ARE BAN ESTIMATORS THE PITMAN-CLOSEST ONES TOO?

by

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1468

October 1984

ARE BAN ESTIMATORS THE PITMAN-CLOSEST ONES TOO ? *

By PRANAB KUMAR SEN

University of North Carolina, Chapel Hill

For estimators admitting an asymptotic representation, it is shown that asymptotic efficiency in the usual sense leads to the 'closest' character in the Pitman-sense. The results are based on the asymptotic distribution theory of regular estimators along with some elementary inequalities on non-central chi squared distributions.

1. Introduction. Let $(\mathcal{X}, \mathcal{A}, \mu)$ be an arbitrary measure space with μ sigma-finite, and let $\{X_i; i \geq 1\}$ be a sequence of independent and identically distributed (i.i.d.) random vectors (r.v.) with X_i taking values in $\mathcal{X}$ with a probability distribution $P_\theta(dx) = f_\theta(x)d\mu(x)$, $x \in \mathcal{X}$, $\theta = (\theta_1, \ldots, \theta_p)' \in \Theta \subset E^p$, the p-dimensional Euclidean space, for some $p \geq 1$. In the sequel, $\mathcal{X}$ will be taken as E^t for some $t \geq 1$ and $\mathcal{A}$ as the Borel field in $\mathcal{X}$. It may not be actually necessary to assume that these r.v.'s are i.d., and the necessary modifications can be done quite routinely. However, for the sake of simplicity of presentation, we shall stick to the i.i.d. setup. We define the log-likelihood function by

$$\text{(1.1)} \qquad \log L_n(\theta) = \Sigma_{i=1}^n \log f_\theta(X_i) \;, \; \theta \in \Theta \;, \; n \geq 1 \;.$$

We assume that the usual regularity conditions on the density f (and its partial derivatives) needed in the context of the asymptotic distribution theory of *maximum likelihood estimators (MLE)* [viz., Hájek(1970), Inagaki (1970,1973) and others] hold. Also, we define

* Work partially supported by the National Heart, Lung and Blood Institute, Contract: NIH-NHLBI-71-2243-L from the National Institutes of Health.

AMS Subject Classification Numbers: 62 G 20, 62 F 12.

Key Words and Phrases. Asymptotic efficiency, BAN estimators, closest estimators, correlated quadratic forms, information matrix, Mahalanobis norm, M-estimators, mle, regular estimators, R-estimators.

(1.2) $\xi_n(y) = n^{-\frac{1}{2}} (\partial/\partial\theta)\log L_n(\theta)|_{\theta=y}$, $y \in \Theta$,

(1.3) $I(\theta) = E\{ -(\partial^2/\partial\theta\partial\theta')\log f_\theta(X_1) \}$, $\theta \in \Theta$.

Then, there exists a sequence $\{\hat{\theta}_n\}$ of MLE of θ , such that

(1.4) $\xi_n(\hat{\theta}_n) \to 0$, in P_θ , as $n \to \infty$,

and the celebrated Hájek-Inagaki theorem also asserts that as $n \to \infty$,

(1.5) $n^{\frac{1}{2}}(\hat{\theta}_n - \theta) - I^{-1}(\theta)\xi_n(\theta) \to 0$, in P_θ .

Note that by virtue of (1.5), we also have

(1.6) $n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \sim N(0, I^{-1}(\theta))$, as $n \to \infty$.

In the sequel, any sequence of estimators satisfying (1.5) will be termed (in the usual sense) *BAN* estimators of θ .

Let $\rho(a,b)$ be a metric defined on $E^p x E^p$, and let $\{T_n\}$ and $\{T_n^*\}$ be two sequence of estimators of θ . If then for every T_n^* ,

(1.7) $P_\theta\{ \rho(T_n,\theta) \leq \rho(T_n^*, \theta) \} \geq 1/2$,

the estimator T_n is termed the *closest estimator* of θ , in the sense of Pitman (1937) . $\{T_n\}$ is termed an *asymptotically closest estimator* of θ if

(1.8) $\liminf_{n\to\infty} P_\theta\{ \rho(T_n,\theta) \leq \rho(T_n^*,\theta)\} \geq 1/2$, for every other $\{T_n^*\}$.

