THE PITMAN CLOSENESS OF STATISTICAL ESTIMATORS: LATENT YEARS AND THE RENAISSANCE

Pranab Kumar Sen, University of North Carolina at Chapel Hill

Abstract

The Pitman closeness criterion is an intrinsic measure of the comparative behavior of two estimators (of a common parameter) based solely on their joint distribution. It generally entails less stringent regularity conditions than in other measures. Although there are some undesirable features of this measure, the past few years have witnessed some significant developments on Pitman-closeness in its tributaries, and a critical account of the same is provided here. Some emphasis is placed on nonparametric and robust estimators covering fixed-sample size as well as sequential sampling schemes.

Introduction

In those days prior to the formulation of statistical decision theory (Wald, 1949), the reciprocal of variance [or mean square error (MSE)] of an estimator (T) used to be generally accepted as an universal measure of its precision (or efficiency). The celebrated Cramér-Rao inequality (Rao, 1945) was not known that precisely although Fisher (1938) had a fair idea about such a lower bound to the variance of an estimator. The use of mean absolute deviation (MAD) criterion as an alternative to the MSE was not that popular (mainly because its exact evaluation often proved to be cumbersome), while other loss functions (convex or not) were yet to be formulated in a proper perspective. In this setup, Pitman (1937) proposed a novel measure of closeness (or nearness) of statistical estimators, quite different in character from the MSE, MAD and other criteria. Let T_1 and T_2 be two rival estimators of a parameter θ belonging to a parameter space $\Theta \subseteq R$. Then T_1 is said to be closer to θ than T_2, in the Pitman sense, if

$$P_\theta\{|T_1 - \theta| \leq |T_2 - \theta|\} \geq 1/2, \forall \theta \in \Theta, \tag{1}$$

with strict inequality holding for some θ. Thus, the Pitman-closeness criterion (PCC) is an intrinsic measure of the comparative behavior of two estimators. Note that in terms of the MSE, T_1 is better than T_2, if

$$E_\theta(T_1 - \theta)^2 \leq E_\theta(T_2 - \theta)^2, \forall \theta \in \Theta, \tag{2}$$

with strict inequality holding for some θ; for the MAD criterion, we need to replace $E_\theta(T - \theta)^2$ by $E_\theta|T - \theta|$. In general, for a suitable nonnegative loss function $L(a, \theta) : R \times R \to R^+$, T_1 dominates T_2 if

$$E_\theta[L(T_1, \theta)] \leq E_\theta[L(T_2, \theta)], \forall \theta \in \Theta, \tag{3}$$

with strict inequality holding for some θ. We represent (1), (2) and (3) respectively as

$$T_1 \succ_{PC} T_2, T_1 \succ_{MSE} T_2 \text{ and } T_1 \succ_L T_2. \tag{4}$$

It is clear from the above definitions that for (2) or (3), one needs to operate the expectations (or moments), while (1) involves a distributional operation only. Thus, in general, (2) or (3) may entail more stringent regularity conditions (pertaining to the existence of such expectations) than needed for (1). In this sense, the PCC is solely a distributional measure while the others are mostly moment based ones, and hence, from this perspective, the PCC has a greater scope of applicability (and some other advantages too). On the other hand, other conventional measures, such as (2) or (3), may have some natural properties which may not be shared by the PCC. To illustrate this point, note that if there are three estimators, say, T_1, T_2 and T_3, of a common parameter θ, such that

$$E_\theta(T_1 - \theta)^2 \leq E_\theta(T_2 - \theta)^2$$

and

$$E_\theta(T_2 - \theta)^2 \leq E_\theta(T_3 - \theta)^2, \forall \theta \in \Theta, \tag{5}$$

then, evidently, $E_\theta(T_1 - \theta)^2 \leq E_\theta(T_3 - \theta)^2$, $\forall \theta \in \Theta$. Or, in other words, the MSE criterion has the transitivity property, and this is generally the case with (3). However, this transitivity property may not always hold for the PCC. That is, T_1 may be closer to θ than T_2, and T_2 may be closer to θ than T_3 (in the Pitman sense), but T_1 may not be closer to θ than T_3 in the same sense!. Although a little artificial, it is not difficult to construct suitable examples testifying the intransitivity of the PCC (Blyth, 1972). Secondly, the measure in (2) or (3) involves the marginal distributions of T_1 and T_2, while (1) involves the joint distribution of (T_1, T_2). Hence, the task of verifying the dominance in (1) may require more elaborate analysis. This was perhaps the main reason why in spite of a good start and notable contributions by Geary (1944) and Johnson (1950), the use of PCC remained somewhat skeptical for more than thirty years! In fat, the lack of transitivity of the PCC in (1) caused some difficulties in extending the pairwise dominance in (1) to that within a suitable class of estimators. Only recently, such results have been obtained by Ghosh and Sen (1989) and Nayak (1990) for suitable *families of equivariant estimators*. We shall comment on them in a later section. Thirdly, in (1), when both T_1 and T_2 have continuous distributions and $T_2 - T_1$ has a non-atomic distribution, the $\leq$ sign may as well be replaced by $<$ sign, without affecting the probability inequality. However, if $T_2 - T_1$ has an atomic distribution, the two probability statements involving $\leq$ and $<$ signs, respectively, may not agree, and somewhat different conclusions may crop up in the two cases. Although this anomaly can be eliminated by attaching suitable probability (viz., 1/2) for the tie

($|T_1 - \theta| = |T_2 - \theta|$), the process can be somewhat arbitrary and less convincing in general. Fourthly, the definitions in (1) through (3) need some modifications in the case where $\underset{\sim}{\theta}$ (and $\underset{\sim}{T}$) are p-vectors, for some $p > 1$. The MSE criterion lends itself naturally to an appropriate *quadratic error loss*, where for some chosen positive definite (p.d.) matrix $\underset{\sim}{Q}$, the distance function is taken as $\| \underset{\sim}{T} - \underset{\sim}{\theta} \|_{\underset{\sim}{Q}}$, and given by

$$\| \underset{\sim}{T} - \underset{\sim}{\theta} \|^2_{\underset{\sim}{Q}} = (\underset{\sim}{T} - \underset{\sim}{\theta})' \underset{\sim}{Q} (\underset{\sim}{T} - \underset{\sim}{\theta}) \tag{6}$$

The use of the Fisher Information matrix $\underset{\sim}{\mathfrak{J}}_{\underset{\sim}{\theta}}$ as $\underset{\sim}{Q}$ leads to the so-called Mahalanobis distance. Recall that

$$E_{\underset{\sim}{\theta}} \| \underset{\sim}{T} - \underset{\sim}{\theta} \|^2_{\underset{\sim}{Q}} = Trace \left[\underset{\sim}{Q} E_{\underset{\sim}{\theta}} \{ (\underset{\sim}{T} - \underset{\sim}{\theta})(\underset{\sim}{T} - \underset{\sim}{\theta})' \} \right], \tag{7}$$

so that (2) entails only the computation of the mean product error (or dispersion) matrix of $\underset{\sim}{T}_1$ and $\underset{\sim}{T}_2$. On the other hand, if instead of $|T_1 - \theta|$ and $|T_2 - \theta|$, in (1), we use $\| \underset{\sim}{T}_1 - \underset{\sim}{\theta} \|_{\underset{\sim}{Q}}$ and $\| \underset{\sim}{T}_2 - \underset{\sim}{\theta} \|_{\underset{\sim}{Q}}$, the probability statement may be a more involved function of the actual distribution of ($\underset{\sim}{T}_1$, $\underset{\sim}{T}_2$) and of the $\underset{\sim}{Q}$. Although in some special cases this can be handled without too much of complications (see for example, Sen, 1989a), in general, we may require more stringent regularity conditions to verify (1) in the vector case. In the asymptotic case, however, an equivalence of BAN estimators and Pitman-closest ones may be established under very general regularity conditions (viz., Sen, 1986), so that (1) and (2) may have asymptotic equivalence. But, in the multiparameter case, best estimators, in the sense of having a minimum value of (7) may not be BAN. A natural reference is the so-called Stein paradox (viz, Stein, 1956) for the estimation of the mean vector of a multivariate normal distribution. For p, the dimension of the multivariate normal law, greater than 2, Stein (1956) showed that the sample mean vector [although being the *maximum likelihood estimator* (MLE)] is not *admissible*, and later on, James and Stein (1962) constructed some other estimators which dominate the MLE in the light of (2) [as amended in (7)]. Such *Stein-rule* or *shrinkage estimators* are typically non-linear and are non-normal, even asymptotically. Thus, they are not BAN. So, a natural question arose: Do the Stein-rule estimators dominate their classical counterparts in the light of the PCC? An affirmative answer to this question has recently been provided by Sen, Kubokawa and Saleh (1989), and we shall discuss this in a later section. Fifthly, we have tacitly assumed so far that we have a conventional *fixed-sample size case*. There are, however, some natural situations calling for suitable *sequential schemes*, so that one may also like to inquire how far the PCC remains adoptable in such a sequential scheme. Some studies in this direction have been made very recently by Sen (1989a), and we shall discuss some of these results in a later section. Another direction in which the PCC has proven to be a very useful avenue for comparing estimators is the employment of more general loss functions (instead of the Euclidean norm or the usual quadratic norm) in the definition in (1). In the context of estimation of the dispersion matrix of a multivariate normal distribution and parameters in some other distributions belonging to the exponential family, one may adopt the entropy (or some related) loss functions which when incorporated in (1) lead to a more general formulation. This has been termed the generalized Pitman nearness criterion (GPNC) (viz., Khattree, 1987, for the dispersion matrix estimation problem). We shall review some of the developments in this area in the last section.

