ASYMPTOTIC THEORY OF SEQUENTIAL SHRUNKEN ESTIMATION OF STATISTICAL FUNCTIONALS

PRANAB KUMAR SEN

#1766

1988

ASYMPTOTIC THEORY OF SEQUENTIAL SHRUNKEN ESTIMATION OF STATISTICAL FUNCTIONALS

Pranab Kumar Sen

Departments of Biostatistics and Statistics
University of North Carolina
Chapel Hill, NC 27599 USA

Based on the sample (empirical) distribution function F_n, let $\mathbf{T}_n = (T_1(F_n),\ldots,T_p(F_n))'$ be an estimator of a smooth statistical functional $\boldsymbol{\tau} = (T_1(F),\ldots,T_p(F))'$. Incorporating cost of sampling and a quadratic error loss, the risk of $\mathbf{T}_n$ is taken as $E_F\|\mathbf{T}_n-\boldsymbol{\tau}\|^2_{\mathbf{W}} + cn$, $c > 0$, where $\mathbf{W}$ is a given positive definite matrix. Employing a jackknifed estimator of $E_F\|\mathbf{T}_n-\boldsymbol{\tau}\|^2_{\mathbf{W}}$, a stopping number N_c can be so formulated that $\mathbf{T}_{N_c}$ has asymptotically (as $c \downarrow 0$) the minimum risk. For $p \geq 3$, $\mathbf{T}_n$ may be dominated by a Stein-rule version $\mathbf{T}^S_n$, an asymptotic treatment of this dominance is presented. The theory is extended to the sequential case of $\mathbf{T}^S_{N_c}$ and some plausible forms of stopping numbers are discussed in this context.

INTRODUCTION

Let $\{\mathbf{X}_i;\ i \geq 1\}$ be a sequence of independent and identically distributed random vectors (i.i.d.r.v.) with a distribution function (d.f.) F, defined on E^q, for some $q \geq 1$. Consider a transformation: $\mathbf{X} \rightarrow \mathbf{Y} = (Y^{(1)},\ldots,Y^{(p)})'$ where $Y^{(j)}$ has a marginal d.f. G_j, $1 \leq j \leq p$, and define a smooth statistical functional $\boldsymbol{\tau} = (\tau_1(G_1),\ldots,\tau_p(G_p))'$, $p \geq 1$, so that

$$T_j(F) = \tau_j(G_j),\ j = 1,\ldots,p. \tag{1}$$

We are primarily interested in an (asymptotically) optimal estimation of $\boldsymbol{\tau}$. Based on a sample $(\mathbf{X}_1,\ldots,\mathbf{X}_n)$ of size $n(\geq 1)$, let

$$F_n(\mathbf{x}) = n^{-1} \Sigma_{i=1}^n I(\mathbf{X}_i \leq \mathbf{x}),\ \mathbf{x} \in E^q, \tag{2}$$

be the sample (empirical) d.f., where I(A) stands for the indicator function of the set A; the empirical d.f.'s $G_{n1},\ldots,G_{np}$ are defined in a similar fashion [replacing the $\mathbf{X}_i$ by $Y_i^{(j)}$, $1 \leq j \leq p$]. Then, granted some mild regularity conditions, a natural estimator of $\underset{\sim}{\tau}$ is $\underset{\sim}{T}_n = (T_{n1},\dots,T_{np})'$, where

$$T_{nj} = \tau_j(G_{nj}), \quad \text{for} \quad j = 1,\dots,p. \tag{3}$$

Consider a quadratic error loss incurred due to estimating $\underset{\sim}{\tau}$ by $\underset{\sim}{T}_n$ and a cost $c(>0)$ per unit sampling, so that the risk in estimating $\underset{\sim}{\tau}$ by $\underset{\sim}{T}_n$ is given by

$$\rho_c(\underset{\sim}{T}_n, \underset{\sim}{\tau}) = cn + E_F \|\underset{\sim}{T}_n - \underset{\sim}{\tau}\|^2_{\underset{\sim}{W}} = cn + E_F\{(\underset{\sim}{T}_n - \underset{\sim}{\tau})'\underset{\sim}{W}(\underset{\sim}{T}_n - \underset{\sim}{\tau})\}, \tag{4}$$

where $\underset{\sim}{W}$ is a given positive definite (p.d.) matrix. Note that (4) depends on c, n, $\underset{\sim}{W}$ as well as on F through the (joint) distribution of $\underset{\sim}{T}_n$. It may not be unreasonable to assume that $E_F\|\underset{\sim}{T}_n - \underset{\sim}{\tau}\|^2_{\underset{\sim}{W}}$ is $\downarrow$ in n, although its precise form may depend on F. Thus, (4) is the sum of two nonnegative terms, one $\uparrow$ in n and the other one is $\downarrow$ in n. Hence, we may assume that there exists a positive integer n_c^0 $(= n^0(c,F,\underset{\sim}{W}))$, such that

$$\rho_c^0(F) = \rho_c(\underset{\sim}{T}_{n_c^0}, \underset{\sim}{\tau}) = \min_{m \geq n_0} \rho_c(\underset{\sim}{T}_m, \underset{\sim}{\tau}), \tag{5}$$

where n_0 is the minimum sample size for which ρ_c in (4) exists. Thus, $\rho_c^0(F)$ is the minimum risk and $\underset{\sim}{T}_{n_c^0}$ is the minimum risk estimator (MRE) of $\underset{\sim}{\tau}$. Note that n_c^0, as defined by (5), may generally depend on the unknown F, so that for any chosen n, $\underset{\sim}{T}_n$ may not be MRE when F is allowed to vary within a class $\mathcal{F}$. For this reason sequential estimation rules are generally advocated to achieve the MRE property (simultaneously for all $F \in \mathcal{F}$), at least, in an asymptotic set up (where $c \downarrow 0$). A systematic account of this sequential estimation problem (in a nonparametric set up) is given in Sen (1981, Ch. 10). Before extending this theory to the current problem, we present some related developments which have far reaching effects in this MRE problem. For the multivariate normal mean vector estmation problem, for $p \geq 3$, the inadmissibility of the classical maximum likelihood estimator (MLE) has been established by Stein (1956). Later on, James and Stein (1962) constructed an alternative (Stein-rule or shrinkage) estimator which dominates the MLE (in quadratic error loss). In the sequential case, parallel results are due to Ghosh, Nickerson and Sen (1987) and others. In the general nonparametric case, an exact treatment of this Stein-phenomenon becomes difficult. Nevertheless, the asymptotic theory follows the same track too: See Sen (1984) and Sen and Saleh (1985, 1987), among others. Motivated by this, we are naturally tempted in considering an appropriate Stein-rule version ($\underset{\sim}{T}_n^S$) of $\underset{\sim}{T}_n$ and in establishing the asymptotic dominance (of $\underset{\sim}{T}_n^S$ over $\underset{\sim}{T}_n$) in some well defined manner. Our main objective is to formulate along with $\underset{\sim}{T}_n^S$ its sequential version $\underset{\sim}{T}_{N_c}^S$, where N_c is an appropriate stopping number, and to establish the asymptotic MRE property of $\underset{\sim}{T}_{N_c}^S$ in a systematic manner. These versions of $\underset{\sim}{T}_n$ are introduced in the next section. The main results are then presented in the third section, and their derivations in the following one. Some general discussions are made in the concluding section.