The object of the present study is to show that for T_n^* belonging to a general class, the BAN estimators in (1.5) are asymptotically closest . In this context, we may remark that for the uniparameter case (i.e., for $p=1$), $\rho(a,b)$ is isomorphic to $|a-b|$. However, for $p > 1$, various norms of a-b may be chosen for $\rho(a,b)$. We have chosen particularly the *Mahalanobis norm* involving the *information matrix* $I(\theta)$, i.e.,

(1.9) $\rho(a,b) = (a-b)'(I(\theta))(a-b)$.

This norm enables us to incorporate some probability inequality (on the ordering of two quadratic forms in correlated normal variables) in the derivation of the main result. As such, this inequality will be considered first in Section 2. The main result is then presented in Section 3. Some general remarks are also listed there.

2. A probability inequality on correlated quadratic forms. The main result of this section is the following.

Theorem 2.1. Let $U = (U_1', U_2')'$ be a (2p-) vector having a multinormal distribution with mean vector 0 and dispersion matrix

$$(2.1) \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad \text{with} \quad \begin{matrix} \Sigma_{12} = \Sigma_{11} \text{ positive definite (p.d.),} \\ \Sigma_{22} - \Sigma_{11} \text{ positive semi-definite (p.s.d.).} \end{matrix}$$

Then,

$$(2.2) \qquad P^* = P\{ U_1'\Sigma_{11}^{-1}U_1 \le U_2'\Sigma_{11}^{-1}U_2 \} \quad \ge \quad 1/2 ,$$

where the equality sign holds only when $\Sigma_{22} = \Sigma_{11}$.

Proof. There exists a non-singular B such that

$$(2.3) \quad B'\Sigma_{11}^{-1}B = I_p \text{ and } B'\Sigma_{22}^{-1}B = L = \text{Diag}(\ell_1, \ldots, \ell_p) ,$$

where ℓ_j is the jth characteristic root of $\Sigma_{11}\Sigma_{22}^{-1}$, $1 \le j \le p$; $0 \le \ell_p \le \ldots \le \ell_1 \le 1$. Letting $U_j = BV_j$, j=1,2 and noting that $V = (V_1', V_2')'$ has a normal distribution with null mean vector and dispersion matrix $\begin{pmatrix} I_p & I_p \\ I_p & L^{-1} \end{pmatrix}$, we have

$$(2.4) \qquad P^* = P\{ V_1'V_1 \le V_2'V_2 \}$$
$$= E[\; P\{ V_1'V_1 \le v'v \mid V_2 = v \} \;].$$

Now, given $V_2 = v$, V_1 is normal with mean Lv and dispersion matrix $I_p - L$, so that $V_1'(I_p - L)^{-1}V_1$ has the non-central chi squared distribution function (d.f.) , $H_p(x; \Delta_v)$, with p degrees of freedom (DF) and noncentrality parameter

$$(2.5) \qquad \Delta_v = v'L(I_p-L)^{-1}Lv \le \ell_1(1-\ell_1)^{-1}v'Lv = \Delta_v^* \text{ , say,}$$

where the equality sign in (2.5) holds only when $\ell_1 = \ldots = \ell_p$. Also, note that

$$(2.6) \qquad V_1'V_1 = V_1'(I_p-L)(I_p-L)^{-1}V_1 \ge (1-\ell_1)V_1'(I_p-L)^{-1}V_1 ,$$

where the equality sign holds when the characteristic roots are all equal. Therefore, we obtain from (2.4) and (2.6) that

$$(2.7) \qquad P^* \ge E[P\{ V_1'(I_p-L)^{-1}V_1 \le (1-\ell_1)^{-1}v'v \mid V_2 = v \}]$$
$$= \int H_p((1-\ell_1)^{-1}v'v; \Delta_v) \, dP(v) ,$$

with the equality sign holding when the ℓ_j are all equal. Since $V_2'LV_2$ has the d.f. $H_p(x;0)$ (= $H_p^o(x)$, say), and