As has been mentioned earlier, for nearly four decades, there were not much activities in this general arena, while the past ten years have witnessed a remarkable growth of the literature on the PCC. This renaissance is partly due to the work of C.R. Rao (1981) who clearly pointed out the shortcomings of the MSE or the quadratic error loss and explained the rationality of the PCC (which attaches less importance to *large deviations*). The work of Efron (1975) also deserves a special mention: the feasibility of an estimator dominating the classical MLE of the univariate normal mean in the light of the PCC clearly points out the adaptability of the PCC in a more general situation where other forms of admissibility criteria may not work out well. A somewhat comparable picture in both the works of Efron (1975) and Rao (1981) might have been based on the MAD criterion which attaches less importance to large deviations than the MSE criterion. However, in the general multiparameter case, the MAD criterion may lose its appeal to a greater extent. This is mainly due to the following factors: (i) lack of invariance under suitable groups of transformations usually employed in multiparameter estimation problems, (ii) complexity of the definitions and (iii) need for the estimation of nuisance parameters (such as the reciprocal of the density functions) in the definition of the norm itself which usually requires really large sample sizes! One might also argue in favor of some other criteria. Hwang (1985) has considered the *stochastic dominance* criterion based on the marginal distributions with an arbitrarily chosen cut-off point, and this in turn introduces some arbitrariness in the adaptation of his measure; the dominance may not hold uniformly in the choice of such a cut-off point. Brown, Cohen and Strawderman (1976) advocated the use of some non-convex loss functions. We have no definite prescription in favor of the PCC, MAD, such non-convex loss functions or the stochastic dominance criterion, although the PCC may have some natural appeal. In passing, we may remark that some controversies have been reported in Roberts and Hwang (1988), although it is very hard to endorse fully the views expressed in this report. We would like to bypass these by adding that *let the cliff fall where it belongs to*! In our opinion, in spite of some of the shortcomings of the PCC, as have been mentioned earlier, the developments in the past decade have, by far, been much more encouraging to advocate in favor of the use of PCC (or the GPNC) in a variety of statistical models which will be considered here in the subsequent sections. We also refer to a recent Panel Discussion on Pitman nearness of statistical estimators at the International Conference on Recent Developments in Statistical Data Analysis and Inference (in honor of C.R. Rao) at Neuchatel, Switzerland (August 24, 1989), where some of these issues have been discussed critically, and a report of these findings is accounted in Mason, Keating, Sen and Rao (1990). As with any other measure, there are pathological examples where the PCC may not appear to be that rational, but in real applications, we will rarely be confronted with such artificial cases. On the other hand, in the conventional linear models and in multivariate analyses, some theoretical studies (supplemented by numerical investigations) made by Mason, Keating, Sen and Blaylock (1990) justify the appropriateness of the PCC, even when a dominance may not hold for the entire parameter space. Finally, in the asymptotic case where the sample size is large enough to justify the usual regularity conditions needed to use simplified distribution theory for the estimators, for a wider class of nonparametric and robust estimators, we may justify the adaptation of the PCC on a very broad ground. We shall stress this point in the subsequent sections. All in all, we welcome the renaissance of the PCC and look forward to further developments in this fruitful area of statistical research.

PCC in the Single Parameter Case

In this section, we stick to the basic definition in (1) and examine the Pitman-closeness of a general class of statistical estimators. According to (1), rival estimators are compared two at a time, while (2) or (3) lends itself readily to suitable classes of estimators. This prompted Ghosh and Sen (1989) to consider Pitman closest estimators within reasonable classes of estimators. In this context, we may remark that under (2), the celebrated Rao-Blackwell theorem depicts the role of *unbiased, sufficient statistics* in the construction of such optimal estimators. Ghosh and Sen (1989) have shown that under appropriate regularity conditions, a *median unbiased* (MU) estimator is Pitman-closest within an appropriate class of estimators. Recall that an estimator T of θ is MU if

$$P_\theta\{T \le \theta\} = P_\theta\{T \ge \theta\}, \forall \theta \in \Theta, \tag{8}$$

and T_0 is Pitman-closest within a class of estimators ($\mathfrak{C}$), if (1) holds for $T_1 = T_0$ and every $T_2 \in \mathfrak{C}$. In many applications, T_0 is a function of a (*complete*) *sufficient statistic* and $T_2 = T_0 + Z$, where Z is *ancillary*. Then, note that

$$\begin{aligned}\left[|T_0 - \theta| \le |T_2 - \theta|\right] &\Leftrightarrow \left[|T_0 - \theta|^2 \le (T_0 - \theta + Z)^2\right]\\ &\Leftrightarrow \left[2Z(T_0 - \theta) + Z^2 \ge 0\right],\end{aligned} \tag{9}$$

while by Basu's (1955) theorem, T_0 and Z are independently distributed. Since Z^2 is a nonnegative random variable, the MU character of T_0 ensure that the right hand side of (9) has probability $\ge 1/2$, $\forall \theta \in \Theta$. This explains the role of MU sufficient statistics in the characterization of the Pitman-closest estimators.

However, the following theorem due to Ghosh and Sen (1989) presents a broader characterization.

Theorem 1.

Let T be MU-estimator of θ and let $\mathcal{C}$ be the class of all estimators of the form $U = T + Z$, where T and Z are independently distributed. Then $P_\theta\{|T-\theta| \leq |U-\theta|\} \geq 1/2$, for all $\theta \in \Theta$ and $U \in \mathcal{C}$.

Theorem 1 typically relates to the estimation of location parameter (θ) in the usual location-scale model where the class $\mathcal{C}$ relates to suitable *equivariant estimators* (relative to appropriate groups of transformation). Various examples of this type have been considered by Ghosh and Sen (1989). In the context of the estimation of the scale parameter, the PCC has been studied in a relatively more detailed manner. Keating (1985) considered a general scale family of distributions, and confined himself to the class ($\mathcal{C}^0$) of all estimators which are scalar multiple of the usual MLE; however, he did not enforce any equivariance considerations to clinch the desired Pitman-closest property. Keating and Gupta (1984) considered various estimators of the scale parameter of a normal distribution, and compared them in the light of the PCC. Again in the absence of any equivariance considerations, their result did not lead to the desired Pitman-closest characterization. The following theorem due to Ghosh and Sen (1989) provides the desired result.

Theorem 2.

Let $\mathcal{C}^*$ be the class of all estimators of the form $U = T(1 + Z)$, where T is MU for θ and is nonnegative, while T and Z are independently distributed. Then, $P_\theta\{|T-\theta| \leq |U-\theta|\} \geq 1/2$, $\forall\, \theta \in \Theta$, $U \in \mathcal{C}^*$.

Both these theorems have been incorporated in the PC characterization of BLUE (*best linear unbiased estimators*) of location and scale parameters in the complete sample as well as censored cases (Sen, 1989b); equivariance plays a basic role in this context too. Further note that if T has a distribution symmetric about θ, then T is MU for Θ. This sufficient condition for T is easy to verify in many practical applications. Similarly, if the conditional distribution of T, given Z, is symmetric about θ, then in Theorem 1, we may not need the independence of T and Z. The uniform distribution on $[\theta - \frac{1}{2}\delta, \theta + \frac{1}{2}\delta]$, $\delta > 0$, provides a simple example of the latter (Ghosh and Sen, 1989).

