No. 2002

STATISTICAL FUNCTIONALS, STOPPING TIMES AND ASYMPTOTIC MINIMUM RISK PROPERTY

PRANAB KUMAR SEN

Departments of Biostatistics & Statistics
University of North Carolina,
Chapel Hill, NC 27599-3260, USA

ABSTRACT

In a nonparametric estimation problem, for a statistical functional, the risk (defined suitably) depends on n, the sample size, as well as F, the underlying distribution function. Sequential estimators based on appropriate stopping times possess the (first order) asymptotic minimum risk property, although the picture is somewhat different for shrinkage versions of such sequential statistical functionals. In this context, the role of stopping times is explored in a systematic manner.

AMS (1980) Subject Classification Nos: 62 L12, 62 L15

Key Words and Phrases: Asymptotic distributional risk; asymptotically minimum risk; differentiable statistical functions; jackknifing; minimum risk; quadratic error loss; optimal sample size; sequential estimation; shrunken estimators; statistical functionals.

Short Title: STOPPING TIMES FOR STATISTICAL FUNCTIONALS

1. INTRODUCTION

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, defined on E^r, for some $r \geq 1$. The functional form of F is not assumed to be given, and it is only assumed that $F \in \mathcal{F}$, where $\mathcal{F}$ is a suitable class of d.f.'s, defined on E^r. In this nonparametric setup, an estimable parameter θ is regarded as a (possibly vector valued) functional of the d.f. F, i.e., we set

$$\theta = (\theta_1,\dots,\theta_p)' = \underset{\sim}{T}(F) = (T_1(F),\dots,T_p(F))', \tag{1.1}$$

where $p \geq 1$, and $T_1(\cdot),\dots,T_p(\cdot)$ are suitable functionals. Based on a sample $(X_1,\dots,X_n)$ of size n (≥ 1), let

$$F_n(x) = n^{-1} \Sigma_{i=1}^n I(X_i \leq x),\ x \in E^r \tag{1.2}$$

be the empirical (sample) d.f. Then, a natural nonparametric estimator of θ is

$$T_n = T(F_n) = (T_{n1} = T_1(F_n),\dots,\ T_{np} = T_p(F_n))',\quad n \geq n_0. \tag{1.3}$$

In estimating θ by T_n, we conceive of a plausible loss function

$$L(T_n,\theta) = L(\|T_n-\theta\|), \tag{1.4}$$

where $L(t)$, $t > 0$ is nonnegative and nondecreasing in t and $\|T_n-\theta\|$ stands for a suitable norm of $T_n-\theta$. For example, we may consider a quadratic error loss $L(\underset{\sim}{a},\underset{\sim}{b}) = (a-b)'Q^{-1}(a-b) = \|a-b\|_Q^2$, where Q is a given positive definite (p.d.f) matrix. For p=1, another possibility is to consider the absolute error loss $L(a,b) = |a-b|$, and this definition can also be extended for $p \geq 1$. Also, let c (> 0) be the cost of sampling per unit observation, so that incorporating the cost of sampling along with the chosen loss function, we may formulate the following risk function for estimating θ by T_n:

$$\rho(n;c,F) = E_F\{L(T_n,\theta) + cn\},\ F \in \mathcal{F},\ c > 0,\ n \geq n_0. \tag{1.5}$$

Note that, in general, the risk function may depend on θ as well. But, by (1.1), $\theta = T(F)$ is itself a functional of the d.f. F, and hence, the risk $\rho(n;c,F)$ is regarded as a function of (n,c) and a functional of the d.f. F. Invoking the consistency of T_n (requiring usually some mild regularity conditions on $T(\cdot)$ and F), we may argue that $\|T_n-\theta\|$ becomes stochastically smaller as n increases, so that incorporating appropriate integrability condition on $L(\|T_n-\theta\|)$, it may be quite reasonable to assume that

$$E_F\, L(T_n,\theta) = \upsilon_n(F) \quad \text{is} \quad \downarrow \text{ in } n\ (\geq n_0); \quad \forall\, F \in \mathcal{F}, \tag{1.6}$$

where $\upsilon_n(F)$ may depend in general on the unknown F ($\in \mathcal{F}$). Thus, by (1.5) and (1.6), we have

$$\rho(n;c,F) = \upsilon_n(F) + cn,\ c > 0,\ n \geq n_0,\ F \in \mathcal{F}, \tag{1.7}$$

where $\upsilon_n(F)$ is $\downarrow$ in n, while cn is $\uparrow$ in n. Therefore, there exists an n_c^0 (depending generally on c and F), such that

$$\rho(n_c^0;c,F) = \inf\{\rho(n;c,F) : n \geq n_0\}; \quad F \in \mathcal{F},\ c > 0. \tag{1.8}$$

[If there are more than one such n_c^0, we choose the smallest one among them as the desired solution.] In this formulation, n_c^0 stands for the optimal samples size (for a given F and c) and $\rho_c^0(F) = \rho(n_c^0;c,F)$ stands for the minimum risk attainable for the estimation of θ when one confines oneself to the class of nonparametric estimators

$$\mathcal{C} = \{T_n = T(F_n); \quad n \geq n_0\}, \tag{1.9}$$

and adopt the risk function in (1.5).

