SEQUENTIAL STEIN-RULE MAXIMUM LIKELIHOOD ESTIMATION: GENERAL ASYMPTOTICS

by

P.K. Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1802

May 1986

SEQUENTIAL STEIN-RULE MAXIMUM LIKELIHOOD ESTIMATION : GENERAL ASYMPTOTICS *

PRANAB KUMAR SEN

University of North Carolina, Chapel Hill

For the multi-parameter minimum risk sequential estimation problem, Stein-rule and preliminary test versions of maximum likelihood estimators are considered. Under the usual regularity conditions, general asymptotics on the distributional risks of these estimators are studied, and these are incorporated in the study of the related asymptotic risk-efficiency results.

1. INTRODUCTION

In the classical parametric estimation theory, the *maximum likelihood estimators (MLE)* play a central role. Asymptotic properties and optimality of the MLE have been studied extensively in the literature; we may refer to the classical text by Ibragimov and Hasminskii (1981) for an excellent account of this asymptotic theory. In a general multi-parameter model, however, the classical MLE may not be fully *risk-efficient* (with respect to some specific *risk functions*), and suitable *shrinkage versions* dominating the classical MLE (known as the *Stein-rule estimators)* have been worked out for various models ; we may refer to Berger (1985) for a detailed account of this theory (mostly, for the multi-normal family of distributions) . In this context, both the *preliminary test estimation (PTE)* and Stein-rule estimation theory provide alternative estimators possessing some desirable properties. The Stein-rule estimators generally dominates the classical versions, but neither the PTE nor the Stein-rule estimator dominates the other.

AMS Subject Classifications.: 62F12, 62L12 , 62L99.

Key Words & Phrases : Asymptotic distributional risk; likelihood ratio statistics; local alternatives; preliminary test estimator; risk; risk-efficiency; shrinkage estimation; stopping rule.

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

RUNNING TITLE : Sequential Stein-rule MLE.

For the MLE, the asymptotic theory of PTE has been considered in a general framework in Sen (1979). The asymptotic distribution theory of the Stein-rule MLE has also been recently studied [Sen (1986)] in a systematic manner, and the relative performance of these versions of the MLE has been considered. All these relate to a traditional non-sequential (i.e., fixed sample size) setup. In the *minimum risk estimation* problem, generally, the risk function involves other unknown (nuisance) parameters, and hence, *sequential estimation* procedures are generally advocated to provide (at least, asymptotically) efficient solutions. For the multivariate normal mean vector, sequential shrinkage estimation procedures have recently been considered by Ghosh, Nickerson and Sen (1986). Under the protection of the multi-normality, these procedures exhibit dominance of the sequential shrinkage estimators over the classical sequential estimator, considered by Ghosh, Sinha and Mukhopadhyay (1976) and others. Improvements over an alternative shrinkage (sequential) procedure by Takada (1984) and two-stage sequential Stein-rule estimators by Ghosh and Sen (1983) have also been discussed there. The current study is centered around the Stein-rule sequential MLE in a general setup, and, extends the results of Sen (1986) to the sequential case. The main focus is on the asymptotic risk-efficiency of the sequential Stein-rule MLE , and in this context, a comparative picture of the sequential PT MLE has also been drawn. Our study is greatly facilitated by the incorporation of the notion of *asymptotic distributional risk (ADR)* of plausible estimators, and in this vein, the results considered here are the direct extensions of the asymptotic efficiency results on MLE (in the fixed sample or sequential case) to the shrinkage estimation problem.

Along with the preliminary notions, the proposed PT MLE and Stein-rule MLE are considered in Section 2 ; the notion of ADR is also introduced in the same section. Section 3 deals with the ADR of the different versions of the MLE in the sequential case. Allied asymptotic risk-efficiency results are presented in Section 4. Some general remarks are made in the concluding section.

2. THE PRELIMINARY NOTIONS AND THE PROPOSED SEQUENTIAL ESTIMATORS

Let X_i, $i \geq 1$ be a sequence of independent and identically distributed (i.i.d.) random vectors (r.v.) with a distribution function (d.f.) $F(x,\underset{\sim}{\theta})$, $x \in E^m$ and $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_r)' \in \Omega \quad E^r$, for some $m \geq 1$ and $r \geq 1$. We assume that $F(x,\underset{\sim}{\theta})$ admits a density function $f(x,\underset{\sim}{\theta})$ (with respect to some sigma-finite measure μ). Then the *log-likelihood function* is defined by

$$\log L_n(\underset{\sim}{\theta}) = \Sigma_{i=1}^{n} \log f(X_i, \underset{\sim}{\theta}) \ , \quad \underset{\sim}{\theta} \in \Omega \ , \ n \geq 1. \tag{2.1}$$

We define the restricted parameter space ω as

$$\omega = \{\underset{\sim}{\theta}: \underset{\sim}{h}(\underset{\sim}{\theta}) = (h_1(\underset{\sim}{\theta}),\ldots,h_p(\underset{\sim}{\theta})) = \underset{\sim}{0}\}, \text{ for some } p \leq r \ . \tag{2.2}$$

Then, an *unrestricted MLE (UMLE)* and a *restricted MLE (RMLE)* of the true parameter $\underset{\sim}{\theta}$, denoted by $\tilde{\underset{\sim}{\theta}}_n$ and $\hat{\underset{\sim}{\theta}}_n$, respectively, are defined as the solutions of

$$\log L_n(\tilde{\underset{\sim}{\theta}}_n) = \sup_{\underset{\sim}{\theta} \in \Omega} \log L_n(\underset{\sim}{\theta}) \ ; \quad \log L_n(\hat{\underset{\sim}{\theta}}_n) = \sup_{\underset{\sim}{\theta} \in \omega} \log L_n(\underset{\sim}{\theta}). \tag{2.3}$$

