

SOME RECENT DEVELOPMENTS IN NONPARAMETRIC AND ROBUST
SHRINKAGE ESTIMATION THEORY

by

Pranab Kumar Sen and A. K. MD. Ehsanes Saleh

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1842

December 1987

Department of Statistics Library

Some Recent Developments in Nonparametric and Robust Shrinkage Estimation Theory

PRANAB KUMAR SEN AND A. K. MD. EHSANES SALEH

In the context of multivariate (or multi-parameter) estimation problems, the Stein-rule or shrinkage estimators are generally admissible and minimax, and they dominate their classical counterparts (in the light of quadratic error risks). However, this glittering picture pertains only to the classical multi-normal and some other special types of exponential families of distributions. During the past five years, attempts have been made to establish this Stein phenomenon in a much wider setup encompassing various nonparametric and robust estimators for possibly non-normal (or non-exponential families of) distributions and for risk functions non necessarily of the quadratic error type. Several interesting features of these recent developments include the following :

- (i) Incorporation of simple and physically meaningful asymptotic considerations leading to solutions closely resembling the classical parametric forms;
- (ii) Development of the concept of asymptotic distributional risk enabling the adaptation of the asymptotic theory for the study of the dominance picture under much less restrictive regularity conditions (than in the parametric cases);
- (iii) Employment of the Pitman closeness criterion in the formulation of alternative risk functions leading to the expected dominance picture for lower dimensional problems as well;
- (iv) Exploitation of the inter-relationships of preliminary test estimation and shrinkage estimation in the study of the related dominance pictures.

All these aspects are systematically reviewed here. Special emphasis has been laid on the robust-efficiency aspects of some of these recent developments.

KEY WORDS: Asymptotic distributional risk; asymptotic dominance; asymptotic risk; Pitman alternatives; Pitman closeness; positive rule shrinkage estimator; preliminary test estimation; risk efficiency.

Pranab K. Sen is Cary C. Boshamer Professor of Biostatistics and Adjunct Professor of Statistics, University of North Carolina, Chapel Hill, NC 27514, USA, and A.K.M. Ehsanes Saleh is Professor of Mathematics & Statistics, Carleton University, Ottawa, Canada K1S 5B6. This work was supported partially by the Boshamer Funds at the University of North Carolina, and partially by the National Science and Engineering Research Council, Canada, Grant No. A 3088.

1. INTRODUCTION

In the classical parametric models, maximum likelihood estimators (MLE) are known to have various optimality properties (mostly, in an asymptotic setup, but, often, for finite sample sizes as well). In linear models with normally distributed errors, often, the MLE reduce to the classical least squares estimators (LSE). In a general multi-parameter model (not necessarily linear), however, the classical MLE/LSE may not be optimal and their exist alternative versions of these estimators (called the shrinkage or Stein-rule estimators (SRE)) which may dominate the MLE/LSE (under appropriate quadratic error or related risk functions). In the literature this is known as the Stein Phenomenon. Research work in this vital area was sparked by Stein(1956) and James and Stein (1961), and the last twenty five years have witnessed a fundamental growth of literature in this fruitful area of research. We may refer to Berger (1985) for a detailed account of some of these developments, mostly relating to the classical multinormal and some specific types of exponential families of distributions.

Generally, the MLE or LSE are not very robust against plausible departures from the assumed (distributional) models, and the Stein-rule versions of the MLE/LSE may suffer from the same drawback. Robust estimation of location/scale/regression parameters has been considered in increasing generality during the past twenty five years, and during the same time period, a significant development has taken place in the area of nonparametric estimation of similar parameters. Optimal unbiased estimation of regular functions [viz., Hoeffding (1948)] is the precursor

of all these developments. Robust estimation of location/scale and regression parameters based on appropriate rank statistics has also been considered by a host of workers' some of these developments are reported in Chapters 6 and 7 of Puri and Sen (1985). Robust M-estimation of location and regression parameters is treated in Huber (1981) and others. Generally, these M-estimators are very robust to local departures from the assumed model, while the rank based estimators are globally robust. Somewhere in between are the L-estimators (based on appropriate linear combinations of functions of order statistics). Modern developments in the area of nonparametric and robust estimation theory are mainly centered around these M-, R- and L-estimators. These estimators together with the U-statistics [Hoeffding (1948)] encompass a wide area, and our main interest centers in this arena.

Let us step into the multi-parameter estimation theory from the robustness point of view. A natural question may arise immediately in this context: Is the Stein phenomenon relevant in robust estimation? The answer is, of course, in the affirmative [viz., Sen (1984), Sen and Saleh (1985, 1987) and others]. Before we proceed to elaborate this issue, we may point out that the success of the dominance of the SRE over their classical counterparts, at least in the multi-normal models, rests heavily of some fundamental identities of Stein (1984), and these may not hold for non-normal distributions. Even so, the Stein phenomenon is largely confined to a neighborhood of the assumed model (pivot)! For example, for the estimation of the mean (vector) $\underline{\theta}$ of a multinormal distribution (with a quadratic error loss), for a sample of size n , the improvement of the SRE over the MLE is significant only when $n \|\underline{\theta} - \underline{\theta}_0\|^2$ is finite (where $\underline{\theta}_0$ is the assumed pivot). Thus, for any (fixed) $\underline{\theta}$ different from $\underline{\theta}_0$, as n increases, $n \|\underline{\theta} - \underline{\theta}_0\|^2 \rightarrow \infty$ and as a result, this improvement becomes less

and less perceptible, and the analytic dominance result on the SRE remains largely of academic interest only. A natural and physically meaningful way of eliminating this drawback of the SRE is to take recourse to appropriate sequences of local alternatives for which the dominance of the SRE holds even in a well defined asymptotic setup. In Section 2, we shall review some recent developments in this direction.