(2.8) $\quad (1-\ell_1)^{-1}V_2'V_2 = (1-\ell_1)^{-1}V_2'LL^{-1}V_2 \geq \ell_1^{-1}(1-\ell_1)^{-1}V_2'LV_2$

(with the equality sign holding for $\ell_1 = \dots = \ell_p$), on noting that for every $x' \geq x \geq 0$ and $\delta' > \delta$, $p \geq 1$,

(2.9) $\quad H_p(x';\delta) \geq H_p(x;\delta) \geq H_p(x;\delta')$,

we conclude that

(2.10) $$P^* \geq \int_0^\infty H_p(\, x/\ell_1(1-\ell_1);\ \ell_1 x/(1-\ell_1))\, dH_p^o(x)$$

$$= \Sigma_{r=0}^{\infty} (r!)^{-1}(\ell_1/2(1-\ell_1))^r \int_0^\infty x^r \exp(-\ell_1 x/2(1-\ell_1))H_{p+2r}^o(x/\ell_1(1-\ell_1)).\, dH_p^o(x)$$

where we have made use of the identity that

(2.11) $\quad H_p(y;a) = \exp(-a/2)\ \Sigma_{r\geq 0}\ (a/2)^r(r!)^{-1}\ H_{p+2r}^o(y)$, $\forall\ y \geq 0$ and $a \geq 0$,

and , where in (2.10), the equality sign holds when all the ℓ_j are equal. Writing $y = x/(1-\ell_1)$ and following some standard steps, it follows that the right hand side of (2.10) is equal to

(2.12) $$(1-\ell_1)^{p/2}\ \Sigma_{r\geq 0}\ \ell_1^r\{r!\,\Gamma(p/2)\,2^{r+p/2}\}^{-1}\int_0^\infty e^{-y/2}y^{p/2+r-1}H_{p+2r}^o(y/\ell_1)dH_p^o(y)$$

$$= (1-\ell_1)^{p/2}\ \Sigma_{r\geq 0}\ \ell_1^r(\,\Gamma(p/2+r)/r!\,\Gamma(p/2)\,)\int_0^\infty H_{p+2r}^o(y/\ell_1)dH_{p+2r}^o(y)\ .$$

Now, by the hypothesis of the theorem, $\ell_1 \leq 1$, so that $H_{p+2r}^o(y/\ell_1) \geq H_{p+2r}^o(y)$, for every $y \in (0,\infty)$, where the equality sign holds when $\ell_1 = 1$, while, $\int_0^\infty H_{2r+p}^o(y)dH_{2r+p}^o(y) = 1/2$, for every $p \geq 1$ and $r \geq 0$. Thus, the right hand side of (2.12) is bounded from below by

(2.13) $$(1/2)(1-\ell_1)^{p/2}\ \Sigma_{r\geq 0}\ \ell_1^r\ (\Gamma(p/2+r)/r!\,\Gamma(p/2)) = 1/2,$$

where the equality sign holds when $\ell_1 = 1$; in the ultimate step, we have made use of the well known identity that $(1-a)^{-p/2} = \Sigma_{r\geq 0}\ a^r\,\Gamma(p/2+r)/(r!\,\Gamma(p/2)\,)$. Since the equality sign in (2.10) holds when all the ℓ_j are equal, while the lower bound in (2.13) holds when $\ell_1 = 1$, the proof of the theorem is completed by noting that $P^* \geq 1/2$, where the equality sign holds when $\ell_1 = \dots = \ell_p = 1$, i.e., $\Sigma_{11} = \Sigma_{22}$.

