we conclude that for every $c (> 0)$, there exists a positive integer $n_c^0 (> m^*)$, such that

$$\rho(U_{n_c^0}, \theta; c) = \min_{n > m^*} \rho(U_n, \theta; c), \quad (2.16)$$

and moreover, for small values of $c (> 0)$,

$$n_c^0 \sim c^{-\frac{1}{2}} [\text{Tr}(Q \Gamma)]^{\frac{1}{2}}. \quad (2.17)$$

However, generally, Γ is unknown, so that this optimal sample size n_c^0 is also unknown. Thus, in general, there may not be a fixed-sample size solution (even asymptotically, for $c > 0$) for the minimum-risk estimation of θ based on $\{U_n\}$. However, the sequential procedure considered in Sen and Ghosh (1981) can be adapted to derive some asymptotically optimal solutions.

Keeping in mind, the estimator $\hat{\Gamma}$ in (2.10) and the optimal sample size n_c^0 in (2.17), and proceeding as in Sen and Ghosh (1981), we define the stopping number

$$\begin{aligned} N_c &= \text{smallest positive integer } n (> m^*), \text{ such that} \\ n^2 &\geq c^{-1} \{ \text{Tr}(Q \hat{\Gamma}_n) + n^{-a} \}, \quad c > 0, \end{aligned} \quad (2.18)$$

where $a (> 0)$ is a suitable constant, which will be defined more formally later on. The main purpose of introducing the term n^{-a} in (2.20) is to eliminate a very early stopping when $\hat{\Gamma}_n$ may be nearly singular or "small" in a suitable norm. It is clear from the above definition that N_c is nonincreasing in $c (> 0)$,

and by using the a.s. convergence of $\hat{\Gamma}$ to Γ [viz., Sen (1977)], it follows that for square integrable kernels,

$$\lim_{c \rightarrow 0} N_c = +\infty, \text{ with probability one,} \quad (2.19)$$

while for every (fixed) $c > 0$,

$$P\{N_c < \infty\} = 1. \quad (2.20)$$

Based on this well-defined stopping rule, we consider the following sequential shrinkage U-statistic:

$$U_{N_c}^S, \text{ defined by (2.12), for } n = N_c, \quad c > 0. \quad (2.21)$$

Side by side, corresponding to the PTE version in (2.15), we consider the sequential PTE U-statistic

$$U_{N_c}^{PT}, \text{ defined by (2.13), for } n = N_c, \quad c > 0. \quad (2.22)$$

In passing, we may remark that for both the sequential versions in (2.21) and (2.22), we have a sequential test statistic L_{N_c} based on the stopping rule in (2.18), where N_c is not the usual stopping rule associated with the sequential probability ratio test (SPRT) (or other related ones) for $H_0: \theta = 0$. Rather, it is formulated with the minimum risk estimation in mind, and adapted here for the Stein-rule estimation problem.

3. ADR IN THE SEQUENTIAL CASE

For the usual sequential U-statistic U_{N_c} , defined by (2.3) for $n = N_c$, where N_c is defined by (2.18), it follows from Sen and Ghosh (1981) that if $\theta(\bar{F})$ is stationary of order 0, the kernel $\phi_j \in L^r$, for some $r > 2$ and a , in (2.20), is less than $(r-2)^2/2r$, then

$$\lim_{c \rightarrow 0} \rho(U_{N_c}, \theta; c) / \rho(U_n, \theta; c) = 1, \quad (3.1)$$

so that U_{N_c} is asymptotically (first-order) risk-efficient. On the other hand, from the results of Sen (1984), it follows that the asymptotic risk of U_n^S is smaller than that of U_n , for every $p \geq 3$, so that if we are able to

verify that (3.1) holds even when $U_{n_c^0}$ and U_{N_c} are replaced by $U_{n_c^0}^S$ and $U_{N_c}^S$, respectively, then we are able to improve the sequential estimator U_{N_c} by the corresponding shrinkage estimator $U_{N_c}^S$. Side by side, we also want to compare the performance of the two sequential versions $U_{N_c}^S$ and $U_{N_c}^{PT}$, when c is small.

Towards these objectives, first, we note that computation of the exact risk of the shrinkage on the PTE U-statistic, even in the non-sequential case, is an immense job. For this reason, the concept of asymptotic risk (by allowing $c \downarrow 0$) has been incorporated, in the literature, to draw such a picture. Even so, the presence of the reciprocal of \mathcal{L}_n in (2.12), demands more stringent regularity conditions on the kernel and the underlying d.f. F , under which the expectations in (2.11) exist and have nice asymptotes. This technicality may largely be avoided by appealing to the notion of asymptotic distributional risks (ADR) which are solely computed by using the asymptotic distributions of suitably normalized forms of the estimates. ADR provides a meaningful picture too.

