Generalized Bayes Estimators in Multivariate Problems

by

James O. Berger[1] and C. Srinivasan
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #481

March 1977

[1]Research supported by the National Science Foundation under grant #MCS 76-06627.

Abstract

Several problems involving multivariate generalized Bayes estimators are investigated. First, a characterization of admissible estimators as generalized Bayes estimators is developed for certain multivariate exponential families and quadratic loss. The problem of verifying whether an estimator is generalized Bayes is then considered. Conditions are also developed for an estimator to be "nearly" generalized Bayes, in the sense that if it were properly smoothed it would be generalized Bayes. These conditions demand a proper "orientation" of the estimator. Applications are given to a number of minimax, empirical Bayes, and ridge regression estimators of current interest.

The paper concludes with the development of an asymptotic approximation to generalized Bayes estimators for general losses and location vector densities. Using this approximation, weakened versions of the above results are obtained for general losses and densities.

Section 1. Introduction

Let $X = (X_1, X_2, \ldots, X_p)^t$ be a random vector from a p-dimensional density $f(x,\theta)$ with respect to some σ-finite Borel measure μ on R^p. It is desired to estimate $\theta \in R^p$ by an estimator $\delta(X) = (\delta_1(X), \ldots \delta_p(X))^t$, under a non-negative loss $L(\delta-\theta)$. If G is a nonnegative σ-finite Borel measure on R^p, the generalized Bayes estimator $\delta^G(x)$, with respect to G, is defined as the vector $c = (c_1, \ldots c_p)^t$ which minimizes

$$(1.1) \qquad \int L(c-\theta) f(x,\theta) G(d\theta).$$

(Only situations where (1.1) has a unique minimum will be considered.)

Section 2 deals with the important situation where f is from an exponential family with natural parameter θ, and where L is a quadratic loss. Results of Sacks (1963) and Brown (1971) are generalized to show that, under certain conditions, an admissible estimator of θ must be generalized Bayes. Easily verifiable necessary and sufficient conditions for an estimator to be generalized Bayes are also developed. These provide a fairly quick check on the potential admissibility of a proposed estimator. These conditions generalize certain results of Strawderman and Cohen (1971) (which dealt with the normal distribution for spherically symmetric estimators).

Inadmissibility due to a lack of being generalized Bayes is, unfortunately, not necessarily a compelling criticism of an estimator. For example, when f is a normal density and L a quadratic loss, it is well known that the positive part Stein estimator

$$(1.2) \qquad \delta(x) = (1-(p-2)/|x|^2)^+ x$$

(where "+" stands for the positive part and $|x|$ denotes the Euclidean norm of

x) is inadmissible because it cannot be generalized Bayes. Nevertheless, significantly better estimators appear unlikely to exist (Efron and Morris (1973)), so the simplicity of (1.2) makes it attractive. In a sense, the difficulty with (1.2) is simply a lack of smoothness. It would be useful to distinguish between estimators which are simply not quite smooth enough, and between those which in a more fundamental way are not generalized Bayes.

This problem assumes major importance due to the recent large literature on estimation in location vector problems (particularly the normal). Many alternatives to the usual best invariant estimator $\delta(x) = x + c^o$ (where c^o is chosen to minimize $\int L(c^o + \theta)f(\theta)d\theta$) have been proposed. Typically, these estimators are of the form