As has been mentioned earlier, $\upsilon_n(F)$ in (1.6) generally depends on the unknown $F(\in \mathcal{F})$. The current situation is more complex than a parametric model where $\mathcal{F}$ is of a given functional form (involving some unknown parameters appearing as algebraic constants), so that $\upsilon_n(F)$ is also of a known functional form (involving the same parameters (or a subset of them) as algebraic constants). In our nonparametric formulation, $\upsilon_n(F)$ is a functional of the unknown F ($\in \mathcal{F}$), and it satisfies (1.6). It is therefore clear from the above discussion that n_c^0 generally depends on F ($\in \mathcal{F}$) as well as on $c(>0)$, so that for any chosen $n_c^0 = n^0$, say, T_{n^0} may not have the minimum risk property, for all $F \in \mathcal{F}$. Thus, a fixed sample size (n) estimator T_n may not be a minimum risk estimator (MRE) of θ, simultaneously for all $F \in \mathcal{F}$.

In the negation of the MRE property of fixed sample size estimators, it is of genuine interest to consider suitable sequential estimators $\{T_{N_c};\ c > 0\}$ based on appropriate stopping times $\{N_c;\ c > 0\}$, such that in some meaningful way (viz., asymptotically as $c \downarrow 0$), T_{N_c} has the desired MRE property. Or, in other words, we intend to formulate suitable stopping times $\{N_c;\ c>0\}$, where for every $c > 0$, N_c is a positive integer valued r.v., such that

$$\lim_{c \downarrow 0} \{E_F[L(T_{N_c},\theta) + cN_c] \,/\, \rho_c^0(F)\} = 1, \quad \forall\, F \in \mathcal{F}. \tag{1.10}$$

In the literature, (1.10) relates to the first order asymptotic minimum risk property, and any $\{T_{N_c}\}$ satisfying (1.10) is termed an asymptotically (first order) minimum risk estimator (AMRE) of θ.

Section 2 is devoted to the formulation of such AMRE of θ. In this context, asymptotic expansions for $\nu_n(F)$ and sequential estimation of the leading terms are of prime importance, and these will be discussed there. For $p \geq 3$, for any fixed n, T_n may not have the smallest risk: There exist suitable shrinkage (or Stein-rule) estimators which may dominate T_n (in an exact or asymptotic sense) with respect to the risk in (1.4). This Stein-phenomenon also holds in the sequential case [viz., Ghosh, Nickerson and Sen (1987) dealing with a specific parametric model and Sen (1989), for a general nonparametric setup]. Thus, based on the stopping time $\{N_c\}$ leading to the AMRE property of T_{N_c}, it may be possible to construct a suitable shrinkage version of T_{N_c} (say, $T_{N_c}^S$), such that the (asymptotic) risk of $T_{N_c}^S$ is smaller than (or equal to) that of T_{N_c}, uniformly in θ. This relative risk dominance picture is depicted in Section 3. In passing, it may be remarked that the usual shrinkage versions of the T_{N_c} may not belong to the class $\mathcal{C}$. As such, there remains some open questions regarding the (asymptotic) optimality of the stopping times in the context of such shrinkage estimators. The main objective of the current study is to focus on this issue.

2. ASYMPTOTICALLY OPTIMAL STOPPING TIMES FOR STATISTICAL FUNCTIONALS

Formulation of stopping times and sequential estimators of θ depends very much on the nature of $\upsilon_n(F)$ in (1.6). Let us assume that for some positive number q,

$$\upsilon_n(F) = n^{-q}\{\upsilon^{(1)}(F) + n^{-1}\,\upsilon^{(2)}(F) + \ldots\} \tag{2.1}$$

where the $\upsilon^{(j)}(F)$ are suitable functionals of the d.f. F. Typically, if we consider the case of a quadratic error loss with a given (p.d.) Q, then $\upsilon_n(F) = \text{trace}(Q\, E_F[T_n-\theta][T_n-\theta]')$, so that invoking the usual asymptotic expansions for the second order moment of $n^{1/2}(T_n-\theta)$, we obtain that (2.1) holds with q=1. Actually, for Hoeffding's (1948) U-statistics and von Mises' functionals such an expansion follows readily from the so called Hoeffding-decomposition (into orthogonal components); for more general forms of such statistics, a Hoeffding-type decomposition has been neatly worked out by van Zwet (1984). For differentiable statistical functions, parallel results (up to the order n^{-2}) have been considered by Sen (1988). Note that by (1.7) and (2.1), we have

$$\rho(n;c,F) = cn + n^{-q}\,\upsilon^{(1)}(F) + n^{-q-1}\,\upsilon^{(2)}(F) + \ldots . \tag{2.2}$$

Thus, if we define

$$n_c^* = [c^{-1}\, q\, \upsilon^{(1)}(F)]^{(1+q)^{-1}}, \quad c > 0, \tag{2.3}$$

then, it is easy to show that as $c \downarrow 0$,

$$\rho_c^0(F) = \{c^q\, \upsilon^{(1)}(F)\}^{(1+q)^{-1}}\, \{q^{(1+q)^{-1}} + q^{-q(1+q)^{-1}}\} + o(c^{q(1+q)^{-1}}); \tag{2.4}$$

$$\rho(n_c^*;c,F) - \rho_c^0(F) = o(c^{q(1+q)^{-1}}), \tag{2.5}$$

$$n_c^* - n_c^0 = o(c^{-(1+q)^{-1}}); \quad n_c^0 = (c^{-1}q\, \upsilon^{(1)}(F))^{(1+q)^{-1}} + o(c^{-(1+q)^{-1}}). \tag{2.6}$$

Thus, for $c \downarrow 0$, n_c^* provides a first order approximation for n_c^0, and the asymptotic risk of $T_{n_c^*}$ satisfies the limit in (1.10). For this reason, in the sequel, we shall replace n_c^0 by n_c^* for our subsequent analysis. It is worth mentioning in this context that for this first order equivalence results, we may not even need an asymptotic expansion (as in (2.1)); it suffices to assume that there exist a positive q and a positive $\nu^{(1)}(F)$, such that as $n \to \infty$,

$$n^q \nu_n(F) \to \nu^{(1)}(F), \quad \forall F \in \mathcal{F}, \tag{2.7}$$

and this can be justified even under less stringent regularity conditions. Also, (2.3) clearly reveals the dependence of n_c^* (or n_c^0) on F through $\nu^{(1)}(F)$ (and the $\nu^{(j)}(F)$). As such, a sequential estimator of $\nu^{(1)}(F)$ can be incorporated in the formulation of an appropriate stopping rule.