We shall now discuss some further results on PCC in the single parameter case pertaining to the asymptotic case and to sequential sampling plans. The current literature on theory of estimation is flooded with asymptotics. Asymptotic normality, asymptotic efficiency and other asymptotic considerations play a vital role in this context. An estimator (T_n) based on a sample of size n is termed a BAN (*best asymptotically normal*) estimator of θ if the following two conditions hold:

$$n^{\frac{1}{2}}(T_n - \theta) \text{ is asymptotically normal } (0, \sigma_T^2) \tag{10}$$

[which is the AN (*asymptotically normal*) criterion], and

$$\sigma_T^2 = \frac{1}{\mathfrak{J}_\theta}, \text{ where } \mathfrak{J}_\theta \text{ is the Fisher information of } \theta \tag{11}$$

[which is the B (*bestness*) criterion]. Let us now consider the class $\mathcal{C}_A$ of estimation $\{U_n\}$ which admit an asymptotic representation of the form:

$$U_n - \theta = n^{-1}\sum_{i=1}^{n}\psi_\theta(x_i) + o_p\left(n^{-\frac{1}{2}}\right), \text{ as } n \to \infty, \tag{12}$$

where the *score function* $\psi_\theta(\cdot)$ may depend on the method of estimation and the model; $E_\theta\psi_\theta(X_i) = 0$ and $E_\theta\psi_\theta^2(x_i) = \sigma_U^2 < \infty$. Recall that for a BAN estimator of θ, we would have a representation of the form (12) where $\mathfrak{J}_\theta \cdot \psi_\theta(x_i) = f'_\theta(x_i, \theta)/f(x_i, \theta)$, $f(\cdot)$ is the probability density function and f'_θ is its first order derivative w.r. to θ. Note further that $E_\theta\{[f'_\theta(x_1; \theta)/f(x_1; \theta)]^2\} = \mathfrak{J}_\theta$, so that for a BAN estimator, $E_\theta\{\psi_\theta(x_1)f'_\theta(x_1)/f(x_1; \theta)\} = 1$, $\forall\ \theta$. Thus, if we let

$$\xi_n = n^{-\frac{1}{2}}\sum_{i=1}^{n}(\partial/\partial\theta)\ log\ f(X_i; \theta), \tag{13}$$

then for a BAN estimator T_n, we have under the usual regularity conditions that as $n \to \infty$,

$$\left(n^{\frac{1}{2}}(T_n - \theta), \xi_n\right) \xrightarrow[\mathfrak{D}]{} \mathcal{N}_2\left((0, 0), \begin{bmatrix} \mathfrak{J}_\theta^{-1} & 1 \\ 1 & \mathfrak{J}_\theta \end{bmatrix}\right). \tag{14a}$$

Consider now the class $\mathcal{C}^0$ of all estimators $\{U_n\}$, such that as $n \to \infty$,

$$\left(n^{\frac{1}{2}}(U_n - \theta), \xi_n\right) \xrightarrow[\mathfrak{D}]{} \mathcal{N}_2\left((0, 0), \begin{bmatrix} \sigma_U^2 & 1 \\ 1 & \mathfrak{J}_\theta \end{bmatrix}\right), \tag{14b}$$

where $\sigma_U^2 \geq \mathfrak{J}_\theta^{-1}$, and the equality sign holds whenever U_n is a BAN estimator of θ. Note that the $\sqrt{n}$-consistency of U_n entails the unit covariance term. As such, by an appeal to Theorem 1 of Sen (1986) we conclude that the BAN estimator satisfying (14a) is asymptotically (as $n \to \infty$) a Pitman-closest estimator of θ (within the class $\mathcal{C}^0$).

Note that this characterization is localized to the class of asymptotically normal estimators. In the context of estimation of location (or simple regression) parameter, incorporating robustness considerations (either on a local or global basis), various other estimators have been considered by a host of workers.

Among those, the *M*-, *L*- and *R-estimators* deserve special mention. The *M-estimators* are especially advocated for plausible local departures from the assumed model, and they retain high efficiency for the assumed model and at the same time possess good local robustness properties. The *R*-estimators are based on appropriate rank statistics and possess good global robustness properties. *L*-estimators are based on linear functions of order statistics with a similar robustness consideration in mind. In general, these *M*-, *L*- and *R*-estimators satisfy the AN condition in (10) through appropriate representations of the type (12), where $\psi_\theta(x) = \psi(x-\theta)$; see for example, Sen (1981, Ch. 8). From considerations of bestness based on the minimum (asymptotic) MSE, the optimal *M*-, *L*- and *R*-estimators all satisfy the bestness condition in (11). Hence, we conclude that an *M*-, *L*- or *R*- estimator of θ having the BAN character in the usual sense is also asymptotically Pitman-closest. This places the PCC in a very comparable stand in the asymptotic case. Note that being a completely distributional measure, the PCC does not entail the computation or convergence of the actual MSE of the estimators, and hence (14a) requiring the usual conditions needed for the BAN property, also leads to the desired PCC property.

We consider now some recent results on PCC in the sequential case (Sen, 1989a). Note that for the estimation of the mean of a normal distribution with unknown variance σ^2, generally a sequential sampling plan is advocated to ensure some control on the performance characteristics (which can not be done in a fixed sample procedure). In this setup, the *stopping number* N is a positive integer valued random variable such that for every $n \geq 2$, the event $[N = n]$ depends only on $\{s_k^2,\ k \leq n\}$, where s_k^2 is the sample variance for the sample size k, $k \geq 2$. It is known that $\{\bar{X}_k,\ k \geq 1\}$ and $\{s_k^2,\ k \leq n\}$ are mutually independent, and hence, given $N = n$ (i.e., the s_k^2, $k \leq n$), $T_n = \bar{X}_n$ satisfies the conditions of Theorem 1, so that $\bar{X}_N$ has the Pitman-closest character. This simple observation can be incorporated in a formulation of the PC characterization of sequential estimators. Let $\{X_i,\ i \geq 1\}$ be a sequence of independent and identically distributed (i.i.d.) random variables (r.v.) with a distribution function (d.f.) $F_\theta(x)$, $x \in R$, $\theta \in \Theta \subset R$. For every $n \geq 1$, consider the transformation:

$$(X_1,\ldots,X_n) \rightarrow (T_n,\ \underset{\sim}{V}_n,\ \underset{\sim}{W}_n) \qquad (\underset{\sim}{V}_n \text{ could be vacuous}) \tag{15}$$

$\underset{\sim}{W}_n$ is a $(n-k-1)$-vector and $\underset{\sim}{V}_n$ is a k-vector, where k is a nonnegative integer. Let $\mathfrak{B}_T^{(n)}$ and $\mathfrak{B}_W^{(n)}$ be the sigma sub-fields generated by T_n and $\underset{\sim}{W}_n$, respectively, for $n \geq 1$,

$$[N = n] \text{ is } \mathfrak{B}_W^{(n)}\text{-measurable,} \tag{16}$$

$$T_n \text{ is MU for } \theta, \tag{17}$$

$$\begin{aligned} Z_n = \ & v_n(\underset{\sim}{W}_n) \text{ is } \mathfrak{B}_W^{(n)}\text{-measurable and} \\ & T_n \text{ and } \underset{\sim}{W}_n \text{ are independently distributed.} \end{aligned} \tag{18}$$

As in Theorem 1, let $\mathcal{C}^0$ be the class of all (sequential) estimators of the form $U_N = T_N + Z_N$. Then, under (16), (17) and (18),

$$P_\theta\{|T_N - \theta| \leq |U_N - \theta|\} \geq 1/2, \quad \forall\, U_N \in \mathcal{C}^0 \text{ and } \theta \in \Theta. \tag{19}$$

A similar extension of Theorem 2 to the sequential case works out under (16) - (18).

The characterization of PC of sequential estimators made above is an exact one, in the sense that it holds for an arbitrary stopping number (N) so long as N satisfies (16). In the context of bounded-width confidence intervals for θ or minimum risk (point) estimation of θ (and in some other problems too), the stopping number N is indexed by a positive real number d (i.e., $N = N_d$), such that N_d is well defined for every $d > 0$ (and N_d is usually $\downarrow$ in d). In this setup, one considers an asymptotic model where $d \downarrow 0$. Often, there exists a sequence $\{n_d^0\}$ of positive integers (n_d^0 is $\downarrow$ in d), such that $n_d^0 \to \infty$, as $d \downarrow 0$, and further,

$$(n_d^0)^{-1} N_d \xrightarrow{p} 1, \text{ as } d \downarrow 0.$$

In such a case, we may extend the PC characterization to the class of BAN (sequential) estimators, without necessarily requiring (16). Consider the BAN estimators treated in (10) through (14), but now adapted to the stopping number $\{N_d\}$. Suppose that the U_n [in (12)] satisfy an Anscombe-type condition [Anscombe (1952)] that for every $\epsilon > 0$ and $\eta > 0$, there exist a $\delta > 0$ and an integer n_0, such that

$$P\left\{\max_{m:|m/n-1| \leq \delta} n^{\frac{1}{2}} |U_m - U_n| > \epsilon\right\} < \eta, \forall\, n \geq n_0. \tag{20}$$

This Anscombe-condition holds for the ξ_n in (13) under no extra regularity conditions. On the other hand, (20) is also a byproduct of (weak) invariance principles for the U_n, which have been studied extensively in the literature [viz., Sen (1981), Ch. 3-8]. Thus, we may replace $\{U_{N_d}, T_{N_d}\}$ by $\{U_{n_d^0}, T_{n_d^0}\}$, as $d \downarrow 0$, and then make use of (14) to characterize the desired PC property of the sequential BAN estimators. Note that, in general, M-estimators of locations are not scale-equivariant (so as to qualify for the class $\mathcal{C}$ in Theorem 1), and L- and R-estimators of location may not also belong to this class. Thus, in finite sample case, the PC characterization may not apply to these estimators. But, in the asymptotic case (sequential or fixed-sample size setup), the PC characterization holds in spite of the fact that these estimators may not belong to the class $\mathcal{C}$ or that (16) may not hold.