We assume that Ω is a convex, compact subset of E^r, and $\log f(x,\underset{\sim}{\theta})$ is almost everywhere (a.e.) differentiable with respect to $\underset{\sim}{\theta}$(at least twice), these derivatives are dominated by some integrable functions, and the second order derivative matrix has the continuity property in the first mean. Further, the density f has a finite Kullback-Leibler information. Let then

$$\underset{\sim}{B}_{\underset{\sim}{\theta}} = ((\iint_{E^m} (\partial/\partial\theta_j)\log f(x,\underset{\sim}{\theta})(\partial/\partial\theta_\ell)\log f(x,\underset{\sim}{\theta})dF(x,\underset{\sim}{\theta})))_{j,\ell=1,\ldots,r}. \tag{2.4}$$

Then, we assume that $\underset{\sim}{B}_{\underset{\sim}{\theta}}$ is continuous in $\underset{\sim}{\theta}$ in some neighbourhood of the true parameter point $\underset{\sim}{\theta}_o$ and $\underset{\sim}{B}_{\underset{\sim}{\theta}_o}$ is positive definite (p.d.). We also assume that $\underset{\sim}{h}(\underset{\sim}{\theta})$ possesses continuous first and second order derivatives with respect to $\underset{\sim}{\theta}$, for every $\underset{\sim}{\theta} \in \Omega$. Let then

$$\underset{\sim}{H}_{\underset{\sim}{\theta}} = (((\partial/\partial\underset{\sim}{\theta})\underset{\sim}{h}(\underset{\sim}{\theta}))) \quad (\text{ of order } r\times p), \tag{2.5}$$

and assume that

$$\underset{\sim}{H}_{\underset{\sim}{\theta}_o} \text{ is of rank } p \ (\leqslant r) \ . \tag{2.6}$$

Further, we define the *likelihood score function* as

$$\underset{\sim}{\Lambda}_n(\underset{\sim}{\theta}) = n^{-\frac{1}{2}}(\partial/\partial\underset{\sim}{\theta})\log L_n(\underset{\sim}{\theta}) \ ; \quad \underset{\sim}{\Lambda}_n^o = \underset{\sim}{\Lambda}_n(\underset{\sim}{\theta}_o) \ . \tag{2.7}$$

Regularity conditions pertaining to the asymptotic normality and optimality of the classical MLE have been studied in increasing generality in Cramér (1946), Huber (1967), Hájek (1970), LeCam (1970), Inagaki (1970) and others. We shall mainly follow here the treatment of Sen (1979, 1986) and omit details of these discussions. Then, we have for large n,

$$n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_o) = B_{\theta_o}^{-1} \Lambda_n^o + o_p(1) \overset{\mathcal{D}}{\sim} \mathcal{N}(0, B_{\theta_o}^{-1}). \tag{2.8}$$

Suppose now that having recored $X_1,\dots,X_n$, the loss incurred in estimating θ by an estimator T_n is

$$u(T_n, \theta; A, c, n) = (T_n - \theta)'A(T_n - \theta) + cn, \tag{2.9}$$

where A is a known $r \times r$ p.d. matrix and $c(>0)$ is the known cost per unit sample. Then, the <u>risk</u> is given by

$$R_n(c,A) = Eu(T_n, \theta; A, c, n) = \mathrm{Tr}(A[E\{(T_n - \theta)(T_n - \theta)'\}]) + cn. \tag{2.10}$$

For the UMLE, we gather from (2.8) and (2.10) that under appropriate regularity conditions, for large n, the risk is given by

$$R_n(\tilde{\theta}_n; c, A) = n^{-1}\mathrm{Tr}(AB_{\theta_o}^{-1}) + cn + o(n^{-1}). \tag{2.11}$$

Thus, if B_{θ_o} were known and c is small, the risk in (2.11) would have a minimum value $(2cn_c^*)$, where the *optimal sample size* n_c^* is given by

$$n_c^* = [\mathrm{Tr}(AB_{\theta_o}^{-1})/c]^{\frac{1}{2}}. \tag{2.12}$$

In practice, neither θ_o nor the information matrix B_{θ_o} is generally known. Hence, this optimal sample size is not known. We may, however, define

$$\hat{B}_n = -n^{-1}(\partial^2/\partial\theta\partial\theta')\log L_n(\theta)\Big|_{\theta = \tilde{\theta}_n}, \tag{2.13}$$

and formulate a *stopping variable* N_c, by letting

$$N_c = \inf\{n \geq n_o : n^2 \geq \{\mathrm{Tr}(A\hat{B}_n^{-1}) + n^{-a}\}/c\}, \quad c > 0, \tag{2.14}$$

where $a(>0)$ is a suitable positive number and n_o is a positive integer $(> r)$. For the multi-normal mean estimation problem, this type of stopping number has already been considered by Ghosh, Sinha and Mukhopadhyay (1976) and others. Then a sequential version of the UMLE is given by $\tilde{\theta}_{N_c}$. We shall see later on that under appropriate conditions, the asymptotic risk of this sequential UMLE may also be approximated by $(2cn_c^*)$, so that it is asymptotically risk-efficient.

