OPTIMALITY OF BLUE AND ABLUE IN THE LIGHT OF THE PITMAN CLOSENESS OF STATISTICAL ESTIMATORS

by
Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1863

June 1989

OPTIMALITY OF BLUE AND ABLUE IN THE LIGHT OF THE PITMAN CLOSENESS OF STATISTICAL ESTIMATORS

PRANAB K. SEN

1. INTRODUCTION

For location-scale families of distributions, the sample order statistics provide the best linear unbiased estimators (BLUE) of the parameters, where the bestness is judged by the minimum variance (or mean square error) criterion. The exact theory stumbles into tremendous computational complexities for large sample sizes, and asymptotically best (ABLUE) estimators have been considered to facilitate these computations; an ABLUE leads to the attainment of the minimum asymptotic variance of estimators. Considerations of the smallness of mean square error or quadratic error loss dominate the scenario of BLUE and ABLUE. Other forms of linear (L-)estimators have been prompted by some robustness considerations, and their (asymptotic) minimax characters have often been studied for suitable local departures from an assumed model. The recent past has also witnessed a spur of activities on L_1-norm estimators based on order statistics , and the main idea is to enhance robustness by curbing the influence of outliers through attachment of smaller weights to larger deviations than in the case of least squares or L_2-norm estimators. In this sense, L_1-norm estimators are quite comparable to the Pitman closest ones : The Pitman closeness (or nearness) criterion is an intrinsic measure of the comparative behavior of two estimators without requiring the existence of their first or second order moments. This pair-wise comparison can be extended to suitable classes of estimators under additional regularity conditions. Since the BLUE or the ABLUE have been mostly motivated by the minimum variance considerations, it is of natural interest to inquire how these estimators behave in the light of the Pitman closeness measure ?

In the asymptotic case, there are some implications of the Pitman closeness of a general class of BAN (best asymptotically normal) estimators, and the ABLUE usually belongs to this class. Thus, Pitman closeness of ABLUE can be established under quite general regulatity conditions. On the other hand, in the multiparameter case, such BAN estimators may not enjoy the Pitman closeness property. In the multinormal models, suitable shrinkage or Stein-rule estimators are known to dominate the classical maximum likelihood estimators (MLE) in the light of the quadratic error loss as well as the Pitman closeness measure. Such shrinkage versions of L-estimators have also been considered in the literature in the recent past, and it has been observed that similar asymptotic dominance results can be obtained in some well defined sense. The current investigation focuses on this broad picture and pertains to the general limit theorems governing the study of the dominance of BLUE and ABLUE in the light of the Pitman closeness measure.

Along with the preliminary notions, some finite sample results are presented in Section 2. Asymptotic theory in the case of a single parameter is considered in Section 3. Multiparameter estimation problems are considered in Section 4, and the relation between asymptotic distributional risk (ADR) and asymptotic Pitman closeness (APC) is studied thoroughly. The concluding section deals with some general remarks .

2. PRELIMINARY NOTIONS

Let θ ($\varepsilon\ \Theta$) be a parameter of interest, and let T_1 and T_2 be two competing estimators of θ (based on a common sample). Then, we say that T_1 is Pitman-closer than T_2 if

(2.1) $$P_\theta\{\ |T_1 - \theta| \leq |T_2 - \theta|\ \} \geq 1/2, \text{ for all } \theta \,\varepsilon\, \Theta ,$$

(with the strict inequality holding for some $\theta\ \varepsilon\ \Theta$); see Pitman (1937). Note that (2.1) involves a pairwise comparison , and to extend this definition to a suitable class of estimators, we may need to consider a class $\mathcal{C}$ of estimators (T) of θ . If there exists an estimator T^o belonging to the class $\mathcal{C}$, such that

(2.2) $\quad P_\theta\{\ |T^o - \theta| \ \leq \ |T - \theta|\ \} \ \geq 1/2,\ \forall\ T \in \mathcal{C}$ and $\theta \in \Theta$,

then T^o is the **Pitman-closest** estimator of θ (within the class $\mathcal{C}$). The nice feature of this measure is that it does not require the existence of the first or second moments of the estimators (as is generally needed in L_1 or L_2-norm estimation theory). On the other hand, often, it may be difficult to construct such a class ($\mathcal{C}$) for which (2.2) holds. Moreover, the Pitman-closeness measure may not have the **transitivity** in the sense that an estimator T_1 may be Pitman-closer than T_2 and T_2 may be Pitman-closer than T_3, but T_1 may not be Pitman-closer than T_3 ; we may refer to Blyth(1972) for some nice discussions on this aspect of the Pitman closeness. For the location-scale family of distributions (where the BLUE are particularly useful), it may not be very difficult to construct such a class ($\mathcal{C}$) of estimators ; mostly, equivariance under suitable group of transformations leads to such a formulation. We may refer to Ghosh and Sen (1989) and Nayak (1989) for such characterizations. It turns out that inspite of having some limitations, the Pitman closeness measure has some natural appeal for being an being an intrinsic measure of the comparative behavior of estimators within appropriate classes.