$$(1.3) \qquad \delta(x) = x + c^o - h(x^tCx)Bx + \tilde{o}(x)|x|h(x^tCx),$$

where B and C are $(p\times p)$ matrices, C being positive definite, h is a positive real valued function, and $\tilde{o}(x)$ is some $(p\times 1)$ vector whose norm goes to zero as $|x| \to\infty$. For example, the minimax estimators in Hudson (1974), Bock (1975), and Berger (1976d); the empirical Bayes estimators in Efron and Morris (1973), Rolph (1976), Rao (1977) and elsewhere; the adaptive ridge regression estimators in Lawless and Wang (1976), Thisted (1976), and Casella (1977); and the tail minimax estimators in Berger (1976a), can all be shown to be of the form (1.3) with $h(y) = 1/y$. A significant question one can ask about such estimators is - how must B and C be related in order for the estimator to be generalized Bayes? This considers not the smoothness of the estimator, but instead the "directional" orientation of the estimator. Estimators which cannot be smoothed to make them generalized Bayes usually suffer from some such orientation problem. With this motivation, we define an estimator $\delta(x) = x + c^o + \gamma(x)$ to be <u>directionally consistent ($\mathcal{G}$)</u> if there exists a generalized Bayes estimator δ^G, with $G \in \mathcal{G}$

(some appropriate class of generalized priors), such that

(1.4) $$\delta^G(x) = \delta(x) + \tilde{o}(x)|\gamma(x)| ,$$

where $\tilde{o}(x)$ has the same meaning as in (1.3). In other words, the estimator δ is directionally consistent if it can be smoothed by the addition of an error term (of smaller order than the correction $\gamma(x)$ which δ makes to the usual estimator $(x + c^o)$) to make it generalized Bayes. An estimator δ is directionally inconsistent ($\mathcal{G}$) if there does not exist a generalized Bayes estimator δ^G (with $G \in \mathcal{G}$) such that (1.4) holds.

In Section 3 it is shown for the normal-quadratic loss problem, that an estimator of the form (1.3) (with h decreasing as a polynomial or faster) is directionally inconsistent (for an appropriate $\mathcal{G}$) unless $B = k \ddagger C$ for some constant k. Estimators violating this condition are thus inadmissible in a manner more serious than mere lack of smoothness.

In Section 4, similar results will be obtained for nonquadratic losses and for general location vector densities, although in a weakened version. The weakening is due to the fact that only appropriately smooth generalized priors can be handled, and also to the lack of a characterization of admissible estimators as generalized Bayes. The results in Section 4 will depend on an approximation for generalized Bayes estimators which is developed along the lines of the heuristic argument in Brown (1974). This approximation is of independent interest, as it appears to be a necessary component of investigations into admissibility of generalized Bayes estimators. (See Brown (1974).)

In the remainder of the paper, the following notation will be used. If $r(x)\colon R^p \to R^1$ is an appropriately differentiable function, let

$$r^{(i)}(x) = \frac{\partial}{\partial x_i} r(x), \; r^{(i,j)}(x) = \frac{\partial^2}{\partial x_i \partial x_j} r(x), \text{ etc.}$$

Also let $\nabla r(x) = (r^{(1)}(x), \ldots r^{(p)}(x))^t$ denote the gradient of r. Finally, let E_θ stand for expectation with respect to X. If the argument of the expectation is a vector or matrix, the expectation is to be taken componentwise.

Section 2. Exponential Families and Quadratic Loss

Assume $f(x,\theta)$ is from the exponential family of probability distributions with respect to μ, i.e. X has density $f(x,\theta) = \beta(\theta)\exp(\theta^t x)$ with respect to μ. Let $\Theta = \{\theta: \int \exp(\theta^t x)\mu(dx) < \infty\}$ denote the convex natural parameter space, S denote the support of μ, K denote the convex hull of S, ∂A denote the boundary of a convex set A, int A denote the interior of a convex set A, and $\bar{A}$ denote the closure of a set A. Also, if G is a σ-finite Borel measure on R^p, let

$$\hat{G}(x) = \int \exp(y^t x) G(dy)$$

denote the p-dimensional Laplace transform of G. It is well known that $\hat{G}(x)$ is infinitely differentiable on the interior of the convex set $T(G) = \{x: \hat{G}(x) < \infty\}$, and that the partial derivatives can be taken inside the integral sign.