For the computation of ADR, the first and foremost requirement is that the asymptotic distribution of the normalized form of an estimator exists and admits finite second order moments. This requirement is generally satisfied by the estimators when $\theta = 0$ (i.e., the null hypothesis H_0 holds). However, if $\theta \neq 0$, then by virtue of the consistency of the test based on L_n , L_n becomes large (in probability), as $n \rightarrow \infty$. Looking at (2.12) and (2.13), we gather that for any (fixed) $\theta \neq 0$, U_n^S and U_n^{PT} both asymptotically become equivalent (in probability) to U_n . Thus, neither shrinkage nor preliminary testing has any profound effect for large n when $\theta \neq 0$. The situation is different when θ is made to coverage to 0 as $n \rightarrow \infty$. Recall that n_c^0 in (2.17) is $O(c^{-\frac{1}{2}})$, as $c \downarrow 0$. As such, if we consider a sequence $\{K_c\}$ of local

alternatives:

$$K_c: \theta = c^{\frac{1}{2}}\lambda, \lambda(\text{fixed}) \text{ in } E^p; c > 0, \quad (3.2)$$

then under $\{K_c\}$, the estimators U_{N_c} , $U_{N_c}^S$ and $U_{N_c}^{PT}$ all have nondegenerate asymptotic distributions, and these may then be incorporated in a meaningful setup for the computation and comparison of the ADR of these estimates. Naturally, the null hypothesis case is included in this setup (by allowing λ to be 0). In passing, we may remark that consideration of such local alternatives is also inherent in the classical parametric shrinkage estimators (even in the non-sequential case), as otherwise there is no effective reduction in the risk due to shrinkage estimation (or PTE). For more detailed discussions of such local alternatives in shrinkage estimation, in the non-sequential case, we may refer to Sen (1984) and Sen and Saleh (1985).

For a well defined sequential estimator $\theta_{N_c}^*$ of θ , we assume that under $\{K_c\}$ in (3.2), the asymptotic (as $c \rightarrow 0$) distribution of $c^{-\frac{1}{2}}(\theta_{N_c}^* - \theta)$ exists (and is non-degenerate), and denote it by

$$G^*(x) = \lim_{c \rightarrow 0} P\{c^{-\frac{1}{2}}(\theta_{N_c}^* - \theta) \leq x \mid K_c\}, x \in E^p. \quad (3.3)$$

We also denote by

$$\Gamma^* = \int \int x x^T dG^*(x) \quad (3.4)$$

(the dispersion matrix of the r.v. X following the d.f. G^*), and let \tilde{E} stand for the expectation based on the d.f. G^* . Let then

$$\xi^* = \lim_{c \rightarrow 0} \{c^{\frac{1}{2}} \tilde{E}(N_c \mid K_c)\}. \quad (3.5)$$

We shall see later on that by virtue of (2.18), we may even take $\xi^* = \{\text{Tr}(\Omega\Gamma)\}^{\frac{1}{2}}$, where Γ is defined in (2.9). Then, corresponding to the loss function $L(\theta_{N_c}^*, \theta; c)$ in (2.14), the ADR (i.e., the risk computed from using the d.f. G^* in (3.3), instead of the actual d.f. of $\theta_{N_c}^*$) is given by

$$\rho^*(\theta_{N_c}^*, \theta; c) = c^{\frac{1}{2}} [\text{Tr}(\Omega\Gamma^*) + \xi^*]. \quad (3.6)$$

This ADR based on the asymptotic d.f. G^* of the estimator $\theta_{N_c}^*$ (as well as on

the asymptotic distribution of the stopping number N_c) provides a meaningful and adaptable means for comparing the performance characteristics of the sequential PTE and shrinkage versions of U-statistics, and this will be fully explored here.

In passing, we may remark here that for the conventional non-sequential case, treated in Sen (1984), the ADR relates to $c_n^{\frac{1}{2}} \text{Tr}(Q\Gamma^*)$ (as in (2.11), the factor c_n is not taken into account). As remarked earlier, ξ^* may have a different limit (than $\text{Tr}(Q\Gamma^*)$), and hence, the ADR results in the sequential case may not totally agree with their counterparts in the non-sequential case. Nevertheless, the asymptotic dominance picture remains very much comparable in the two cases. As such, by suitable cross reference to Sen (1984), we shall try to omit the mathematical manipulations to a certain extent, and concentrate on the motivations to a larger extent.