Basically, the SRE are testimators; appropriate test statistics for testing the adequacy of the assumed pivot are incorporated in the actual formulation of the SRE. Thus, the SRE may have generally a factor involving the reciprocal of a test statistic. In the conventional multinormal models, these test statistics may have usually noncentral chi square or noncentral variance ratio distributions, and the Stein identities pave the way for the computation of the (quadratic) risk of the SRE. Unfortunately, for non-normal distributions, these Stein identities may not generally hold, and hence, the exact evaluation of the risk of SRE or the proof of its dominance may be considerably difficult, if not impracticable. Even in an asymptotic setup, the intensity of this difficulty is not lessened to any extent. Fortunately, the asymptotic setup of Section 2 permits the computation of the risk of SRE from their asymptotic distributions; these are termed the asymptotic distributional risks (ADR). In Section 3, we shall exploit fully the role of ADR in Stein-rule estimation theory. We shall see that this provides an easy and meaningful access to the study of the asymptotic dominance of SRE in a much broader setup.

Although a quadratic error loss is a natural contender in the formulation of risk of an estimator, it may lead to considerable complications for SRE and/or related estimators; See Section 3 for some details. Thus, one may also seek for alternative loss functions which lead

to less complicated mathematical manipulations towards the common goal of establishing the desired dominance results. The Pitman (1937) closeness criterion has recently received considerable attention in this context. In Section 4, we shall exploit fully the role of the Pitman Closeness criterion in SRE. This seems to be a viable alternative way to studying the asymptotic dominance results on SRE in robust and nonparametric setups.

In a sense, the preliminary test estimators (PTE) may be regarded as the precursors of the SRE. The inter-relationships of the PTE and SRE are studied in Section 5. This study leads to a natural adaptation of the so called positive rule SRE (PRSRE) in a broad nonparametric setup. The results in Section 2, 3 and 4 are then incorporated in the study of the related dominance properties of the PTE, SRE and PRSRE.

The concluding section deals with a glossary of SRE problems in nonparametric and robust setups, and a list of pertinent references to the relevant works is also provided.

2. ASYMPTOTIC DOMINANCE IN SRE: RELEVANCE OF PITMAN ALTERNATIVES

To illustrate the role of the classical Pitman (local) alternatives in SRE, we start with the following simple model. Let X_1, \dots, X_n be n independent and identically distributed (i.i.d.) random vectors (r.v.) drawn from a multivariate normal population with mean vector θ and covariance matrix Σ (positive definite). For simplicity, suppose that Σ is known, and, without any loss of generality, we may set $\Sigma = I_p$, where p is ≥ 3 . The MLE of θ is $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$, and with the quadratic error loss $\|T - \theta\|^2$, the risk of \bar{X}_n is given by

$$E\|\bar{X}_n - \theta\|^2 = n^{-1}p \quad (2.1)$$

Side by side, we consider the Stein-rule estimator:

$$\bar{X}_n^S = \theta_0 + (1 - cn^{-1} \|\bar{X}_n - \theta_0\|^{-2}) (\bar{X}_n - \theta_0), \quad (2.2)$$

where $0 < c < 2(p-2)$ is the shrinkage factor and θ_0 is the assumed pivot.

For simplicity, we may set $\theta_0 = 0$. Then

$$nE\|\bar{X}_n^S - \theta\|^2 = p-c[2(p-2)-c]E[(p-2+2K_n)^{-1}] \quad (2.3)$$

where K_n has the Poisson distribution with parameter $\Delta_n = EK_n = n\|\theta\|^2$.

The right hand side of (2.3) is a minimum for $c = p-2$. But as n increases,

$\Delta_n \rightarrow \infty$, and hence,

$$E[(p-2+2K_n)^{-1}] \rightarrow 0 \text{ as } n \rightarrow \infty \quad (\forall \theta \neq 0). \quad (2.4)$$

Consequently, for any $\theta \neq 0$,

$$\lim_{n \rightarrow \infty} \{nE\|\bar{X}_n^S - \theta\|^2\} = \lim_{n \rightarrow \infty} \{nE\|\bar{X}_n - \theta\|^2\} = p. \quad (2.5)$$

Thus, the dominance of \bar{X}_n^S over \bar{X}_n is confined to compact intervals for Δ_n i.e., for $\|\theta\| = O(n^{-1/2})$. This is not surprising, as the test statistic $(n\|\bar{X}_n\|^2)$ is consistent against any $\|\theta\| > 0$, and it has a non-degenerate asymptotic distribution for Pitman alternatives $\{H_n\}$ where

$$H_n: \theta = n^{-1/2} \gamma, \quad \gamma \in \Gamma, \text{ compact.} \quad (2.6)$$

Motivated by this simple example, we may now pose the SRE theory in a wider spectrum of robust and nonparametric setups incorporating the same asymptotics in a general mold.