It will be assumed that L is the quadratic loss $L(\delta-\theta) = (\delta-\theta)^t Q(\delta-\theta)$, Q being positive definite. A straightforward calculation then verifies that if δ is generalized Bayes with respect to a σ-finite measure G', and $x \in T(G)$ where $G(d\theta) = \beta(\theta)G'(d\theta)$, then

(2.1) $$\delta(x) = \frac{\nabla \hat{G}(x)}{\hat{G}(x)} = \nabla \log \hat{G}(x)$$

for almost all x. ("almost all x" is, of course, with respect to μ.)

The following theorem gives the major necessary condition for admissibility of an estimator δ. The theorem is analagous to results in Sacks (1963) (which dealt with p=1) and to results in Brown (1971) (which considered the normal density). See also Farrell (1966). The proof is very similar to the proof in Brown (1971).

THEOREM 2.1. Let $\delta(x)$ be an admissible estimator. Then there exists a σ-finite measure G_o (with support in $\bar{\Theta}$) such that if $x \in \text{int } K$ then $\hat{G}_o(x) < \infty$, and for almost all $x \in \text{int } K$

$$\delta(x) = \nabla \log \hat{G}_o(x). \tag{2.2}$$

PROOF. If δ is admissible, it follows from Farrell (1968) (see also Stein (1955)) that there exists a sequence of finite measures G_n' with compact support such that $G_n'(\Gamma) \geq 1$ for some compact $\Gamma \subset \Theta$, and such that

$$\int_{\Theta}\int \{ L(\delta^n(x)-\theta)-L(\delta(x)-\theta)\}\beta(\theta)\exp(\theta^t x)\mu(dx)G_n'(d\theta) \to 0, \tag{2.3}$$

where δ^n is the Bayes estimator with respect to G_n'. Defining $G_n(d\theta) = \beta(\theta)G_n'(d\theta)$, it follows from the fact that G_n' has compact support that $T(G_n) = R^p$. Hence by (2.1), for almost all x

$$\delta^n(x) = \frac{\nabla \hat{G}_n(x)}{\hat{G}_n(x)}. \tag{2.4}$$

Using (2.3) and (2.4), a calculation as in Stein (1960) and Brown (1971) shows that

$$\int | \delta^n(x) - \delta(x) |^2 \hat{G}_n(x)\mu(dx) \to 0. \tag{2.5}$$

Since $G_n'(\Gamma) \geq 1$, it follows for some $K_1 > 0$ and K_2, that

$$\hat{G}_n(x) = \int \beta(\theta)\exp(\theta^t x)G_n'(d\theta) > K_1\exp(-K_2|X|).$$

Hence (2.4) and (2.5) imply that there exists a subsequence n' such that except for $x \in N$ (where $\mu(N) = 0$),

$$\frac{\nabla\hat{G}_{n'}(x)}{\hat{G}_{n'}(x)} \to \delta(x). \tag{2.6}$$

Define $M = \{x \in S\colon \delta(x) \text{ is finite}\}$. Clearly $\mu(M^c) = 0$, and hence int K is contained in the convex hull of (M-N). It thus follows that a sequence $\{p_m\}$ of closed convex polyhedra with vertices in (M-N) can be chosen, such that $\bigcup_{m=1}^{\infty} p_m \supset \text{int } K$. Furthermore, since $A_m = \{\text{vertices of } p_m\}$ is finite, there exist constants $b_m < \infty$ such that $|\delta(x)| < b_m$ for $x \in A_m$. From (2.6) it can be concluded that if $x \in A_m$, then

$$\frac{|\nabla\hat{G}_{n'}(x)|}{\hat{G}_{n'}(x)} < b_m' < \infty.$$

Theorem 2.2.1 of Brown (1971) (a continuity theorem for Laplace transorms) can thus be applied, and together with a standard diagonal argument shows that there exists a subsequence n'' of n' and a measure $G_o(d\theta)$ (with support in $\bar{\Theta}$) such that for all $x \in \text{int } K$, $\hat{G}_o(x) < \infty$ and

$$\frac{\nabla\hat{G}_{n''}(x)}{\hat{G}_{n''}(x)} \to \frac{\nabla\hat{G}_o(x)}{\hat{G}_o(x)}.$$

Together with (2.6), this completes the proof.||

The following corollaries are immediate.