4. ADR OF SEQUENTIAL U-STATISTICS

The stopping number N_c , defined in (2.18), plays a vital role in the computation and interpretation of the ADR results for the various sequential versions. Towards this study, first, we note that by virtue of the basic representation of $\hat{\Gamma}_n$ in terms of a linear combination of U-statistics [c.f. Sen and Ghosh (1981)] and almost sure convergence properties of U-statistics (under the minimal assumption of finiteness of their first moment only), it readily follows that whenever the kernels ϕ_j are square integrable,

$$\hat{\Gamma}_n \rightarrow \Gamma \text{ a.s.}, \text{ as } n \rightarrow \infty. \quad (4.1)$$

In fact, if $E\|\phi\|^{2r} < \infty$ for some $r > 2$, then

$$E\|\hat{\Gamma}_n - \Gamma\|^r < K_r n^{-r^*}, \quad (4.2)$$

where $K_r (< \infty)$ does not depend on n and $r^* = r/2$ for $r \geq 2$ and $r^* = r - 1$, for $1 \leq r \leq 2$. As such, from (2.17) and (2.18), we may conclude (by using essentially the same steps as in Sen and Ghosh (1981) that

$$N_c/n_c^0 \rightarrow 1 \text{ a.s.}, \text{ as } c \downarrow 0. \quad (4.3)$$

On the other hand, for U_n , general invariance principles (relating to the weak convergence of partial sequences or tail sequences to processes of Brownian motions) have been studied by a host of workers; most of these developments has been systematically presented in Chapter 3 of Sen (1981). A direct consequence of such weak invariance principles (actually, the 'tightness' part) is that the classical (Anscombe-) condition on the uniform continuity, in probability, (with respect to $c^{\frac{1}{2}}$) holds. As such, using (4.3) and Theorem 3.3.3 of Sen (1981), we obtain that for $E\|\phi\|^2 < \infty$,

$$c^{-\frac{1}{2}} \|U_{N_c} - U_{n_c^0}\| \rightarrow 0, \text{ in probability, as } c \downarrow 0. \quad (4.4)$$

Let us next look at the shrinkage U-statistics in (2.12). Note that by (4.1), as $c \downarrow 0$,

$$d_n = \text{ch}_p(Q \hat{\Gamma}_n) \xrightarrow{\text{a.s.}} \delta = \text{ch}_p(Q \Gamma), \quad (4.5)$$

$$Q^{-1} \hat{\Gamma}_n^{-1} \xrightarrow{\text{a.s.}} Q^{-1} \Gamma^{-1} \text{ whenever } \Gamma \text{ is p.d.} \quad (4.6)$$

Further, using (4.4) and keeping in mind (3.2), we conclude that under $\{K_c\}$, as $c \downarrow 0$,

$$\begin{aligned} L_{N_c} &= N_c (U_{N_c}^T \hat{\Gamma}_{N_c}^{-1} U_{N_c}) \\ &= L_{n_c^0} + o_p(c^{\frac{1}{2}} n_c) \\ &= L_{n_c^0} + o_p(1). \end{aligned} \quad (4.7)$$

Finally, under $\{K_c\}$ in (3.2), $\|U_{n_c}\| = O_p(c^{\frac{1}{2}})$. As such, by (2.12) and (4.4) through (4.7), we obtain that under $\{K_c\}$,

$$c^{-\frac{1}{2}} \|U_{N_c}^S - U_{n_c^0}^S\| \rightarrow 0, \text{ in probability, as } c \downarrow 0. \quad (4.8)$$

Similarly, looking at (2.13), we write

$$\begin{aligned} U_{N_c}^{PT} - U_{n_c^0}^{PT} &= U_{N_c} I(L_{N_c} \geq \chi_{p,\alpha}^2) - U_{n_c^0} I(L_{n_c^0} \geq \chi_{p,\alpha}^L) + o_p(c^{\frac{1}{2}}) \\ &= [(U_{N_c} - U_{n_c^0}) I(L_{n_c^0} \geq \chi_{p,\alpha}^L)] + \end{aligned}$$

$$\begin{aligned}
& U_{N_c} [I(L_{N_c} \geq \chi_{p,\alpha}^2) - I(L_{n_c^0} \geq \chi_{p,\alpha}^2)] + o_p(c^{\frac{1}{2}}) \\
& = o_p(c^{\frac{1}{2}}),
\end{aligned} \tag{4.9}$$

where the last step follows from (4.4) and (4.7), after noting that $\|U_{N_c}\| = o_p(c^{\frac{1}{2}})$, under $\{K_c\}$. Consequently,

$$c^{-\frac{1}{2}} \|U_{N_c}^{PT} - U_{n_c^0}^{PT}\| \rightarrow 0, \text{ in probability, as } c \rightarrow 0. \tag{4.10}$$