SHRUNKEN ESTIMATION OF $\underset{\sim}{\tau}$

We assume that for each $j(=1,\ldots,p)$, $\tau_j(\cdot)$ is Hadamard-continuous at G_j, so that

$$|\tau_j(H) - \tau_j(G_j)| \to 0 \quad \text{with} \quad \|H-G_j\| \to 0 \qquad \text{on} \quad H \in \mathscr{A}, \tag{6}$$

where $\mathscr{A}$ is a topological vector space. Actually, we assume more: For each j, $\tau_j(\cdot)$ is Hadamard (or compact) differentiable at G_j, so that

$$\tau_j(H) = \tau_j(G_j) + \int \tau_{j1}(G_j;x)d[H(x) - G_j(x)] + R_j(G_j;H-G_j), \tag{7}$$

where

$$|R(G_j;H-G_j)| = o(\|H-G_j\|), \text{ uniformly in } H \in \mathscr{K}, \quad 1 \le j \le p, \tag{8}$$

and $\mathscr{K} \in \mathscr{C}$, a class of compact subsets of $\mathscr{A}$, $\|G-F\|$ refers to the usual sup-norm and $\tau_{j1}(G_j;\cdot)$ is the compact derivative (or influence function) of $\tau_j(\cdot)$ at G_j, which can be so normalized that

$$\int \tau_{j1}(G_j;x)\, dG_j(x) = 0, \quad 1 \le j \le p. \tag{9}$$

For the shrunken estimators (to be considered here), we need to estimate $E_F(\underset{\sim}{T}_n-\underset{\sim}{\tau})(\underset{\sim}{T}_n-\underset{\sim}{\tau})'$. For this purpose, we shall employ the classical jackknifing method which rests on the construction of pseudovariables, and we formulate them first. For each $j(=1,\ldots,p)$, we denote by

$$G^{(i)}_{(n-1)j}(x) = (n-1)^{-1} \sum_{r=1(\neq i)}^{n} I(Y_r^{(j)} \le x), \quad i=1,\ldots,n, \quad x \in E; \tag{10}$$

$$T^{(i)}_{(n-1)j} = \tau_j(G^{(i)}_{(n-1)j})$$

and

$$T_{nj,i} = nT_{nj} - (n-1)T^{(i)}_{(n-1)j}, \quad 1 \le i \le n, \ 1 \le j \le p. \tag{11}$$

Then the $T_{nj,i}$ are the pseudovaribles and the classical jackknifed estimator $\underset{\sim}{T}^*_n = (T^*_{n1},\dots,T^*_{np})'$ is given by

$$T^*_{nj} = n^{-1} \Sigma^n_{i=1} T_{nj,i}, \quad \text{for} \quad j=1,\dots,p. \tag{12}$$

Let $\underset{\sim}{T}_{n,i} = (T_{n1,i},\dots,T_{np,i})'$, $1 \le i \le n$, and let

$$\underset{\sim}{V}^*_n = (n-1)^{-1} \Sigma^n_{i=1} (\underset{\sim}{T}_{n,i} - \underset{\sim}{T}^*_n)(\underset{\sim}{T}_{n,i} - \underset{\sim}{T}^*_n)'. \tag{13}$$

Also, let $\underset{\sim}{\tau}_1(\underset{\sim}{G},\underset{\sim}{Y}) = (\tau_{11}(G_1;Y^{(1)}),\dots,\tau_{p1}(\underset{\sim}{G}_p;\underset{\sim}{Y}^{(p)}))'$ and

$$\underset{\sim}{\Sigma}^* = E\{[\underset{\sim}{\tau}_1(\underset{\sim}{G};\underset{\sim}{Y})][\underset{\sim}{\tau}_1(\underset{\sim}{G};\underset{\sim}{Y})]'\}. \tag{14}$$

We assume that $\underset{\sim}{\Sigma}^*$ is p.d. and has finite elements. Further, we assume that $\underset{\sim}{T}^{**}_1(\underset{\sim}{H}) = ((\int\int \tau_{j1}(H_j;x)\tau_{\ell 1}(H_\ell;y)dH_{j\ell}(x,y))$ is Hadamard-continuous at $\underset{\sim}{G}$. Then proceeding as in Sen (1988), it follows that

$$\underset{\sim}{V}^*_n \to \underset{\sim}{\Sigma}^* \text{ almost surely (a.s.), as } n \to \infty. \tag{15}$$

The Stein-rule estimators rest on the choice of a suitable <u>pivot</u> $\underset{\sim}{\tau}^0$ (known) whose plausibility plays a vital role in their effective dominance; without any loss of generality, we may let $\underset{\sim}{\tau}^0 = \underset{\sim}{0}$. Let then

$$d_n = \text{smallest characteristic root of } \underset{\approx}{W}\underset{\sim}{V}^*_n; \tag{16}$$

$$\mathcal{L}_n = n\|\underset{\sim}{T}_n - \underset{\sim}{0}\|^2_{\underset{\sim}{V}^{*-1}_n}, \tag{17}$$

and let k : 0 < k < 2(p-2), $p \ge 3$) be a shrinkage factor. Then we may proceed as in Sen (1986) and consider the following shrinkage version of $\underset{\sim}{T}_n$:

$$\underset{\sim}{T}^S_n = (I - kd_n \mathcal{L}^{-1}_n \underset{\sim}{W}^{-1} \underset{\sim}{V}^{*-1})\underset{\sim}{T}_n. \tag{18}$$