In passing, we may remark that in (2.2), we may replace Σ_{11}^{-1} by Σ_{22}^{-1} or Σ^{-1}, where Σ is any convex combination of Σ_{11} and Σ_{22}, and the inequality $P^* \geq 1/2$ (along with the condition for the attainment of the lower bound) remain in tact.

3. Pitman-closest characterization. We consider here BAN estimators of for which (1.5)-(1.6) hold, though $\hat{\theta}_n$ need not be the mle. Also, let $\mathcal{T}$ be the class of all estimators $\{T_n\}$, for which

$$(3.1) \qquad n^{\frac{1}{2}}\begin{pmatrix} T_n - \theta \\ \xi_n(\theta) \end{pmatrix} \sim N_{2p}\left(0 , \begin{pmatrix} \Sigma & I_p \\ I_p & \mathcal{I}(\theta) \end{pmatrix} \right); \ \Sigma\mathcal{I}(\theta)-I_p = \text{p.s.d.}$$

Then, we have the following.

Theorem 3.1. For $\rho(a,b)$ defined by (1.9) and for every $\{T_n\} \in \mathcal{T}$,

$$(3.2) \qquad \liminf_{n\to\infty} \ P_\theta\{ \rho(\hat{\theta}_n,\theta) \leq \rho(T_n,\theta) \} \geq 1/2 ,$$

where $\{\hat{\theta}_n\}$ is any BAN estimator of θ, and the equality sign in (3.2) holds when $\Sigma\mathcal{I}(\theta) = I_p$.

Proof. Note that by virtue of (1.5) and (3.1), for any BAN estimator $\hat{\theta}_n$ and T_n belonging to the class $\mathcal{T}$,

$$(3.3) \qquad n^{\frac{1}{2}}\begin{pmatrix} \hat{\theta}_n - \theta \\ T_n - \theta \end{pmatrix} \sim N_{2p}\left(0 , \begin{pmatrix} \mathcal{I}^{-1}(\theta) & \mathcal{I}^{-1}(\theta) \\ \mathcal{I}^{-1}(\theta) & \Sigma \end{pmatrix} \right) ; \ \Sigma\mathcal{I}(\theta)-I_p = \text{p.s.d.}$$

Therefore, the proof follows directly by appealing to Theorem 2.1 [along with the Sverdrup (1952) theorem , which justifies the asymptotic treatment], and hence, the details are omitted.

As has been noted in Section 2, in the definition of $\rho(a,b)$ in (1.9), one may replace $\mathcal{I}(\theta)$ by Σ^{-1} or any convex combination of $\mathcal{I}(\theta)$ and Σ^{-1}. However, the choice of $\mathcal{I}(\theta)$ appears to be the most natural one (in view of the basic assumption that for $T_n \in \mathcal{T}$, $\Sigma\mathcal{I}(\theta) - I_p$ is p.s.d.). In the uniparameter case (i.e., p=1), one may take $\rho(a,b) = |a-b|$, as was originally considered by Pitman (1937).

To appreciate fully the scope of this characterization, it may be quite appropriate to examine the regularity conditions concerning the class $\mathcal{T}$ of estimators. Note that if T_n is an unbiased estimator of θ , then

(3.4) $\int (T_n - \theta)L_n(\theta)d\mu_n = 0$,

so that if differentiation (with respect to the elements of θ) is permissible under integration (with respect to the product measure μ_n), we have

(3.5) $I_p \{ -\int L_n(\theta)d\mu_n\} + \int n^{\frac{1}{2}}(T_n-\theta)\xi_n(\theta)L_n(\theta)d\mu_n = 0.$

Therefore, the actual covariance of $n^{\frac{1}{2}}(T_n-\theta)$ and $\xi_n(\theta)$ is the identity matrix I_p. In fact, the unbiasedness of T_n is not very crucial in the asymptotic setup. If $ET_n = b_n(\theta) + \theta$, and the bias (vector) $b_n(\theta)$ is differentiable (with respect to θ) and the pxp matrix