Looking at (3.3) and (4.4), (4.8) and (4.10), we gather that for each of these sequential versions of U-statistics, for the asymptotic distribution, it suffices to consider the corresponding non-sequential versions based on the sample size n_c^0 . Further, note that by (2.17), (3.5) and (4.3), we have

$$\xi^* = [\text{Tr}(Q\Gamma)]^{\frac{1}{2}}. \tag{4.11}$$

Also, by (2.17), as $c \rightarrow 0$,

$$c^{-\frac{1}{2}} E(U_{n_c^0} - \theta)(U_{n_c^0} - \theta)^T \rightarrow [\text{Tr}(Q\Gamma)]^{\frac{1}{2}} \cdot \Gamma. \tag{4.12}$$

Consequently, we have under $\{K_c\}$,

$$\rho^*(U_{N_c}, \lambda; c) = 2[c\text{Tr}(Q\Gamma)]^{\frac{1}{2}}, \quad \forall \lambda \in E^P. \tag{4.13}$$

Similarly, using the non-sequential ADR results in Sen (1984) along with (4.10) and (4.11), we obtain that

$$\begin{aligned}
\rho^*(U_{N_c}^{PT}, \lambda; c) &= [c\text{Tr}(Q\Gamma)]^{\frac{1}{2}} \{2 - \Pi_{p+2}(\chi_{p,\alpha}^2; \Delta) \\
&\quad + \Delta^* (2\Pi_{p+2}(\chi_{p,\alpha}^2; \Delta) - \Pi_{p+4}(\chi_{p,\alpha}^2; \Delta))\},
\end{aligned} \tag{4.14}$$

where $\Pi_q(x; \delta)$ stands for the noncentral chi squared d.f. with q DF and non-centrality parameter δ (so that $\Pi_p(\chi_{p,\alpha}^2; 0) = 1 - \alpha$) and, defining λ as in (3.2),

$$\Delta = \lambda^T \Gamma^{-1} \lambda \text{ and } \Delta^* = \lambda^T Q \lambda. \tag{4.15}$$

In this context, it may be noted that by the Courant Theorem,

$$\Delta^*/\Delta \leq \text{ch}_1(Q\Gamma) = [\text{ch}_p(Q^{-1}\Gamma^{-1})]^{-1}, \quad \forall \lambda \in E^P. \tag{4.16}$$

Finally, by (4.8), (4.11) and the non-sequential ADR results in Sen (1984), we have

$$\begin{aligned} \rho^*(U_{N_c}^S, \lambda; c) &= [c\text{Tr}(Q\Gamma)]^{\frac{1}{2}} \{2 - 2s(p-2) [\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)] E(\chi_p^{-2}(\Delta)) \\ &\quad + s^2 [\text{ch}_p(Q\Gamma)]^2 [\text{Tr}(Q\Gamma)]^{-1} [\text{Tr}(Q^{-1}\Gamma^{-1}) E(\chi_{p+2}^{-4}(\Delta)) \\ &\quad + (\lambda^T \Gamma^{-1} Q^{-1} \Gamma^{-1} \lambda) E(\chi_{p+4}^{-4}(\Delta))] \} \end{aligned} \quad (4.17)$$

where the shrinkage factor $s(0 < s < 2(p-2))$ is defined in (2.12),

$$E(\chi_q^{-2r}(\Delta)) = \int_0^\infty x^{-r} d\Pi_q(x; \Delta), \quad (4.18)$$

and we denote by

$$\Delta^0 = \lambda^T \Gamma^{-1} Q^{-1} \Gamma^{-1} \lambda. \quad (4.19)$$

Note that by the Courant Theorem, here also, we have

$$\Delta^0/\Delta \leq \text{ch}_1(\Gamma^{-1} Q^{-1}) = [\text{ch}_p(\Gamma Q)]^{-1}, \quad \forall \lambda \in E^p. \quad (4.20)$$

In the special case of $Q = \Gamma^{-1}$ (4.13) reduces to $2(cp)^{\frac{1}{2}}$, while (4.17) reduces to $(cp)^{\frac{1}{2}} \{2 - 2(s/p) E(\chi_p^{-2}(\Delta)) + s^2 [E(\chi_{p+2}^{-4}(\Delta)) + p^{-1} \Delta E(\chi_{p+4}^{-4}(\Delta))] \}$. Also, for the special choice of $s = p-2$, the last expression reduces to $(cp)^{\frac{1}{2}} \{2 - 2(1-2/p) E(\chi_p^{-2}(\Delta)) + (p-2)^2 [E(\chi_{p+2}^{-4}(\Delta)) + p^{-1} \Delta E(\chi_{p+4}^{-4}(\Delta))] \}$. We shall incorporate these, in the next section, in the study of the related asymptotic dominance results.