To sum up the main findings on PCC in the uniparameter case, we observe that the MU property (along with ancillarity and sufficiency) provide us with the desired tool for finding the Pitman-closest estimators in the fixed-sample as well as sequential cases. In the asymptotic case, BAN estimators enjoy the PC-property, and this naturally raises the question: What is the relationship of the PCC and the (asymptotic) variance of an estimator? Following the lead of Rao et al. (1986) and Keating and Mason (1985), Peddada and Khattree (1986) studied this problem; however, their main results pertain to two estimators, say, T_1 and T_2, which are distributed independently of each other, and hence, the conclusions derived from these results may not apply to an usual situation where the two rival estimators of a common parameter θ are not independently distributed. Moreover, as they were assuming normality in most of the cases (treated by them), more general results for such models can be obtained from Sen (1986).

PCC in the Multiparameter Case

There has been a lot of research work on the PCC in the multiparameter case, including *shrinkage* and sequential estimators. Let us consider the case of a vector $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)'$ of parameters, where $\underset{\sim}{\theta} \in \Theta \subset R^p$, for some $p \geq 1$. Let $\underset{\sim}{T} = (T_1,\ldots,T_p)'$ be an estimator of θ. First, we need to extend the definition of the distance $|T - \theta|$ in (1) to the multiparameter case. Although the Euclidean norm is a possibility, since the different components of $\underset{\sim}{T}$ may have different importance (and they are generally not independent), a more general quadratic norm is usually adopted. We may define

$$\| \underset{\sim}{d} \|^2_{\underset{\sim}{Q}} = \underset{\sim}{d}' \underset{\sim}{Q} \underset{\sim}{d}, \; \underset{\sim}{d} \in R^p, \tag{21}$$

where $\underset{\sim}{Q}$ is a given p.d. matrix. It is not uncommon to use some other metric (viz., entropy, etc.), so that we may as well take a general

$$L(\underset{\sim}{T}, \underset{\sim}{\theta}), \text{ satisfying the usual properties of a 'norm'.} \tag{22}$$

In the last section, in the context of estimation of dispersion matrices of multivariate normal distributions, we shall use such norms. As an extension of (1) and following the lead of Peddada (1985), we consider the following *generalized Pitman nearness criterion* (GPNC): An estimator $\underset{\sim}{T}_1$ is GPN closer than $\underset{\sim}{T}_2$ if

$$P_{\underset{\sim}{\theta}}\left\{L(\underset{\sim}{T}_1, \underset{\sim}{\theta}) \leq L(\underset{\sim}{T}_2, \underset{\sim}{\theta})\right\} \geq 1/2, \; \forall \, \underset{\sim}{\theta} \in \Theta. \tag{23}$$

In the context of multivariate location models and in other situations too, it is quite possible to identify a class of estimators similar to that in Theorem 1. However, this would rest on plausible extensions of the notion of median unbiasedness in the multiparameter case. Since the components of $\underset{\sim}{T}$ may not be all independent and $\underset{\sim}{Q}$ in (21) may not be a diagonal matrix, the MU property for each coordinate of $\underset{\sim}{T}$ may not suffice. For our purpose, under (21), it seems that the following definition of multivariate MU property may suffice.

We say that $\underset{\sim}{T}$ is MU for $\underset{\sim}{\theta}$, if

$$\underset{\sim}{\ell}'(\underset{\sim}{T} - \underset{\sim}{\theta}) \text{ is MU for 0, for every } \underset{\sim}{\ell} \in R^p, \underset{\sim}{\theta} \in \Theta. \tag{24}$$

In passing, we may remark that if $\underset{\sim}{T}$ has a distribution *diagonally symmetric* about $\underset{\sim}{\theta}$, then (24) holds, although the converse is not necessarily true. Recall that $\underset{\sim}{T}$ has a diagonally symmetric d.f. around $\underset{\sim}{\theta}$ if $\underset{\sim}{T} - \underset{\sim}{\theta}$ and $\underset{\sim}{\theta} - \underset{\sim}{T}$ both have the same d.f.

Theorem 3.

Let $\underset{\sim}{T}$ be a MU-estimator of $\underset{\sim}{\theta}$ [in the sense of (24)], and let $\mathcal{C}$ be the class of all estimators of the form $\underset{\sim}{U} = \underset{\sim}{T} + \underset{\sim}{Z}$, where $\underset{\sim}{T}$ and $\underset{\sim}{Z}$ are independently distributed. Then for any arbitrary p.d. $\underset{\sim}{Q}$,

$$P_{\theta}\left\{\| \underset{\sim}{T} - \underset{\sim}{\theta} \|_{\underset{\sim}{Q}} \leq \| \underset{\sim}{U} - \underset{\sim}{\theta} \|_{\underset{\sim}{Q}}\right\} \geq 1/2, \quad \forall\, \underset{\sim}{\theta} \in \underset{\sim}{\Theta}, \underset{\sim}{U} \in \mathcal{C}. \tag{25}$$

The proof is simple (Sen, 1989a) and is omitted. As a simple example illustrating (25), consider the case where $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ are i.i.d. r.v.'s having the multinormal distribution with mean vector $\underset{\sim}{\theta}$ and dispersion matrix $\underset{\sim}{\Sigma}$. Then $\underset{\sim}{T}_n = N^{-1}\Sigma_{i=1}^n \underset{\sim}{X}_i$ is MU in the sense of (24). Further, for known $\underset{\sim}{\Sigma}$, $\underset{\sim}{T}_n$ is sufficient for $\underset{\sim}{\theta}$, and the class $\mathcal{C}$ consists here of all estimators of the form $\underset{\sim}{T}_n + \underset{\sim}{Z}_n$ where $\underset{\sim}{Z}_n$ is ancillary; this rests on the group of affine transformations $\underset{\sim}{X}_i \rightarrow \underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{X}_i$, $\underset{\sim}{B}$ nonsingular and $\underset{\sim}{a}$ arbitrary. Thus, by Theorem 3, within the class of such equivariant estimators of $\underset{\sim}{\theta}$, the sample mean $\underset{\sim}{T}_n$ (MLE) is the Pitman-closest one. By using the classical Helmert transformation for the multivariate normal vectors, it can be shown that the conclusion remains true in the case of unknown (but nonsingular) $\underset{\sim}{\Sigma}$. Moreover, the interesting feature of this example [or (25)] is that the construction of $\underset{\sim}{T}$ or the class $\mathcal{C}$ does not depend on $\underset{\sim}{Q}$ in (21). In the multiparameter case, we shall study the GPNC for the *Stein-rule* or *shrinkage estimators*, and in that context, it will be seen that neither these estimators belong to the class $\mathcal{C}$ nor their dominance may hold for all $\underset{\sim}{Q}$ (i.e., for a given $\underset{\sim}{Q}$, the construction of PC $\underset{\sim}{T}_n$ may generally depend on $\underset{\sim}{Q}$, and this $\underset{\sim}{T}_n$ may not retain its optimality simultaneously for all $\underset{\sim}{Q}$, possibly different from the adapted one). For the time being, we refrain ourselves from generalizing Theorem 2 to the vector-case; we shall make comments on it in the last section. Perhaps, it will be to our advantage to discuss the sequential analogue of Theorem 3, i.e., a multi-parameter extension of (19). Let us consider the same model as in (14) – (18) with the exception that in (15), $\underset{\sim}{T}_n$ is a vector and in (18), $\underset{\sim}{Z}_n$ is a vector too. Then the following result is proved in Sen (1989a):

Under (16), (18) and (24), for the class $\mathcal{C}^0$ of (sequential) estimators of the form $\underset{\sim}{U}_N = \underset{\sim}{T}_N + \underset{\sim}{Z}_N$, we have

$$P_{\theta}\left\{\| T_N - \theta \|_Q \leq \| U_N - \theta \|_Q\right\} \geq 1/2, \quad \forall\, \theta \in \Theta,\ U_N \in \mathcal{C}^0, \tag{26}$$

for any arbitrary (p.d.) Q.

Again as an illustration, we may consider the multinormal mean vector (θ) estimation problem when the covariance matrix (Σ) is arbitrary and unknown. Ghosh, Sinha and Mukhopadhyay (1976) and others have considered suitable stopping numbers (N) which are based solely on the sample covariance matrices $\{S_n;\ n > p\}$, so that (16) and (18) hold (for $T_n = \bar{X}_n$, $n \geq 1$). Further, (24) follows from the diagonal symmetry of the d.f. of $\bar{X}_n$ (around θ), $\forall$ $n \geq 1$. Hence, (26) holds.