(3.6) $(\partial/\partial\theta)b_n(\theta)$ converges to a null matrix , as $n \to \infty$,

then the right hand side of (3.5) , instead of being 0, will also converge to a null matrix, and hence, the covariance of $n^{\frac{1}{2}}(T_n -\theta)$ and $\xi_n(\theta)$ converges to I_p, as $n \to \infty$. Also, for any unbiased estimetor T_n of θ , it follows from a general result in Rao (1973,p.265) that $nE[(T_n-\theta)(T_n -\theta)'] - \mathcal{I}^{-1}(\theta)$ is p.s.d., and under (3.6), the same result extends to possibly biased estimators in an asymptotic setup. Hence, the two conditions on the covariance matrix of the asymptotic multinormal law in (3.2) hold under the usual regularity conditions pertaining to the validity of (1.5). However, to justify these two regularity conditions, it is not necessary to impose the existence of the first and second order moments of $\{T_n\}$. It may be enough to assume that there exists a sequence $\{T_n^*\}$ of random vectors and a sequence $\{\upsilon_n\}$of stochastic vectors, such that υ_n converges in probability to 0 as $n \to \infty$, and

(3.7) $n^{\frac{1}{2}}(T_n - \theta) = T_n^* + \upsilon_n$ where ET_n^* exists and $ET_n^*T_n^{*'}$ satisfy (3.1).

Indeed, a representation of $\{T_n^*\}$ in terms of independent summands leads to an easy avenue towards the verification of the joint normality in (3.2). Hence, we proceed on to make further comments on such a representation.

Hájek(1970), Inagaki(1970,1973) and others considered a general class of estimators which may be presented in terms an *estimating function* $\xi_n^*(\theta) = n^{-\frac{1}{2}}\Sigma_{i=1}^n \eta(X_i,\theta)$ where $\lambda(\theta) = E\eta(X_i,\theta)$ has continuously differentiable (with respect to θ) elements and the $\eta(X_i,\theta)$ have a finite, p.d. covariance matrix. If T_n be a system of solution of this system of equations, i.e.,

(3.8) $\qquad \xi_n^*(T_n) \rightarrow 0$, in P_θ ,

then, by appealing to the Hájek-Inagaki theorem, we conclude that under quite general regularity conditions,

(3.9) $\qquad n^{\frac{1}{2}}(T_n - \theta) = -\Lambda^{-1}(\theta)\,\xi_n^*(\theta) + o_p(1)$, as $n \rightarrow \infty$, (in P_θ)

where $\Lambda(\theta) = (\partial/\partial\theta)\lambda(\theta)$ is assumed to be of full rank (p). In particular, this representation applies to the estimators obtained by the method of least squares as well as method of moments. More commonly, the *M-estimators* [c.f. Huber (1981)] also satisfy (3.9) under quite general conditions ; we may refer to Sen (1981, Ch.8) and Jurečková (1984) for some specific models. For ranked based (i.e. R-) estimators, though the estimating function is of a form different from $\xi_n^*(\theta)$, an asymptotic representation similar to (3.9) holds, i.e., we have

(3.10) $\qquad n^{\frac{1}{2}}(T_n - \theta) = n^{-\frac{1}{2}}\Sigma_{i=1}^n\ g_\theta(X_i) + o_p(1)$,