Based on a set $\{X_1, \dots, X_n\}$ of n observations, we conceive of a vector $\underline{T}_n = (T_{n1}, \dots, T_{np})'$, of estimator of a suitable parametric vector $\underline{\theta} = (\theta_1, \dots, \theta_p)'$, where $p \geq 3$. Keeping in mind, the various nonparametric and robust estimators, we may assume that as n increases,

$$n^{1/2}(\underline{T}_n - \underline{\theta}) \xrightarrow{D} N_p(0, \underline{v}), \quad (2.7)$$

where \underline{v} is a positive definite (p.d.) matrix (possibly unknown). We also assume that there exists a sample counterpart of \underline{v} , namely, a stochastic matrix \underline{V}_n , such that

$$\underline{V}_n \xrightarrow{P} \underline{v}, \text{ in probability, as } n \rightarrow \infty. \quad (2.8)$$

For an arbitrary estimator $(\hat{\delta}_n)$ of $\underline{\theta}$, we consider a quadratic error loss function

$$\begin{aligned} L_n(\hat{\delta}_n, \underline{\theta}) &= n(\hat{\delta}_n - \underline{\theta})' \underline{W} (\hat{\delta}_n - \underline{\theta}) \\ &= n \left\| \hat{\delta}_n - \underline{\theta} \right\|_{\underline{W}}^2, \end{aligned} \quad (2.9)$$

where \underline{W} is a given p.d. matrix. Further, we conceive of a pivot $\underline{\theta}_0$ for $\underline{\theta}$; side by side, we consider a suitable test statistic L_n for testing the adequacy of the pivot $\underline{\theta}_0$. We assume that when $\underline{\theta}_0$ holds,

$$L_n \xrightarrow{D} \chi_p^2, \text{ as } n \rightarrow \infty \quad (2.10)$$

Mostly, robust or nonparametric tests can be incorporated in this context. However, the estimator \underline{T}_n (of $\underline{\theta}$) and the test statistic L_n should be

inter-related in a manner such that the Stein phenomenon can be preserved by their simultaneous use. Towards this we assume that as n increases

$$|L_n - n(\underline{T}_n - \underline{\theta}_0)' \underline{V}_n^{-1} (\underline{T}_n - \underline{\theta}_0)| \xrightarrow{P} 0, \quad (2.11)$$

under the null hypothesis $H_0: \underline{\theta} = \underline{\theta}_0$ as well as for local alternatives. For M-, R- and L- estimators of $\underline{\theta}$, (2.11) holds via the asymptotic (first order) representation results of Jurečková (1977), while for U-statistics or von Mises' functionals, \underline{V}_n is taken as the jackknifed estimator of \underline{V} and $L_n = n(\underline{T}_n - \underline{\theta}_0)' \underline{V}_n^{-1} (\underline{T}_n - \underline{\theta}_0)$, so that (2.11) holds by virtue of (2.8) and the Courant theorem. For MLE, L_n is the classical likelihood ratio test and (2.11) holds [see Sen (1986)]. For LSE, \underline{V}_n is the sample covariance matrix and hence, (2.8) holds and this, in turn, implies (2.11).

Now, for a given $(\underline{\theta}_0, W)$, based on the triplet $(\underline{T}_n, \underline{V}_n, L_n)$, the SRE version of \underline{T}_n may be posed as

$$\underline{T}_n^S = \underline{\theta}_0 + [I - cd_n L_n^{-1} W^{-1} V_n^{-1}] (\underline{T}_n - \underline{\theta}_0) \quad (2.12)$$

where

$$d_n = \text{ch}_{\min} (WV) \quad \text{and} \quad 0 < c \leq 2(p-2). \quad (2.13)$$

It is possible to choose some alternative versions paralleled to the normal theory case treated in Berger (1980), and we may refer to them later on.

Our main contention is to study the relative risk picture of \underline{T}_n and \underline{T}_n^S when $\underline{\theta}$ is "close to" $\underline{\theta}_0$. Note that in most of the cases, by virtue of (2.11) or by actual construction of L_n , the test for $H_0: \underline{\theta} = \underline{\theta}_0$ based on L_n is consistent against $\underline{\theta} \neq \underline{\theta}_0$, and as such for any (fixed) $\underline{\theta} \neq \underline{\theta}_0$, $L_n^{-1} \rightarrow 0$,

in probability, as $n \rightarrow \infty$, so that by (2.8) and (2.12): $T_n^S - T_n \rightarrow 0$ in probability, as $n \rightarrow \infty$. This limiting degeneracy can be avoided when L_n^{-1} is itself a nondegenerate r.v., and for this, a natural setup is to consider Pitman-type alternatives:

$$H_n: \theta = \theta_0 + n^{1/2} \underline{\gamma}, \quad \underline{\gamma} \in \Gamma \text{ (compact)}. \quad (2.14)$$

Under (2.14), L_n has asymptotically non-central χ^2 distribution with p degrees of freedom and noncentrality parameter $\Delta = \underline{\gamma}' \underline{V}^{-1} \underline{\gamma}$ ($< \infty$), and this permits us to draw the picture of the relative risk of T_n and T_n^S in a meaningful way. The case of $H_0: \theta = \theta_0$ is included in this setup as a particular one (where $\underline{\gamma} = 0$). This explains the relevance of Pitman alternatives in SRE; for finite sample sizes, this also explains the role of n in identifying the effective zone of dominance of the SRE over their classical counterparts. For more details, we may refer to Sen (1984) and Sen and Saleh (1985) where this concept was explored systematically.