5. ASYMPTOTIC DOMINANCE AND RISK-EFFICIENCY RESULTS

We study here the asymptotic risk-efficiency (ARE) result in the light of the ADR results, studied in Section 4. First, comparing (4.13) and (4.14), we obtain that

$$\begin{aligned} \rho^*(U_{N_c}^{PT}, \lambda; c) / \rho^*(U_{N_c}, \lambda; c) \\ = 1 - \frac{1}{2} \{ \Pi_{p+2}(\chi_{p,\alpha}^2; \Delta) - \Delta^* \{ 2\Pi_{p+2}(\chi_{p,\alpha}^2; \Delta) - \Pi_{p+4}(\chi_{p,\alpha}^2; \Delta) \} \}. \end{aligned} \quad (5.1)$$

Under $H_0: \lambda = 0$, i.e., $\Delta = 0$, the right hand side of (5.1) reduces to

$$1 - \frac{1}{2} \Pi_{p+2}(\chi_{p,\alpha}^2; 0) \in \left(\frac{1+\alpha}{2}, 1 \right). \quad (5.2)$$

Actually, for λ in a closed neighborhood of 0, (5.1) is less than 1, while the opposite inequality holds when $\Delta^* \geq \Delta_0$, where $\Delta_0 < 1$. Thus, in the light of their ADR, none of the sequential PTE U-statistic and U_{N_c} fails to dominate the other. On the other hand, as λ moves away from 0, i.e., Δ and $\Delta^* \rightarrow +\infty$, the right hand side of (5.1) converges to 1, so that for significant departures from the pivot, the PTE and classical U_{N_c} (in the sequential case) become risk-equivalent, as $c \rightarrow 0$. Note that

$$\begin{aligned}
 1 - \frac{1}{2} \Pi_{p+2}(\chi_{p,\alpha}^2; 0) &= \inf_{\lambda} \rho^*(U_{N_c}^{PT}, \lambda; c) / \rho^*(U_{N_c}, \lambda; c) \\
 &< \sup_{\lambda} \rho^*(U_{N_c}^{PT}, \lambda; c) / \rho^*(U_{N_c}, \lambda; c) \\
 &= 1 + \sup_{\lambda} \frac{1}{2} \{ \Delta^* \{ 2 \Pi_{p+2}(\chi_{p,\alpha}^2; \Delta) - \Pi_{p+4}(\chi_{p,\alpha}^2; \Delta) \} \\
 &\quad - \Pi_{p+2}(\chi_{p,\alpha}^2; \Delta) \}
 \end{aligned} \tag{5.3}$$

where, in the right hand side of (5.3), the excess over 1 is generally very moderate while, for small values of α ($0 < \alpha < 1$) and not so small values of p (> 1), the lower bound is close to $\frac{1}{2}$. Thus, the sequential PTE U-statistic may have a natural appeal when λ is close to 0, i.e., one has high confidence in the tenacity of H_0 .

Comparing (4.13) and (4.17), we have

$$\begin{aligned}
 &\rho^*(U_{N_c}^S, \lambda; c) / \rho^*(U_{N_c}, \lambda; c) \\
 &= 1 - (p-2) s [ch_p(Q\Gamma) / \text{Tr}(Q\Gamma)] E(\chi_p^{-2}(\Delta)) + \\
 &\quad \frac{1}{2} s^2 [ch_p(Q\Gamma)]^2 [\text{Tr}(Q\Gamma)]^{-1} \{ \text{Tr}(Q^{-1}\Gamma^{-1}) E(\chi_{p+2}^{-4}(\Delta)) + \Delta^0 E(\chi_{p+4}^{-4}(\Delta)) \}.
 \end{aligned} \tag{5.4}$$