The RMLE, defined in (2.3), generally, behaves better (at least asymptotically) than the UMLE when $\underset{\sim}{\theta}_o$ is contained in ω. However, this picture does not hold when $\underset{\sim}{\theta}_o \notin \omega$. Towards this study, we assume that the following matrix (of order $(p+r)\times(p+r)$)

$$\begin{pmatrix} \underset{\sim}{B}_{\underset{\sim}{\theta}_o} & -\underset{\sim}{H}_{\underset{\sim}{\theta}_o} \\ -\underset{\sim}{H}'_{\underset{\sim}{\theta}_o} & \underset{\sim}{0} \end{pmatrix} \quad \text{is p.d. ,} \tag{2.15}$$

and let

$$\begin{pmatrix} \underset{\sim}{P}_{\underset{\sim}{\theta}_o} & \underset{\sim}{Q}_{\underset{\sim}{\theta}_o} \\ \underset{\sim}{Q}'_{\underset{\sim}{\theta}_o} & \underset{\sim}{R}_{\underset{\sim}{\theta}_o} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{B}_{\underset{\sim}{\theta}_o} & -\underset{\sim}{H}_{\underset{\sim}{\theta}_o} \\ -\underset{\sim}{H}'_{\underset{\sim}{\theta}_o} & \underset{\sim}{0} \end{pmatrix}^{-1} . \tag{2.16}$$

Then [viz., Sen (1986)], for $\underset{\sim}{\theta}_o \in \omega$, as $n \to \infty$,

$$n^{\frac{1}{2}}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o) = \underset{\sim}{P}_{\underset{\sim}{\theta}_o} \underset{\sim}{\Lambda}_n^o + \underset{\sim}{o}_p(1) \;\; \underset{\sim}{\mathcal{D}} \;\; \mathcal{N}(\underset{\sim}{0} , \underset{\sim}{P}_{\underset{\sim}{\theta}_o}) , \tag{2.17}$$

where

$$\underset{\sim}{I}_r - \underset{\sim}{B}_{\underset{\sim}{\theta}_o} \underset{\sim}{P}_{\underset{\sim}{\theta}_o} \quad \text{is idempotent (of rank } p(\leqslant r) \text{) .} \tag{2.18}$$

Thus, for $\underset{\sim}{\theta}_o \in \omega$, the asymptotic risk of the RMLE is given by

$$\begin{aligned} R(\hat{\underset{\sim}{\theta}}_n ; c,\underset{\sim}{A}) &= n^{-1}\mathrm{Tr}(\underset{\sim}{A}\underset{\sim}{P}_{\underset{\sim}{\theta}_o}) + cn + o(n^{-1}) \\ &= n^{-1}\mathrm{Tr}(\underset{\sim}{A}\underset{\sim}{B}^{-1}_{\underset{\sim}{\theta}_o}) - n^{-1}\mathrm{Tr}(\underset{\sim}{A}\underset{\sim}{B}^{-1}_{\underset{\sim}{\theta}_o}(\underset{\sim}{I}-\underset{\sim}{B}_{\underset{\sim}{\theta}_o}\underset{\sim}{P}_{\underset{\sim}{\theta}_o})) + cn + o(n^{-1}) \\ &= R(\tilde{\underset{\sim}{\theta}}_n ;c,\underset{\sim}{A}) - n^{-1}\mathrm{Tr}(\underset{\sim}{A}\underset{\sim}{B}^{-1}_{\underset{\sim}{\theta}_o}(\underset{\sim}{I}-\underset{\sim}{B}_{\underset{\sim}{\theta}_o}\underset{\sim}{P}_{\underset{\sim}{\theta}_o})) + o(n^{-1}) \\ &\leq R(\tilde{\underset{\sim}{\theta}}_n ;c,\underset{\sim}{A}) + o(n^{-1}). \end{aligned} \tag{2.19}$$

This picture may not hold when $\underset{\sim}{\theta}_o \notin \omega$. We shall make more comments on that in a later section. With the stopping time N_c defined in (2.14), we denote the sequential version of the RMLE by $\hat{\underset{\sim}{\theta}}_{N_c}$.

For testing the null hypothesis H_0: $\underset{\sim}{\theta}_o \in \omega$, the classical (log-) likelihood ratio test statistic is given by

$$\mathcal{L}_n = -2\log\{ L_n(\hat{\underset{\sim}{\theta}}_n)/L_n(\tilde{\underset{\sim}{\theta}}_n)\} . \tag{2.20}$$

The null hypothesis is then accepted or rejected according as $\mathcal{L}_n$ is $\leq$ or $>$ $\ell_{n,\alpha}$, the level $\alpha(0 < \alpha < 1)$ critical value of $\mathcal{L}_n$; $\ell_{n,\alpha}$ is generally taken as $\chi^2_{p,\alpha}$, the upper $100\alpha\%$ point of the central chi square d.f. with p degrees of freedom (DF). The PT MLE $\hat{\theta}_n^{PT}$ is defined then as

$$\hat{\theta}_n^{PT} = \begin{cases} \hat{\theta}_n \,, & \text{if } \mathcal{L}_n \le \ell_{n,\alpha} \\ \tilde{\theta}_n \,, & \text{if } \mathcal{L}_n > \ell_{n,\alpha} \,. \end{cases} \tag{2.21}$$

With the stopping number N_c defined as in (2.14), we denote the sequential version of the PT MLE by $\hat{\theta}_{N_c}^{PT}$.

To introduce the Stein-rule MLE, we consider the usual James-Stein (1961) rule and define [as in Sen (1986)]

$$\hat{\theta}_n^S = \tilde{\theta}_n + (\hat{\theta}_n - \tilde{\theta}_n)(p-2)/\mathcal{L}_n \ . \tag{2.22}$$

It is possible to consider a more general form of shrinkage MLE (and we shall comment on that in the concluding section). However, for the risk-efficiency picture, this simple form provides a basis for other complicated forms, and hence, for the sake of simplicity of presentation, we shall mainly consider this form. The sequential version of (2.22) with the stopping number defined in (2.14) is denoted by $\hat{\theta}_{N_c}^S$.