where $E_\theta g_\theta(X_i) = 0$ and $g_\theta(X_i)$ has a p.d. covariance matrix for which (3.1) holds. We may again refer to Sen (1981, Ch.8) for some specific models. For estimators based on linear functions of combinations of order statistics, an asymptotic representation similar to (3.10) holds under very general regularity conditions [see Chapter 7 of Sen(1981)]. Note that the first term on the right hand side of (3.9) can also be expressed as in the first term on the right hand side of (3.10). Hence, for all such estimates, we may interpret (3.10) as an asymptotic representation in terms of average over independent summands. We denote by $\mathcal{T}_0$ the entire class of estimators for which (3.10) holds , where the dispersion matrix of the $g_\theta(X_i)$ satisfies the condition in (3.1). Note that for any estimator T_n belonging to $\mathcal{T}_0$, $(\, g_\theta(X_i),\ (\partial/\partial\theta)\log f_\theta(X_i)\,)'$, i=1,...,n,.. are i.i.d.r.v., so that by virtue of the classical central limit theorem (in the multivariate case) and (1.1), (1.2) and (3.9), we conclude that the asymptotic (joint) normality in (3.1) holds. Thus, we have $T_0 \subset T$, and this leads to the following.

Theorem 3.2. The asymptotic Pitman-closest characterization in Theorem 3.1 holds for the class of estimators (T_0) for which an asymptotic representation as in (3.10) holds.

In a finite sample setup, mostly, for location and scale parameters, Pitman (1937) characterized the closest property in terms of sufficient statistics (assuming the existence of the latter), while, the current results provide some extension of his characterization in an asymptotic setup where the existence of sufficient statistics is not required , nor θ be necessarily a location/scale parameter ; the asymptotic sufficiency of mle or BAN estimators provides the link in this context.

We conclude this section with a remark that Kaufman(1966), Hájek(1970) and Inagaki(1970,1973) considered the 'closeness' of an estimator in the following sense: Let S be any symmetric (about the origin) and convex subset in R^p and let C be a class of estimators of θ , such that

$$(3.11) \qquad \lim_{n\to\infty} P_\theta\{n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \in S\} \geq \lim_{n\to\infty} P_\theta\{n^{\frac{1}{2}}(T_n - \theta) \in S\} ,$$

for every $\{T_n\} \in C$. Then, $\{\hat{\theta}_n\}$ has the asymptotic closest character (relative to the class C). The Hájek-Inagaki theorem characterizes the class C for which the mle enjoys the property in (3.11). Their C and our T are not strictly the same, though for both of them, $\Sigma I(\theta) - I_p$ is p.s.d., and this , in view of the multivariate Cramér-Rao inequality, is not a very restrictive condition. While the proof of (3.11) exploits the basic Anderson (1955) inequality, we are not in a position to use the same directly in the proof of our theorems. However, our Theorem 2.1 provides a visible and directly verifiable proof of the Pitman-closest characterization in the asymptotic case.

REFERENCES

[1] Anderson, T.W. (1955). The integral of a symmetric unimodal function. *Proc. Amer. Math. Soc. 6*, 170-176.

[2] Hajek, J. (1970). A characterization of limiting distributions of regular estimates. *Zeit. Wahrsch. verw. Geb. 14*, 323-330.

[3] Huber, P.J. (1981). *Robust Statistics*. New York : John Wiley.

[4] Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniform property. *Ann. Inst. Statist. Math. 22*, 1-13.

[5] Inagaki, N. (1973). Asymptotic relations between the likelihood estimating function and the maximum likelihood estimator. *Ann. Inst. Statist. Math. 25*, 1-26.

[6] Jureckova, J. (1984). Robust estimators of location and their second order asymptotic relations. *Internat. Statist. Rev. (Centennial Vol.), in press*

[7] Kaufman, S. (1966). Asymptotic efficiency of the maximum likelihood estimator. *Ann. Inst. Statist. Math. 18*, 155-178.

[8] Pitman, E.J.G. (1937). The "closest" estimates of statistical parameters. *Proc. Cambridge Phil. Soc. 33*, 212-217.

[9] Rao, C.R. (1973). *Linear Statistical Inference and its Applications*. New York : John Wiley.

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

[11] Sverdrup, E. (1952). The limit distribution of a continuous function of random variables. *Skand. Aktuar. 35*, 1-10.