Note that the last term on the right hand side of (5.4) is bounded from above by

$$\begin{aligned}
 &\frac{1}{2} s^2 [ch_p(Q\Gamma)]^2 [\text{Tr}(Q\Gamma)]^{-1} p ch_1(Q^{-1}\Gamma^{-1}) E(\chi_{p+2}^{-4}(\Delta)) + \\
 &\quad ch_1(Q^{-1}\Gamma^{-1}) \Delta E(\chi_{p+4}^{-4}(\Delta)) \\
 &= \frac{1}{2} s^2 [ch_p(Q\Gamma)] [\text{Tr}(Q\Gamma)]^{-1} \{ p E(\chi_{p+2}^{-4}(\Delta)) + \Delta E(\chi_{p+4}^{-4}(\Delta)) \}
 \end{aligned}$$

$$= \frac{1}{2}s^2 [\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)] E(\chi_p^{-2}(\Delta)), \quad (5.5)$$

as $\text{Tr}(Q^{-1}\Gamma^{-1}) \leq p \text{ch}_1(Q^{-1}\Gamma^{-1})$, $\text{ch}_1(Q^{-1}\Gamma^{-1}) = [\text{ch}_p(Q\Gamma)]^{-1}$, $\Delta^0 \leq \text{ch}_1(Q^{-1}\Gamma^{-1})\Delta$

and $pE(\chi_{p+2}^{-4}(\Delta)) + \Delta E(\chi_{p+4}^{-4}(\Delta)) = E(\chi_p^{-2}(\Delta))$. Thus (5.4) is bounded from above by

$$\begin{aligned} & 1 - (p-2)s[\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)]E(\chi_p^{-2}(\Delta)) + \\ & \quad \frac{1}{2}s^2[\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)]E(\chi_p^{-2}(\Delta)) \\ & = 1 - [\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)]E(\chi_p^{-2}(\Delta))s\{(p-2) - \frac{1}{2}s\}, \end{aligned} \quad (5.6)$$

and the right hand side of (5.6) is less than one for every s : $0 < s < 2(p-2)$.

An optional choice of s in this content is $s = p-2$, and in this case, (5.6)

reduces to

$$1 - \frac{1}{2}[\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)](p-2)^2 E(\chi_p^{-2}(\Delta)). \quad (5.7)$$

Note that for the most natural choice of Q (i.e., $Q = \Gamma^{-1}$), (5.4) is equal to (5.6) with the further simplification that $\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma) = p^{-1}$. Also, note that $E(\chi_p^{-2}(\Delta)) \leq \frac{1}{p-2}$, so that under $H_0: \lambda = 0$, (5.7) reduces to $1 - \frac{1}{2p}(p-2) = \frac{1}{2} + \frac{1}{p} < 1$, for every $p > 2$. For general Q , $\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma) \leq \frac{1}{p}$, so that the amount of shrinkage in (5.7) may be somewhat smaller, depending on $Q\Gamma$. In any case, (5.7) is strictly less than 1, for all $p > 2$, Q such that $Q\Gamma$ is $p \cdot d \cdot$ and finite Δ . As $\Delta \rightarrow \infty$, $E(\chi_p^{-2}(\Delta)) \rightarrow 0$, so that (5.7) converges to 1. This clearly shows that in the light of the ADR, the shrinkage U-statistic dominate the classical U-statistic, in the sequential case. This provides a strong justification for the use of the sequential shrinkage U-statistics, when $p > 2$, and, also to choose, in (2.12), $s = (p-2)$.

Finally, comparing (4.14) and (4.17), we obtain that

$$\begin{aligned} & \rho^*(U_{N_c}^S, \lambda; c) / \rho^*(U_{N_c}^{PT}, \lambda; c) \\ & = (1 - \frac{1}{2}[\Pi_{p+2}(\chi_{p,\alpha}^2; \Delta) - \Delta\{2\Pi_{p+2}(\chi_{p,\alpha}^2; \Delta) - \Pi_{p+4}(\chi_{p,\alpha}^2; \Delta)\}])^{-1} \end{aligned}$$

$$(1 - (p-2)s[\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)]E(\chi_p^{-2}(\Delta)) + \frac{1}{2}s^2[\text{ch}_p(Q\Gamma)]^2 \cdot [\text{Tr}(Q\Gamma)]^{-1}\{\text{Tr}(Q^{-1}\Gamma^{-1})E(\chi_{p+2}^{-4}(\Delta)) + \Delta^0 E(\chi_{p+4}^{-4}(\Delta))\}) \quad (5.8)$$