Let us next consider the asymptotic case parallel to that in the previous section. As in (10) – (11), a BAN estimator T_n is characterized by its asymptotic (multi-) normality along with the fact that the dispersion matrix of this asymp totic distribution is equal to $\mathfrak{J}_\theta^{-1}$, where $\mathfrak{J}_\theta$ is the Fisher information matrix. The representation in (12) also extends readily to this multiparameter case, and (13) relates to a stochastic p-vector which has the dispersion matrix $\mathfrak{J}_\theta$. Consider then the class $\mathcal{C}^0$ of all estimators $\{U_n\}$ for which

$$\begin{bmatrix} n^{\frac{1}{2}}(U_n - \theta) \\ \xi_n \end{bmatrix} \underset{\mathfrak{D}}{\rightarrow} \mathcal{N}_{2p}\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \nu, & I \\ I, & \mathfrak{J}_\theta \end{bmatrix}\right], \tag{27}$$

where $\nu - \mathfrak{J}_\theta^{-1}$ is positive semi-definite, and the $\sqrt{n}$-consistency of U_n entails the identity matrix I in (27); for a BAN estimator T_n, $\nu = \mathfrak{J}_\theta^{-1}$. Finally, in (21), it seems quite appropriate to let $Q = \mathfrak{J}_\theta$. Then, by Theorem 1 of Sen (1986) we conclude that within the class $\mathcal{C}^0$ of estimators which are asymptotically multi-normal and for which (27) holds [with $\mathfrak{J}_\theta^{-1}$, being replaced by the asymptotic dispersion matrix of $n^{1/2}(U_n - \theta)$], the BAN estimators are Pitman-closest with respect to the norm in (21), where $Q = \mathfrak{J}_\theta$.

The interesting feature is that we are no longer restricting ourselves to the class $\mathcal{C}$ of estimators (which are generally equivariant), but the Pitman-closest property depends on the adaptation of $Q = \mathfrak{J}_\theta$. For an arbitrary Q, this property may not hold. The asymptotic theory of Pitman-closeness of sequential estimators runs parallel to that in the concluding part of last section, and hence, we do not repeat these details.

In multiparameter estimation problems, the usual MLE may not be *admissible* (in the light of quadratic error loss functions). Stein (1956) considered the simple model that X has a multi-normal distribution with mean vector θ and dispersion matrix, say, I_p, for some $p \geq 1$. He showed that though X is the

MLE of $\underset{\sim}{\theta}$ for all $p \geq 1$, it is inadmissible for $p \geq 3$. James and Stein (1962) constructed a shrinkage version which dominates $\underset{\sim}{X}$ in quadratic error loss. Sparked by this *Stein-phenomenon*, during the past twenty-five years, a vast amount of work has been done in improving the classical estimators in various multiparameter estimation problems by suitable shrinkage versions; these improvements being judged by the smallness of appropriate quadratic error loss function based *risks*. Coming back to the multivariate normal law, such shrinkage or Stein-rule estimators do not belong to the class $\mathfrak{C}$ considered in Theorem 3! Thus, the characterization of PC made in Theorem 3 is not applicable to such shrinkage estimators. This raises the question: Does the usual Stein-rule estimator have the PC property too? The answer is affirmative in a variety of situations, and moreover, this PC dominance may hold even under less restrictive regularity conditions.

Rao (1981) initiated renewed interest in the PCC by showing that some simple shrinkage estimators may not be the Pitman closest ones! He actually argued that the usual quadratic error loss function places undue emphasis on large deviations which may occur with small probability, and hence, minimizing the mean square error may insure against large errors in estimation occurring more frequently rather than providing greater concentration of an estimator in neighborhoods of the true value. Since, typically, a Stein-rule estimator is non-linear and may not have (even asymptotically) multi-normal law, Rao's criticism is more appropriate in this context. Actually, Rao, Keating and Mason (1986) and Keating and Mason (1988) have shown by extensive numerical studies that for the p-variate normal distribution, for $p \geq 2$, the James-Stein estimator is closer (in the Pitman sense) than the MLE $\underset{\sim}{X}$. The quadratic error loss criterion may also cause some difficulties in the usual linear models when the incidence (design) matrix is nearly singular; in such a case, a *ridge regression* estimator is generally preferred. In this context too, one may enquire whether such ridge regression estimators have the Pitman closeness property. This issue has been taken up by Mason, Keating, Sen and Blaylock (1990), and both theoretical and numerical studies are made. So long as the incidence matrix is non-singular, a ridge estimator may not dominate the classical least square estimator in the PCC, although it fares well over a greater part of Θ. The lack of dominance arises mainly due to the fact that as $\underset{\sim}{\theta}$ moves away from the pivot, the performance of a ridge estimator may deteriorate, so that the inequality in (23) may not hold for all θ belonging to Θ, although it generally holds for all $\underset{\sim}{\theta} : \|\underset{\sim}{\theta}\| < C$, where C is related to the factor k (> 0) arising in the construction of a ridge estimator. Their study also covers the comparison of two arbitrary linear estimators in the light of the PCC.

The interesting fact is that the PCC may not even need that p is ≥ 2 (comparable to $p \geq 3$ for the quadratic error loss)! Even for $p = 1$, $X \sim \mathcal{N}(\theta, 1)$, Efron (1975) showed that for

$$\delta = X - \Delta(X); \quad \Delta(x) = \frac{1}{2}\left[min\{x, \Phi(-x)\}\right], \; x \geq 0, \tag{28}$$

$[\Delta(-x) = -\Delta(x),\ x \geq 0$ and $\Phi(\cdot)$ is the standard normal d.f.], (1) holds for $T_1 = \delta$ and $T_2 = X$. He made some conjectures for $p \geq 2$. For the multivariate normal mean estimation problem, a systematic account of the PC dominance of Stein-rule estimators is given by Sen, Kubokawa and Saleh (1989). Consider first the model that for some positive integer p, $\underset{\sim}{X}$ has a p-variate normal distribution with mean vector $\underset{\sim}{\theta}$ and dispersion matrix $\sigma^2 \underset{\sim}{V}$, where $\underset{\sim}{V}$ is known (and p.d.), while $\underset{\sim}{\theta}$ and σ^2 are unknown. Also assume that s^2 is an estimator of σ^2, such that (i) $ms^2/\sigma^2 \underset{\mathfrak{D}}{=} \chi^2_m$, a r.v. having the central chi square distribution with m (≥ 1) degrees of freedom (DF), and (ii) s^2 is distributed independently of $\underset{\sim}{X}$. [In actual application, $\underset{\sim}{X}$ may be the sample mean vector or a suitable linear estimator (of regression parameters, for example) and s^2 is the residual mean square (with $m = n - q$, for some $q \geq 1$]. Keeping in mind the loss function in (21), we may consider a Stein-rule estimator of the form

$$\underset{\sim}{\delta}_\phi = \left[\underset{\sim}{I} - \phi(\underset{\sim}{X}, s^2)s^2 \| \underset{\sim}{X} \|^{-2}_{\underset{\sim}{Q}, \underset{\sim}{V}} \underset{\sim}{Q}^{-1} \underset{\sim}{V}^{-1}\right]\underset{\sim}{X}, \tag{29}$$

where $\phi(\underset{\sim}{x}, s^2)$ is a nonnegative r.v. bounded from above by a constant c_p (depending on p) (with probability one), and $\| \underset{\sim}{X} \|^{-2}_{\underset{\sim}{Q}, \underset{\sim}{V}} = \underset{\sim}{X}' \underset{\sim}{V}^{-1} \underset{\sim}{Q}^{-1} \underset{\sim}{V}^{-1} \underset{\sim}{X}$. Note that estimators of this type with a different bound for $\phi(\cdot)$ (and for $p \geq 3$) were considered by Stein (1981), and hence, we regard them as Stein-rule estimators. Then, we have the following result due to Sen et al. (1989).

Theorem 4.

Assume that $p \geq 2$, and

$$0 \leq \phi(\underset{\sim}{X}, s^2) \leq (p-1)(3p+1)/(2p), \text{ for every } (\underset{\sim}{X}, s^2) \text{ a.e.} \tag{30}$$

Then $\underset{\sim}{\delta}_\phi$, given by (29), is closer than $\underset{\sim}{X}$ in the Pitman sense [i.e., (23) holds for $\underset{\sim}{T}_1 = \underset{\sim}{\delta}_\phi$, $\underset{\sim}{T}_2 = \underset{\sim}{X}$ and $L(\underset{\sim}{T}, \underset{\sim}{\theta}) = \| \underset{\sim}{T} - \underset{\sim}{\theta} \|^2_{\underset{\sim}{Q}}$].