Borrowing the basic ideas from the **weighted least squares (WLS)** theory, in BLUE, one seeks to have a linear combination of the sample order statistics having the basic requirement of unbiasedness and the minimum variance (within this class) ; thus, the existence of the second moment is a prerequisite for the BLUE. For a sample $X_1,\ldots,X_n$ of size n drawn at random from a population having a distribution function (d.f.) F , we denote the sample order statistics by $X_{n:1},\ldots,X_{n:n}$ respectively. Thus, typically a BLUE has the form

(2.3) $\quad T_n = \Sigma_{k=1}^{n} c_{nk} X_{n:k}$,

where the c_{nk} are (non-stochastic) constants (depending on the underlying form of F and the sample size n), chosen in such a way that $E_F(T_n) = \theta$ and $Var_F(T_n)$ is a minimum among all possible members of this class. Naturally, the computation of the second moment of T_n requires the computation of the variance-covariances of the $X_{n:k}$ and incorporating them in the minimization of the variance of T_n . Either task may become quite involved for large values of n. There is a more intriguing question (even for a finite sample size case) : How to formulate the BLUE in the light of the Pitman closeness criterion in (2.2) ? Or, in other words, how to choose the c_{nk} in (2.3) so that (2.2) holds ? Excepting in certain simple cases (where sufficient statistics exist), a simple answer to this question is not known. We may refer to Ghosh and Sen (1989) and Nayak (1989) for some general formulations of Pitman-closeness measures for equivariant-estimators in the location-scale families. Although this equivariance may hold for BLUE in the usual location-scale family of distributions, minimal sufficiency may not always be taken for granted ! The sufficiency and completeness of the set of order statistics (in a nonparametric setup) may not be enough to clinch the desired Pitman closeness in a parametric setup. For this reason, we shall first look into the Pitman closeness of BLUE of a single parameter θ in a simple fixed sample size case.

As pointed out in Ghosh and Sen (1989), median unbiasedness (MU) property of estimators have an important role to play in the context of the Pitman-closeness . Recall that [viz., Lehmann (1983,p.6)] an estimator T is MU for θ if

(2.4) $$P_\theta\{ T \leq \theta \} = P_\theta\{ T \geq \theta \} \quad , \text{ for all } \theta \in \Theta \quad .$$

Then, it follows from Theorem 1 of Ghosh and Sen (1989) that if C be the class of all estimators of the form $U = T + Z$, where T is a MU-estimator of θ,and T and Z are independently distributed, then

(2.5) $$P_\theta\{ |T - \theta| \leq |U - \theta| \} \geq 1/2 \text{ , for all } U \in C \text{ , and } \theta \in \Theta.$$

Typically, T is a MU-sufficient statistic, Z is ancillary , and the class C relates to suitable families of equivariant estimators. With this prescription in hand, let us examine the role of BLUE in the Pitman-closeness dominance of estimators. Typically, a BLUE is unbiased, but may not be MU (unless its distribution is symmetric about θ). For example, in the case of a normal distribution with (unknown) mean μ and variance σ^2 , the BLUE of σ is unbiased for σ , but is not generally MU . Moreover, the sample variance (a quaratic function of the order statistics) is sufficient, so that a BLUE (being a linear one) may not be so. Thus, although the BLUE of σ is unbiased and has the smallest variance among all linear unbiased estimators of σ (based on order statistics), it is not a function of the minimal sufficient statistic and in general it is not MU. Hence, the BLUE is not necessarily the Pitman closest estimator of σ (within the class of scale-equivariant estimators). On the other hand, for the mean μ , the sample mean is the BLUE and it enjoys the Pitman-closeness character too. Let us then consider a simple formulation of the Pitman closeness characterization of BLUE in the presence of sufficient statistics.

Let S_n be a sufficient statistic for the parameter θ , such that there exists a function h(.), for which $T_n = h(S_n)$ is MU for θ . Then (2.5) may be applicable to a class of estimators containing T_n . Suppose now that the BLUE of θ is also a function $g(S_n)$ of the sufficient statistic S_n ; by definition $E[g(S_n)] = \theta$. In this case, if $h(g^{-1}(.))$ is well defined then $T_n = h(g^{-1}(\text{BLUE}))$ is MU for θ and hence will have the Pitman closest character. Thus, the problem reduces to that of finding the right transformation to induce the MU property and clinch the desired Pitman closeness. As an example, we consider the following life-testing model :

Let $X_1,\ldots,X_n$ be independent and identically distributed (i.i.d.) r.v.'s with the simple exponential probability density function (p.d.f.) $f_\theta(x) = \theta^{-1}e^{-x/\theta}$, for $x \geq 0$, where $\theta \in R^+$. In the context of life-testing, Type II censored model, we observe the order statistics $X_{n:1},\ldots,X_{n:r}$, for some prefixed r ($\leq$ n) and the remaining (n-r) observations are censored. Thus, the joint distribution of the observable r.v.'s is given by

(2.6) $$\theta^{-r} \exp\{-\Sigma_{j=1}^{r} X_{n:j}\}. \exp\{-(n-r)X_{n:r}\}.[n\ldots(n-r+1)].$$

Since the $(n-i+1)[X_{n:i} - X_{n:i-1}]$, i=1,...,r (where $X_{n:0} = 0$) are i.i.d.r.v. each having the simple exponential law, it follows that the BLUE of θ is given by

(2.7) $$T_n = r^{-1}\{ X_{n:1} + \ldots + X_{n:r-1} + (n-r+1)X_{n:r} \},$$

and this is a sufficient statistic too. T_n is unbiased for θ , and moreover, for any given r($\leq$ n), rT_n/θ has the central chi square distribution with 2r degrees of freedom. Thus, there exists a positive scaler constant c_r , depending on r, such that

(2.8) c_rT_n is MU for θ .

Actually, rc_r is the median of the central chi square d.f. with 2r degrees of freedom. Thus, the minimal sufficiency of T_n and the exact distribution of rT_n/θ enable us to derive a version of the BLUE which would have the Pitman closest character. This result holds for r = n as well.

Another simple example, considered by Ghosh and Sen (1989) relates to the two-parameter uniform distribution with the p.d.f.