In the simplest situation, we may assume that the functional form of $\nu^{(1)}(\cdot)$ is known, and define

$$V_n = \nu^{(1)}(F_n), \ n \geq n_0, \tag{2.8}$$

as the sample counterparts of $\nu^{(1)}(F)$; it is tactily assumed that

$$V_n \to \nu^{(1)}(F) \text{ a.s (almost surely), as } n \to \infty. \tag{2.9}$$

Then (2.3) and (2.9) lead us consider a stopping time

$$N_c = \inf\{n \geq n_0 : n^{q+1} \geq (c^{-1} q)[V_n + n^{-h}]\}, \quad c > 0, \tag{2.10}$$

where $h > 0$ (specified), and n^{-h} protects against too early termination (if V_n is too small).

It may be remarked that $\{[F_n(x) - F(x)]: \ x \in E^r\}$, $n \geq 1$ (or $\{[F_n(x), x \in E^r], n \geq 1\}$) is a reverse martingale (process) and $\nu^{(1)}(\cdot)$ is nonnegative. So whenever $\nu^{(1)}(\cdot)$ is a convex functional, such that $E \ \nu^{(1)}(F_n)$ exists for some $n = n_0$ (≥ 1), then

$$\{V_n;\ n \geq n_0\} \text{ is a nonnegative reverse submartingale.} \tag{2.11}$$

Incorporating this reverse sub-martingale property along with appropriate moment convergence results for $\{V_n\}$, one may virtually repeat the steps in Sen and Ghosh (1981) and conclude that the AMRE property in (1.10) holds for $\{T_{N_c}\}$. Thus, in this sense, $\{N_c, c > 0\}$ is asymptotically (as $c \downarrow 0$) an optimal stopping time. In the negation of such a reverse sub-martingale property of the V_n, one may require more stringent moment convergence properties of the V_n for the derivation of the AMRE property of $\{T_{N_c}\}$. In dealing with rank based (R-) estimators of location, such an approach has been systematically explored in Sen (1980,1981) [see also Sen (1984) for the multiparameter case], and similar results can be derived for other cases too.

In some situations (particularly in the multiparameter case), $\upsilon^{(1)}(F)$ may be of quite complicated form, so that the formulation of V_n may involve rather complex computational schemes. For example, suppose that $T(\cdot)$ is differentiable at F, so that $T(F_n)$ can be expanded around $T(F)$. In this expansion, the first term is $\int T^{(1)}(F;x)\, dF_n(x) = n^{-1} \Sigma_{i=1}^{n} T^{(1)}(F;X_i)$, where $T^{(1)}(F;x)$ is the so called influence function (at F). If we choose a quadratic error loss function (with a given p.d. matrix Q), then $\upsilon^{(1)}(F) = \text{trace}(Q\, \Sigma^*(F))$, where $\Sigma^*(F)$ is the dispersion matrix of $T^{(1)}(F;X_1)$. Usually, the functional form of $T^{(1)}(F;x)$ depends on the unknown F, so that for the estimation of $\Sigma^*(F)$ (or $\upsilon^{(1)}(F)$), one may require the estimation of $T^{(1)}(F;x)$ as well. Jackknifing methods can be used with advantage in such a case. From the base sample $(X_1,\ldots,X_n)$ of size n, we delete the ith observation (X_i) and denote the corresponding empirical d.f. by $F_{n-1}^{(i)}$, so that $T_{n-1}^{(i)} = T(F_{n-1}^{(i)})$ is an estimator of θ, for $i=1,\ldots,n$. Let then

$$T_{n,i} = n\, T_n - (n-1)T_{n-1}^{(i)}, \quad i=1,\ldots,n; \tag{2.12}$$

$$T_n^* = \frac{1}{n} \Sigma_{i=1}^{n} T_{n,i} \quad \text{and} \quad V_n^* = \frac{1}{n-1} \Sigma_{i=1}^{n} (T_{n,i}-T_n^*)(T_{n,i}-T_n^*)'. \tag{2.13}$$

The $T_{n,i}$ are so called pseudovariables, T_n^* is the classical jackknifed estimator of θ and V_n^* is the jackknifed estimator of $\Sigma^*(F)$. Under quite general regularity conditions [viz., Sen (1988)], we have

$$n\|T_n - T_n^*\| = O(1) \quad \text{almost surely (a.s.), as } n \to \infty, \tag{2.14}$$

$$V_n^* \to \Sigma^*(F) \quad \text{a.s.,} \quad \text{as } n \to \infty, \tag{2.15}$$

so that

$$V_n = \text{trace}(Q\, V_n^*) \to \upsilon^{(1)}(F) = \text{trace}\,(Q\, \Sigma^*(F)) \quad \text{a.s.,} \quad \text{as } n \to \infty. \tag{2.16}$$

Note that for every n $(\geq n_0)$, $T_{n,1},\ldots,T_{n,n}$ are exchangeable r.v.'s, but their dependence structure changes with n. As such, in general, it may be difficult to obtain a reverse sub-martingale characterization for $\{V_n\}$. However, using the differentiability approach in Sen (1988), moment convergence properties of V_n can be studied under appropriate regularity conditions on $T(\cdot)$ and F, and hence, the asymptotic optimality of the stopping time $\{N_c,\ c>0\}$ can be established as in before.