COROLLARY 2.2. In the situation of Theorem 2.1 if $G_o(\partial\Theta) = 0$ or Θ is closed, then for almost all $x \in \text{int } K$, $\delta(x)$ is generalized Bayes with respect to the measure $g_o(d\theta) = G_o(d\theta)/\beta(\theta)$.

COROLLARY 2.3. If $\mu(\partial K) = 0$ and Θ is closed, then every admissible estimator is generalized Bayes and is given by (2.1) for some measure G. In particular, when $K = \Theta = R^p$, the result holds.

COROLLARY 2.4. Assume X has density $f(x,\theta) = \beta(\theta)\exp(\theta^t \Sigma^{-1} x)$ with respect to μ, where Σ is a known positive definite matrix. If $\delta(x)$ is an admissible estimator of θ, then for almost all $x \in \text{int } K$

(a.) $\delta(x) = \Sigma \nabla \log \hat{G}(x)$ for some σ-finite measure G with support in $\overline{\Theta}^*$ ($\Theta^* = \{\theta: \int \exp(\theta^t \Sigma^{-1} x)\mu(dx) < \infty\}$).

(b.) $\delta(x) = x + \Sigma \nabla \log g(x)$, where $g(x) = \exp(-x^t \Sigma^{-1} x/2)\hat{G}(x)$.

PROOF. Part (a.) follows immediately from Theorem 2.1 and the observation that the natural parameter is $\Sigma^{-1}\theta$. ($G(d\theta)$ will be $G_o(d(\Sigma^{-1}\theta))$.) Part (b.) is an easy calculation from (a.) .||

In applications, the above corollary is very convenient to work with, especially when f is normal with mean θ and covariance matrix Σ. In verifying whether or not a given estimator is admissible, there are two necessary conditions implied by Corollary 2.4.

Condition 1. There exists an infinitely differentiable function h such that for almost all $x \in \text{int } K$, $\delta(x) = \Sigma \nabla h(x)$.

Condition 2. $\exp(h(x))$ is a Laplace transform of some measure G on $\overline{\Theta}^*$.

The verification of Condition 2 is a well known mathematical problem. (See Widder (1946).) When f is normal, more explicit means of verification are given in Hirschmann and Widder (1955). (See also Strawderman and Cohen (1971).) Condition 2 can loosely be interpreted as requiring that the estimator δ be smooth enough. It is Condition 1 which demands a proper "orientation" of the estimator as discussed in Section 1. Note that Condition 1 is easily verified when p=1 or δ is spherically symmetric, as $\delta(x)$ can then be simply integrated to obtain h. The verification of Condition 1 turns out to be relatively simple also when int $K \subset S$, as the following development shows.

Assume $x^o \in S$, and let $A_n = \{x: |x-x^o| < \frac{1}{n}\}$. Define (if it exists)

$$\delta^*(x^o) = \lim_{n\to\infty} \frac{1}{\mu(A_n)} \int_{A_n} \delta(x)\mu(dx). \tag{2.7}$$

If Condition 1 is satisfied, then for $x^o \in (\text{int } K) \cap S$,

$$\delta^*(x^o) = \lim_{n\to\infty} \frac{1}{\mu(A_n)} \int_{A_n} \Sigma\nabla\, h(x)\mu(dx) = \Sigma\nabla\, h(x^o) \tag{2.8}$$

by the continuity of ∇h. Clearly δ^* is equivalent to δ, so it suffices to consider whether or not δ^* satisfies Condition 1. The following well known calculus lemma is needed.