Note that when $\lambda = 0$, (5.8) reduces to

$$(1 - \frac{1}{2}\Pi_{p+2}(\chi_{p,\alpha}^2;0))^{-1}\{1 - (p-2)s[\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)]E(\chi_p^{-2}(\Delta)) + \frac{1}{2}s^2[\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)]\text{ch}_p(Q\Gamma)\text{Tr}(Q^{-1}\Gamma^{-1})E(\chi_{p+2}^{-4}(\Delta))\} \\ = (1 - \frac{1}{2}\Pi_{p+2}(\chi_{p,\alpha}^2;0))^{-1}\{1 - [\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)]s + \frac{1}{2}s^2[\text{ch}_p(Q\Gamma)/\text{Tr}(Q\Gamma)]\text{ch}_p(Q\Gamma)\text{Tr}(Q^{-1}\Gamma^{-1})(p(p-2))^{-1}\} \quad (5.9)$$

For the optimal choice of $s(=p-2)$ and for $Q = \Gamma^{-1}$, (5.9) reduces to

$$(1 - \frac{1}{2}\Pi_{p+2}(\chi_{p,\alpha}^2;0))^{-1}\{1 - \frac{p-2}{p} + \frac{1}{2}\frac{p-2}{p}\} \\ = (\frac{1}{2} + \frac{1}{p})/(\frac{1}{2} + \frac{1}{2}[1 - \Pi_{p+2}(\chi_{p,\alpha}^2;0)]) \quad (5.10)$$

Now (5.10) is greater than 1 when

$$\Pi_{p+2}(\chi_{p,\alpha}^2;0) > (p-2)/p = 1 - 2/p. \quad (5.11)$$

Generally, for small values of α , (5.11) holds. So that under H_0 , the PTE $U_{N_c}^{PT}$ has a smaller ADR than the shrinkage $U_{N_c}^S$. The same picture holds for other choices of $s(\in(0, 2(p-2)))$ and Q . Thus, the sequential shrinkage $U_{N_c}^S$ may not dominate the PTE $U_{N_c}^{PT}$, particularly when α is small. Actually, in a PTE, generally, α is chosen to be small, and hence, (5.11) holds. On the other hand, as λ moves away from 0, the ADR of the PTE $U_{N_c}^{PT}$ increases, attains a maximum value (greater than $2[\text{cTr}(Q\Gamma)]^{\frac{1}{2}}$) at some intermediate value of Δ , and then tapers off to $2[\text{cTr}(Q\Gamma)]^{\frac{1}{2}}$ as $\Delta \rightarrow \infty$ (although continuing to stay above this asymptote, for all finite $\Delta \geq \Delta_0$). On the other hand, by (5.7), the ADR of the sequential shrinkage $U_{N_c}^S$ monotonically increases (as Δ increases) and is always $\leq 2[\text{cTr}(Q\Gamma)]^{\frac{1}{2}}$, where the upper bound is attained for $\Delta \rightarrow \infty$. Therefore, there exists a closed region \mathcal{E} , with center 0, such that

for $\lambda \in \mathcal{E}$, (5.8) is greater than one, while for $\lambda \notin \mathcal{E}$, (5.8) is less than one. Further, as $\Delta \rightarrow \infty$, (5.8) converges to 1. Thus, in the light of the ADR, the PTE and shrinkage U-statistics, in the sequential case, compare quite favorably to each other (none dominates the other and both fare well relative to U_{N_c}). The shrinkage U-statistic dominates U_{N_c} and may also have the asymptotic minimax character (in the light of the ADR), which is not shared by the PTE $U_{N_c}^{PT}$. Hence, there may be some point in favor of the shrinkage $U_{N_c}^S$. However, for the shrinkage estimator, we need $p \geq 3$, while for the PTE, $p \geq 1$ suffices. Hence, if p is not so large, the PTE may have some distinct advantages over the shrinkage estimator; this is certainly the case for $p = 1$ or 2 .

6. SOME GENERAL REMARKS

In earlier sections, we have mainly considered the case of U-statistics. In (2.12), (2.13) as well as in L_n , we may virtually replace U_n by V_n and define the shrinkage and PTE versions of von Mises' functionals. By virtue of (2.7), (2.17) and (4.3), it follows that as $c \rightarrow 0$,

$$\|U_{N_c} - V_{N_c}\| = o(c^{\frac{1}{2}}) \text{ a.s.} \quad (6.1)$$

As such, (4.8) and (4.10) hold for these von Mises' functionals too. Hence, the theory developed in Sections 3 through 5 applies to the parallel (sequential) versions of von Mises' functionals too.