(2.9) $$f_{\theta,\delta}(x) = \delta^{-1} I(\theta - \delta/2 \leq x \leq \theta + \delta/2), \; \theta \text{ real and } \delta > 0.$$

In this case, the BLUE of θ is the mid-range $M_n = [X_{n:1} + X_{n:n}]/2$, and letting $R_n = [X_{n:n} - X_{n:1}]$, we know that (M_n, R_n) is a complete (and hence, minimal) sufficient statistic for (θ,δ), and within the class of translation-equivariant estimators of θ, M_n is the Pitman-closest estimator. As in the case of the first example, here also, we may consider a censoring scheme where the smallest r and largest r observations in the sample are censored, for some nonnegative r (< n/2). The Pitman closest characterization remains in tact.

The basic difference between the two examples considered above is that in the first one, the regularity assumptions pertain to the attainability of the Cramér-Rao lower bound to the variance of any estimator, while in the second case, these regularity conditions are not all tenable. However, in each case, sufficiency and MU property play the basic role. In the regular case, if the BLUE is not a sole function of a sufficient statistic, the transformation worked out before (2.6) may not be applicable, and hence, the Pitman-closest characterization of the BLUE may not hold. The normal variance model pertains to this situation.

We may conclude this section with the remark that in the finite sample case, the construction of the BLUE is so much dependent on the underlying d.f. that in the absence of sufficient statistics, the minimum variance property may not generally lead to the Pitman closest one. The transformation suggested is crucially dependent on such sufficient statistics, and in the negation of this, a case by case study may be needed to examine the desired Pitman-closeness property of BLUE. The situation is vastly different in the asymptotic case, and this will be treated in Section 3.

3. ABLUE AND PITMAN CLOSENESS PROPERTY

As has been noted earlier, the construction of BLUE demands the computation of the variances and covariances of the set of order statistics, so that for moderate to large sample sizes, one encounters a tremendous task. On the other hand, for large sample sizes, the weights in the linear combination of order statistics for the BLUE can be approximated very conveniently by some smooth 'scores', and such an estimator has been termed an ABLUE. Going back to (2.3), the basic idea is to use a smooth score function $h = \{ h(u),\ u \in (0,1)\}$, and express

(3.1) $$c_{nk} = n^{-1}h(k/(n+1)) \text{ , for } k = 1,\dots,n.$$

For this choice of the c_{nk} , we may rewrite T_n in (2.3) as a **L-estimator**

(3.2) $$L_n = n^{-1}\Sigma_{k=1}^{n} h(k/(n+1))X_{n:k} = \int_R h(\tfrac{n}{n+1}F_n(x))x\, dF_n(x) \text{ ,}$$

where

(3.3) $$F_n(x) = n^{-1}\Sigma_{i=1}^{n} I(X_i \leq x) \text{ , } x \in R$$

is the sample (or empirical) d.f. It is natural to define the parameter of interest as

(3.4) $$\theta = \int_R h(F(x))xdF(x) \text{ .}$$

Granted the usual regularity conditions under which an '**expansion**' of $h(F_n)$ around $h(F)$ is permissible, we may obtain that as n increases,

(3.5) $$n^{\frac{1}{2}}[\, L_n - \theta \,] \xrightarrow{\mathcal{D}} \mathcal{N}(\, 0 \,,\, \sigma^2(F,h) \,),$$

where

(3.6) $$\sigma^2(F,h) = 2 \int\int_{-\infty<x\leq y<\infty} F(x)[1-F(y)]h(F(x))h(F(y))dxdy.$$

There is a vast literature on this asymptotic normality of L-estimators ; we may refer to Sen (1981,Ch.7) for some accounts. [Note that the actual convergence of the mean square error of $n^{\frac{1}{2}}(L_n - \theta)$ to $\sigma^2(F,h)$ may generally require more stringent regularity conditions.] Thus, the crux of the problem is to choose the function h(.) in such a way that (3.4) holds and further that $\sigma^2(F,h)$ in (3.6) is a minimum. Such a solution leads to the so called ABLUE estimators . For the location-scale family of distributions, Jung (1955,1962) obtained explicitly the form of h(.). For example, for the location parameter θ , if F possess an absolutely continuous density function f(.) a.e. (whose first derivative is f'(.)), then defining

$K = \{K(u), 0<u<1\}$ by $(d/dx)K(F(x)) = h(F(x))$, $x \in R$, we have an optimal choice

(3.7) $$K(F(x)) = [-f'(x)/f(x)]/I(f) \ , \ x \in R \ ; \quad I(f) = \int_R (f'/f)^2 dF \ ,$$

that is, $K(F(x))$ is the usual Fisher score function. This solution is motivated by the usual minimum (asymptotic) variance criterion. There is a natural question : How this ABLUE behaves in the light of the Pitman closeness criterion ?