If σ^2 were known, then in (29) and (30), we would have taken $\phi(\underset{\sim}{X}, \sigma^2)$ instead of $\phi(\underset{\sim}{X}, s^2)$. In this sense, the classical James-Stein (1962) estimator is a special case of (29). We may take $\phi(\underset{\sim}{X}, s^2) = a : 0 < a < (p-1)(3p+1)/2p$, and consider the following versions:

$$\underset{\sim}{\delta}_a = \underset{\sim}{X} - as^2 \| \underset{\sim}{X} \|^{-2}_{\underset{\sim}{Q}, \underset{\sim}{V}} \underset{\sim}{Q}^{-1} \underset{\sim}{V}^{-1} \underset{\sim}{X}, \tag{31}$$

$$\underset{\sim}{\delta}^+_a = \underset{\sim}{X} - min\left\{as^2 \| \underset{\sim}{X} \|^{-2}_{\underset{\sim}{Q}, \underset{\sim}{V}}, \underset{\sim}{X}' \underset{\sim}{V}^{-1} \underset{\sim}{X} \| \underset{\sim}{X} \|^{-2}_{\underset{\sim}{Q}, \underset{\sim}{V}}\right\} \underset{\sim}{Q}^{-1} \underset{\sim}{V}^{-1} \underset{\sim}{X}, \tag{32}$$

so that $\underset{\sim}{\delta}_a$ is a James-Stein estimator and $\underset{\sim}{\delta}^+_a$ is the so-called *positive-rule* version. Then again (23) holds with $\underset{\sim}{T}_1 = \underset{\sim}{\delta}^+_a$, $\underset{\sim}{T}_2 = \underset{\sim}{\delta}_a$, $L(\underset{\sim}{T}, \underset{\sim}{\theta}) = \| \underset{\sim}{T} - \underset{\sim}{\theta} \|^2_{\underset{\sim}{Q}}$ and

$0 < a < (p-1)(3p+1)/2p$. Thus, the positive rule version dominates the classical James-Stein version in the light of the PCC as well. It may be remarked that for the quadratic error loss dominance, Stein (1981) had $p \geq 3$ and $0 \leq a \leq 2(p-2)$, while here $p \geq 2$ and $0 \leq a \leq (p-1)(3p+1)/2p$. For $p \in [2, 5]$, $(p-1)(3p+1)/2p > 2(p-2)$. For $p \geq 6$, in (30), we may as well replace $(p-1)(3p+1)/2p$ by $2(p-2)$. The main motivation of the upper bound in (30) was to include the case $p = 2$ and to have a larger shrinkage factor for smaller values of p.

The proof of Theorem 4 depends on some intricate properties of noncentral chi square densities which may have some interest on their own. Basically, to verify (23) for $\underset{\sim}{T}_1 = \underset{\sim}{\delta}_\phi$ and $\underset{\sim}{T}_2 = \underset{\sim}{X}$, it follows through some standard steps that a sufficient condition is

$$P_\lambda\left\{\chi^2_{p,\lambda} \geq \lambda + c\chi^2_m\right\} \geq 1/2, \ \forall \lambda \geq 0, m \geq 1, p \geq 2, \tag{33}$$

where $c = (p-1)(3p+1)/(4pm)$, $\chi^2_{p,\lambda}$ has the noncentral chi square d.f. with p DF and noncentrality parameter λ (≥ 0), and χ^2_m has the central chi square d.f. with m DF, independently of $\chi^2_{p,\lambda}$. The trick was to show that the left hand side of (33) in λ (≥ 0) and that as $\lambda \to \infty$, it converges to 1/2. Sen et al. (1989) also considered the case of $\underset{\sim}{X} \sim \mathcal{N}_p(\underset{\sim}{\theta}, \underset{\sim}{\Sigma})$, $\underset{\sim}{\Sigma}$ arbitrary (p.d.), $\underset{\sim}{S} \sim \mathit{Wishart}(\underset{\sim}{\Sigma}, p, m)$ independently of $\underset{\sim}{X}$ with $m \geq p$, and considered the usual shrinkage estimator

$$\underset{\sim}{\delta}^*_\phi = X - (m-p+1)^{-1}\phi(\underset{\sim}{X}, \underset{\sim}{S})d_m\|X\|^{-2}_{\underset{\sim}{S}^{-1}}\underset{\sim}{Q}^{-1}\underset{\sim}{S}^{-1}\underset{\sim}{X}, \tag{34}$$

where $d_m = ch_{min}(\underset{\sim}{Q}\,\underset{\sim}{S})$ and $\phi(\underset{\sim}{x}, \underset{\sim}{S})$ has the same bound as in (30). Then, for every $p \geq 2$, (23) holds for $\underset{\sim}{T}_1 = \underset{\sim}{\delta}^*_\phi$ and $\underset{\sim}{T}_2 = \underset{\sim}{X}$.

Let us now consider the asymptotic picture relating to the Stein-rule estimators under the PCC. Generally, we have a sequence $\{\underset{\sim}{T}_n\}$ of estimators, such that as $n \to \infty$,

$$n^{\frac{1}{2}}(\underset{\sim}{T}_n - \underset{\sim}{\theta}) \xrightarrow{\mathfrak{D}} \mathcal{N}_p(\underset{\sim}{0}, \underset{\sim}{\Sigma}), \underset{\sim}{\Sigma} \text{ p.d.}, \tag{35}$$

and, also, we have a sequence $\{\underset{\sim}{S}_n\}$ of stochastic matrices, such that

$$\underset{\sim}{S}_n \to \underset{\sim}{\Sigma}, \text{ in probability, as } n \to \infty. \tag{36}$$

Thus, a suitable test statistic for testing the hypothesis of a null pivot is

$$\mathcal{L}_n = n\underset{\sim}{T}'_n\underset{\sim}{S}^{-1}_n\underset{\sim}{T}_n, \tag{37}$$

so that an asymptotic version of (34) is

$$\underset{\sim}{\delta}^{0*}_{\phi,n} = \phi(\underset{\sim}{T}_n, \underset{\sim}{S}_n)d_m\mathcal{L}^{-1}_n\underset{\sim}{Q}^{-1}\underset{\sim}{S}^{-1}_n\underset{\sim}{T}_n. \tag{38}$$

This form is of sufficient generality to cover a large class of $\{\underset{\sim}{T}_n\}$, both of parameter and nonparameter forms. In particular, for R- and M-estimators, for $\mathcal{L}_n$ in (37), instead of $\underset{\sim}{T}_n$, suitable rank or M-statistics may also be used. Also, in (38), a null pivot has been used; the modifications for a general $\underset{\sim}{\theta}_0$ are straightforward. Now, if $\underset{\sim}{\theta} \neq \underset{\sim}{0}$, then $n^{-1}\mathcal{L}_n \xrightarrow{p} \underset{\sim}{\theta}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{\theta}$, as $n \to \infty$, so that $\mathcal{L}_n^{-1} \xrightarrow{p} 0$, as $n \to \infty$. Thus, for any fixed $\underset{\sim}{\theta} \neq \underset{\sim}{0}$,

$$\sqrt{n} \left\| \underset{\sim}{T}_n - \underset{\sim}{\delta}^{0*}_{\phi,n} \right\|_{\underset{\sim}{Q}} \xrightarrow{p} 0, \text{ as } n \to \infty, \tag{39}$$

so that asymptotically the Stein-rule version becomes stochastically equivalent to the classical version. For this reason, the asymptotic dominance picture has been considered in the case where $\underset{\sim}{\theta}$ belongs to a Pitman-neighborhood of the assumed pivot ($\underset{\sim}{0}$). Thus, we may consider a sequence $\{K_n\}$ of local (Pitman-) alternatives

$$K_n : \underset{\sim}{\theta} = \underset{\sim}{\theta}_{(n)} = n^{-\frac{1}{2}}\underset{\sim}{\lambda}, \ \underset{\sim}{\lambda} \in R^p. \tag{40}$$

Further, by virtue of (36), we may replace $\underset{\sim}{S}_n$ by $\underset{\sim}{\Sigma}$, and appeal to Theorem 4 (where s^2 is taken as 1 and $\underset{\sim}{V} = \underset{\sim}{\Sigma}$). As such, we obtain that for every $\phi(\cdot)$, satisfying (40),

$$\lim_{n\to\infty} P\left\{\sqrt{n} \left\| \underset{\sim}{\delta}^{0*}_{\phi,n} - \underset{\sim}{\theta} \right\|_{\underset{\sim}{Q}} \leq \sqrt{n} \left\| \underset{\sim}{T}_n - \underset{\sim}{\theta} \right\|_{\underset{\sim}{Q}} \middle| K_n \right\} \geq 1/2. \tag{41}$$

Thus, the usual robust and nonparametric Stein-rule estimators enjoy the Pitman closeness property in the asymptotic case (and for Pitman-alternatives) under less restrictive regularity conditions (than in the conventional case of quadratic error losses).