[If WV^*_n is not of full rank, we may use a generalized inverse of $\underset{\approx}{W}\underset{\sim}{V}^*_n$.] Similarly, in (18), replacing $\underset{\sim}{T}_n$ by its jackknifed version ($\underset{\sim}{T}^*_n$), we may consider a Stein-rule estimator $\underset{\sim}{T}^{*S}_n$. Our first objective is to study the asymptotic dominance of $\underset{\sim}{T}^S_n$ (or $\underset{\sim}{T}^{*S}_n$) over $\underset{\sim}{T}_n$ (or $\underset{\sim}{T}^*_n$). But given the picture in (5), we intend to consider their sequential versions. For this purpose, first, we need to introduce suitable stopping numbers. In this context, we may note that [viz., Sen (1988)] under the assumed regularity conditions,

$$n^{1/2}(\mathbf{T}_n - \boldsymbol{\tau}) \underset{\mathcal{D}}{\rightarrow} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}^*), \quad \text{as} \quad n \to \infty, \tag{19}$$

(and the "convergence in law" may as well be replaced by "convergence in second mean"), so that (at least) for large n, (4) looks like

$$\rho_c(\mathbf{T}_n, \boldsymbol{\tau}) = cn + n^{-1} \text{ trace}(\mathbf{W}\boldsymbol{\Sigma}^*) + o(n^{-1}) \tag{20}$$

Thus, if $tr(\mathbf{W}\boldsymbol{\Sigma}^*)$ were known and $c(>0)$ is small, then the optimal sample size n_c^0 in (5) [for $\{\mathbf{T}_n\}$] is given by

$$n_c^0 \sim \{c^{-1} \text{ tr}\{\mathbf{W}\boldsymbol{\Sigma}^*)\}^{1/2} \quad (\text{as } c \downarrow 0), \tag{21}$$

and as a result,

$$\rho_c^0(F) \sim 2\{c \text{ tr}(\mathbf{W}\boldsymbol{\Sigma}^*)\}^{1/2}, \quad \text{as} \quad c \downarrow 0. \tag{22}$$

Keeping (15) in mind, we formulate a stopping number N_c by letting

$$N_c = \inf\{n \geq n_0 : n^2 \geq \{tr(\mathbf{W}\mathbf{V}_n^*) + n^{-c}\}/c\}, \quad c > 0 \tag{23}$$

where $n_0(\geq p)$ is a positive integer and a (> 0) is a suitable number; the factor n^{-a} eliminates a too early stopping (for small c) and thereby helps us in the manipulations with the asymptotic theory of N_c or $\mathbf{T}_{N_c}$. The stopping number N_c may also be defined in some alternative ways, and we will briefly discuss this at the end. Combining (18) and (23), we consider the following sequential shrunken version of the estimator $\mathbf{T}_n$

$$\mathbf{T}_{N_c}^S = (\mathbf{I} - k\, d_{N_c} \mathcal{L}_{N_c}^{-1} \mathbf{W}^{-1} \mathbf{V}_{N_c}^{*-1})\mathbf{T}_{N_c}, \quad k > 0. \tag{24}$$

Using the same risk function as in (4), but adapted to the sequential case, we may define now

$$\rho_c^*(F) = CEN_c + E_F(\|\mathbf{T}_{N_c} - \boldsymbol{\tau}\|_{\mathbf{W}}^2), \tag{25}$$

$$\rho_c^{*S}(F) = CEN_c + E_F(\|\mathbf{T}_{N_c}^S - \boldsymbol{\tau}\|_{\mathbf{W}}^2). \tag{26}$$

Then, our main interest lies in the comparative shedy of $\rho_c^0(F)$ in (5) and $\rho_c^*(F)$ and $\rho_c^{*S}(F)$, when $c \downarrow 0$. In this context, we shall find it convenient to make use of the concept of asymptotic distributional risk (ADR). We may recall that the Stein phenomenon is essentially a local one only in a neighborhood of the pivot. A Stein-rule estimator usually dominates its classical counterpart, and this dominance becomes imperceptible as the true parameter point $\underset{\sim}{\tau}$ moves away from the pivot (here $\underset{\sim}{0}$). Also, note that by (21), $n_c^0(\sim O(c^{-1/2})) \to \infty$ as $c \downarrow 0$, and hence, the effective domain of this dominance of the Stein-rule version shrinks to the pivot as $c \downarrow 0$. The situation is comparable to the case of (sequential) Stein-rule MLE's treated in detail in Sen (1987a). As such, we consider the following sequence $\{K_c\}$ of (local) alternatives:

$$k_c : \underset{\sim}{\tau} = \underset{\sim}{\tau}(F) = c^{1/4} \underset{\sim}{\gamma}, \ \underset{\sim}{\gamma} \in E^p \text{ (fixed)}, \ c \downarrow 0. \tag{27}$$

In (4), (5), (25) and (26), F is no longer treated as a (fixed) d.f., rather, F is replaced by a sequence $\{F_c\}$, where F_c satisfies K_c, so that the desired expectations are all computed under $\{K_c\}$ in (27). As we shall see later on, under $\{K_c\}$ (containing $H_0 : \underset{\sim}{\tau} = \underset{\sim}{0}$ as a particular case), $c^{-1/4}(\underset{\sim}{T}_{N_c} - \underset{\sim}{\tau})$ (or $c^{-1/4}(\underset{\sim}{T}^S_{N_c} - \underset{\sim}{\tau})$) has a limiting (as $c \downarrow 0$) distribution which may be incorporated in the evaluation of the (asymptotic) expectations in (4), (25) and (26). There are two main advantages of this adaptation: (i) Use of such an asymptotic distribution of the estmator and the stopping number leads to considerably simpler asymptotic expressions, and (ii) this approach may generally require less stringent regularity conditions than in the usual case. We shall elaborate these two points in the current context in the concluding section. However, we shall adapt this "asymptotic distributional risk" (ADR) approach and present our main results in the next section. For a suitable sequential estimation $\underset{\sim}{U}_{N_c}$ of $\underset{\sim}{\tau}$, we denote by

$$F^{0*}(\underset{\sim}{x}) = \lim_{c \downarrow 0} P\{c^{-1/4}(\underset{\sim}{U}_{N_c} - \underset{\sim}{\tau}) \le \underset{\sim}{x} \mid K_c\}, \ \underset{\sim}{x} \in E^p, \tag{28}$$

and assume that F^{0*} is nondegenerate with a finite (p.d.)