3. ASYMPTOTIC RISK VS. ADR

Note that (2.7), (2.8) and (2.9) may not ensure that

$$E\{L_n(T_n, \theta)\} \rightarrow \text{Trace}(\underline{W}\underline{V}), \text{ as } n \rightarrow \infty, \quad (3.1)$$

although (3.1) holds under appropriate moment convergence properties of T_n and \underline{V}_n . The situation is far more complicated with the SRE in (2.12) (or with its other variant forms). The shrinkage factor (i.e., $cd \frac{L_n^{-1} \underline{W}^{-1} \underline{V}_n^{-1}}{L_n^{-1}}$) is highly nonlinear in form, and, further, the factor L_n^{-1} makes it necessary to bring in the existence (and convergence) of the negative moments of L_n (up to a certain order). For the multi-normal mean problem,

the ingenious Stein identities come to the rescue of this technical difficulty (for $p \geq 3$), although these identities may not suffice for other distributions. In fact, if L_n can be equal to 0 with a positive probability (however small it may be), L_n^{-1} assumes the value $+\infty$ with a non-zero probability, and hence, the SRE may not have a finite quadratic error risk. This case may typically arise in the case of discrete distributions (for L_n) and is commonly encountered in rank based tests for H_0 . This technical difficulty remains in tact in the asymptotic case as well: The computation of the asymptotic risk (AR) of the SRE in a general (non-normal or arbitrary) distribution may require more than the convergence of L_n^{-1} in the q th mean, for some $q \geq 1$. Verification of this [even, under the sequence of local alternatives in (2.14)] for general nonparametric or robust statistics may be quite involved and may thereby require quite stringent regularity conditions on the underlying distributions and/or the adapted score functions. There are alternative ways of eliminating this drawback:

(i) Use of some other form of risk which does not require the L_n^{-q} -convergence of L_n^{-1} , for some $q \geq 1$. In this context, the "Pitman closeness" criterion to be discussed in the next section is a viable alternative.

(ii) SRE adapted to left truncation of L_n . Note that the above mentioned difficulty arises mainly when L_n is close to zero. Thus, the problem may be avoided to a greater extent by modifying the SRE for L_n "close to" 0. One possibility is to consider the modified SRE:

$$T_n^{*s} = \begin{cases} T_n^s, & \text{if } L_n \geq a_n (> 0) \\ \theta_0 \text{ (or } T_n), & \text{if } L_n < a_n \end{cases} \quad (3.2)$$

So long as a_n is (uniformly in n) bounded away from 0, the uniform integrability of L_n^{-1} (over $\{L_n > a_n\}$) does not pose any serious problem, and hence, the computation of the asymptotic risk of T_n^* can be made under standard regularity conditions; for some related developments, we may refer to Sen (1984) and Sen and Saleh (1985). In the particular case $V_n = s_n^2 W_n^{-1}$, for some (possibly stochastic) nonnegative s_n^2 , the shrinkage factor $c d_n L_n^{-1} W_n^{-1} V_n^{-1}$ reduces to $I(1 - c L_n^{-1})$, so that the left truncation on L_n is similar, in principle, to the positive rule version of SRE. In fact, the PRSRE then corresponds to the special case of (3.2) when $a_n = c$. Thus, for the PRSRE, asymptotic risk can be computed under regularity conditions weaker than those in the case of the SRE. For the multivariate normal mean estimation problem, the PRSRE is known to perform better than the SRE, and hence, by drawing analogy with this model (through (2.7), (2.8) and (2.10)), in the general case too, we may advocate the use of this left truncation on L_n .

(iii) Incorporation of ADR. For a moment, let us go back to the classical case where the variance (or the dispersion matrix) associated with the asymptotic (multi-) normal distribution of an estimator (in its normalized form) is generally used in the definition of its asymptotic efficiency. The main advantage of this approach is that it does not require any convergence result stronger than the asymptotic normality (or multi-normality), and thereby, avoids the necessity of making stronger regularity assumptions (which are otherwise pertinent to the AE). The same feature is true for SRE.