Let us now consider sequential Stein-rule estimators and discuss their dominance in the light of the PCC. Consider a simple model: $\{\underset{\sim}{X}_i, 1 \geq 1\}$ are i.i.d.r.v. with $\mathcal{N}_p(\underset{\sim}{\theta}, \sigma^2\underset{\sim}{I}_p)$ d.f.; $\underset{\sim}{\theta}$ and σ^2 are unknown. Let $s_n^2 = (np)^{-1}\Sigma_{i=1}^n(\underset{\sim}{X}_i - \bar{\underset{\sim}{X}}_n)' \times (\underset{\sim}{X}_i - \bar{\underset{\sim}{X}}_n)$; $\bar{\underset{\sim}{X}}_n = n^{-1}\Sigma_{i=1}^n \underset{\sim}{X}_i$, and consider a stopping number N, such that for every $n \geq 2$, $[N = n]$ depends only on $\{s_k^2, k \leq n\}$. Let then

$$\underset{\sim}{\delta}_N^b = \left\{1 - bs_N^2\left(N\| \bar{\underset{\sim}{X}}_N \|^2\right)^{-1}\right\} \bar{\underset{\sim}{X}}_N, \tag{42}$$

where

$$0 < b \leq (p-1)(3p+1)/(2p), \ p \geq 2. \tag{43}$$

We may even allow b to be replaced by $\phi(\bar{\underset{\sim}{X}}_N, s_N^2)$, where $\phi(\cdot)$ satisfies (40). Again note that $[N = n] \Leftrightarrow [s_k^2, k \leq n]$, so that by virtue of the independence of

$\{\bar{X}_n\}$ and $\{s_n^2\}$, given $[N = n]$, $\bar{X}_n$ has a multinormal distribution $(\theta, \frac{1}{n}\Sigma)$, independently of the s_k^2, $k \geq 2$. However, the shrinkage factor $\left(bs_N^2\left(N\|\bar{X}_N\|^2\right)^{-1}\right)$ in (42) depends on all the r.v.'s (N, $\bar{X}_N$ and s_N^2). Hence, the simple proof for (26) may not be adaptable in this more complex situation. Nevertheless, it has been shown by Sen (1989a) that by virtue of certain log-concavity property of the noncentral chi square density and the non-sequential results in Sen, Kubokawa and Saleh (1989) the following result holds.

Theorem 5.

For the class of Stein-rule estimators in (42), whenever the stopping number N satisfies (16) [with $W_n = (s_2^2, \ldots, s_n^2)$, $n \geq 2$], for every $b \in (0, (p-1)(3p+1)/2p]$,

$$P_{\theta}\left\{\left\|\delta_N^b - \theta\right\|_Q \leq \left\|\bar{X}_N - \theta\right\|_Q\right\} \geq 1/2, \forall\, \theta\, \sigma. \tag{44}$$

In passing, we may remark that a parallel dominance result under a quadratic error loss has been proved by Ghosh, Nickerson and Sen (1987). In the fixed-sample size case, the PC dominance of δ_ϕ^* in (44) has been established for an arbitrary (p.d.) Σ. On the other hand, for arbitrary Σ, the sequential case either in terms of the PCC or a quadratic error loss has not yet been resolved.

The asymptotic theory of sequential shrinkage estimation in the light of the PCC has been worked out systematically in Sen (1987a, b; 1989c, d). The basic idea is to incorporate (19) for the proposed stopping rules, verify (20) as amended in the multivariate case, and then by appeal to (35) through (41) completing the proof. Although, in the cited references, suitable quadratic error losses were used, our (35) through (41) ensure that the results remain adaptable in the PCC as well. Further, in this asymptotic setup, the covariance matrix Σ can be quite arbitrary (p.d.). In the case of a quadratic error loss, the actual asymptotic risk functions were replaced by asymptotic distributional risk functions, so that the desired dominance results could be obtained under less restrictive regularity conditions. In the case of PCC, this replacement makes no difference in the asymptotic picture, and therefore, there is no need to assume additional regularity conditions under which the asymptotic limits of the actual quadratic error loss based risks exist. In the case of shrinkage estimation, there is a technical problem in finding an asymptotically optimal stopping time, and this has been discussed in detail in Sen (1989d).

GPNC and Estimation of a Dispersion Matrix

To motivate, let us consider the problem of estimating the dispersion matrix Σ (p.d. but arbitrary) of a multinormal distribution. An unbiased estimator of Σ is $S = (n - 1)^{-1}\Sigma_{i=1}^{n}(X_i - \bar{X}_n)(X_i - \bar{X}_n)'$, where $X_1, \ldots, X_n$ are i.i.d.r. vectors and $\bar{X}_n = n^{-1}\Sigma_{i=1}^n X_i$. Note that $A = (n-1)S \sim Wishart(\Sigma, n-1, p)$. One possibility is to take $\theta = vec(\Sigma)$ and the class $\mathfrak{C}_1$ of equivariant estimators $T = vec(cA)$, $c > 0$, under the quadratic error loss function $L(T, \theta)$ as in (21) – (23). But the natural appeal for such a quadratic error loss function is not so convincing in this setup, and other forms of loss functions have been considered by various workers (viz., Haff, 1980, Sinha and Ghosh, 1987, and others). A popular choice is the so-called *entropy loss function*:

$$L(S, \Sigma) = tr(S\Sigma^{-1}) - log \mid S\Sigma^{-1}\mid - p; \tag{45}$$

a second one

$$L(S, \Sigma) = tr(S\Sigma^{-1} - I)^2 \tag{46}$$

also deserves mention. [For the estimation of the precision matrix Σ^{-1}, S^{-1} is a natural choice, and in (45) or (46), we may replace S and Σ^{-1} by S^{-1} and Σ, respectively.] Consider the class of estimation ($\mathfrak{C}_1$) of the form

$$\{cA : c > 0 \text{ and } (n-1)S \sim W(\Sigma, n-1, p)\}. \tag{47}$$

Also, consider the GPNC in (23). Then the following result is due to Khattree (1987).

Theorem 6.

Let $0 < a_2 < a_1 < 1$ and $a_iA \in \mathfrak{C}_1$, $i = 1, 2$. Also, let $c_{p,n} = med\{\chi^2_{p(n-1)}\}$. Then $a_1A >_{GPN} a_2A$ under the loss function in (45) if and only if

$$p\ log(a_1/a_2) > (a_1 - a_2)c_{p,n}. \tag{48}$$

Also, let $c^*_{p,n} = med\{\tau_p\}$ where $\tau_p = [tr(WW')] \div [tr(W)]$ and $W \sim Wishart(I, n-1, p)$. Then, under (46), $a_1A >_{GPN} a_2A$ iff

$$c^*_{p,n} < 2(a_1 + a_2)^{-1}. \tag{49}$$

Thus, if we let $a_0 = p/c_{p,n}$ and $a_0^* = 1/c^*_{p,n}$, then within the class $\mathfrak{C}_1$ of estimators of Σ, a_0A (or a_0^*A) is a unique best (in the GPNC sense) estimator of Σ under the entropy loss [or (46)], and this can not be improved within this class $\mathfrak{C}_1$.

It may be noted that $\mathfrak{C}_1$ is the class of equivariant estimators under the (full affine) group of transformations:

$$X \to a + BX,\ A \to BAB',\ B \text{ nonsingular, } a \text{ arbitrary.} \tag{50}$$

Sinha and Ghosh (1987) and Sinha (1988) also considered a larger class $\mathfrak{C}_2$ of the form:

$$\mathcal{C}_2 = \Big\{ \underset{\sim}{T}\underset{\sim}{Q}\underset{\sim}{T}' : A = \underset{\sim}{T}\underset{\sim}{T}' \sim W(\Sigma, n-1, p);$$

$$\underset{\sim}{Q} = Diag(q_1,\ldots,q_p),\ q_j > 0, \text{ for } j = 1,\ldots,p \Big\}, \tag{51}$$

and established the inadmissibility of the class $\mathcal{C}_1$ relative to the class $\mathcal{C}_2$, under various loss functions. A natural question arises in this context: Are the estimators in the class $\mathcal{C}_2$ admissible in the GPN sense? To address this problem properly, we may note that the entropy loss in (45) was first introduced in the univariate case by James and Stein (1961); in this special case, $\mathcal{C}_1 \equiv \mathcal{C}_2$ contains the class of scalar multiples of the sample variance, and hence, the PC of an estimator can as well be judged by using the usual quadratic error loss. This was accomplished by Ghosh and Sen (1989) (from the PCC point of view). This equivalence result does not, however, hold generally for the multivariate case, and hence, a different approach is needed. The class $\mathcal{C}_2$ is too big, and although for suitable subclasses of $\mathcal{C}_2$ (defined by imposing additional partial ordering), admissibility of estimators in the GPN sense can be established, such a result may not generally hold for the entire class $\mathcal{C}_2$. This is being explored in detail (viz., Sen, Nayak and Khattree, 1990). The following results are worth mentioning in this context:

(i) Within the class $\mathcal{C}_2$, no estimator of $\underset{\sim}{\Sigma}$ is GPN-optimal!