While all these estimators are comparable under the null hypothesis H_0: $\theta_o \in \omega$, when θ_o lies outside ω, asymptotically , the PT MLE and the Stein-rule MLE both become equivalent (in probability) to the UMLE. Thus, to have meaningful comparisons, we confine ourselves to local alternatives, for which the different versions have possibly non-equivalent asymptotic risks. Note that n_c^* defined in (2.12) is $O(c^{-\frac{1}{2}})$ and it goes to ∞ as $c \downarrow 0$. To retain the asymptotic structure of the proposed sequential versions of the MLE , we therefor consider the case where c goes to 0, and , for this case, we consider the following sequence $\{K_c\}$ of local alternatives :

$$K_c \ : \ h(\theta_o) = c^{\frac{1}{4}} \gamma \ , \ \gamma \text{ a real vector in } E^p, \ c \downarrow 0 \ . \tag{2.23}$$

It is well known that in the usual (parametric or nonparametric) case, a shrinkage estimator leads to an effective reduction of risk only when the actual parameter point lies near the pivot, and the choice of (2.22) is geared to this consideration too. 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). (2.23) is an analogue of the usual Pitman alternative in the sequential case.

In passing, we may note that we can conceive of a sequence $\{\theta_o^{(c)}\}$ of parametric points, such that $h(\theta_o^{(c)}) = 0$ and

$$\lim_{c\downarrow 0} c^{-1/4} (\theta_o^{(c)} - \theta_o) = \gamma^* \text{ exists, where } \gamma = H_{\theta_o} \gamma^* . \tag{2.24}$$

In general, the MLE are non-linear functions of the sample observations, and hence, the computation of the moment in (2.10) may need additional regularity conditions. On the top of that in the PT MLE and the Stein-rule MLE, the likelihood ratio statistic introduces further complications for such moment computations. To overcome this problem, we take recourse to the *asymptotic distributional risk (ADR)* of the sequential versions of the MLE. For a suitable sequential estimator $\theta^*_{N_c}$ of θ_o, we denote the asymptotic d.f. (under $\{K_c\}$ in (2.23)) by

$$G^*(x) = \lim_{c\downarrow 0} P\{ c^{-1/4} (\theta^*_{N_c} - \theta_o) \le x \mid K_c \} , \tag{2.25}$$

and assume that G^* exists and is non-degenerate. Then, corresponding to the loss function in (2.9), the ADR of $\theta^*_{N_c}$, at the point θ_o, is defined by

$$\begin{aligned} R^*(\theta^*_{N_c} ;c,A) &= c^{1/2}[\mathrm{Tr}(A[\textstyle\iint xx'dG^*(x)]) + \lim_{c\downarrow 0} c^{1/2}\tilde{E}(N_c \mid K_c)] \\ &= c^{1/2}[\mathrm{Tr}(AV^*) + \xi^*] , \text{ say} , \end{aligned} \tag{2.26}$$

where V^* stands for the mean product matrix for the d.f. G^* and $\tilde{E}(.)$ refers to the expected value computed from the asymptotic distribution of $c^{1/4}(N_c - n_c^*)$ under $\{K_c\}$, so that $\xi^* = \lim_{c\downarrow 0} c^{1/2}\tilde{E}(N_c \mid K_c) = [\mathrm{Tr}(AB^{-1}_{\theta_o})]^{1/2}$. The risk thus computed from the asymptotic distributions of the (sequential) estimator as well as the stopping number provides a meaningful and adaptable way for comparing the performance characteristics of the proposed sequential versions of the MLE under the usual regularity conditions. For parallel results in the non-sequential case, we may refer to Sen (1986).

3. ADR OF THE PROPOSED SEQUENTIAL VERSIONS OF THE MLE

Keeping in mind the *Mahalanobis distance* for the UMLE and looking at (2.8), it seems quite plausible to choose $A = B_{\theta_o}$. Of course, in (2.14), A is assumed to be a given p.d. matrix, while B_{θ_o} is generally unknown. This apparent anomaly may easily be eliminated by considering an arbitrary A. However, since our basic goal is to study the related *asymptotic risk-efficiency (ARE)* results (in Section 4), this particular choice of $\underset{\sim}{A}$ would greatly simplify the mathematical manipulations. However, the results with minor modifications hold for an arbitrary $\underset{\sim}{A}$ too.

First, we show that under the assumed regularity conditions, as $c \downarrow 0$,

$$(c/r)^{\frac{1}{2}} N_c \rightarrow 1 \text{ almost surely (a.s.)} . \tag{3.1}$$

We denote the characteristic roots of $\underset{\sim}{B}_{\underset{\sim}{\theta}_o}^{-1}\hat{\underset{\sim}{B}}_n$ by b_{nj} , $j=1,\ldots,r$, respectively, so that $\mathrm{Tr}(\underset{\sim}{B}_{\underset{\sim}{\theta}_o} \underset{\sim}{B}_n^{-1}) = \Sigma_{j=1}^r b_{nj}^{-1}$. Then, note that by (2.14), for every $c > 0$ and $n \geq n_o$,

$$P\{ N_c > n\} = P\{ cm^2 - m^{-a} \leq \Sigma_{j=1}^r b_{mj}^{-1} , \text{ for every } m \in [n_o,n] \}$$

$$\leq P\{ cn^2 - n^{-a} \leq \Sigma_{j=1}^r b_{nj}^{-1} \}, \tag{3.2}$$

where we note that $(n_c^*)^2 \sim (r/c)$, as $c \downarrow 0$. Then, for every $\varepsilon > 0$, whenever $n \geq n_c^*(1 + \varepsilon)$, $cn^2 - n^{-a} > r(1+2\varepsilon) + o(1)$. Thus, if we are able to show that

$$\max\{ |b_{nj} - 1| : 1\leq j\leq r \} \rightarrow 0 \text{ a.s., as } n \rightarrow \infty , \tag{3.3}$$

then, $P\{N_c > n_c^*(1+ \varepsilon)\} \rightarrow 0$ as $c \downarrow 0$ (for every $\varepsilon > 0$). A similar treatment holds for the lower tail $P\{N_c < n_c^*(1-\varepsilon)\}$. Thus, the problem reduces to showing that (3.3) holds. Towards this note that on defining $\hat{\underset{\sim}{B}}_n^o$ as in (2.13) with $\tilde{\underset{\sim}{\theta}}_n$ being replaced by $\underset{\sim}{\theta}_o$, we have $\hat{\underset{\sim}{B}}_n = \hat{\underset{\sim}{B}}_n^o + (\hat{\underset{\sim}{B}}_n - \hat{\underset{\sim}{B}}_n^o)$, so that on $[||\tilde{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_o|| < \delta]$, $\delta > 0$,