Judged by the asymptotic normality in (3.5) and the minimum (asymptotic) variance property of the ABLUE, we may characterize them as BAN estimators in the usual sense. As such, we may be able to use the Pitman-closest characterization of BAN estimators [viz., Sen (1986)] under additional (mild) regularity conditions. In this context, we may note that for an arbitrary $h(.)$ [not necessarily the optimal one], under suitable 'smoothness' conditions on $h(.)$ and the d.f. F , as $n \to \infty$,

(3.8) $$L_n - \theta = n^{-1}\Sigma_{i=1}^n \int_R [I(X_i \le x) - F(x)]h(F(x))dx + o_p(n^{-\frac{1}{2}}) \ ;$$

we may refer to Theorem 7.5.1 of Sen (1981) , and further relaxation of the regularity conditions has been in effect during the past few years [see Jurečková, Saleh and Sen (1989)]. As such, if we write

(3.9) $$Z_i = \int_R [I(X_i \le x) - F(x)]h(F(x))dx \ , \text{ for } i \ge 1,$$

then, we obtain that the Z_i are i.i.d.r.v's with mean zero and variance $\sigma^2(F,h)$, defined by (3.6). Moreover, we are in a parametric model, so that we may define the usual 'efficient score statistic' by

(3.10) $$S_n = \Sigma_{i=1}^n \{f_\theta'(X_i;\theta)/f(X_i;\theta)\} = \Sigma_{i=1}^n W_i, \text{ say}$$

where f_θ' refers to the derivative with respect to θ . Recall that $E[W_i] = 0$ and

(3.11) $$E[W_i^2] = \int_R [f_\theta'(x;\theta)/f(x;\theta)]^2 dF(x;\theta) = \text{Fisher information } I(\theta).$$

Moreover, defining $K(.)$ by $(d/dx)K(F(x)) = h(F(x))$ [as in before (3.7)], we have by partial integration [on (3.9)] $Z_i = K(F(X_i))$, for every $i \ge 1$, so that

(3.12) $$E[W_i Z_i] = E[K(F(X_i))f_\theta'(X_i;\theta)/f(X_i;\theta)] = \int K(F(x))f_\theta'(x;\theta)dx \ .$$

Note that for the location model, by (3.7) and (3.12), $E[W_i Z_i] = 1$ [as $f_\theta' = -f'$]. For the general parametric model, $K(F(x)) = [f_\theta'(x;\theta)/f(x;\theta)]/I(\theta)$, so that (3.12) is also equal to one. Thus, for the optimal $h(.)$, (3.12) is equal to 1. Finally, by the ABLUE character of L_n , $\sigma^2(F,h) = [I(\theta)]^{-1}$ for the optimal $h(.)$, so that

$$(3.13) \qquad \begin{pmatrix} n^{\frac{1}{2}}(L_n - \theta) \\ n^{-\frac{1}{2}} S_n \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_2\left(\underset{\sim}{0}, \begin{pmatrix} [I(\theta)]^{-1} & 1 \\ 1 & I(\theta) \end{pmatrix} \right).$$

As a result, by Theorem 2.1 of Sen (1986), we conclude that ABLUE are Pitman-closest ones when the representation in (3.8) is valid.

In the above discussion, we have made use of (3.8) which in turn may require sufficient smoothness conditions on the c_{nk} . There are two variants of the ABLUE which deserve some attention in this context. Firstly, based on a subset of selected (fixed k) order statistics, say $X_{n:m_1}, \ldots, X_{n:m_k}$ (where $m_j \simeq np_j$, $j=1,\ldots,k$) corresponding to $0 < p_1 <\ldots<p_k < 1$, there has been a lot of attempts to find out the optimal values of $p_1,\ldots,p_k$ and the corresponding coefficients $c_{n,m_1},\ldots,c_{n,m_k}$ for which an L-statistic $\Sigma_{j=1}^{k} c_{n,m_j} X_{n:m_j}$ has the smallest (asymptotic) mean square error (as an estimator of θ). Secondly, instead of all the n order statistics or a subset of selected ones , there is an intermediate stage where a subset of k_n order statistics (with $k_n \to \infty$ with $n \to \infty$) is so selected that the corresponding percentile values become dense on (0,1) as $n \to \infty$. For this situation, one may again compile an asymptotically best linear combination, and under very general regularity conditions, such an estimator shares the ABLUE properties (in terms of the asymptotic mean square error and related criteria). We would like to discuss the Pitman closeness of such estimators in general.

For the case of a selected subset of order statistics $X_{n:m_1},\ldots,X_{n:m_k}$, k fixed, we denote by

$$(3.14) \qquad \xi_j = F^{-1}(p_j) \quad , \; f_j = f(F^{-1}(p_j)), \; j= 1,\ldots,k,$$

and assume that the density f is continuous and positive at each of the points ξ_j , for $j=1,\ldots,k$. Then, by using the weaker form of the classical **Bahadur(1966) representation of sample quantiles** [due to Ghosh (1971)], we obtain that

$$(3.15) \qquad [\, X_{n:m_j} - \xi_j \,] = -f_j^{-1}[\, F_n(\xi_j) - p_j \,] + o_p(n^{-\frac{1}{2}}), \text{ for every } j =1,\ldots,k,$$

where $F_n(.)$ is the sample d.f., defined by (3.3). Thus, we have on writing

$$(3.15) \qquad T_n = \Sigma_{j=1}^{k} c_{nj} X_{n:j} \quad \text{and} \quad \theta_n = \Sigma_{j=1}^{k} c_{nj}\xi_j ,$$

that $\theta_n - \theta$ should be at least $o(n^{-\frac{1}{2}})$ (for this estimator to be asymptotically normal), and under this additional condition, we have

$$(3.16) \qquad (T_n - \theta) = - \Sigma_{j=1}^{k} c_{nj} f_j^{-1} [F_n(\xi_j) - p_j] + o_p(n^{-\frac{1}{2}})$$

$$= - n^{-1} \Sigma_{i=1}^{n} \{ \Sigma_{j=1}^{k} c_{nj} f_j^{-1} [I(X_i \leq \xi_j) - p_j] \} + o_p(n^{-\frac{1}{2}})$$

$$= n^{-1} \Sigma_{i=1}^{n} U_{ni} + o_p(n^{-\frac{1}{2}}),$$

where the U_{ni} are (row-wise) independent and i.d. r.v.'s with zero mean and finite variance. This shows that a representation similar to (3.8) holds in this case too. We may therefore repeat the arguments following (3.9) and arrive at the following:

Define the W_i as in (3.10) and $I(\theta)$ as in (3.11). Then if the following two conditions hold :

$$(3.17) \qquad \lim_{n\to\infty} \mathrm{Var}(U_{ni}) = [I(\theta)]^{-1} ,$$

$$(3.18) \qquad \lim_{n\to\infty} E[U_{ni} W_i] = 1 ,$$

the estimator T_n in (3.15) is BAN and it satisfies the conditions of Theorem 2.1 of Sen (1986), so that T_n is asymptotically Pitman closest.