We conclude this section with a remark that the asymptotic optimality of the stopping time $\{N_c,\ c \downarrow 0\}$ in (2.10) is relative to the AMRE property in (1.10). In this context it is tactily assumed that $\rho_c^0(F)$ is the minimum risk attainable for the class ($\mathscr{C}$) of estimators in (1.9). For the multiparameter model (i.e., $p \geq 3$), it is possible to construct some shrinkage versions of T_n which have smaller (asymptotic) risk. This leads to the basic question: Is N_c, $c \downarrow 0$, an asymptotically optimal stopping time for shrunken estimators too? We study this problem in the next section.

3. STOPPING-TIMES AND SHRUNKEN STATISTICAL FUNCTIONALS

Shrunken versions of statistical functionals have been considered in Sen (1989) where the asymptotic dominance of sequential shrunken functionals over their classical counterparts have been discussed. To motivate our main results, we recapitulate briefly the basic results on shrunken functions.

Let us define T_n, V_n^* and Q as in earlier sections, and let

$$d_n = \text{smallest characteristic root of } QV_n^*,\ n \geq n_0. \tag{3.1}$$

Side by side, let δ = smallest characteristic root of $Q\Sigma^*(F)$, where $\Sigma^*(F)$ is the dispersion matrix of the asymptotic (multi-normal) distribution of $n^{1/2}(T_n-\theta)$. Then

$$d_n \rightarrow \delta \quad \text{a.s.,} \quad \text{as } n \rightarrow \infty. \tag{3.2}$$

Next, note that Stein-rule (or shrinkage) estimators are basically testimators: An appropriate test statistic (say, $\mathscr{L}_n$) is incorporated in the estimator itself, and this test statistic is primarily a measure of divergence of θ from the assumed pivot θ_0. Without any loss of generality, we may set $\theta_0=0$ (otherwise, substract θ_0 from T_n and reduce the case to a null pivot). Then an appropriate $\mathscr{L}_n$ to test the adequacy of the pivot (0) is

$$\mathcal{L}_n = nT_n' V_n^{*-1} T_n, \tag{3.3}$$

where under the null hypothesis (θ=0), $\mathcal{L}_n$ has asymptotically (central) chi square distribution with p degrees of freedom (DF). Larger values of $\mathcal{L}_n$ depict the inadequacy of the assumed null pivot. We consider the following type of shrinkage functionals:

$$T_n^S = \{I - k\, d_n\, \mathcal{L}_n^{-1}\, Q^{-1}\, V_n^{*-1}\}T_n \tag{3.4}$$

where p is taken greater than 2, the shrinkage factor k is a positive number ($0 < k < 2(p-2)$) and, in this context, the loss function $L(a,b) = (a-b)'Q(a-b)$ is quadratic with the p.d. matrix Q as its discriminant.

In the parametric (normal theory) model, T_n has a multinormal distribution, $(n-1)V_n^* \sim \text{Wishart}(\Sigma, p, n-1)$, and T_n and V_n^* are independent. Thus, $\mathcal{L}_n$ has the Hotelling T^2-distribution (noncentral when $\theta \neq 0$). This enables one to compute the dispersion matrix of T_n^S and incorporate the same in the computation of the risk of T_n^S. But, in a nonparametric model, the exact distribution of $\mathcal{L}_n^{-1}$ (or the independence of T_n and V_n^*) may be extremely difficult to comprehend, and moment-convergence results for T_n^S may require quite stringent regularity conditions. This poses a serious problem in the computation of the (asymptotic) risk of T_n^S (with an unbounded loss function), and the problem becomes harder in the sequential case. In the normal theory model, the technique employed by Ghosh, Nickerson and Sen (1987) rests heavily on the stochastic independence of $\{T_n; n \geq 1\}$ and $\{V_n^*;\ n \geq 2\}$ and their joint sufficiency. Parallel results may not hold for the nonparametric case. Hence, the theory developed so far [Sen (1989)] relates to the asymptotic case only.

There are two easy ways eliminating the technical difficulty caused by the presence of $\mathcal{L}_n^{-1}$ in (3.4). First, note that

$$d_n\, Q^{-1}\, V_n^{*-1} = \{\text{largest root of } Q^{-1}\, V_n^{*-1}\}^{-1}\, Q^{-1}\, V_n^{*-1}, \tag{3.5}$$

so that the characteristic roots of $d_n\, Q^{-1}\, V_n^{*-1}$ are all less than or equal to 1 (these are nonnegative too). As such, on the set $\{\mathcal{L}_n > k\}$, $\{I-k\, d_n\, \mathcal{L}_n^{-1}\, V_n^{*-1}\}$

is positive semi-definite (actually p.d. with probability one) and has characteristic roots all nonnegative and bounded by 1. The picture can be quite different on the set $\{\mathcal{L}_n \leq k\}$, particularly when $\mathcal{L}_n$ is close to zero. This prompts us to consider a "positive rule" version of T_n^S in (3.4). We denote by

$$U_n^* = \begin{cases} 0, & \text{if } \mathcal{L}_n \leq k \\ I-k\, d_n\, \mathcal{L}_n^{-1}\, Q^{-1}\, V^{*-1}, & \text{if } \mathcal{L}_n > k, \end{cases} \tag{3.6}$$