$$||\hat{\underset{\sim}{B}}_n - \hat{\underset{\sim}{B}}_n^o || \leq \sup_{||\underset{\sim}{t}||<\delta} \left|\left| \begin{array}{l}\text{difference of the second derivative} \\ \text{matrices evaluated at } \underset{\sim}{\theta}_o \text{ and } \underset{\sim}{\theta}_o + \underset{\sim}{t}\end{array} \right|\right|$$

$$\leq n^{-1}\Sigma_{i=1}^n \{ \sup_{||\underset{\sim}{t}||<\delta} ||(\partial^2/\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}')\log f(X_i,\underset{\sim}{\theta})\Big|_{\underset{\sim}{\theta}_o} - (\partial^2/\partial\underset{\sim}{\theta}\partial\underset{\sim}{\theta}')\log f(X_i,\underset{\sim}{\theta})\Big|_{\underset{\sim}{\theta}_o+\underset{\sim}{t}} || \}$$

$$= n^{-1}\Sigma_{i=1}^n U_i(\delta), \text{ say,} \tag{3.4}$$

where the $U_i(\delta)$ are independent r.v., and, by the assumed continuity property (in the first mean) of the second order derivative matrix , we have

$$E[U_i(\delta) \mid \underset{\sim}{\theta}_o] \downarrow 0 \text{ as } \delta \downarrow 0. \tag{3.5}$$

Now, by the Khintchine strong law of large numbers and (3.5), we conclude that $n^{-1}\Sigma_{i=1}^n U_i(\delta)$ converges a.s. to 0 as $n \rightarrow \infty$ (and δ is made to converge to 0). On the other hand, the assumed regularity conditions insure the strong consistency of the UMLE, so that $\tilde{\underset{\sim}{\theta}}_n \rightarrow \underset{\sim}{\theta}_o$ a.s., as $n \rightarrow \infty$. Thus, $\hat{\underset{\sim}{B}}_n - \hat{\underset{\sim}{B}}_n^o \rightarrow \underset{\sim}{0}$ a.s., as $n \rightarrow \infty$.

On the other hand, by the Khintchine strong law of large numbers, as $n \to \infty$, $\hat{B}^o_n = -\frac{1}{n}\Sigma_{i=1}^n (\partial^2/\partial\theta\partial\theta')\log f(X_i,\theta)|_{\theta_o}$ converges a.s. to its expectation i.e., B_{θ_o} . Therefore , $\hat{B}_n \to B_{\theta_o}$ a.s., as $n \to \infty$, so that $B^{-1}_{\theta_o}\hat{B}_n \to I_r$ a.s., as $n \to \infty$. This ensures that all the charachteristic roots of $B^{-1}_{\theta_o}\hat{B}_n$ converge a.s. to 1 as $n \to \infty$, and this proves (3.3).

Now, in the context of repeated significance testing, <u>local asymptotic Wiener</u> property of the MLE's (in the unrestricted as well as restricted cases) and the associated likelihood ratio test statistics have been studied in detail by So and Sen (1981) and others. A direct consequence of this local asymptotic Wiener property is that for each of these sequence of estimators and statistics , the classical Anscombe (1952) condition (on uniform continuity in probability with respect to $\{n^{-\frac{1}{2}}\}$) holds. As such, by an appeal to (3.1) and the results in Sen (1986) [dealing with the non-sequential case], we conclude that under the assumed regularity conditions, as $c \downarrow 0$,

$$c^{-\frac{1}{4}}||\ \tilde{\theta}_{N_c} - \tilde{\theta}_{n^*_c}\ || \to 0 \text{, in probability,} \tag{3.6}$$

$$c^{-\frac{1}{4}}||\ \hat{\theta}_{N_c} - \hat{\theta}_{n^*_c}\ || \to 0 \text{, in probability ,} \tag{3.7}$$

$$\mathcal{L}_{N_c} - \mathcal{L}_{n^*_c} \to 0\text{, in probability.} \tag{3.8}$$

Note that (3.6)-(3.8) hold under $\{K_c\}$ (and H_0 as well). As such, using (2.21), (2.22), (3.6)-(3.8) and the Slutzky theorem, it follows that as $c \downarrow 0$,

$$c^{-\frac{1}{4}}\ ||\ \hat{\theta}^{PT}_{N_c} - \hat{\theta}^{PT}_{n^*_c}\ || \to 0 \text{, in probability,} \tag{3.9}$$

$$c^{-\frac{1}{4}}\ ||\ \hat{\theta}^{S}_{N_c} - \hat{\theta}^{S}_{n^*_c}\ || \to 0\text{, in probability.} \tag{3.10}$$

This shows that for the proposed sequential versions of the MLE, the asymptotic distribution theory agrees with the non-sequential counterparts when for the sample size n, we take $n = n^*_c$. This asymptotic equivalence makes it easier for us to use the results in Sen (1986), and with the modifications in the sample size n^*_c and the parallel formulation of the local alternatives, we arrive at the following ADR results. Note that in these expressions, we have made use of the choice $A = B_{\theta_o}$.