$$\underset{\sim}{V}^{0*} = \iint \underset{\sim}{x}\underset{\sim}{x}' d\, F^{0*}(\underset{\sim}{x}). \tag{29}$$

Then based on the same risk as in (4) (but incorporating the asymptotic distributional approach), the ADR of $\underset{\sim}{U}_{N_c}$ is defined by

$$\rho^{0*}_{c,\underset{\sim}{\gamma}}(F) = c\,\tilde{E}(N_c) + c^{1/2}\,\mathrm{tr}(\underset{\sim}{W}\underset{\sim}{V}^{0*}), \tag{30}$$

where $\tilde{E}$ refers to the expectation with respect to the asymptotic distribution of N_c (and $\underset{\sim}{U}_{N_c}$). Note that $N_c = n^0_c + (N_c - n^0_c)$, where n^0_c is defined by (21), so that whenever $(N_c - n^0_c)/n^0_c \overset{P}{\to} 0$ (as we shall see later on), $\tilde{E}N_c$ can as well be replaced by $n^0_c + o(c^{-1/2})$, and hence, (30) can be expressed as

$$\rho^{0*}_{c,\underset{\sim}{\gamma}}L(F) = c^{1/2}\{(\mathrm{tr}(\underset{\sim}{W}\underset{\sim}{\Sigma}^{*})^{1/2} + \mathrm{tr}(\underset{\sim}{W}\underset{\sim}{V}^{0*})\} + o(c^{1/2}),\ \text{as } c \downarrow 0. \tag{31}$$

We shall find this expression suitable for our subsequent analysis. Also, for the jackknifed versions, $\underset{\sim}{T}^{*}_{n}$, $\underset{\sim}{T}^{*}_{N_c}$, $\underset{\sim}{T}^{*S}_{N_c}$ etc., we may need a second order Hadamard-differentiability of $\underset{\sim}{\tau}(\cdot)$ of $\underset{\sim}{G}$; this is an one-step extension of (7) and (8) which incorporates a quadratic term in (7) and for which (8) is $o(\|H-G_j\|^2)$, uniformly in $H \in \mathcal{X}$, $1 \le j \le p$. We omit the details of these manipulations here and refer to Sen (1988) for some detailed accounts.

ADR RESULTS: SEQUENTIAL CASE

The ADR versions of (25) and (26) [in the light of (30)] will be denoted by $\tilde{\rho}^{*}_{c,\underset{\sim}{\gamma}}(F)$ and $\tilde{\rho}^{*S}_{c,\underset{\sim}{\gamma}}(F)$ respectively; the parallel measures for $\underset{\sim}{T}^{*}_{N_c}$ and $\underset{\sim}{T}^{*S}_{N_c}$ are denoted by $\tilde{\rho}^{*0}_{c,\underset{\sim}{\gamma}}(F)$ and $\tilde{\rho}^{*0S}_{c,\underset{\sim}{\gamma}}(F)$ respectively. Then, we have the following.

THEOREM 1. *If $\underset{\sim}{\tau}(\cdot)$ is first order Hadamard differentiable at $\underset{\sim}{G}$ and $\underset{\sim}{T}^{**}_{1}(\cdot)$ is Hadamard continuous at $\underset{\sim}{G}$, then in the light of ADR, $\underset{\sim}{T}_{N_c}$ is asymptotically (as* $c \downarrow 0$) MRE, i.e., for $p \ge 1$,

$$\lim_{c\downarrow 0}\{\tilde{\rho}^{*}_{c,\underset{\sim}{\gamma}}(F)/\rho^{0}_{c}(F)\} = 1,\quad \forall\, \underset{\sim}{\gamma} \in E^p. \tag{32}$$

If, in addition, $\underset{\sim}{\tau}(\cdot)$ is second order Hadamard differentiable at $\underset{\sim}{G}$, then the same AMRE property holds for $\underset{\sim}{T}^{*}_{N_c}$.

THEOREM 2. *Suppose that $p \ge 3$ and the hypothesis of Theorem 1 holds. Then*

$$\lim_{c\downarrow 0}\{\tilde{\rho}^{*S}_{c,\underset{\sim}{\gamma}}(F)/\tilde{\rho}^{*}_{c,\underset{\sim}{\gamma}}(F)\} \le 1 \quad\text{and}\quad \lim_{c\downarrow 0}\{\tilde{\rho}^{*0S}_{c,\underset{\sim}{\gamma}}(F)/\tilde{\rho}^{*0}_{c,\underset{\sim}{\gamma}}(F)\} \le 1, \tag{33}$$

for all $\underset{\sim}{\gamma} \in E^p$, where the strict inequality signs hold for all $\underset{\sim}{\gamma}$ close to the *pivot* ($\underset{\sim}{0}$) *Thus*, $\underset{\sim}{T}^{S}_{N_c}$ (*or* $\underset{\sim}{T}^{*S}_{N_c}$) *has the desired dominance property.*

Before we proceed to sketch the proofs of Theorems 1 and 2, we present the following remarks:

(i) In (33), the equality sign is attained in the limit $\|\underset{\sim}{\tau}\|^2_{\underset{\sim}{W}} \to \infty$. The implication of this result is that for any significant detour from the pivot, the shrinkage effect is asymptotically negligible. Thus, the Stein-rule estimators (in the sequential case too) are advocated only when one has a prior belief that for the true parameter point $\underset{\sim}{\tau}$ (though unknown) and the chosen pivot $\underset{\sim}{\tau}^0$, $\|\underset{\sim}{\tau}-\underset{\sim}{\tau}^0\| = O(c^{1/4})$ where the cost per unit sampling, c, is small.

(ii) For both the theorems, the ADR measures may be replaced by their <u>asymptotic risk</u> (<u>AR</u>) counterparts (computed from (25) and (26)) under the same asymptotic setup in (27). The conclusions would have been the same. However, this could require more stringent regularity conditions on $\underset{\sim}{\tau}(\cdot)$. We shall discuss these in the concluding section.

(iii) Often, it is of interest to study the asymptotic distribution theory (viz., normality) of the stopping number N_c (in a standardized form: $(n_c^0)^{1/2}\{N_c-n_c^0\}$, as $c \downarrow 0$). This is also possible under additional regularity conditions, and we shall discuss them in the last section.