LEMMA 2.5. A continuously differentiable vector field $F(x) = (F_1(x),\ldots F_p(x))^t$ defined on an open convex subset of R^p is a gradient field (i.e. $F(x) = \nabla h(x)$ for some h) if and only if the Jacobian matrix J(F) of F is symmetric. (J(F) has (i,j) element $F_i^{(j)}(x)$.)

The following theorem is an immediate consequence of (2.7), (2.8), and Lemma 2.5.

THEOREM 2.6. If int $K \subset S$, then Condition 1 is satisfied if and only if

$(\Sigma^{-1}\delta^*)$ is continuously differentiable in int K and has a symmetric Jacobian matrix.

Theorem 2.6 gives an easily verifiable necessary condition for an estimator δ to be admissible. Providing δ (or δ^*) **satisfies** the conditions in Theorem 2.6, one can construct h(x) by simply integrating $(\Sigma^{-1}\delta^*)$ along any path from a fixed point $x^o \in$ int K to x.

The following corollary gives an example of the use of Theorem 2.6.

COROLLARY 2.7. Assume f is normal with mean θ and known covariance matrix Σ. Consider estimators of the form

$$\delta(x) = x - h(x^tCx)Bx,$$

where C is positive definite and B is nonsingular. Assume (w.l.o.g.) that δ equals δ^*. For δ to be admissible, h must be continuously differentiable. If in addition h'(y) (the derivative of h) is nonzero for some $y_o > 0$, then a necessary condition for δ to be admissible is that $B = k\,\Sigma\,C$ for some constant k.

PROOF. By Theorem 2.6 it is only necessary to determine if $J(\Sigma^{-1}\delta)$ is symmetric. A calculation shows that

$$J(\Sigma^{-1}\delta) = \Sigma^{-1} - h(x^tCx)\Sigma^{-1}B - 2h'(x^tCx)\Sigma^{-1}B(xx^t)C.$$

Let $\Gamma = \{x\colon\ x^tCx = y_o\}$. Clearly for $x \in \Gamma$, $\Sigma^{-1}B(xx^t)C$ is not constant, and $h'(x^tCx) \neq 0$. Hence for J to be symmetric, it must be true that $\Sigma^{-1}B$ is symmetric and $[\Sigma^{-1}B(xx^t)C]$ is symmetric. $\Sigma^{-1}B$ is symmetric only if $B = k\,\Sigma\,A$ for some symmetric matrix A. Then $[A(xx^t)C]$ is symmetric for all $x \in \Gamma$ only if $A = C$.||

Section 3. Directional Inconsistency

In this section it will be assumed that f is a normal density with mean θ and known covariance matrix $\ddagger$, and that L is a quadratic loss as in Section 2. In this situation, as mentioned in Section 1, estimators of the form

$$\delta(x) = x - h(x^tCx)Bx + \tilde{o}(x)|x|h(x^tCx) \tag{3.1}$$

(where B, C, h, and $\tilde{o}(x)$ are as in (1.3)) have come under considerable scrutiny as competitors to the usual best invariant estimator $\delta^o(x) = x$. The results of Section 2 can be used to show whether or not such estimators are generalized Bayes (and hence potentially admissible), provided the $\tilde{o}(x)$ term is known explicitely. Section 2 is not useful, however, in determining directional consistency (i.e. approximability by a generalized Bayes estimator as defined in Section 1). The following theorem gives conditions under which an estimator of the form (3.1) is directionally inconsistent ($\mathcal{G}_o$), where

$$\mathcal{G}_o = \{G: \ g^*(x) = \int f(x,\theta)G(d\theta) < \infty\}.$$

(All admissible estimators are generalized Bayes with respect to some $G \in \mathcal{G}_o$ (Brown (1971)), so this is the natural class of G to consider.)