In this context, we may note that (3.17) entails (3.18) (via the Fisher information), so that essentially, we have the single condition (3.17) which is also the condition for T_n to be asymptotically best in the classical sense of minimum mean square error. Thus, the bestness in the classical sense entails the Pitman-closest character in the asymptotic case. However, looking at the U_{ni} in (3.16) and W_i in (3.10), we may observe that in order that (3.18) holds , the Fisher score function $f'_\theta(X_i;\theta)/f(X_i;\theta)$ should it self be expressible as a linear combination of the indicator functions $I(X_i \leq \xi_j)$, $j=1,\ldots,k$. This is true,for example,in the case of the double exponential distribution, but, in general, for strictly monotone Fisher score functions this will not be true. For non-monotone score functions which are not constants on disjoint intervals, (3.18) may not hold [viz., Cauchy distribution], so that the use of a fixed number of sample quantiles may not lead to the Pitman closest character for the estimator T_n (nor its bestness in terms of the asymptotic mean square error).

Let us next consider the case of an increasing number of quantiles. In this situation, we would have a representation similar to (3.16) where k has to be replaced by k_n where k_n goes to ∞ as $n \to \infty$, and further, the p_j are to be replaced by a double sequence $\{p_{nj}, j \leq k_n\}$, such that

(3.19) $\max\{ p_{nj} - p_{n(j-1)} : 1 \leq j \leq k_n+1 \} \to 0$ as $n \to \infty$,

[where $p_{n0} = 0$ and $p_{n(k_n+1)} = 1$]. Although in this case, the Bahadur-representation on the whole line [0,1] may be used to obtain the parallel form of (3.16), that may entail extra regularity conditions (particularly for the p_{nj} close to 0 or 1]. One simple way of avoiding some of these extra conditions is to express T_n as a statistical functional [as in (3.2)] . In that case, the corresponding h(.) may not be that smooth as to allow the representation in (3.8) by direct "Taylor" type expansions. Nevertheless, the projection technique for linear functions of order statistics worked out by Stigler (1969) can be used with advantage, and this will lead to a form comparable to (3.8) with an appropriate $h_n(.)$ where $h_n(.)$ converges a.e. [on (0,1)] to a smooth h(.). With this, we may repeat the same arguments as in (3.17) and (3.18) and obtain the asymptotic Pitman-closest characterization of T_n under its ABLUE characterization.

In the rest of this section, we illustrate the use of this characterization in a very special case : Estimation of the scale parameter σ in a normal distributional model. Let X_i, $i \geq 1$ be i.i.d.r.v. with the normal (μ, σ^2) distribution. Based on the order statistics $X_{n:1},\ldots,X_{n:n}$ (corresponding to $X_1,\ldots,X_n$) , we consider the BLUE of σ . This is typically given by (3.2), where the c_{nk} are to be determined by the variance-covariance matrix of the $X_{n:j}$ [see for example, Sarhan and Greenberg (1962, pp.206-272)]. Note that for a fixed n, L_n is a linear combination of the $X_{n:j}$ while the optimal estimator (s_n^2) is a quadratic function of these $X_{n:j}$. Thus, although the BLUE of σ is translation-invariant and scale-equivariant, it is not a version of the minimal sufficient statistic , and hence, the characterization of the Pitman-closeness (in the exact sense) made in Section 2 is not tenable here. However, in the asymptotic case, the representation in (3.8) holds, and in this case (3.12) is equal to 1, so that the ABLUE of σ is also the Pitman-closest one. If we use a fixed subset of sample quantiles to estimate σ , although the efficiency can be quite close to 1 (depending on the appropriate p_j), neither (3.17) nor (3.18) may hold. As a result, we may not be able to claim that this estimator has asymptotically the Pitman closeness property. On the other hand, if we allow $k = k_n$ to increase with the sample size, then the Pitman-closeness property remains true in the asymptotic case (for the ABLUE). A very similar picture holds for the L-estimators of the scale parameter in the Cauchy distribution .

4. PITMAN CLOSENESS OF BLUE : MULTIPARAMETER CASE

In the univariate case, BLUE are generally considered for both the location and scale parameters in the location-scale model. In the multivariate case, for the coordinatewise location parameters,often, BLUE are considered based on the coordinatewise order statistics. In either case, the bestness of the estimators is judged in terms of the dispersion matrix : usually the 'generalized variance' criterion is used, although the **trace of the dispersion matrix** also provides a satisfactory tool. In any case, one needs a real valued function of the dispersion matrix as an yardstick , and both the generalized variance and trace criteria are members of this class and they depend only on the characteristic roots of the dispersion matrix. However, the optimal estimator may depend on the particular criterion used for defining the bestness. As for the Pitman-closeness criterion in (2.1), one needs to introduce a norm for the vector case. Here $\underset{\sim}{T}_1$ and $\underset{\sim}{T}_2$ are both p-vectors, so also is $\underset{\sim}{\theta}$. A suitable norm is the quadratic one : $||\underset{\sim}{d}||_{\underset{\sim}{Q}} = \underset{\sim}{d}'\underset{\sim}{Q}\underset{\sim}{d}$, where $\underset{\sim}{Q}$ is a given positive definite matrix. Thus, $\underset{\sim}{T}_1$ is Pitman-closer than $\underset{\sim}{T}_2$ in the norm $||\ ||_{\underset{\sim}{Q}}$ if