(ii) Let $\underset{\sim}{D}_2 = Diag(d_{21},\ldots,d_{2p})$ with $d_{2j}^{-1} = med(\chi^2_{n+p-2j})$, for $j = 1,\ldots,p$, and let $\hat{\underset{\sim}{\Sigma}}_2 = \underset{\sim}{T}\underset{\sim}{D}_2\underset{\sim}{T}'$. Also, let

$$\mathcal{C}_3 = \{\underset{\sim}{A} \in \mathcal{C}_2 : \underset{\sim}{Q} - \underset{\sim}{D}_2 = \text{positive semi-definite (p.s.d.)}\}; \tag{52}$$

$$\mathcal{C}_4 = \{\underset{\sim}{A} \in \mathcal{C}_2 : \underset{\sim}{D}_2 - \underset{\sim}{Q} = \text{p.s.d.}\}. \tag{53}$$

Then, within the subclass $\mathcal{C}_3$, $\hat{\underset{\sim}{\Sigma}}_2$ is GPN-optimal. Within the subclass $\mathcal{C}_4$, no estimator of Σ is GPN-optimal.

(iii) Let $\underset{\sim}{D}_1 = Diag(d_{11},\ldots,d_{1p})$ with $d_{1j} = (n + p - 2j)^{-1}$, for $j = 1,\ldots,p$, and let $\hat{\underset{\sim}{\Sigma}}_1 = \underset{\sim}{T}\underset{\sim}{D}_1\underset{\sim}{T}'$. Then, $\hat{\underset{\sim}{\Sigma}}_1$ is the James-Stein estimator of $\underset{\sim}{\Sigma}$, and its properties have already been studied by Sinha (1988). The usual estimator of $\underset{\sim}{\Sigma}$ is $\hat{\underset{\sim}{\Sigma}}_0 = (n-1)^{-1}\underset{\sim}{A}$. Then, although there is no GPN-optimal estimator of $\underset{\sim}{\Sigma}$ within the class $\mathcal{C}_2$, both $\hat{\underset{\sim}{\Sigma}}_1$ and $\hat{\underset{\sim}{\Sigma}}_2$ dominate the classical estimator $\hat{\underset{\sim}{\Sigma}}_0$ in the GPN-sense.

Acknowledgements

Thanks are due to the reviewer for his very critical reading of the manuscript leading to various improvements in the presentation.

References

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

Basu, D. (1955): On statistics independent of a complete sufficient statistic, *Sankhya* 15, 377-380.

Blyth, C. R. (1972): Some probability paradoxes in choice from among random alternatives, *J. Amer. Statist. Assoc.* 67, 366-373.

Brown, L. D., Cohen, A. and Strawderman, W. E. (1976): A complete class theorem for strict monotone likelihood ratio with applications, *Ann. Statist.* 4, 712-722.

Efron, B. (1975): Biased versus unbiased estimation, *Adv. Mathem.* 16, 259-277.

Fisher, R. A. (1938): Statistical Theory of Estimation, Calcutta University Readership Lectures.

Geary, R. C. (1944): Comparison of the concepts of efficiency and closeness for consistent estimates of a parameter, *Biometrika* 33, 123-128.

Ghosh, M. and Sen, P. K. (1989): Median unbiasedness and Pitman closeness, *J. Amer. Statist. Assoc.* 84, 1089-1091.

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

Haff, L. R. (1980): Empirical Bayes estimation of the multivariate normal covariance matrix, *Ann. Statist.* 8, 586-597.

Hwang, J. T. (1985): Universal domination and stochastic domination: Estimation simultaneously under a broad class of loss functions, *Ann. Statist.* 13, 295-314.

James, W. and Stein, C. (1962): Estimation with quadratic loss, *Proc. 4th Berkeley Symp. Math. Statist. Prob.* 1, 361-379.

Johnson, N. L. (1950): On the comparison of estimators, *Biometrika* 37, 281-287.

Keating, J. P. (1985): More on Rao's phenomenon, *Sankhya, Ser. A*, 47, 18-21.

Keating, J. P. and Gupta, R. C. (1984): Simultaneous comparison of scale parameters, *Sankhya, Ser. B*, 46, 275-280.

Keating, J. P. and Mason, R. L. (1985a): Practical relevance of an alternative criterion in estimation, *Amer. Statist.* 39, 203-205.

Keating, J. P. and Mason, R. L. (1985b): Pitman's measure of closeness, *Sankhya, Ser. B.*, 47, 22-32.

Keating, J. P. and Mason, R. L. (1988): James-Stein estimation from an alternative perspective, *Amer. Statist.* 42, 160-164.

Khattree, R. (1987): On comparison of estimators of dispersion using generalized Pitman nearness criterion, *Commun. Statist. Theor. Meth.* 16, 263-274.

Mason, R. L., Keating, J. P., Sen, P. K. and Blaylock, N. W. (1990): Comparison of linear estimators using Pitman's measure of closeness, *J. Amer. Statist. Assoc.* 85, in press.

Mason, R. L., Keating, J. P., Sen, P. K. and Rao, C. R. (1990): Pitman nearness of statistical estimators: A panel discussion on some recent developments, *Comput. Statist. & Data Anal.* 8, in press.

Nayak, T. (1990): Estimation of location and scale parameters using generalized Pitman's measure of closeness, *J. Statist. Plan. Infer.* 24, 259-268.

Peddada, S. D. (1985): A short note on Pitman measure of nearness, *Amer. Statist.* 39, 298-299.

Peddada, S. D. and Khattree, R. (1986): On Pitman nearness and variance of estimators, *Commun. Statist. Theor. Meth.* 15, 3005-3017.

Pitman, E. J. G. (1937): The closest estimates of statistical parameters, *Proc. Cambridge Phil. Soc.* 33, 212-222.

Rao, C. R. (1945): Information and accuracy attainable in the estimation of statistical parameters, *Bull. Calcutta Math. Soc.* 37, 31.

Rao, C. R. (1981): Some comments on the minimum mean square as a criterion of estimation, in *Statistics and Related Topics*, eds. M. Csorgo, D. A. Dawson, J. N. K. Rao and A. K. Md. E. Saleh, North Holland, Amsterdam, pp. 123-143.

Rao, C. R., Keating, J. P. and Mason, R. L. (1986): The Pitman nearness criterion and its determination, *Commun. Statist. Theor. Meth.* 15, 3173-3191.

Robert, C. and Hwang, J. T. (1988): A dubious criterion: Pitman closeness, preprint.

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

Sen, P. K. (1986): Are BAN estimators the Pitman closest ones too?, *Sankhya, Ser. A.*, 48, 61-68.

Sen, P. K. (1987a): Sequential Stein-rule maximum likelihood estimators: General asymptotics, in *Statistical Decision Theory and Related Topics*, IV, eds. S. Gupta and J. O. Berger, Springer-Verlag, Vol. 2, 195-208.

Sen, P. K. (1987b): Sequential shrinkage U-statistics: General asymptotics, *Rev. Brasileira Prob. Estatst.* 1, 1-21.

Sen, P. K. (1989a): On the Pitman closeness of some sequential estimators, Inst. Statist. Univ. North Carolina Mimeo Rep. No. 1788.

Sen, P. K. (1989b): Optimality of the BLUE and ABLUE in the light of the Pitman closeness of statistical estimators, Inst. Statist. Univ. North Carolina Mimeo Rep. No. 1863. To appear in *Colloq. Limit Theo. Probab. Statist.* (ed. P. Révesz).

Sen, P. K. (1989c): Asymptotic theory of sequential shrunken estimation of statistical functionals, *Proc. 4th Prague Conference Asymp. Meth. Statist.*, eds. P. Mandel and M. Hušková, Charles Univ., Prague, pp. 83-100.

Sen, P. K. (1989d): Statistical functionals, stopping times and asymptotic minimum risk property, Inst. Statist. Univ. North Carolina Mimeo Rep. No. 2002.

Sen, P. K., Kubokawa, T. and Saleh, A. K. M. E. (1989): The Stein paradox in the sense of the Pitman measure of closeness, *Ann. Statist.* 17, 1375-1386.

Sen, P. K., Nayak, T. and Khattree, R. (1990): Comparison of equivariant estimators of a dispersion matrix under generalized Pitman nearness criterion, preprint.

Sinha, B. K. (1988): Inadmissibility of the best equivariant estimators of the variance-covariance matrix, the precision matrix and the generalized variance: A survey, in *Advances in Multivariate Statistical Analysis*, eds. S. Dasgupta and J. K. Ghosh, Statist. Pub. House, Calcutta, pp. 483-497.

Sinha, B. K. and Ghosh, M. (1987): Inadmissibility of the best equivariant estimator of the variance covariance matrix and the generalized variance under entropy loss, *Statist. Dec.* 5, 201-227.

Stein, C. (1956): Inadmissibility of the usual estimator of the mean of a multivariate normal distribution, *Proc. 3rd Berkeley Symp. Math. Statist. Prob.* 1, 197-209.

Stein, C. (1981): Estimation of the mean of a multivariate normal distribution, *Ann. Statist.* 9, 1135-1151.

Wald, A. (1949): Statistical decision functions, *Ann. Math. Statist.* 20, 165-20.