THEOREM 3.1. Assume

$$\delta(x) = x - h(x^tCx)Bx + \tilde{o}(x)|x|h(x^tCx), \tag{3.2}$$

where C is positive definite, B is any $(p\times p)$ matrix, and h satisfies the following condition: there exist $K > 0$, $0 < \rho < 1$, and $0 < \tau < 1$, such that $h(z) \le \tau h(y)$ for any positive numbers y and z satisfying $y \ge K$ and

$z > (1+\rho)y$, and $h(y)$ is continuous and positive for $y \geq K$. Then δ is directionally inconsistent $(\mathcal{G}_o)$ unless $B = k \ddagger C$ for some constant k. (Remark: The condition on h essentially says that $h(z)$ must be decreasing at least as fast as z^{-a} for some $a > 0$.)

PROOF. By definition, $\delta(x) = x + \gamma(x)$ is directionally consistent $(\mathcal{G}_o)$ only if there exists a σ-finite measure G such that

$$(3.3) \qquad \delta^G(x) = x + \gamma(x) + \tilde{o}(x)|\gamma(x)|.$$

It is easy to check that for $G \in \mathcal{G}_o$, δ^G can be written

$$(3.4) \qquad \delta^G(x) = x + \ddagger\nabla\log g^*(x).$$

Combining (3.2), (3.3) and (3.4), gives the necessary condition

$$(3.5) \qquad -\nabla \log g^*(x) = h(x^tCx)\ \ddagger^{-1}Bx + \tilde{o}(x)|x|h(x^tCx).$$

Consider the transformed problem induced by the transformation $Y = C^{1/2}X$ (where $(C^{1/2})^tC^{1/2} = C$). Defining $r(y) = -\log g^*(C^{-1/2}y)$, it is easy to check that (3.5) becomes

$$(3.6) \qquad \nabla r(y) = h(|y|^2)(C^{-1/2})^t\ddagger^{-1}BC^{-1/2}y + \tilde{o}(y)|y|h(|y|^2).$$

It will be shown that a necessary condition for the equation

$$(3.7) \qquad \nabla r(y) = h(|y|^2)Ay + \tilde{o}(y)|y|h(|y|^2)$$

to have a solution, is $A = kI$ for some constant k. This will imply for (3.6) that

$$(C^{-1/2})^t\ddagger^{-1}BC^{-1/2} = kI,$$

or that $B = k \ddagger C$ as was desired.

Consider (3.7) when $p = 2$. Both sides of the equation will be integrated along the rectangular paths from the point (c,c) to the point $\alpha(c,c)$, where $\alpha = [(1+\rho)/(1-\rho)]^{1/2}$ and $c > K$. Letting P_1 be the path consisting of the line segments $L_1 = \{(sc,c)\colon\ 1 \leq s \leq \alpha\}$ and $L_2 = \{(\alpha c, sc)\colon\ 1 \leq s \leq \alpha\}$, calculation using (3.7) gives

(3.8)
$$\int_{P_1} [\nabla r(y)]^t \cdot dy = \int_{P_1} \{h(|y|^2)Ay + \tilde{o}(y)|y|h(|y|^2)\}^t \cdot dy$$
$$= \int_1^\alpha h([s^2+1]c^2)\{A(sc,c)^t + \tilde{o}([sc,c]^t[s^2+1]^{1/2}c\}^t \cdot (c,0)ds$$
$$+ \int_1^\alpha h([s^2+\alpha^2]c^2)\{A(\alpha c,sc)^t + \tilde{o}([\alpha c,sc]^t)[s^2+\alpha^2]^{1/2}c\}^t \cdot (0,c)ds,$$

where "$\cdot$" is the usual dot product.