(4.1) $\quad P_{\underset{\sim}{\theta}}\{\ ||\underset{\sim}{T}_1 - \underset{\sim}{\theta}||_{\underset{\sim}{Q}} \leq ||\underset{\sim}{T}_2 - \underset{\sim}{\theta}\ ||_Q\ \} \geq 1/2$, for all $\underset{\sim}{\theta}$.

In this case, the characterization in (2.5) can be extended to the vector case, but that would require an extended definition of the median unbiasedness in the vector case. A sufficient condition in this respect is the following :

(4.2) $\quad \underset{\sim}{\ell}'(\underset{\sim}{T} - \underset{\sim}{\theta})$ is MU for 0, for any arbitrary p-vector $\underset{\sim}{\ell}$.

In the context of multivariate location models, the **diagonal symmetry** of the d.f. of $\underset{\sim}{T}$ (around $\underset{\sim}{\theta}$) ensures (4.2). However, this may not generally be the case. Thus, a characterization of the exact Pitman-closeness of BLUE in the multiparameter case may require rather strong regularity assumptions ; the multivariate normal mean model is a notable case. However, the asymptotic case may be worked out very conveniently.

Keeping in mind (3.2) and (3.8), we now assume that an estimator $\underset{\sim}{T}_n = (T_{n1}, \ldots, T_{np})'$ [where each component is a BLUE for the corresponding component of $\underset{\sim}{\theta}$] admits a representation of the type (3.8) [coordinatewise]. Moreover, in this case, the efficient score statistic in (3.10) is also a p-vector ($\underset{\sim}{S}_n$), as $\dot{f}_{\underset{\sim}{\theta}}$ is a p-vector. So that in (3.11) we would have a p×p matrix $\underset{\sim}{I}(\underset{\sim}{\theta})$, the Fisher information matrix. Also, in (3.12), we would have a p×p matrix, denoted by $\underset{\sim}{\Delta}$. Finally, $n^{\frac{1}{2}}(\underset{\sim}{T}_n - \underset{\sim}{\theta})$ would have a p×p dispersion matrix, denoted by $\underset{\sim}{\nu}_n$. Thus, we have a 2p×2p matrix

(4.3) $$\begin{pmatrix} \underset{\sim}{\nu}_n , & \underset{\sim}{\Delta} \\ \underset{\sim}{\Delta}' , & \underset{\sim}{I}(\underset{\sim}{\theta}) \end{pmatrix}$$

According to the multiparameter Rao-Cramér information limit, T_n will be asymptotically efficient (in the quadratic error sense) if

(4.4) $$\underset{\sim}{\nu}_n \rightarrow [\underset{\sim}{I}(\underset{\sim}{\theta})]^{-1} , \text{ as } n \rightarrow \infty .$$

Finally, by virtue of the representation [of the type (3.8)], we assume that

(4.5) $$n^{\frac{1}{2}}(\underset{\sim}{T}_n - \underset{\sim}{\theta}) \quad \overrightarrow{\mathcal{D}} \quad \mathcal{N}_p(\underset{\sim}{0} , \underset{\sim}{\nu}_n).$$

In fact, we have more :

(4.6) $$\begin{pmatrix} n^{\frac{1}{2}}(\underset{\sim}{T}_n - \underset{\sim}{\theta}) \\ n^{-\frac{1}{2}} \underset{\sim}{S}_n \end{pmatrix} \quad \overrightarrow{\mathcal{D}} \quad \mathcal{N}_{2p}\left(\underset{\sim}{0} , \begin{pmatrix} \underset{\sim}{\nu}_n , & \underset{\sim}{\Delta} \\ \underset{\sim}{\Delta}' , & \underset{\sim}{I}(\underset{\sim}{\theta}) \end{pmatrix} \right).$$

Then, by an appeal to Theorem 2.1 of Sen (1986), we may conclude that the following holds :

Suppose that (4.4) and (4.6) holds with $\underset{\sim}{\Delta} = \underset{\sim}{I}_p$, then $\underset{\sim}{T}_n$ is asymptotically **a** Pitman closest estimator of $\underset{\sim}{\theta}$. Recall that $\underset{\sim}{T}_n$ ia a BAN estimator of $\underset{\sim}{\theta}$ in the usual sense , and this characterization is confined to the class of ABLUE which are **asymptotically normally distributed.**

At this stage, we refer to the classical Stein-rule or shrinkage estimators of the multinormal mean vector. It has been pointed out by Stein (1956) that for $p \geq 3$, the sample mean vector [which is the maximum likelihood as well as the BLU estimator] of the population mean is not admissible, and during the past thirty years, scores of estimators have been considered which dominate the sample mean vector in quadratic loss. In the literature, these are known as the Stein-rule or shrinkage estimators. Such Stein-rule estimators have also been considered in nonparametric models [see Sen and Saleh (1985,1987)]. Recently, Sen, Kubokawa and Saleh (1989) have shown that for the multivariate normal mean estimation problem, the Stein-rule estimators dominate the classical BLUE in the light of the Pitman measure of closeness as well, and this dominance holds even for $p \geq 2$. Given this picture, it is of natural interest to inquire about the dominance of ABLUE in the multiparameter case in the light of the Pitman closeness criterion taking into accout plausible shrinkage factors. We shall make some concrete suggestions in this respect.