For an estimator T_n^* of θ , assume that under (2.14), the asymptotic distribution of $\sqrt{n}(T_n^* - \theta)$ is nondegenerate; we denote it by

$$G^*(x) = \lim_{n \rightarrow \infty} P\{\sqrt{n}(T_n^* - \theta) \leq x \mid H_n\}, \quad x \in R^p. \quad (3.3)$$

For the classical estimator, (2.7) ensures that G^* is a multinormal distribution function. For various nonparametric and robust estimation problems, for the SRE, the form of G^* has been studied in detail by Sen (1984), Sen and Saleh (1985, 1987) and others. This asymptotic distribution conforms to the case of the Stein-rule estimator for the multivariate normal mean problems, and, as such, G^* has finite second order moments too (when $p \geq 3$). Let then

$$\underline{v}^*(\underline{\gamma}) = \int \dots \int \underline{xx}' / dG^*(x) , \quad (3.4)$$

be the dispersion matrix for G^* ; since G^* may depend on $\underline{\gamma}$ (through $\{H_n\}$), $\underline{v}^*(\underline{\gamma})$ is also generally dependent on $\underline{\gamma}$. Then, corresponding to the loss function in (2.9), the ADR of \underline{T}_n^* is defined by

$$p^*(\underline{T}_n^*, \underline{\gamma}) = \text{Trace} (\underline{W} \underline{v}^*(\underline{\gamma})) . \quad (3.5)$$

On the other hand, granted the existence of the AR, we have

$$\begin{aligned} p(\underline{T}_n^*, \underline{\gamma}) &= \lim_{n \rightarrow \infty} E \{ L_n(\underline{T}_n^*, \underline{\theta}_0 + n^{-1/2} \underline{\gamma}) \} \\ &= \text{Trace} (\underline{W} \underline{v}^{0*}(\underline{\gamma})) \end{aligned} \quad (3.6)$$

where $\underline{v}^{0*}(\underline{\gamma}) = \lim_{n \rightarrow \infty} E \{ [(\underline{T}_n^* - \underline{\theta}_0 - n^{-1/2} \underline{\gamma})(\underline{T}_n^* - \underline{\theta}_0 - n^{-1/2} \underline{\gamma})'] \}$ and the existence of this limit may generally require more stringent regularity conditions. However, granted these regularity conditions, (3.5) and (3.6) both have the same expression. Thus, the main advantage of using the ADR

in (3.5) is to bypass these extra regularity conditions (needed for (3.6)) and to arrive at a comparable expression. For the classical estimator $\{T_n\}$, satisfying (2.7), we readily obtain from (3.3), (3.4) and (3.5) that

$$p^*(T, \underline{\gamma}) = p^*(T, 0) = \text{Trace}(\underline{W}\underline{v}), \quad \forall \underline{\gamma} \in \Gamma. \quad (3.7)$$

where \underline{v} appears in (2.7). For the SRE T_n^s in (2.12), we obtain (on proceeding as in Sen (1984)) that

$$\begin{aligned} p^*(T^s, \underline{\gamma}) &= \text{Trace}(\underline{W}\underline{v}) - 2c \{ch_p(\underline{W}\underline{v})\} \{1 - \Delta E(\chi_{p+2, \Delta}^{-2})\} \\ &\quad + c^2 \{ch_p(\underline{W}\underline{v})\}^2 \cdot \{\text{Trace}(\underline{W}\underline{v})^{-1} E(\chi_{p+2, \Delta}^{-4}) + \Delta^* E(\chi_{p+4, \Delta}^{-4})\} \end{aligned} \quad (3.8)$$

where

$$\Delta = \underline{\gamma}' \underline{v}^{-1} \underline{\gamma} = \text{Trace}(\underline{v}^{-1} \underline{\gamma} \underline{\gamma}'), \quad \Delta^* = \underline{\gamma}' \underline{v}^{-1} \underline{W}^{-1} \underline{v}^{-1} \underline{\gamma}, \quad (3.9)$$

and $\chi_{q, \delta}^2$ stands for a random variable having the noncentral chi-square distribution with q th degrees of freedom and noncentrality parameter δ ; $\chi_{q, \delta}^{-2r} = (\chi_{q, \delta}^2)^{-r}$, $r = 1, 2$. It is easy to verify that the right hand side of (3.8) is less than $\text{trace}(\underline{W}\underline{v})$, for every $\underline{\gamma} \in \Gamma$, whenever $0 < c < 2(p-2)$. This establishes the asymptotic dominance of the SER in the light of the ADR, and this also illustrates the utility of the ADR criterion in this context.

4. DOMINANCE IN THE LIGHT OF PITMAN-CLOSENESS

In the multi-parameter (θ) case, for two competing estimators $T_{\sim n}$ and $T_{\sim n}^*$, with the norm $\|\cdot\|_{\sim W}$ defined in (2.9) [for a given p.d. $\sim W$], we say that $T_{\sim n}$ dominates $T_{\sim n}^*$ in the light of the Pitman-closeness, if

$$P_{\sim \theta} \{ \|\tilde{T}_{\sim n} - \tilde{\theta}\|_{\sim W} \leq \|\tilde{T}_{\sim n}^* - \tilde{\theta}\|_{\sim W} \} \geq \frac{1}{2}, \forall \tilde{\theta} \quad (4.1)$$

with the strict inequality sign holding at least for some $\tilde{\theta}$. In this definition, we do not need to compute the first or second order moments of $\|\tilde{T}_{\sim n} - \tilde{\theta}\|_{\sim W}$ or $\|\tilde{T}_{\sim n}^* - \tilde{\theta}\|_{\sim W}$, and hence, we may need less stringent regularity conditions to verify (4.1). There has been some recent development in this area. Some justifications for using (4.1) for SRE have also been laid down both on empirical and analytical grounds. For some of the related developments, we may refer to Kubokawa, Sen and Saleh (1987), where other references have also been cited.