and consider the modified estimator

$$T_n^{S+} = U_n^*\, T_n. \tag{3.7}$$

Note that $T_n^{S+} = 0$ on the set $\{\mathcal{L}_n \leq k\}$, while for $\mathcal{L}_n > k$, $\mathcal{L}_n^{-1} \leq k^{-1}$, so that moment-convergence results for T_n^{S+} may not require any stringent regularity conditions (pertaining otherwise to the integrability of $\mathcal{L}_n^{-1}$). In fact, such a positive-rule version is known to possess better dominance properties than T_n^S itself [see Sclove, Morris and Radhakrishnan (1972) for the normal theory model]. The other possibility is to use the original estimator T_n^S in (3.4), but to adopt a modified definition of the asymptotic risk. Towards this, we may note tht if the true parameter point θ is different from the pivot (0), then under quite general regularity conditions, as $n \to \infty$

$$n^{-1}\, \mathcal{L}_n \to \Delta = \theta'[\Sigma^*(F)]^{-1}\, \theta > 0, \text{ in probability}, \tag{3.8}$$

so that $\mathcal{L}_n = 0_p(n)$ when $\theta \neq 0$. In this context, by (3.4), we have

$$\begin{aligned} n\|T_n^S - T_n\|_Q^2 &= k^2 n\, \mathcal{L}_n^{-2}\, T_n'\, V_n^{*-1}\, Q^{-1}\, V_n^{*-1}\, T_n \\ &= (k^2\, \mathcal{L}_n^{-1})\, (n\, \mathcal{L}_n^{-1})(T_n'\, V_n^{*-1}\, Q^{-1}\, V_n^{*-1}\, T_n) \\ &= 0_p(n^{-1}), \text{ as } n \to \infty. \end{aligned} \tag{3.9}$$

Thus, $n^{1/2}(T_n^S - \theta)$ and $n^{1/2}(T_n - \theta)$ are stochastically equivalent and they share the common asymptotic properties. Or in other words, there is no asymptotic improvement due to shrinkage if the actual θ (fixed) is different from the assumed pivot. The situation is different when θ is "close to" the assumed pivot. In that case, there is some improvement due to shrinkage, and this will be considered here. We take into account such a 'small neighborhood' of the pivot (in conjunction with $c \downarrow 0$) in our formulation of the "asymptotic risk" function. For simplicity, we confine ourselves to a quadratic error loss, so that as in Section 2, we have

$$n_c^0 = O(c^{-1/2}), \text{ as } c \downarrow 0. \tag{3.10}$$

It will be seen later on that the class of stopping times $\{N_c;\ c > 0\}$ may also be so chosen that

$$N_c = O_p(c^{-1/2}), \text{ as } c \downarrow 0. \tag{3.11}$$

As such, looking at (3.8) and (3.9) along with (3.10) and (3.11), we may conceive of a sequence $\{A(c);\ c>0\}$ of nested neighborhoods.

$$A(c) = \{\underset{\sim}{\theta} : \|\theta-0\| \le A\, c^{-1/4}\},\ c > 0, \tag{3.12}$$

where $A (0 < A < \infty)$ is a fixed positive number (otherwise, arbitrary). Within this domain $(A(c))$ and for N_c satisfying (3.11), $c^{1/4}[T_{N_c}^S - T_{N_c}]$ has a nondegenerate limit distribution (as $c \downarrow 0$), and hence, $T_{N_c}^S$ and T_{N_c} are not generally asymptotically equivalent, while outside $A(c)$, as $c \downarrow 0$, (3.9) holds [with $O_p(n^{-1})$ replaced by $o_p(1)$]. As such, we may as well consider the asymptotic (as $c \downarrow 0$) distribution of $c^{-1/4}(T_{N_c} - \theta)$ and $c^{-1/4}(T_{N_c}^S - \theta)$, denote then G and $G^{(S)}$ respectively, and compute the asymptotic risk by reference to such asymptotic distributional risk (ADR) [see, Sen (1987a, 1989), for example]. The main advantage of using the ADR criterion is that we do not have to consider more stringent regularity conditions pertaining to the asymptotic value of the actual risk $\rho(n;c,F)$, and whenever such a limit exists, it would also agree with the ADR. For the positive rule estimator $T_{N_c}^{S+}$ the asymptotic risk and ADR would have a common value (under the usual regularity conditions). The important fact is that the results based on these ADR comparisons remain applicable to the ones based on the asymptotic risks, when additional regularity (i.e., integrability) conditions are met.

It follows along the lines of Sen (1989) that in the light of the ADR and adapted to the stopping time $\{N_c,\ c>0\}$ in (2.10), $T_{N_c}^S$ dominates T_{N_c}. A very similar conclusion holds for $T^{S+}_{N_c}$ (vs T_{N_c}). Thus, with respect to the asymptotically optimal stopping times pertaining to the classical case (treated in Section 2), further reduction in the asymptotic risk is possible due to shrinking the functionals towards a pivot, when the true parameter lies "close to" the pivot. This is in agreement with the general results on sequential shrinkage estimation for the normal theory case [Ghosh, Nickerson and Sen (1987)], where also the dominance is perceptible only in a 'small' neighborhood of the pivot. However, our conclusions remain applicable to a much larger class of estimators and a wider class of nonparameter models.

With these results at hand, we now consider the final question regarding asymptotically optimal stopping times relating to shrinkage estimators. In this context, first, we provide an answer to the query posed at the end of Section 2. An affirmative answer to the asymptotic optimality of N_c, $c > 0$ in (2.10) for shrinkage estimators can be made when the true θ is away from the pivot. Or, in other words, if a pivot is so chosen that it is not likely to be 'very close' to the actual parameter point θ, then shrinkage estimators fail to provide any perceptible improvement over their classical counterparts, and hence, their asymptotic risk equivalence entails the asymptotic optimality of the stopping time N_c, $c>0$, in (2.10). However, we may have an altogether different answer when θ lies "close to" the assumed pivot. We explore this situation in the remaining of this section.