Define

(3.9)
$$R_1 = \int_1^\alpha s\, h([s^2+1]c^2)ds,\quad R_2 = \int_1^\alpha s\, h([s^2+\alpha^2]c^2)ds,$$
$$T_1 = \int_1^\alpha h([s^2+1]c^2)ds, \text{ and } T_2 = \alpha\int_1^\alpha h([s^2+\alpha^2]c^2)ds.$$

Note that

$$\inf_{1 \leq s \leq \alpha} \frac{s^2+\alpha^2}{s^2+1} = \frac{2\alpha^2}{\alpha^2+1} = \frac{2(1+\rho)/(1-\rho)}{(1+\rho)/(1-\rho)+1} = 1+\rho.$$

Hence, by the assumption on h it is clear that

(3.1) $$R_2 \leq \tau R_1 \quad\text{and}\quad T_2 \leq \alpha\tau T_1.$$

Note also that for $1 \leq s \leq \alpha$ both $\tilde{o}([sc,c]^t)$ and $\tilde{o}([c,sc]^t)$ have norms which are $o(c^{-1})$ (i.e. go to zero faster than c^{-1} as $c\to\infty$). Using (3.9), equation (3.8) can thus be rewritten

$$(3.11) \qquad \int_{P_1} [\nabla r(y)]^t \cdot dy = c^2(A_{11}R_1+A_{12}T_1+A_{21}T_2+A_{22}R_2) + o(c^{-1})c^2T_1,$$

where A_{ij} is the (i,j) element of A.

Consider next the path P_2 from (c,c) to $\alpha(c,c)$, consisting of the line segments $L_3 = \{(c,sc): 1 \leq s \leq \alpha\}$ and $L_4 = \{(sc, c): 1 \leq s \leq \alpha\}$. As above it can be shown that

$$(3.12) \qquad \int_{P_2} [\nabla r(y)]^t \cdot dy = c^2(A_{11}R_2 + A_{12}T_2 + A_{21}T_1 + A_{22}R_1) + o(c^{-1})c^2T_1.$$

Since P_1 and P_2 have the same endpoints, the expressions in (3.11) and (3.12) must be equal. It follows that

$$(3.13) \qquad (A_{11}-A_{22})(R_1-R_2) = (A_{21}-A_{12})(T_1-T_2) + o(c^{-1})T_1.$$

The above analysis can also be conducted on paths between $(c,-c)$ and $\alpha(c,-c)$. The conclusion of such an analysis is that

$$(3.14) \qquad (A_{11}-A_{22})(R_1-R_2) = (A_{12}-A_{21})(T_1-T_2) + o(c^{-1})T_1.$$

Adding (3.13) and (3.14) gives

$$(3.15) \qquad (A_{11}-A_{22})(R_1-R_2) = o(c^{-1})T_1.$$

Clearly $R_1 > T_1$, so (3.10) implies that

$$(R_1-R_2) \geq (1-\tau)R_1 > (1-\tau)T_1.$$

Thus (3.15) can be satisfied as $c\to\infty$ only if $A_{11} = A_{22}$.

To complete the argument, consider transformations $z = \Gamma y$, where Γ is an orthogonal (2×2) matrix, Equation (3.7) becomes

$$\nabla r(z) = h(|z|^2)\Gamma A \Gamma^t z + \tilde{o}(z)|z|h(|z|^2).$$

The identical argument now shows that $\Gamma A \Gamma^t$ must also have equal diagonal elements. It is easy to check that if $\Gamma A \Gamma^t$ has equal diagonal elements for all orthogonal Γ, then A must be a multiple of the identity. This completes the proof for p=2.

The generalization to $p > 2$ follows from consideration of (3.7) in the subspace

$$\Omega_{ij} = \{x \in R^p: \ x_k = 0 \text{ for } k \neq i \text{ or } k \neq j\}.$$

For (3.7) to be satisfied when $y \in \Omega_{ij}$, it must in particular be true that

$$(r^{(i)}(y), r^{(j)}(y))^t = h(|y|^2)([Ay]_i, [Ay]_j)^t + \tilde{o}(y)|y|h(|y|^2).$$

Noting that

$$([Ay]_i, [Ay]_j)^t = \begin{pmatrix} A_{ii} & A_{ij} \\ A_{ji} & A_{jj} \end{pmatrix} \begin{pmatrix} y_i \\ y_j \end{pmatrix},$$

it follows from the two dimensional result that $A_{ii} = A_{jj}$ and $A_{ij} = A_{ji} = 0$ $(i \neq j)$.