$$R^*(\tilde{\theta}_{N_c};c,B_{\theta_o}) = 2(\ cr\ )^{\frac{1}{2}}\ ,\ \forall\ \gamma \in E^p\ . \tag{3.11}$$

$$R^*(\hat{\theta}_{N_c};c, B_{\theta_o}) = (c/r)^{\frac{1}{2}}[\,\mathrm{Tr}(B_{\theta_o}P_{\theta_o}) + (\gamma^{*'}B_{\theta_o}\gamma^*)] + (cr)^{\frac{1}{2}}$$

$$= (cr)^{\frac{1}{2}}[\,2 - r^{-1}\{\,\mathrm{Tr}(I - B_{\theta_o}P_{\theta_o}) - (\gamma^{*'}B_{\theta_o}\gamma^*)\}\,]. \qquad (3.12)$$

$$R^*(\hat{\theta}^{PT}_{N_c};c,B_{\theta_o}) = (cr)^{\frac{1}{2}} + (c/r)^{\frac{1}{2}}\{\,r - \mathrm{Tr}(I-B_{\theta_o}P_{\theta_o})\Pi_{p+2}(\chi^2_{p,\alpha};\Delta)$$
$$\Delta[\Pi_{p+2}(\chi^2_{p,\alpha};\Delta) - \Pi_{p+4}(\chi^2_{p,\alpha};\Delta)\,]\,\}$$

$$= (cr)^{\frac{1}{2}}\{\,2 - r^{-1}(\mathrm{Tr}(I-B_{\theta_o}P_{\theta_o})\Pi_{P+2}(\chi^2_{p,\alpha};\Delta) +$$
$$\Delta[\,\Pi_{p+2}(\chi^2_{p,\alpha};\Delta) - \Pi_{p+4}(\chi^2_{p,\alpha};\Delta)])\}, \qquad (3.13)$$

where $\Delta = (\gamma^{*'}B_{\theta_o}\gamma^*)$ and $\Pi_q(x;\delta)$ stands for the non-central chi square d.f. with q DF and non-centrality parameter δ . Finally,

$$R^*(\hat{\theta}^S_{N_c};c,B_{\theta_o}) = (cr)^{\frac{1}{2}} + (c/r)^{\frac{1}{2}}\{r + (p-2)^2[\,rE(\chi^{-4}_{p+2}(\Delta)) + \Delta E(\chi^{-4}_{p+4}(\Delta))]$$
$$- 2(p-2)rE(\chi^{-2}_{p+2}(\Delta)) + 4(p-2)\Delta E(\chi^{-4}_{p+4}(\Delta))\,\}$$

$$= (cr)^{\frac{1}{2}}\,\{2 - 2(p-2)E(\chi^{-2}_{p+2}(\Delta)) + 4r^{-1}(p-2)\Delta E(\chi^{-4}_{p+4}(\Delta))$$
$$+ (p-2)^2[\,E(\chi^{-4}_{p+2}(\Delta)) + r^{-1}\Delta E(\chi^{-4}_{p+4}(\Delta)]\,\}, \qquad (3.14)$$

where $E(\chi_q^{-2r}(\delta))$ stands for the negative rth moment of a random variable having the d.f. $\Pi_q(x,\delta)$. We shall find these expressions convenient for further analysis in the next section.

4. ASYMPTOTIC DOMINANCE AND RISK-EFFICIENCY RESULTS.

In the light of the ADR results in Section 3, we like to study the asymptotic dominance of some versions of the sequential MLE over the others, and also the related asymptotic risk-efficiency results will be considered. In this context, the key role is played by the sequence of local alternatives in (2.23), and this is in line with the usual asymptotic theory of sequential point estimation [as has been treated in detail in Chapter 10 of Sen (1981)]. Comparing (3.11) and (3.14), we obtain that

$$R^*(\hat{\theta}^S_{N_c};c,B_{\theta_o})/R^*(\tilde{\theta}_{N_c};c,B_{\theta_o}) = 1 - (p-2)E(\chi^{-2}_{p+2}(\Delta)) + \tfrac{1}{2}(p-2)^2E(\chi^{-4}_{p+2}(\Delta))$$
$$+\tfrac{1}{2}(p-2)(p+2)r^{-1}\Delta E(\chi^{-4}_{p+4}(\Delta)). \qquad (4.1)$$

Under the null hypothesis H_0 , we have $\Delta = 0$ and (4.1) reduces to $1 - (p-2)/p + \tfrac{1}{2}(p-2)^2/p(p-2) = \tfrac{1}{2}+p^{-1}(< 1)$ for every $p > 2$; the larger is the value of p, the greater is the reduction in the ADR of the Stein-rule MLE (under H_0). It also follows from the results of Sen (1986) [after noting that r > p > 2] that the right hand side of (4.1) is less than 1 for all $\underset{\sim}{\gamma}^*$ such that $\Delta < \infty$, and as $\Delta \to \infty$, (4.1) converges to 1. This shows that in the light of the ADR, the Stein-rule sequential MLE dominates the usual sequential MLE. Comparing (3.12) and (3.11), we obtain that

$$R^*(\hat{\underset{\sim}{\theta}}_{N_c};c,\underset{\sim}{B}_{\theta_o})/R^*(\tilde{\underset{\sim}{\theta}}_{N_c};c,\underset{\sim}{B}_{\theta_o}) \text{ is } \lesseqgtr 1 \text{ according as } \Delta \lesseqgtr \mathrm{Tr}(\underset{\sim}{I}-\underset{\sim}{B}_{\theta_o}\underset{\sim}{P}_{\theta_o}). \quad (4.2)$$

Thus, under the null hypothesis (and for smaller values of Δ), the restricted MLE (in the sequential case) has a smaller ADR than the unrestricted one, although for $\underset{\sim}{\gamma}^*$ away from the origin, the opposite picture holds. Thus, none of the UMLE and RMLE (in the sequential case) dominates the other (in the light of the ADR). The RMLE has the discouraging picture that as $\Delta \to \infty$, the ADR in (3.12) tends to $+\infty$, and this unbounded risk makes it a poor competitor.