In the context of the asymptotic theory of SRE, in a more general nonparametric and robust estimation setup, (4.1) can be incorporated with great advantages. Firstly, we need to consider only the joint (asymptotic) distribution of $(\tilde{T}_{\sim n}, \tilde{T}_{\sim n}^*)$ (properly normalized), and hence, the question of verifying the usual moment convergence results needed for the asymptotic risk computations does not arise here. Secondly, in the context of the study of the AR or ADR [made in Section 3], we generally need that $p \geq 3$ (i.e., there is no improvement due to shrinkage for $p \leq 2$). By an adaptation of (4.1) it can be shown that the dominance picture may as well be studied for $p = 2$. Thirdly, both the ADR and (4.1) are based on distributional properties of the estimators. However, in the ADR, in (3.4), we use the second order moment of G^* , and this in turn, leads to the requirement of $p \geq 3$. We do not need this for (4.1).

Note that by the Courant Theorem,

$$\frac{(T_n/V_n^{-1}W_n^{-1}V_n^{-1}T_n)/(T_n/V_n^{-1}T_n)}{\chi_1^2(V_n^{-1}W_n^{-1})} = d_n^{-1}, \quad (4.2)$$

and hence, using (2.12) and (4.1), to establish the (asymptotic) dominance of the SRE over the classical estimator T_n , it suffices to show that

$$P_\theta \{ (T_n - \theta)' V_n^{-1} T_n \geq c/2 \} \geq 1/2, \quad \forall \theta \in R^p, \quad (4.3)$$

and in this context, θ may as well be replaced by a Pitman sequence [as in (2.14)]. An analytical proof for (4.3) is non-trivial and rests heavily on some intricate properties of noncentral chi square distributions; we may refer to Kubokawa, Sen and Saleh (1987) for some of these details. The result is very prospective in the sense that it applies to almost the entire stock of nonparametric and robust estimators and their shrinkage versions; asymptotic multinormality [(2.7)] and the weak convergence of V_n (to V) suffice for this purpose.

5. RELATIVE PERFORMANCE OF PTE AND SRE

The SRE have been considered primarily on the ground of their dominance over their classical versions and other related minimaxity properties. On the other hand, the PTE have generally been proposed on more intuitive grounds; they are generally neither minimax nor admissible, but they are also generally not dominated by the SRE. Typically, a PTE of θ ,

like the SRE, is a testimator and is based on (T_n, L_n) . If $l_{n,\alpha}$ represents the critical value of L_n (for testing $H_0: \theta = \theta_0$) at the level of significance α ($0 < \alpha < 1$), then the PTE may be written as

$$T_n^{PT} = \theta_0 I(L_{n,\alpha}) + T_n I(L_n > l_{n,\alpha}), \quad (5.1)$$

where $I(A)$ stands for the indicator function of the set A . A PTE may be formulated in a more general way in terms of "restricted" and "unrestricted" versions of natural estimators which are weighed in the light of the sample evidence on the plausibility of the "restraints" governing the restricted estimates. In practice, generally, the significance level α is taken small (i.e., 0.05 or 0.10), although α may also be chosen in such a way that the PTE has a bounded shortcoming over a class of alternatives. A PTE generally performs better than a SRE when θ lies "close to" the pivot, θ_0 , although outside a close neighbourhood of θ_0 the PTE may have risk higher than that of the SRE or the classical version. Nevertheless, the risk of a PTE is not unbounded and for large deviations of θ from θ_0 , it performs closely to the SRE. Thus, generally, a PTE neither dominates a SRE nor it is dominated by a SRE. There may be some advantages for a PTE:

(i) For the dominance of a SRE, it is generally necessary the $p \geq 3$. On the other hand, even for $p = 1$ or 2 , a PTE may have better performance than the classical version in the neighbourhood of the pivot.

(ii) Like the SRE, a PTE generally possesses robustness (in terms of risk) over a range of θ .

(iii) Often we have a partial shrinkage model. For example, for a linear model: $\underline{X} = \underline{A} \underline{\beta} + \underline{e}$ where $\underline{\beta}$ is a p -vector, we may partition $\underline{\beta}' = (\underline{\beta}'_1, \underline{\beta}'_2)$, where the $\underline{\beta}_j$ are p_j -vector, $p_1 + p_2 = p$. Suppose that the

pivot relates to $\beta_2 = 0$, while β_1 may be quite arbitrary. We may refer to Sen and Saleh (1987) for some detailed accounts of related PTE and SRE based on M-statistics and M-estimators. In such a case, the PTE works out well for $p_2 \geq 1$ and $p \geq 1$. On the other hand, for the dominance of the SRE to hold (over the classical estimator of β), we not only need that $\min(p_1, p_2) \geq 3$, but also that some other conditions on the associated design matrix $A'A$ hold. Thus, the PTE may have a greater scope of adaptability than the SRE.