Let us now confine ourselves to the case of $\theta \in A(c)$, $c \downarrow 0$. Towards this, we consider a sequence $\{F_{(c)}, c > 0\}$ of d.f.'s such that

$$\mathcal{K}_c : \theta_c = T(F_{(c)}) = c^{1/4} \lambda, \quad \lambda = (\lambda_1, \ldots, \lambda_p)' \text{ fixed, } c>0. \tag{3.13}$$

This is more effectively conceived by allowing $F_{(c)} \to F$ as $c \downarrow 0$, such that $\theta_0 = T(F) = 0$ (the pivot), so that $\mathcal{K}_c$ corresponds to a sequence of local alternatives. Note that the dispersion matrix of the asymptotic distribution of $n^{1/2}(T_n - \theta)$ is not affected by $\{\mathcal{K}_c\}$ in (3.13), and hence, the ADR of T_{N_c} (when N_c is defined by (2.10)) does not depend on λ (or the pivot)). On the other hand, the distribution of $n^{1/2}(T^S_n - \theta)$ (even asymptotically) may depend on λ, and hence, the ADR of $T^S_{N_c}$ depends on λ as well.

We define by Γ the dispersion matrix of the asymptotic distribution of

$n^{1/2}(T_n-\theta)$, and for $L(\cdot)$, we take a quadratic error loss (with a p.d. Q), so that as $c \downarrow 0$

$$\rho(n;c,F_{(c)}) = n^{-1} \text{ trace } (Q\Gamma) + cn, \tag{3.14}$$

whenever cn^2 is bounded from below by some positive ϵ. For every $n : n^2c = O(1)$ (as $c \downarrow 0$), we denote by

$$\Delta_{nc} = (nc^{1/2})\lambda'\Gamma^{-1} \lambda \quad \text{and} \quad \Delta^0_{nc} = (nc^{1/2})\lambda'\Gamma^{-1} Q^{-1} \Gamma^{-1} \lambda. \tag{3.15}$$

Then, following the lines of Sen (1987a,b), it can be shown that the ADR of T_n^S (whenever $n^2c = O(1)$ and $n^2 c>0$) is given by

$$cn + n^{-1}[\text{tr}(Q\Gamma) - 2k(p-2)\delta\, E(\chi_p^{-2}(\Delta_{nc}) + k^2\delta^2[\text{tr}(Q^{-1}\Gamma^{-1})]E(\chi^{-4}_{p+2}(\Delta_{nc}))$$
$$+ \Delta^0_{nc}[\text{tr}(Q\Gamma)]E(\chi^{-4}_{p+4}(\Delta_{nc}))], \tag{3.16}$$

where k is the shrinkage factor $[0 < k < 2(p-2)]$, δ is defined after (3.1) and denoting by $H_q(x;\Delta)$ the non-central chi-square d.f. with q DF and noncentrality parameter Δ,

$$E(\chi_q^{-2r}(\Delta)) = \int_0^\infty x^{-r}\, dH_q(x;\Delta),\ r \geq 0,\ \Delta \geq 0. \tag{3.17}$$

Since Δ_{nc} and Δ^0_{nc} both depend on λ in (3.13), it is clear that (3.16) depends on (c,n), Q, Γ as well as λ. For any given c, n, Q and Γ, (3.16) attains a minimum at $\Delta_{nc} = \Delta^0_{nc} = 0$ (i.e., $\lambda=0$), and this minimum value is given by

$$cn + n^{-1}\{\text{tr}(Q\Gamma) - k\delta[2-p^{-1}(p-2)^{-1} k\delta\text{tr}(Q^{-1} \Gamma^{-1})]\}, \tag{3.18}$$

where $\delta\ \text{tr}(Q^{-1}\ \Gamma^{-1}) = \text{tr}(Q^{-1}\ \Gamma^{-2})/\text{ch}_{max}(Q^{-1}\Gamma^{-1}) \leq p$ and $0 < k < 2(p-2)$. Hence, (3.18) is less than $cn + n^{-1}\ \text{tr}(Q\Gamma) = \rho(n;c,F)$, so that for $\lambda=0$, the minimum value of (3.18) (over n) is given by

$$2\{\sqrt{c}[\text{tr}(Q\Gamma) - k\delta\{2-p^{-1}(p-2)^{-1} k\delta\text{tr}(Q^{-1}\Gamma^{-1})\}]^{1/2}, \tag{3.19}$$

which is smaller than $\rho^0_c(F) = 2\sqrt{c}\{\text{tr}(Q\Gamma)\}^{1/2}$. Thus, if we consider a stopping time N^o_c, $c > 0$, defined by

$$N^o_c = \inf\{n>n_0 : n^2 > \tfrac{1}{c}\{\text{tr}(QV^*_n) - kd_n[2-p^{-1}(p-2)^{-1}kd_n\ \text{tr}(Q^{-1}V^{*-1}_n)]\}\}, \tag{3.20}$$

(for $c > 0$), then it follows by the same technique as in Sen (1989) that the asymptotic (as $c \downarrow 0$) risk of $T^S_{N^0_c}$, when $\theta=0$, is given by (3.19), and moreover, defining N_c, $c > 0$ as in (2.10) (with $q=1$),

$$\lim_{c\downarrow 0}(N^o_c/N_c) \overset{a.s.}{=} 1 - \frac{k\delta}{tr(Q\Gamma)}[2 - \frac{k\delta}{p(p-2)} tr(Q^{-1} \Gamma^{-1})] < 1. \qquad (3.21)$$