Comparing (3.12) and (3.14), we obtain that under H_0,

$$R^*(\hat{\underset{\sim}{\theta}}^S_{N_c};c,\underset{\sim}{B}_{\theta_o})/R^*(\hat{\underset{\sim}{\theta}}_{N_c};c,\underset{\sim}{B}_{\theta_o}) = \frac{2-2(p-2)/p+(p-2)^2/p(p-2)}{2-[\mathrm{Tr}(\underset{\sim}{I}-\underset{\sim}{B}_{\theta_o}\underset{\sim}{P}_{\theta_o})]/r}$$
$$= [r(p+2)]/[r(p+2)-\{p^2-r(p-2)\}]. \quad (4.3)$$

Thus, (4.3) exceeds one whenever $p^2 > r(p-2)$, and the opposite inequality holds when $p^2 < r(p-2)$. The situation is very much comparable to the non-sequential case treated in Sen (1986), and hence, avoiding details, we conclude that usually, under H_0, the sequential Stein-rule MLE may not dominate the RMLE. When the null hypothesis is not true, the ADR of the RMLE increases with Δ, while the Stein-rule MLE has a bounded ADR, and hence, at least outside a closed ellipsoid (in $\underset{\sim}{\gamma}^*$) with $\underset{\sim}{0}$ as the origin, the Stein-rule MLE would have a smaller ADR than the RMLE. Thus, though the Stein-rule sequential MLE may not dominate the RMLE, it possesses a more robust ARE picture.

Similarly, if we compare (3.13) and (3.14), we obtain that under H_0,

$$R^*(\hat{\underset{\sim}{\theta}}^S_{N_c};c,\underset{\sim}{B}_{\theta_o})/R^*(\hat{\underset{\sim}{\theta}}^{PT}_{N_c};c,\underset{\sim}{B}_{\theta_o}) = \frac{2-2(p-2)/p+(p-2)^2/p(p-2)}{2-(p/r)\Pi_{p+2}(\chi^2_{p,\alpha};0)}$$
$$= [r(p+2)]/[p(2r-p\Pi_{p+2}(\chi^2_{p,\alpha};0))]. \quad (4.4)$$

Now, (4.4) is > or < 1 according as $\Pi_{p+2}(\chi^2_{p,\alpha};0)$ is > or < $r(p-2)/p^2$. Note that $\Pi_{p+2}(\chi^2_{p,\alpha};0) < \Pi_p(\chi^2_{p,\alpha};0) = 1-\alpha$, and larger is the value of p, the smaller is the difference between the two. Thus, generally, when α is small and p is not very small, (4.4) exceeds 1 , so that the Stein-rule sequential MLE may not dominate the PT MLE (under H_0). However, as we move away from the null space ω , the ADR of each of these two sequential estimators becomes larger. In the case of the PT MLE, the ADR first increases, attains a maximum [somewhat larger than $2(cr)^{\frac{1}{2}}$] and then converges to $2(cr)^{\frac{1}{2}}$ as the noncentrality parameter Δ tends to $+\infty$. On the other hand, the ADR of the sequential Stein-rule MLE is monotone in the noncentrality parameter Δ , it never exceeds the value $2(cr)^{\frac{1}{2}}$ and it converges to its upper asymptote $2(cr)^{\frac{1}{2}}$ as $\Delta \to \infty$. Consequently, outside a closed ellipsoid (in $\underset{\sim}{\gamma}^*$) with $\underset{\sim}{0}$ as the origin, the ADR of the sequential Stein-rule MLE is smaller than that of the PT MLE, although their ratio is never too large and it converges to 1 as $\Delta \to \infty$. The situation is quite similar to the non-sequential case, treated in detail in Sen (1986) with the only change that the ARE is more closer to 1 [because of the second term on the right hand side of (2.26) which remains the same for both the versions]. As such, we may refer to Sen(1986) for some details of this ARE study.

We conclude this section with some general remarks on the ADR based ARE results in the sequential case. Under the null hypothesis H_0: $\underset{\sim}{\theta}_o \in \omega$, the ARE results are quite clearly interpretable and reveal a clear picture on the relative performances of the different versions. For a fixed alternative (i.e., $\underset{\sim}{\theta}_o \notin \omega$), the asymptotic distributions of the different versions of sequential MLE's reduce to that of the UMLE (excepting the case of the RMLE which becomes degenerate). As such, there remains very little interest in the study of their ARE. The sequence of alternatives in (2.23) is the right choice for the study of the ADR and ARE results , and under such alternatives, we have discriminating results casting light on the relative performances of the different versions of the sequential MLE. In this context, the interpretation of the 'small' values of c remains the same as in the classical sequential point estimation problem (where, of course, such alternatives do not enter into the formulation of the asymptotic risk, although the expected sample size is $O(c^{-\frac{1}{2}})$). The other difference between the sequential and non-sequential cases is the actual values of the ARE for the different versions. This is due to the second factor (i.e., ξ^*) in (2.26) which is generally the same for all the versions of the MLE. As such, the ADR in the sequential case is the sum of the ADR in the non-sequential case and a second term (constant) of the same order of magnitude. This makes the ARE in the sequential case closer to 1 than in the non-sequential case. Thus, in the non-sequential case, under H_0, the ARE of the Stein-rule MLE with respect to the UMLE is p/2 , and this can be very large for large values of p, while, in the sequential case, the parallel ARE value is 2p/(p+2) which can not exceed 2. Similarly, for the non-null case, the ARE of the sequential Stein-rule MLE with respect to the sequential UMLE (though greater than 1) is closer to 1 than in the non-sequential case. A similar picture holds for the other versions too. This difference between the sequential and non-sequential cases can be attributed to the difference of their corresponding risk functions. However, the relative picture (of the ARE) remains very much comparable in the two cases.