(iv) The asymptotic dominance result in Theorem 2 rests on the particular adaptation of $\underset{\sim}{W}$ in (24). If in (24), one uses a different matrix (say, $\underset{\sim}{W}^*$) while in (4) $\underset{\sim}{W}$ is adapted, then (33) may not hold for every $(\underset{\sim}{W},\underset{\sim}{W}^*)$. Or, in other words, the dominance of (sequential or fixed-sample size) Stein-rule estimator over the classical version may depend very much on the chosen $\underset{\sim}{W}$. In practive $\underset{\sim}{W}$ may not be unique, and hence, there may be an issue regarding the <u>robustness</u> aspect of the Stein-rule estimators (with respect to the variation in the chosen $\underset{\sim}{W}$). In the parametric case, $\underset{\sim}{W}$ may be linked to the Fisher <u>information matrix</u> (say, $\underset{\sim}{\mathcal{I}}$), and a natural analogue of this in the nonparametric case is $\underset{\sim}{W} = (\underset{\sim}{\Sigma}^*)^{-1}$, where $\underset{\sim}{\Sigma}^*$ is the dispersion matrix of the influence functions, defined in (14). In this setup, $\underset{\sim}{W}$ is unknown, so that in the definition of T_n^{*S} etc. we need to make some adjustments. One simple way is to estimate $\underset{\sim}{W}$ by $\underset{\sim}{V}_n^{*-1}$, where $\underset{\sim}{V}_n^*$ is defined in (13); the justification is provided by (15). In this case, $d_n=1$ with probability one and (18) simplifies to

$$\underset{\sim}{T}_n^S = (1 - k\,\mathcal{L}_n^{-1})\underset{\sim}{T}_n, \quad 0 < k < 2(p-2), \quad p \geq 2. \tag{34}$$

This simplified form corresponds to the classical James-Stein (1962) version. In this case, one may even consider a <u>positive-rule version</u>:

$$\underset{\sim}{T}_n^{S+} = (1 - k\,\mathcal{L}_n^{-1})^+\underset{\sim}{T}_n, \tag{35}$$

where $a^+ = \max(a, 0)$. A positive rule version $\underset{\sim}{T}_{N_c}^{S+}$ of $\underset{\sim}{T}_{N_c}^{S}$ may be defined in the same fashion. Also (34) and (35) extend readily to their jackknifed versions $\underset{\sim}{T}_n^*$ and $\underset{\sim}{T}_{N_c}^*$. The positive rule versions have usually smaller ADR than their usual Stein-rule counterparts.

(v). Note that in Theorems 1 and 2, the jackknifed covariance matrix $\underset{\sim}{V}_n^*$ rests on the first order compact differentiability only, while for the jackknifed estimators $\underset{\sim}{T}_n^*$, $\underset{\sim}{T}_{N_c}^{*S}$, $\underset{\sim}{T}_{N_c}^*$ and $\underset{\sim}{T}_{N_c}^{*S+}$, we have invoked the second order differentiability property of $\underset{\sim}{T}(\cdot)$. This subtle point will be made clear in the next section.

OUTLINE OF PROOFS OF (32) AND (33)

First, defining n_c^0 and N_c as in (21) and (23), we show that

$$N_c/n_c^0 \to 1 \quad \text{a.s., as } c \downarrow 0. \tag{36}$$

Note that by definition in (23), for every $n \geq n_0$

$$[N_c > n] \equiv [cm^2 - m^{-a} \leq \operatorname{tr}(\underset{\sim}{W}\underset{\sim}{V}_m^*), \ \forall\, n_0 \leq m \leq n], \tag{37}$$

and letting $n_c^* \sim c^{-(2+a)^{-1}}$, $c > 0$, we obtain from (23) that

$$P\{N_c \geq n_c^*\} = 1, \ \forall\, c > 0, \text{ where } n_c^* \to \infty \ \text{ as } \ c \downarrow 0. \tag{38}$$

Further, by (21), for $c \downarrow 0$, we may replace n_c^0 by $c^{-1/2}\{\operatorname{tr}(\underset{\sim}{W}\underset{\sim}{\Sigma}^*)\}^{1/2}$, i.e., $c(n_c^0)^2$ by $\operatorname{tr}(\underset{\sim}{W}\underset{\sim}{\Sigma}^*)$. Thus, the right hand side of (37) can be expressed as

$$[c(m^2 - (n_c^0)^2 - m^{-1} \leq \operatorname{tr}(\underset{\sim}{W}(\underset{\sim}{V}_m^* - \underset{\sim}{\Sigma}^*]), \ \forall\, n_c^*, \ m \leq n] \tag{39}$$

Let $n_{c,j}^0 = (1+(-1)^j\epsilon)n_c^0$, $j = 1,2$, $\epsilon(>0)$ arbitrary. Then, putting $n = n_{c,1}^0$ in (37) and (39) and noting that $c(m^2 - (n_c^0)^2) \leq c(n_c^0)^2[(1-\epsilon)^2 - 1] = -\epsilon(2-\epsilon)$. $\operatorname{tr}(\underset{\sim}{W}\underset{\sim}{\Sigma}^*) < 0$, $\forall\, n \leq n_{c,1}^0$, we obtain by using (23) and (39) that

$$N_c \geq n_c^0(1-\epsilon) = n_{c,1}^0 \quad \text{a.s., as } c \downarrow 0 \tag{40}$$

Similarly, for $n \geq n_{c,2}^0$, $c[n^2 - (n_c^0)^2] \geq (2-\epsilon)\epsilon \; tr(\underset{\sim}{W}\underset{\sim}{\Sigma}^*) > 0$, so that by (23), (37) and (39), we obtain by a few simple steps that

$$N_c \leq n_{c,2}^0 = n_c^0(1+\epsilon) \quad \text{a.s., as } c \downarrow 0. \tag{41}$$

Then, (36) follows from (40) and (41). It is interesting to note that for this a.s. convergence property of the stopping number, for $\{V_n^*\}$, the first order differentiability of $\underset{\sim}{\tau}(\cdot)$ suffices. For every $n(\geq 1)$ let us denote by $\underset{\sim}{\bar{T}}_n^0 = (\bar{T}_{n,1}^0, \ldots, \bar{T}_{n,p}^0)'$ where

$$\bar{T}_{n,j}^0 = n^{-1} \Sigma_{i=1}^n \tau_{j1}(G_j; Y_i^{(j)}), \quad 1 \leq j \leq p. \tag{42}$$