Thus, at the pivot ($\theta=0$), the stopping time N^o_c is stochastically smaller than N_c and at the same time leads to a smaller asymptotic risk for the shrunken estimator $T^S_{N^o_c}$ (compared to $T^S_{N_c}$) when $\lambda \neq 0$ (i.e. $\theta \neq 0$). To see this, we may note that under (3.13), the ADR of $T^S_{N^o_c}$ is given by

$$c^{1/2}\{tr(Q\Gamma) - k\delta[2 - \frac{k\delta}{p(p-2)} tr(Q^{-1} \Gamma^{-1})]\}^{1/2} + \qquad (3.22)$$

$$c^{1/2}\{tr(Q\Gamma) - k\delta[2 - \frac{k\delta}{p(p-2)} tr(Q^{-1} \Gamma^{-1})]\}^{-1/2} \{Tr(Q\Gamma) - k\delta(2 - \frac{k\delta}{p(p-2)} tr(Q^{-1}\Gamma^{-1}))$$

$$+ h(\delta, Q, \Gamma, \Delta_{n^o_c c}, \Delta^o_{n^o_c c})\}$$

where $n^o_c = c^{-1/2}[tr(Q\Gamma) - k\delta\{2-p^{-1}(p-2)^{-1} k\delta\ tr(Q^{-1} \Gamma^{-1})\}]$ and

$$h(\delta, Q, \Gamma, \Delta_{nc}, \Delta^o_{nc}) = (3.16)-(3.18) \text{ (at n)}; \qquad (3.23)$$

(3.23) is nonnegative and bounded from above by $k\delta[2 - \frac{k\delta}{p(p-2)} tr\ (Q^{-1} \Gamma^{-1})]$ (where the upper bound is attained when $\|\lambda\| \to \infty$.). At this stage, we may note that the ADR for $T^S_{N_c}$ is given by

$$c^{1/2}\{tr(Q\Gamma)\}^{1/2} [1 + [tr(Q\Gamma)]^{-1}\{TR(Q\Gamma) - K\delta(2 - \frac{k\delta}{p(p-2)} tr(Q^{-1} \Gamma^{-1}))$$

$$+ h(\delta, Q, \Gamma, \Delta_{n^*_c c}, \Delta^0_{n^{0*}_c c})\}] \qquad (3.24)$$

where $n^{o*}_c \sim c^{-1/2}(tr(Q\Gamma))^{1/2} > n^0_c$. By using (3.23) along with the upper bound for $h(\cdot)$, we obtain that (3.24) can not be larger than $2\sqrt{c}\{tr(Q\Gamma)\}^{1/2} = \rho^0_c(F)$. Hence, it suffices to show that (3.22) can be larger than $\rho^0_c(F)$, for large values of $\|\lambda\|$. Towards this note that for $0 \le d \le b < a$,

$$(a-b)^{1/2} + (a-b)^{-1/2}(a-b+d) = (a-b)^{-1/2}(2a-2b+d)$$

$$= (2a-b)(a-b)^{-1/2} - (b-d)(a-b)^{-1/2}$$

$$= 2\sqrt{a}\left(1 - \frac{1}{2a}\right)\left(1 - \frac{b}{a}\right)^{-1/2} - (b-d)(a-b)^{-1/2}. \tag{3.25}$$

Since $(1-b/2a)^2 = 1-b/a + b^2/4a^2 > 1-b/a$, $\forall\ b < a$, the right hand side of (3.25) can be made $> 2\sqrt{a}$ for all d sufficiently close to b. Now as $\|\lambda\| \to \infty$, $h(\delta, Q, \Gamma, \Delta_{n_c^0 c}, \Delta^0_{n_c^0 c}) \to d\delta\{2-p^{-1}(p-2)^{-1} k\ \delta\ tr(Q^{-1}\ \Gamma^{-1})\}$, by (3.22) and (3.25), we conclude that the ADR of $T^S_{N_c^0}$ can indeed exceed that of T_{N_c} (and hence, $T^S_{N_c}$) when θ is away from the pivot (0). This exhibits the lack of asymptotic optimality of $\{N_c^0, c > 0\}$, uniformly in $\lambda \in E^p$. A similar criticism applies to $\{N_c, c > 0\}$ [in (2.10)] when one intends to use shrinkage functional [as in a neighborhood of the pivot (i.e., for $\|\lambda\|$ small) (3.24) will be larger than (3.22)], although $\{N_c\ c > 0\}$ has the nice property that the ADR of $T^S_{N_c}$ can't be larger than that of T_{N_c} (while that of $T^S_{N_c^o}$ can be so). This leads us to the question: Does there exist a stopping rule $\{N_c^*, c > 0\}$, such that

(a) N_c^* is stochastically smaller than N_c,

and (b) the ADR of $T^s_{N_c^*}$ is $\le$ the ADR of T_{N_c}, uniformly in λ?

We construct an adaptive stopping time to meet both these requirements. In this context, we may note that there are nice series expansions for $E(\chi_q^{-2}(\Delta))$ and $E(\chi_{q+2}^{-4}(\Delta))$, $q > 2$, $\Delta \ge 0$, in terms of Poisson weights, and hence, $h(d_n, Q, V_n^*, a_n, b_n)$ can be conveniently computed for suitable a_n, b_n, d_n, Q and V_n^*. Let us define then N_c^o, $c > 0$, as in (3.20). Also, for every $n \ge n_0$, let

$$a_n^* = \min\{tr(QV_n^*) - kd_n[2 - \frac{kd_n}{p(p-2)}\ tr(Q^{-1}V_n^{*-1})] + h(d_n, Q, V_n^*, a_{n1}, a_{n2})],$$

$$tr(QV_n^*)\}, \tag{3.26}$$

where

$a_{n1} = \max\{0,\ n\, T_n'\, V_n^{*-1}\, T_n - p\}$, $a_{n2} = \max\{0,\ n\, T_n'\, V_n^{*-1} Q^{-1}\, V_n^{*-1}\, T_n - \mathrm{tr}(Q^{-1}\, V_n^{*-1})\}$

and the other notations have already been introduced earlier. Let then

$$N_c^{(1)} = [c^{-1/2}(a^*_{N_c^0})^{1/2}] + 1, \quad c > 0, \tag{3.27}$$

and by iteration, we let

$$N_c^{(r)} = [c^{-1/2}(a^*_{N_c^{(r-1)}})] + 1, \ c > 0, \quad \forall\ r \geq 1. \tag{3.28}$$