5. SOME GENERAL REMARKS

The sequential Stein-rule MLE $\hat{\underset{\sim}{\theta}}{}^S_{N_c}$, defined by (2.22) and (2.14), is motivated by the particular choice of $\underset{\sim}{A} = \underset{\sim}{B}_{\underset{\sim}{\theta}_o}$. In line with general shrinkage estimators, we may consider a more general shrinkage MLE as

$$\underset{\sim}{\theta}{}^*_n = \hat{\underset{\sim}{\theta}}_n + (\underset{\sim}{I} - q_n d_n \mathcal{L}_n^{-1} \underset{\sim}{A}^{-1} \underset{\sim}{S}_n^{-1})(\tilde{\underset{\sim}{\theta}}_n - \hat{\underset{\sim}{\theta}}_n) , \tag{5.1}$$

where $\{q_n\}$ is a sequence of positive numbers converging to a positive constant q with $q \in (0,2(p-2))$, d_n = smallest characteristic root of $\underset{\sim}{A}\underset{\sim}{S}_n$, and $\underset{\sim}{S}_n$ is a BAN estimator of $\underset{\sim}{B}^{-1}_{\underset{\sim}{\theta}_o} - \underset{\sim}{P}_{\underset{\sim}{\theta}_o}$. Such an estimator has also been studied in Sen (1986), and using our stopping variable N_c in (2.14), we may as well consider the sequential shrinkage MLE $\underset{\sim}{\theta}{}^*_{N_c}$. Like in the case of $\hat{\underset{\sim}{B}}_n$, the strong consistency of $\underset{\sim}{S}_n$ may be established along the same line, so that parallel to (3.10), we have

$$c^{-\frac{1}{4}} || \underset{\sim}{\theta}{}^*_{N_c} - \underset{\sim}{\theta}{}^*_{n^*_c} || \to 0, \text{ in probability, as } c \downarrow 0. \tag{5.2}$$

As such, the ADR and ARE results studied in Sen(1986) can directly be incorporated to show that in the light of the ADR, $\theta^*_{N_c}$ dominates the UMLE $\tilde{\theta}_{N_c}$, where we do not have to take $A = B_{\theta_o}$. The ARE picture of this shrinkage (sequential) MLE with respect to the sequential RMLE and PT MLE remains very much comparable to the one presented in Section 4 [see (4.3), (4.4) and the discussions following those]. However, the expressions for these relative risks , for an arbitrary A , are generally more complicated than the ones considered in Section 4.

The sequential versions of the different MLE's are all based on the common stopping number N_c , defined in (2.14), and this has been motivated by the classical sequential estimation problem (without appealing to the Stein-rule philosophy). The asymptotic theory presented here remains intact for any other well defined stopping variable (say, N^*_c), such that N^*_c increases indefinitely as $c \downarrow 0$. In such a case, we need to find out the corresponding sequence $\{n^*_c\}$, for which,as $c \downarrow 0$, $N^*_c/n^*_c \xrightarrow{P} 1$, and in the ADR results, we have to consider the obvious modifications with the adjustment for this sample size. There remains the issue of finding an (asymptotically) optimal stopping rule in this context, and this will be considered in a future publication.

The choice of the matrix A in (2.9) is of some importance too. In the usual linear models, B_{θ_o} in (2.4) usually involves a known matrix B_o and an unknown scalar factor σ^{-2}. In such a case, A can naturally be taken as B_o, and the stopping variable in (2.14) is then based on the sequence of estimators of the scale factor σ^2 . For the normal theory case, optimal estimators of σ^2 are known, and hence, some further strengthening of the asymptotic theory (for moderately small values of c) can be made with some high power mathematics. However, in the general case, such a simplification may not be possible, and the asymptotic theory presented here may be used to provide reasonable approximations to the actual solutions. Our moderate experience with some Monte Carlo studies reveals that under (2.23), for moderate values of c, the asymptotic theory provides good approximations (much better, if c is still smaller).

R E F E R E N C E S

Anscombe, F.J. (1952). Large sample theory of sequential estimation. Proc. Cambridge Phil. Soc. 48, 600-607.

Berger, J.O. (1985). Statistical Decision Theory, 2nd Ed. Springer-Verlag, New York

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton Univ. Press, New Jersey.

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. Sequential Anal. 2 , 359-367.

Ghosh, M., Sinha, B. and Mukhopadhyay, N. (1976). Multivariate sequential point estimation. Jour. Multivar. Anal. 6 , 281-294.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. Zeit. Wahrsch. verw. Geb. 14, 323-330.

Huber, P.J. (1967). The behavior of maximum likelihood estimators under non-standard conditions. Proc. 5th Berkeley Symp. Math. Statist.Prob. 1, 221-233.

Ibragimov, I.A. and Has'minskii, R.Z. (1981). Statistical Estimation: Asymptotic Theory. Springer-Verlag, New York.

Inagaki, N. (1970). On the limiting distribution of a sequence of estimators under uniformity property. Ann. Inst. Statist. Math. 22, 1-13.

James, W. and Stein, C. (1981). Estimation with quadratic loss. Proc. 4th Berkeley Symp. Math. Statist. Prob. 1 , 361-380.

LeCam, L.M. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimators. Ann.Math. Statist. 41, 802-828.

Sen, P.K. (1979). Asymptotic properties of maximum likelihood estimators based on conditional specifications. Ann. Statist. 7 , 1019-1033.

Sen, P.K. (1981). Sequential Nonparametrics: Invariance Principles and Statistical Inference. John Wiley, New York.

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

Sen, P.K. (1986). On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators. Sankhya, Ser.A 48 (in press).

Sen, P.K. and Saleh, A.K.Md.E. (1985). On some shrinkage estimators of multivariate location. Ann. Statist. 13, 272-281.

So, Y.C. and Sen, P.K. (1981). Repeated significance tests based on likelihood ratio statistics. Commun. Statist. Theor. Meth. A 10, 2149-2176.

Takada, Y. (1984). Inadmissibility of a sequential estimation rule of the mean of a multivariate normal distribution. Sequential Anal. 3 , 267-271.