The ADR results studied in Sections 4 and 5 are quite similar to the non-sequential case, treated in Sen (1984). However, the ARE results are closer to one in the sequential case than in the nonsequential one. This is primarily because of the second term on the right hand side of (3.6), which depends only on N_c , and hence, remains the same for all the estimators; the first term resembles the non-sequential case.

We may comment on the appropriateness of $\{K_c\}$ in (3.2) in the sequential shrinkage estimation problem. Such local alternatives define an

effective neighborhood of the pivot, where shrinkage is effective. Beyond this domain, the PTE and shrinkage (sequential) U-statistics are asymptotically risk-equivalent to the classical U-statistics, and there is not much interest in the asymptotic theory. The asymptotic theory works out well in practice for "small c"; the interpretation of "small c" remains the same as in the classical sequential estimation problem [viz., Sen and Ghosh (1981)].

Note that all the sequential versions of the U-statistics are based on the same stopping variable (N_c) which is defined in (2.18) and is motivated by the classical sequential point estimation problem [viz. Sen and Ghosh (1981)], where the Stein-rule philosophy has not been incorporated. It is possible to consider some other stopping variables (say, N_c^*) where the Stein-rule philosophy may also be incorporated in this formulation. The asymptotic theory presented in earlier sections remains intact so long as there exists another sequence of positive integers $\{n_c^*\}$, such that $n_c^* \rightarrow \infty$ as $c \downarrow 0$ and $N_c^*/n_c^* \rightarrow 1$, in probability, as $c \downarrow 0$. There, we need to make only minor adjustments for this change from n_c^0 to n_c^* in the formulae for the ADR.

The choice of the matrix Q is of some interest too. Use of Mahalanobis distance for the loss function is often recommended on natural grounds. Based on this consideration, $Q = \Gamma^{-1}$ is a natural choice. While this choice does not affect the formulation of the PTE and classical U-statistics, in (2.12), Q plays a basic role. The advantage in the sequential estimation case is that we may as well choose $Q = \hat{\Gamma}^{-1}$, so that (2.12) simplifies to

$$U_n^S = (1 - s(nU_n^T \hat{\Gamma}_n^{-1} U_n)^{-1}) U_n. \quad (6.2)$$

In this, however, (2.18) leads to

$$N_c = \min\{n \geq 2: n \geq c^{-\frac{1}{2}}(p+n^{-a})^{\frac{1}{2}}\}, \quad (6.3)$$

which renders N_c as non-stochastic, so that we would not have a genuine sequential estimation problem. Nevertheless, the results in the present paper

continue to hold.

ACKNOWLEDGEMENT

This work was supported by the Office of Naval Research,
Contract No. N00014-83-K-0387.

R E F E R E N C E S

- BERGER, J.O. (1985). Statistical Decision Theory, 2nd Ed. New York: Springer-Verlag.
- GHOSH, M., NICKERSON, D.M. AND SEN, P.K. (1986). Sequential shrinkage estimation. Ann. Statist.
- GHOSH, M. AND SEN, P.K. (1983). On two-stage James-Stein estimators. Sequen. Anal. 2, 359-367.
- GHOSH, M., SINHA, B.K. AND MUKHOPADHYAY, N. (1976). Multivariate sequential point estimation. J. Multivar. Anal. 6, 281-294.
- HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19, 293-325.
- SEN, P.K. (1977). Some invariance principles relating to jackknifing and their role in sequential analysis. Ann. Statist. 5, 315-329.
- SEN, P.K. (1981). Sequential Nonparametrics. New York: Wiley.
- SEN, P.K. (1984). A James-Stein detour of U-Statistics. Commun. Statist. Theor. Meth. A13, 2725-2747.
- SEN, P.K. (1986a). On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators. Sankhya, Ser. A 48.
- SEN, P.K. (1986b). Sequential Stein-rule maximum likelihood estimation: General asymptotics. Inst. Statist, Univ. N. Carolina, Mim. Rept. No. 1802.

- SEN, P.K. AND GHOSH, M. (1981). Sequential point estimation of estimable parameters based on U-statistics. Sankhya, Ser. A 43, 331-344.
- SEN, P.K. AND SALEH, A.K.Md.E. (1985). On some shrinkage estimators of multivariate location. Ann. Statist. 13, 272-281.
- TAKADA, Y. (1984). Inadmissibility of a sequential estimation rule of the mean of a multivariate normal distribution. Sequen. Anal. 3, 267-271.
- VON MISES, R. (1947). On the asymptotic distribution of differentiable statistical functions. Ann. Math. Statist. 18, 309-348.