Then, under the first order Hadamard differentiability of $\underset{\sim}{\tau}(\cdot)$, we have [viz., Parr (1985), Sen (1988)]

$$\underset{\sim}{T}_n - \underset{\sim}{\tau} = \underset{\sim}{\bar{T}}_n^0 + o((\|\underset{\sim}{G}_n - \underset{\sim}{G}\|), \tag{43}$$

where $\|\underset{\sim}{G}_n - \underset{\sim}{G}\| = \max\{\|G_{nj} - G_j\|, \; 1 \leq j \leq p\}$. From the classical results on the weak convergence of Kolmogorov-Smirnov statistic it follows that under (36), as $c \downarrow 0$,

$$(n_c^0)^{1/2}\|\underset{\sim}{G}_{N_c} - \underset{\sim}{G}\| = O_p(1), \tag{44}$$

so that by (43) and (44), as $c \downarrow 0$

$$(n_c^0)^{1/2}[\underset{\sim}{T}_{N_c} - \underset{\sim}{\tau}] = (n_c^0)^{1/2} \, \underset{\sim}{\bar{T}}_{N_c}^0 + o_p(1). \tag{45}$$

Since the $\tau_{j1}(\cdot)$ are square integrable, we have

$$(n_c^0)^{1/2} \, \{\underset{\sim}{\bar{T}}_{n_c}^0\} \xrightarrow{\mathcal{D}} N_p(\underset{\sim}{0}, \underset{\sim}{\Sigma}^*) \quad \text{as} \quad c \downarrow 0, \tag{46}$$

and the classical Anscombe condition holds, i.e.,

$$\max_{m: |m-n| \leq \delta n} n^{1/2}\|\underset{\sim}{\bar{T}}_m^0 - \underset{\sim}{\bar{T}}_n^0\| \xrightarrow{P} \text{ as } n \to \infty \; (\delta > 0 \text{ small}). \tag{47}$$

From (45), (46) and (47), we directly obtain that as $c \downarrow 0$

$$(n_c^0)^{1/2}[\underset{\sim}{T}_{N_c} - \underset{\sim}{\tau}] \xrightarrow{\mathcal{D}} N_p(\underset{\sim}{0}, \underset{\sim}{\Sigma}^*). \tag{48}$$

$$(n_c^0)^{1/2}[\underset{\sim}{T}_{N_c} - \underset{\sim}{T}_{n_c^0}] \xrightarrow{P} \underset{\sim}{0}, \quad \text{as } c \to 0. \tag{49}$$

Given (48) and (49), we may evaluate (30) or (31) (for $\underset{\sim}{T}_{N_c}$) by using $\{\underset{\sim}{T}_{n_c^0}\}$ (when $c \downarrow 0$) and this leads to (32). To obtain the parallel result for $\{\underset{\sim}{T}^*_{N_c}\}$, we may note that the magnifying factors (n and n-1) in (11) may call for some extra manipulations in the derivations of the desired results. Under the (assumed) second order differentiability condition on $\underset{\sim}{\tau}(\cdot)$ [see Theorem 1], we may proceed as in Sen (1988) and write

$$\underset{\sim}{T}^*_n = \underset{\sim}{T}_n + O(\tfrac{1}{n}) \quad \text{a.s.} \quad (\text{as } n \to \infty)$$

$$= \underset{\sim}{\tau} + \underset{\sim}{\bar{T}}^0_n + o(\|\underset{\sim}{G}_n - \underset{\sim}{G}\|) + O(\tfrac{1}{n}) \quad \text{a.s.} \tag{50}$$

Hence, we may repeat (44) through (49) and obtain (32) (for $\underset{\sim}{T}^*_{N_c}$). Let us proceed to the proof of Theorem 2. Parallel to (17), we write $\mathcal{L}^*_n = n\|\underset{\sim}{T}_n\|^2_{\underset{\sim}{\Sigma}^{*-1}}$. Then by using the Courant Theorem (on the ratio of two quadratic forms), we obtain by (15) that

$$\mathcal{L}_n / \mathcal{L}^*_n \to 1 \quad \text{a.s.,} \quad \text{as } n \to \infty. \tag{51}$$

On the other hand, using (45) and some standard steps, we have

$$|\mathcal{L}^*_{N_c} - \mathcal{L}^*_{n_c^0}| \xrightarrow{P} 0 \quad \text{as} \quad c \downarrow 0 \tag{52}$$

[under $H_0 : \underset{\sim}{\tau} = \underset{\sim}{0}$ as well as $\{K_c\}$ in (27)]. Combining (51) and (52), it follows (by noting that $n_c^0 \sim \{c^{-1} \operatorname{tr}(\underset{\sim}{\Sigma}^* \underset{\sim}{W})\}^{1/2}$) that

$$\mathcal{L}_{N_c} - \mathcal{L}_{n_c^0} \xrightarrow{P} 0 \quad (\text{under } \{K_c\} \text{ or } H_0), \quad \text{as} \quad c \downarrow 0; \tag{53}$$

$$\mathcal{L}_{n_c^0} \xrightarrow{\mathcal{D}} \chi^2_{p,\Delta}; \quad \Delta = [\operatorname{tr}(\underset{\sim}{\Sigma}^* \underset{\sim}{W})]^{1/2} \underset{\sim}{\tau}' \underset{\sim}{\Sigma}^{*-1} \underset{\sim}{\tau}. \tag{54}$$

In fact, we let $\underset{\sim}{Z}_{n_c^0} = (n_c^0)^{1/2}[\underset{\sim}{T}_{n_c^0} - \underset{\sim}{0}]$ and note that under $\{K_c\}$, $\underset{\sim}{Z}_{n_c^0} \xrightarrow{\mathcal{D}} \underset{\sim}{Z} \sim$

$N_p(\tilde{\tau}, \tilde{\Sigma}^*)$, then we have under $\{K_c\}$,

$$n^{1/2}(\tilde{T}^S_{n_c 0}) \overset{\mathcal{D}}{\rightarrow} (\tilde{I} - k\delta \|\tilde{Z}\|^{-1}_{\tilde{\Sigma}^{*-1}} \tilde{W}^{-1} \tilde{\Sigma}^{*-1})\tilde{Z} \tag{55}$$

and parallel to (49), we have then

$$c^{-1/4}\|\tilde{T}^S_{N_c} - \tilde{T}^S_{n_c 0}\| \overset{P}{\rightarrow} 0, \quad \text{as} \quad c \downarrow 0. \tag{56}$$