First, we may note that the Stein-rule estimators are the so called **testimators** where some pivot is conceived as the desired value of the parameter, and the estimator takes into account the divergence of the conventional estimator from the assumed pivot. Secondly, this typically involves the construction of a test statistic for the adequacy of the pivot, and this test statistic is incorporated in a form in the estimator. Because of this, the resulting estimator may not be generally equivariant (under suitable group of transformation), even though the original estimator was so. Thirdly (and most importantly), even if the original estimator had (asymptotically) a multinormal distribution, the shrinkage estimator may not have (asymptotically) multinormal distribution. For this reason, the characterization of the Pitman-closeness made after (4.6) may not be generally applicable to such shrinkage estimators. Finally, the dominance of the Stein-rule or shrinkage estimators over their classical counterparts holds only in a small neighborhood of the pivot ; the diameter of this neighborhood is $O(n^{-\frac{1}{2}})$ where n is the sample size. So in the study of the asymptotic dominance, it is generally taken for granted that the classical Pitman alternatives to the pivot form the core of plausible parameter space.

Let $\underset{\sim}{T}_n$ **be a BLUE** for $\underset{\sim}{\theta}$, such that (4.5) holds. Also, let $\{\underset{\sim}{V}_n ; n \geq n_o\}$ be a sequence of stochastic matrices such that $||\underset{\sim}{V}_n - \underset{\sim}{\nu}_n|| \rightarrow 0$, in probability, where $||.||$ refers to the max-norm. Further, let $\underset{\sim}{\theta}_o$ be the assumed pivot for $\underset{\sim}{\theta}$; without any loss of generality, we may set $\underset{\sim}{\theta}_o = \underset{\sim}{0}$. Then, consider a sequence $\{K_n\}$ of local (Pitman-type) alternatives :

(4.7) $\quad K_n : \underset{\sim}{\theta} = \underset{\sim}{\theta}_{(n)} = n^{-\frac{1}{2}} \underset{\sim}{\lambda}$, $\underset{\sim}{\lambda}$ (fixed) εR^p.

Then, a Stein-rule or shrinkage version of T_n is typically of the form

(4.8) $\quad \underset{\sim}{\delta}_n = [\underset{\sim}{I}_p - cd_n(n\underset{\sim}{T}_n'\underset{\sim}{V}_n^{-1}\underset{\sim}{T}_n)^{-1}\underset{\sim}{Q}^{-1}\underset{\sim}{V}_n^{-1}] \underset{\sim}{T}_n$; $d_n = ch_{min}(\underset{\sim}{Q}\underset{\sim}{V}_n)$,

where c : $0 < c \leq (p-1)(3p+1)/2p$ (or sometimes $2(p-2)$) is the shrinkage factor, the matrix $\underset{\sim}{Q}$ has been introduced in (4.1) and ch_{min} stands for the minimum characteristic root. Note that the asymptotic distribution of $n^{\frac{1}{2}}(\underset{\sim}{\delta}_n - \underset{\sim}{\theta})$ [even under $\{K_n\}$ in (4.7)] is not multinormal, so that for such an estimator (4.5) or (4.6) may not hold. Moreover, in the computation of the quadratic error risk, the presence of the factor $(n\underset{\sim}{T}_n'\underset{\sim}{V}_n^{-1}\underset{\sim}{T}_n)$ in (4.8) is rather disturbing ; it may blow off the asymptotic limit for the risk. For this reason, the risk is often computed from the asymptotic distribution of $n^{\frac{1}{2}}(\underset{\sim}{\delta}_n - \underset{\sim}{\theta})$, and this is termed the **asymptotic distributional risk**. Then, it follows from Sen and Saleh (1987) that in terms of this asymptotic distributional risk, the Stein-rule estimator in (4.8) dominates the classical estimator $\underset{\sim}{T}_n$ when $p \geq 3$ and $0 < c < 2(p-2)$. Further, it follows from Theorem 2 of Sen, Kubokawa and Saleh (1989) [as extended to the asymptotic case] that in the light of the Pitman closeness measure, $\underset{\sim}{\delta}_n$ in (4.8) dominates $\underset{\sim}{T}_n$ for all local alternatives (including the null case) covered under (4.7), when $p \geq 2$, and $0 < c < (p-1)(3p+1)/2$. This clearly shows that a Stein-rule estimator not belonging to the same class of equivariant estimators as $\underset{\sim}{T}_n$ can be so formulated that asymptotically it dominates $\underset{\sim}{T}_n$ in the light of the measure of Pitman-closeness. This result is of particular interest in the location-scale model (univariate case) where one may want to estimate simultaneously both the location and scale parameters (based on order statistics). In such a two-parameter case, the quadratic error risk formulation may not yield the desired dominance result, but the Pitman nearness considerations lead to the desired result. In general, for the multiparameter estimation problems, the use of the Pitman closeness measure may entail less restrictive regularity conditions than the quadratic error loss, and for the BLUE/ABLUE where asymptotic distribution theory can be derived under quite general regularity conditions, this appears to be a plus point in favor of the Pitman measure.