Finally, we define

$$N_c^* = \lim_{r\to\infty} N_c^{(r)}. \tag{3.29}$$

It is clear from the definition in (3.26)-(3.29) that N_c^* is stochastically smaller than N_c. [Note that $a_n^* \leq \{\mathrm{tr}(QV_n^*)\}$, $\forall\ n \geq n_0$, and hence $N_c^{(r)} \leq N_c$ $\forall\ r \geq 0$]. Further note that under $\{K_c\}$ in (3.13),

$$n\, T_n'\, V^{*-1}\, T_n - p = n\, c^{1/2}\, \lambda'\, V^{*-1}\, \lambda + 2\, nc^{1/4}\, \lambda'\, V_n^{*-1}(T_n - c^{1/4}\, \lambda)$$
$$+\ n(T_n - c^{1/4}\, \lambda)'\, V_n^{*-1}(T_n - c^{1/4}\, \lambda) - p \tag{3.30}$$

where for $n \leq n_c^*$, $n\, c^{1/2} < \mathcal{K} < \infty$, the second term on the right hand side of (3.30) has asymptotically a normal distribution with zero mean and variance $4nc^{1/2}\lambda'\Gamma^{-1}\lambda$, while the third term has the central chi squared d.f. with p DF. Thus, $n\, T_n'\, V_n^{*-1}\, T_n - p$ behaves like $\Delta_{nc} + Z_c + (\chi_p^2 - p)$, where $Z_c \sim \mathcal{N}(0, 4\Delta_{nc})$. A similar result holds for a_{n2}. Thus, a_{n1} and a_{n2} are not stochastically equivalent to Δ_{nc} and Δ_{nc}^0, respectively; they are subject to some random variation [due to Z_c and χ_p^2] of the same order. This creates additional trouble for finding out the ADR of $T^S_{N_c^*}$. Fortunately, this relates to the case, where

$$N_c^{(r)}/n_c^* \xrightarrow{P} (1-\beta) + U_r, \text{ as } c \downarrow 0, \tag{3.31}$$

where $0 < \beta < 1$ and U_r has a nondegenerate distribution over $(0,\infty)$. This representation allows us to use the central limit theorem for random sample size [c.f. Billingsley (1968, p. 147)], and justifies the computation of the

ADR of $T^S_{N_c^*}$ from its asymptotic distribution. It follows that the ADR of $T^S_{N_c^*}$ is less than (or equal to) that of T_{N_c}, for all λ. Thus, the adoptation of the stopping time N_c^*, $c > 0$, leads to a stochastically smaller stopping time for $T^S_{N_c^*}$ without violating the asymptotic risk dominance of $T^S_{N_c^*}$ over T_{N_c}. But, viewed from a practical point of view, when θ is not close to the pivot (even under (3.13)), $T^S_{N_c^*}$ and $T^S_{N_c}$ may be very close to each other (with respect to their ADR), and $\{N_c, c > 0\}$ signals a great simplification of the computation of ADR, without changing the stopping time. Hence, one may as well favor N_c, $c > 0$, as a working rule.

REFERENCES

1. Billingsley, P. Convergence of Probability Measures. John Wiley, New York, 1968.
2. Ghosh, M., Nickerson, D.M. and Sen, P.K. Sequential shrinkage estimation. Ann. Statist. 15 (1987), 817-829.
3. Hoeffding, W. A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19 (1948), 293-325.
4. Sclove, S.L., Morris, C. and Radhakrishnan, R. Nonoptimality of preliminary test estimators for the mean of a multivariate normal distribution. Ann. Math. Statist. 43 (1972), 1481-1490.
5. Sen, P.K. On nonparametric sequential point estimation of location based on general rank order statistics. Sankhya Ser. A 42 (1980), 201-219.
6. Sen, P.K. Sequential Nonparametrics: Invariance principles and Statistical Inference. John Wiley, New York, 1981.
7. Sen, P.K. On sequential nonparametric estimation of multivariate location. Proc. Third Prague Conf. Asymp. Meth. (Eds: Huskova, et. al.), 1984, 119-130.
8. Sen, P.K. Sequential Stein-rule maximum likelihood estimation: General asymptotics. Statist. Dec. Th. & Rel. Top. IV (Eds. Gupta, S.S. and Berger, J.O.), 2(1987), 195-208.
9. Sen, P.K. Sequential shrinkage U-statistics: General asymptotics. Rev. Brasileira de Prob. Estatist. 1 (1987), 1-21.
10. Sen, P.K. Functional jackknifing: Rationality and general asymptotics. Ann. Statist. 16 (1988), 450-469.
11. Sen, P.K. Asymptotic theory of sequential shrunken estimation of statistical functionals. Proc. 4th Prague Conf. Asymp. Meth. (Eds. M. Huskova, et al.), 1989, in press.
12. Sen, P.K. and Ghosh, M. Sequential point estimation of estimable parameters based on U-statistics. Sankhya, Ser. A 43 (1981), 331-344.
13. van Zwet. W.R. A Berry-Esseen bound for symmetric statistics. Zeit. Wahrsch. verw. Geb. 66 (1984), 425-440.