Now (55) permits us to evaluate (29) for the Stein-rule version $\tilde{T}^S_{N_c}$, and, as such, using (29)-(31) along with the expression (for the ADR of the Stein-rule estimator in the nonsequential case for M-estimators) in Sen and Saleh (1987), we obtain that

$$\tilde{\rho}^{*S}_{c,\tilde{\tau}}(F) = \tilde{\rho}^{*}_{c,\tilde{\tau}}(F) - \ell(k, \tilde{\tau}, p, F), \quad \forall\, \tilde{\tau} \in E^p, \tag{57}$$

where $\ell(\cdot)$ is nonnegative for every $k : 0 < k < 2(p-2)$, $p \geq 3$ and $\tilde{\tau} \in E^p$; it is strictly positive at $\tilde{\tau} = \tilde{0}$ as well as for $\tilde{\tau}$ in a neighborhood of $\tilde{0}$, and it goes to 0 as $\tilde{\tau}$ moves away from the pivot. Under the second order compact differentiability of $\tilde{\tau}(\cdot)$, using (50), it follows that (55) extends to $\tilde{T}^{*S}_{n_c}$ as well, and hence, (57) also pertains to $\tilde{T}^{*S}_{N_c}$. This shows that (33) holds.

SOME CONCLUDING REMARKS

We make some remarks here on the rationality of the use of ADR (instead of the asymptotic risk (AR)) in the sequential case. If we look at the conventional estimator $\tilde{T}_n$, its AR can be computed under fairly general regularity conditions. Relatively more stringent regularity conditions are needed for the computation of the AR of $\tilde{T}_{N_c}$. This is primarily due to a "uniform integrability condition" on the $n\|\tilde{T}_m - \tilde{T}_n\|^2_{\tilde{V}^{*-1}_m}$ (for $m : |m-n| < \delta n$, $n \rightarrow \infty$, $\delta \downarrow 0$) as well as on the elements of $\tilde{V}^*_m$, $|m-n| < \delta n$. The second uniformly condition demands some L_1-norm approximations for $\tilde{V}^*_m$. In the particular case of U-statistics (i.e., von Mises functionals), by assuming that the kernel has a finite rth absolute) moment for some $r > 4$, Sen and Ghosh (1981) were able to verify this uniform integrability condition. However, in the current context, the assumed first order Hadamard-differentiability of $\tilde{\tau}(\cdot)$ (and the continuity of $T^{**}(\cdot)$) may not suffice. If, we assume that $\tau(\cdot)$ is second order Hadamard differentiable then for the jackknifed covariance matrix $\underset{\sim}{V}_n^*$, the desired uniform (square) integrability can be established along the lines of Sen (1988) when $E\|\underset{\sim}{T}(\underset{\sim}{G}_n)\|^r < \infty$ for some $r \geq 4$. This would also entail the asymptotic normality of the stopping number N_c (i.e., $c^{1/4}(N_c - n_c^0)$ will have asymptotically (as $c \downarrow 0$) a normal distribution with 0 mean and a finite variance). The situation is far more complicated with the sequential shrunken estimators. Even in the fixed sample case, T_n^S in (18) encounters the same difficulty. Not only one would require uniform square integrability for the elements of V_n^{*-1}, but also that $\mathcal{L}_n^{-1}$ has a finite expectation. In the standard normal theory model, $\mathcal{L}_n$ has a (central or noncentral) variance-ratio distribution, and hence, for $p \geq 3$, $E\,\mathcal{L}_n^{-1}$ exists. On the other hand, under H_0 or $\{K_c\}$, $\mathcal{L}_n$ converges in law to a central or noncentral chi square variable (with p degrees of freedom). But, this "convergence in law" does not ensure the "convergence in negative moments" of $\mathcal{L}_n$ to that of a chi square variable. In fact, in some cases, $\mathcal{L}_n$ may be arbitrarily close to '0' with a positive probability (however, small it may be), and this can push up the expected value of $\mathcal{L}_n^{-1}$ much beyond the value as may be obtained by using the appropriate chi square distribution. This technical problem can be taken care of in some ways. First, if one considers a positive rule shrinkage estimator [as in (35) then for $\mathcal{L}_n \leq k$, by forcing $\underset{\sim}{T}_n^{+S}$ to $\underset{\sim}{0}$, one avoids this highly inflated status of $\underset{\sim}{T}_n^S$ near $\mathcal{L}_n \sim 0$ and for $\mathcal{L}_n > k$, $\mathcal{L}_n \xrightarrow{\mathcal{D}} \chi^2_{p,\Delta} \Rightarrow E\{\mathcal{L}_n^{-1}\, I(\mathcal{L}_n > k)\} \rightarrow E(\chi^2_{p,\Delta}\, I(\chi^2_{p,\Delta} > k)\}$. Secondly, as in Sen and Saleh (1985), we may allow a small truncation near the origin, and this will enable us to incorporate the convergence in law of $\mathcal{L}_n^{-1}$ to that of L_1-norm convergence on $\mathcal{L}_n \geq \epsilon > 0$. In either case, if $\underset{\sim}{T}(\cdot)$ is assumed to be second order Hadamard differentiable, so that the $\underset{\sim}{V}_n^*$ has the desired uniform integrability property, then the AE results can be obtained for the parallel sequential versions. Whenever these AE results hold they are in agreement with the corresponding ADR results. Hence, from the interpretational point of view, the ADR results serves the right purpose without unnecessarily calling for these extra regularity conditions or modifications. As has been explained in (34)-(35), there are good points in considering the James-Stein (1962) versions. In such a case, the stopping rule in (23) may as well be replaced by a nonstochastic integer:

$$N_c^{00} \sim (p/c)^{1/2} \quad \text{as} \quad c \downarrow 0, \tag{58}$$

so that the sequential rules may all be replaced by their nonsequential counterparts with n given by (59). However, even in this special case, a genuine stopping time may arise in a natural way. Suppose that in (20), we replace $\mathbf{T}_n$ by $\mathbf{T}_n^S$ and denote the corresponding covariance matrix by $\mathbf{\Sigma}^S$ (instead of $\mathbf{\Sigma}^*$). If it is possible to derive a suitable estimator $\mathbf{V}_p^S$ of $\mathbf{\Sigma}^S$ (and we choose $\mathbf{\Sigma}^* = \mathbf{W}^{-1}$), then parallel to (23), we would have a stopping number

$$N_c^S = \inf\{n \geq n_0 : n^2 \geq \{\mathrm{tr}(\mathbf{V}_n^{*-1}\ \mathbf{V}_n^S) + n^{-a}\}/c\}, \quad c > 0. \tag{59}$$