(iv) In general, the asymptotic distribution of a PTE (for Pitman type alternatives) is of much simpler form than that of a SRE. Further, the AR of a PTE may be computed under much less restrictive regularity conditions than in the case of a SRE. In terms of the ADR, the PTE and SRE are, however, quite comparable with respect to these regularity conditions.

(v) The PTE, as has been formulated in (5.1), may not depend on \underline{W} , whereas in the case of the SRE [in (2.12)], \underline{W} enters explicitly in the form of the estimator. Thus, when there is no general consensus on the choice of \underline{W} , a SRE may not be robust for variations of \underline{W} from a chosen one, while a PTE remains unaffected by this variation. Further, the ADR of a SRE depends explicitly on \underline{W} , and a SRE for a given choice of \underline{W} may not retain its minimaxity or other properties for other choices too. On the other hand, the performance characteristics of a PTE may also be judged by its generalized variance, while with respect to this generalized variance criterion, the dominance of a SRE may not remain intact for a general class of \underline{W} .

Based on all these considerations, it seems quite reasonable to attach good weights to PTE when the prior information on $\underline{\theta}$ is of uncertain nature: We may lean more towards the PTE when $\underline{\theta}$ is suspected to be close to the pivot, while for $\underline{\theta}$ away from the pivot, the SRE may be more attractive. In

any case, if the choice of \tilde{W} is not that unambiguous or if $\tilde{\theta}$ is too far away from the pivot then none of the PTE or SRE may differ perceptibly from their classical versions, and hence, there may not be any point in pursuing either of them.

6. GENERAL REMARKS

With the general principles laid down in Section 2 through 5, we now proceed to summarize the recent developments on SRE in some specific areas.

(i) General estimable parameters. Let $\tilde{\theta} = \tilde{\theta}(F)$ be a functional of a distribution function F and let \tilde{U}_n be the corresponding U-statistics (vector) and let \tilde{V}_n be the jackknifed estimator of the variance-covariance functional of \tilde{U}_n . For any given pivot $\tilde{\theta}_0$ and weight matrix \tilde{W} , the SRE of $\tilde{\theta}$ has been worked out in Sen (1984). The relevance of Pitman-alternatives and the computation of the AR (under a left truncation on L_n) were also stressed in this work.

(ii) Multivariate location model: General F . The model is similar to the multi-normal mean model, but the underlying distribution F is unknown and quite arbitrary. Robust R-estimators and M-estimators of location along with related rank and M-tests for the adequacy of the pivot were incorporated in the formulation of SRE; we may refer to Sen and Saleh (1985) and Saleh and Sen (1985a) for detailed discussions.

(iii) MLE: Multi-parameter case. The ADR of SRE based on MLE and related likelihood ratio test statistics has been studied and the asymptotic dominance results are unified in Sen (1986).

(iv) Multiple regression model: Subset pivot. Consider the usual linear model $\underline{X} = \underline{A} \underline{\beta} + \underline{e}$, where $\underline{\beta}' = (\underline{\beta}'_1, \underline{\beta}'_2)$, $\underline{\beta}_j$ is a p_j -vector, $j = 1, 2$; $p = p_1 + p_2$, $p_1 \geq 0$, $p_2 \geq 0$, and the error \underline{e} need not have normal distributions. The problem is to provide SRE of $\underline{\beta}_1$ (or $\underline{\beta}$) when the pivot

relates to $\beta_2 = 0$. The robustness of the LSE based SRE of β_1 and their ADR were studied by Saleh and Sen (1987a). The case of R-estimators needed a different approach for implementing the required asymptotics, and a systematic account of the related SRE is given in Saleh and Sen (1986b). M-estimators of regression parameters and related M-tests for the pivot have been incorporated by Sen and Saleh (1987) for studying the asymptotic properties of related SRE. An interesting feature of these studies is that for any effective dominance of the SRE, one not only needs that $\min(p_1, p_2) \geq 3$ but also that $A'A$ satisfies an intricate condition on its characteristic roots. Particular cases of practical importance of this model include the so called "parallelism problem", and related SRE based on R- and M-estimators were considered by Saleh and Sen (1985c, 1986a).

Though for each of these problems, the general line of attack is a common one (based on the concepts in Sections 2, 3, and 4), the actual manipulations for adapting the general asymptotics (for obtaining G^* in specific cases) depend very much on the actual problem and on the type of estimators employed. For this reason, the technicalities need to be considered in isolation for the diverse problems. In fact, the simple expression for the ADR of the SRE in (3.8) may not hold for the multiple regression model, and considerable manipulations are needed to yield comparable expressions. Also, in some cases, the weight matrix W can be related to the usual Mahalanobis distance and this simplifies the manipulations considerably. This is particularly true for the multiple regression model where $\underline{v} = v^2 \cdot \underline{c}$ for some given \underline{c} (p.d.) and v^2 is a scalar constant (possibly unknown and dependent on the type of estimators used). Here $\underline{W} = \underline{c}^{-1}$ is a natural choice, so that all the characteristic roots of $\underline{W} \underline{v}$ are equal to v^2 . In this case, (2.12) can be simplified as $\underline{\theta}_0 + (1 - c \underline{L}_n^{-1}) (\underline{T}_n - \underline{\theta}_0)$, and hence, the treatment of its ADR is often

quite simpler compared to the general case. But, this may not be generally true for the multivariate location model, where for different types of estimators, we would have different forms of v (which may not be proportional to each other), and relative performance of the SRE based on these different types of estimators may not be thus studied in a simple manner. There is a good scope for further work in this direction.