5. SOME CONCLUDING REMARKS

Although the BLUE/ABLUE have been mostly justified on the ground of efficiency in a particularly adopted parametric model, there has been a vast literature on L-estimators in the context of nonparametric as well as robust models. In a non-parametric case, we may have the difficulty in justifying the bestness on the attainment of the minimum mean square error , as the underying d.f. F is allowed to vary within a class and the mean square error may as well vary with this d.f. Thus, a particular L-estimator may be BLUE for a particular model but may not behave that well for some other F (imagine a true F as Cauchy vs. the assumed one as normal)! Thus, the emphasis is on the better performance over a class of F's. Even in this setup, Stein-rule or shrinkage estimators are quite attractive, as they would tend to improve the original estimator for all F belonging to a class. Thus, instead of BLUE/ABLUE, if we consider suitable L-functionals, then in the multi-parameter case, Stein-rule estimator would be prefered in the light of the Pitman closeness criterion. In robust estimation problems, often L-estimators have been justified on local robustness considerations, and in some cases, a local asymptotic minimaxity property is attributable to such L-estimators. Again, in the case of two or more parameters, if one wants to have improved estimators, shrinkage ones present better prospects. Basically, asymptotic (multi-)normality of L-estimators (in the vector case) and consistency of their dispersion estimates provide the necessary tools for the construction of such shrinkage estimators. In this context, the choice of a suitable pivot plays a very important role. Note that the dominance of the shrinkage estimators is confined only to a local neighborhood of the pivot. Hence, if an unreasonable pivot is chosen, for all practical purpose such shrinkage will lead to little improvement over their classical versions. In many practical situations, a pivot can be chosen on some natural grounds, and hence, Stein-rule or shrinkage versions remain very pertinent . Between the two criteria : quadratic error risk and Pitman closeness (based on a quadratic norm) , each one has some advantages as well as disadvantages. A quadratic error loss leads to an optimal estimator within a class whereas the Pitman closeness measure is generally a pairwise comparison, and additional regularity conditions may be necessary to extend this comparison to a class of estimators [viz., Ghosh and Sen (1989)]. The latter measure may not be generally transitive. However, from the asymptotic theory point of view, there are some distinct advantages for the PC dominance. First, a quadratic error loss demands the existenxe of second order moments of the estimators. For Stein-rule estimators [viz., (4.8)] , the presence of $(n\underset{\sim}{T}_n'\underset{\sim}{V}_n^{-1}\underset{\sim}{T}_n)^{-1}$ may cause difficulties in ensuring the existence and convergence of second order moments. This is particularly true for nonnormal models, as there $(n\underset{\sim}{T}_n'\underset{\sim}{V}_n^{-1}\underset{\sim}{T}_n)$ may not have exactly the (noncentral) chi square d.f. for which the celebrated Stein-identities would be applicable. In the PC dominance, this is not that needed, and weak convergence suffices. Further, the rate of convergence for the asymptotic distribution is generally mich better than that of the second order moments, and from that stand point, the PC dominance yields asymptotic theory which are likely to be adaptable even for moderate sample sizes. The main advantage of the Pitman measure of closeness is that the asymptotic theory may not need any extra regularity conditions over those ones needed for the asymptotic distribution theory of the usual versions, and in that respect, for BLUE/ABLUE, the PC dominance holds under the same regularity conditions pertaining to their asymptotic normality. Numerical study of the actual extent of improvement due to adoptation of the Stein-rule versions of BLUE/ABLUE in the multi-parameter case needs to be undertaken , and this is contemplated in the future.

REFERENCES

[1] R.R. Bahadur, A note on quantiles in large samples, *Ann. Math. Statist.* 37 (1966), 577-580.

[2] C. Blyth, Some probability paradoxes in choice from among random alternatives (with discussion), *J. Amer. Statist. Asso.* 67 (1972), 366-387.

[3] J.K. Ghosh, A new proof of the Bahadur representation of quantiles and an application, *Ann. Math. Statist.* 42 (1971), 1957-1961.

[4] M. Ghosh - P.K. Sen, Median unbiasedness and Pitman closeness, *J. Amer. Statist. Asso.* 84 (1989), in press.

[5] J. Jung, On linear estimates defined by a continuous weight function, *Ark. Mat.* Band 3, Nr. 15 (1955), 199-209.

[6] J. Jung, Approximation to the best linear estimates, in *Contributions to Order Statistics* (eds. Sarhan and Greenberg;J. Wiley, New York), 1962,pp.28-33.

[7] E.L. Lehmann, *Theory of Point Estimation*, J. Wiely, New york, 1983.

[8] T. Nayak, Estimation of location and scale parameters using generalized Pitman nearness criterion, *J. Statist. Plan. Infer.* (1989), in press.

[9] A.E. Sarhan - B.G. Greenberg (eds.), *Contributions to Order Statistics*, J. Wiley, New York, 1962.

[10] P.K. Sen, *Sequential Nonparametrics : Invariance Principles and Statistical Inference*, J. Wiley, New York, 1981.

[11] P.K. Sen, Are BAN estimators the Pitman-closest ones too ? *Sankhya, Ser.* A 48,(1986), 51-58.

[12] P.K. Sen, T. Kubokawa - A.K.M.E. Saleh, The Stein paradox in the sense of the Pitman measure of closeness, *Ann. Statist.* 17 (1989), in press.

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

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

[15] C. Stein , Inadmissibility of the usual estimator of the mean of a multivariate normal distribution, in *Proc. Third Berkeley Symp. Math. Statist. Probability* (univ. Calif. Press) 1 (1956), 197-206.

[16] S.M. Stigler, Linear functions of order statistics, *Ann. Math. Statist.* 40 (1969), 770-784.

Pranab K. Sen
University of North Carolina, Departments of Biostatistics & Statistics,
Chapel Hill, NC 27599-7400 (3260), USA