The stochastic convergence of $\{N_c^{-S}\}$ to a suitable sequence $\{n_c^{0S}\}$ would naturally depend on (15) as well as the convergence properties of $\{\mathbf{V}_n^S\}$, and granted this, the rest of the results would follow on parallel lines. Since $\mathrm{tr}(\mathbf{W\Sigma}^S) \leq \mathrm{tr}(\mathbf{W},\mathbf{\Sigma}^*)$, it can be shown that $N_c^S/N_c \leq 1$ a.s., as $c \downarrow 0$, and hence, the use of the stopping number leads to an asymptotically smaller ASN too. However, we may remark that the difference $\mathrm{tr}(\mathbf{W\Sigma}^*) - \mathrm{tr}(\mathbf{W\Sigma}^S)$ becomes smaller as the true parameter point $\boldsymbol{\theta}$ moves away from the pivot. Thus, for any (fixed) $\boldsymbol{\theta} \neq \mathbf{0}$, as $c \downarrow 0$, $\mathrm{tr}(\mathbf{W\Sigma}^*) - \mathrm{tr}(\mathbf{W\Sigma}^S)$ may have asymptotically (as $c \downarrow 0$) a positive limit, and in that case, the stopping time in (59) leads to a real reduction in the ASN (in the asymptotic case where $c \downarrow 0$). Finally, we make a general comment on the mapping $\mathbf{X} \rightarrow \mathbf{Y}$ leading to the definition of the functionals in (1). In many situations we may have the $\mathbf{X}_i$ as p-vectors and the $\mathbf{Y}_i$ represent the same vectors i.e., $\mathbf{X}_i \equiv \mathbf{Y}_i$(and p=q). For example, consider the p-variates location model where the d.f.F stands for a p-variate d.f., while $G_1,\ldots,G_p$ are the p univariate marginal d.f. and $\tau_j(G_j)$ is a typical location parameter for the jth marginal, $1 \leq j \leq p$. A very similar case may arise when the $\tau_j(G_j)$ are suitable scale parameters for the marginal d.f.'s. In either case, one may choose the $\tau_j(G_j)$ as suitable L-, M- or R-functionals for which sample estimates $\{T_{nj}\}$ are known to have good robustness and efficiency properties. Functionals corresponding to the trimmed or Winsorized means may also be considered in the same vein. Viewed from this angle, the functionals corresponding to R- or M- estimators may require some regularity conditions more stringent than the usual ones need for a direct approach [See Jurečková and Sen (1982) and Sen (1980)]. There may be other situations where $\mathbf{X} \neq \mathbf{Y}$ and $q \leq p$. For example, consider the case where F, defined on E^4, is the d.f. of $(X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)})$, and we are interested in the pairwise association (correlation) parameters (involving $\binom{4}{2} = 6$ bivariate distributions). Thus, $q = 4 < 6 = p$. For each pair $(X^{(r)}, X^{(s)})$, we may consider an appropriate association parameter τ_{rs}, such as the Kendall tau, Spearman grade correlation coefficient and others [discussed in detail in Hoeffding (1948)], for $1 \leq r < s \leq 4$. The sample counterparts are Hoeffding's U-statistics and/or von Mises' functionals, and hence, they can be treated in the same setup as in the current study. Alternatively, the results of Sen (1987b) may also be used for them. In eithr approach, the functional form of the d.f. F is not assumed to be given, and hence, to retain the nonparametric structure of F, we may need to take recourse to an asymptotic setup (where $c \downarrow 0$), as has been adapted here. A finite $c(>0)$ with an unspecified F may call for an altogether different (and presumably parametric) approach and may be much more complicated. In fact, generating such an optimal solution (for a given (fixed) $c>0$) in a genuine nonparametric formulation is still an open problem!

REFERENCES

[1] Ghosh, M., Nickerson, D.M. and Sen, P.K. Sequential shrinkage estimation, Ann. Statist. 15 (1987), 817-819.

[2] Hoeffding, W. A class of statistics with asymptotically normal distributions, Ann. Math. Statist. 19 (1947), 293-325.

[3] James, W. and Stein, C. Estimation with quadratic loss, in Proc. 4th Berkeley Symp. Math. Statist. Probability 1 (1962), 361-380.

[4] Jurečková, J. and Sen, P.K. M-estimation and L-estimators of location: Uniform integrability and asymptotically risk-efficient sequential versions. Sequen. Anal. 1 (1982), 27-56.

[5] Parr, W.C. Jackknifing differentiable statistical functionals. Jour. Roy. Statist. Soc. B 47 (1985), 56-56.

[6] Sen, P.K. On nonparametric sequential point estimation of location based on general rank order statistics, Sankhya, Ser. A 42 (1980), 201-219.

[7] Sen, P.K. Sequential Nonparametric: Invariance Principles and Statistical inference, John Wiley, New York, 1981.

[8] Sen, P.K. A James-Stein detour of U-statistics, Commun. Statist. Theor. Meth. A 13 (1984), 2725-2747.

[9] Sen, P.K. On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators, Sankhya, Ser. A 48 (1986), 354-371.

[10] Sen, P.K. Sequential Stein-rule maximum likelihood estimation: General asymptotics, In Statistical Decision Theory and Related Topics IV (eds. S.S. Gupta and J.O. Berger), 2 (1987), 195-208.

[11] Sen, P.K. Sequential shrinkage U-statistics: General asymptotics, Rev. Brasileira de Prob. Estatist. 1 (1987), 1-20.

[12] Sen, P.K. Functional jackknifing: Rationality and general asymptotics, Ann. Statist. 16 (1988), 450-469.

[13] Sen, P.K. and Ghosh, M. Sequential point estimation of estimable parameters based on U-statistics, Sankhya, Ser. A 43 (1981), 331-344.

[14] Sen, P.K. and Saleh, A.K.M.E. On some shrinkage estimation of multivariate location, Ann. Statist. 13 (1985), 272-281.

[15] Sen, P.K. and Saleh, A.K.M.E. On preliminary test and shrinkage M-estimation in linear models, Ann. Statist. 15 (1987), 1580-1592.

[16] Stein, C. Inadmissibility of the usual estimator of the mean of a multivariate normal distribution, in Proc. 3rd Berkeley Symp. Math. Statist. Prob. 1 (1956), 197-206.