We conclude this section with some remarks on another related area in SRE where there has been a good spur of recent works: Sequential SRE. The James-Stein (1961) technique has been effectively extended to the sequential case (under normal distributional assumption) by Ghosh, Nickerson and Sen (1987); other references are cited there. Sen (1987a) has considered the asymptotic situation with the SRE based on MLE and likelihood ratio test in a sequential setup, and parallel results for sequential U-statistics and generalized U-statistics have also been obtained by Sen (1987b,c). More work in this line is on progress.

REFERENCES

- Berger, J.O. (1980), "A Robust Generalized Bayes Estimator and Confidence Region for a Multivariate Normal Mean", Annals of Statistics, 8, 716-761.
- Berger, J.O. (1985), Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.
- Ghosh, M., Nickerson, D.M. and Sen, P.K. (1987), "Sequential Shrinkage Estimation," Annals of Statistics, 15, 817-829.
- Hoeffding, W. (1948), "A Class of Statistics with Asymptotically Normal Distribution", Annals of Mathematical Statistics, 19, 293-325.
- Huber, P.J. (1981), "Robust Statistics", John Wiley, New York.
- James, W., and Stein, C. (1961), "Estimation with Quadratic Loss", Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability (Edited by L. LeCam and J. Neyman), University of California Press, Berkeley, 1, 361-379.
- Jurečková, J. (1977), "Asymptotic Relations of M-estimates and R-estimates in Linear Regression Models", Annals of Statistics, 5, 464-472.

- Kubokawa, T., Sen, P.K. and Saleh, A.K.M.E. (1987), "The Stein Paradox in the Pitman Closeness", Annals of Statistics (submitted).
- Pitman, E.J.G. (1937), "The 'Closest' Estimates of Statistical Parameters", Proceedings of the Cambridge Philosophical Society, 33, 212-217.
- Puri, M.L. and Sen, P.K. (1985), Nonparametric Methods in General Linear Models, John Wiley, New York.
- Saleh, A.K.M.E. and Sen, P.K. (1985a), "On Shrinkage M-estimators of Location Parameters", Communications in Statistics: Theory and Methods A14, 2313-2329.
- Saleh, A.K.M.E. and Sen, P.K. (1985b), "Nonparametric Shrinkage Estimators of Location in a Multivariate Simple Regression Model", Proceedings of the 4th Pannonian Conference in Mathematical Statistics, 311-324.
- Saleh, A.K.M.E. and Sen, P.K. (1985c), "Nonparametric Shrinkage Estimation in a Parallelism Problem", Sankhya, Series A47, 156-165.
- Saleh, A.K.M.E. and Sen, P.K. (1986a), "On Shrinkage Least Squares Estimation in a Parallelism Problem", Communications in Statistics: Theory and Methods, 15, 1587-1596.
- Saleh, A.K.M.E. and Sen, P.K. (1986b), "On Shrinkage R-estimation in a Multiple Regression Problem", Communications in Statistics: Theory and Methods 15, 2329-2244.
- Saleh, A.K.M.E. and Sen, P.K. (1986c), "On Shrinkage and Preliminary Test M-estimation in a Parallelism Problem", Joshi Festschrifts (Edited by I. McNeil et al), 2, 105-119.
- Saleh, A.K.M.E. and Sen, P.K. (1987a), "Relative performance of Stein-Rule and Preliminary Test Estimators in Linear Models: Least Squares Theory", Communications in Statistics: Theory and Methods, 16, 461-476.
- Sen, P.K. (1984), "A James-Stein Detour of U-Statistics", Communications in Statistics: Theory and Methods A13, 2725-2747.
- Sen, P.K. (1986), "On the Asymptotic Distributional Risks of Shrinkage and Preliminary Test Versions of Maximum Likelihood Estimators", Sankhya, Series A 48, 354-371.
- Sen, P.K. (1987a), "Sequential Stein-Rule Maximum Likelihood Estimators: General Asymptotics", Statistical Decision Theory and Related Topics (Edited by S.S. Gupta and J.O. Berger), Springer-Verlag, in press.
- Sen, P.K. (1987b), "Sequential Shrinkage U-Statistics: General Asymptotics," Revista Brasileira de Probabilidade e Estatística 1, in press.
- Sen, P.K. (1987c), "Sequential Stein-Rule Generalized U-Statistics", Proceedings of the Sino-American Statistical Meeting, Beijing: Contributed Papers Volume, 381-386.

- Sen, P.K. and Saleh, A.K.M.E. (1985), "On Some Shrinkage Estimators of Multivariate Locations", Annals of Statistics, 13, 272-281.
- Sen, P.K. and Saleh, A.K.M.E. (1987), "On Preliminary Test and Shrinkage M-Estimation in Linear Models", Annals of Statistics 15, in press.
- Stein, C. (1981), "Estimation of the Mean of a Multivariate Normal Distribution", Annals of Statistics, 9, 1135-1151.