Hence A must be a multiple of the identity and the proof is complete.||

Note that directional inconsistency (for simplicity ξ_o will be suppressed) of an estimator, δ, essentially occurs when Condition 1 of Section 2 is violated even for "smooth" versions of the estimator. Although

Condition 1 can be easily checked for a specific estimator δ by verifying the symmetry of the Jacobian of δ, directional consistency cannot be handled in this manner, since there is no control over the derivatives of the "smoothing" error term $\tilde{o}(x)|\gamma(x)|$. Indeed it is easy to construct estimators with nonsymmetric Jacobians, such that when an error term $o(x)|\gamma(x)|$ is added to the estimator, Condition 1 becomes satisfied. Thus it was necessary to resort to the more difficult integration argument of Theorem 3.1 to prove directional inconsistency.

As a simple application of Theorem 3.1, consider adaptive ridge regression estimators of the form

$$\delta(x) = (\Sigma^{-1} + A/(x^t Cx))^{-1}\Sigma^{-1}x.$$

(In terms of the usual regression model, x is the least squares estimator, while $\Sigma = \sigma^2(T^tT)^{-1}$, T being the design matrix and σ^2 the variance of the independent normal errors.) It is easy to check that δ is of the form (3.2) with $B = \Sigma A$ and $h(x^tCx) = 1/(x^tCx)$. This h satisfies the condition of Theorem 3.1 (take $K = 1$, $\rho = .5$, and $\tau = 2/3$ for example). Thus if $A \neq k\,C$ (i.e. $B = \Sigma A \neq k\,\Sigma\,C$) for some constant k, then δ is directionally inconsistent. Two commonly considered choices of A and C are (i) $A = pI$, $C = I$, and (ii) $A = (p/\sigma^2)I$, $C = \Sigma^{-1}$. (These are discussed in Lawless and Wang (1976) and elsewhere). Clearly the first choice satisfies $A = k\,C$, while the second choice does not. Hence the second choice is directionally inconsistent. (It will be shown in Section 4 that the first choice is actually directionally consistent.)

Unfortunately, the problem of trying to find an estimator uniformly better than a directionally inconsistent estimator seems to be very difficult.

It appears that estimators of entirely different functional forms need to be considered. For example, the directionally consistent ridge estimator, given above, is not uniformly better (in terms of expected loss) than the directionally inconsistent one. Indeed the latter estimator is by no means a "bad" estimator (even though directionally inconsistent) in that in many situations it seems to perform considerably better than the usual estimator $\delta^{o}(x) = x$. It is just that there may well be something considerably better still.

Section 4. Results for General Loss and Location Density

The results of Sections 2 and 3 were solely for quadratic loss. It is obviously desirable to obtain some type of extension to other loss functions. In particular, the criticism of directional inconsistency of an estimator becomes more valid if it can be shown to hold for a variety of losses. Unfortunately, general characterizations of admissible estimators as generalized Bayes estimators seem very difficult to obtain for nonquadratic losses. Because of this and certain technical problems, only weakened versions of the results of Sections 2 and 3 will be given.

It will first be necessary to develop an approximate formula for a generalized Bayes estimator. Aside from its use in this paper, the approximation should play an important role in future investigations of admissibility of generalized Bayes estimators. (See Brown (1974).)

Let Ω be a convex subset of R^p, and let $d(x)$ be a real valued function on R^p such that

(i) $d(x) > 0 \Rightarrow x \in \Omega$, and

(ii) for all $0 < K < \infty$, $|\theta - x| < K \Rightarrow d(\theta) > d(x) - K$.

An approximation to a generalized Bayes estimator $\delta^G(x)$ will be established in appropriate regions Ω for large values of $d(x)$. (For most applications it will suffice to have $\Omega = R^p - \Gamma$ (Γ a compact set) and $d(x) = |x| - \rho$, where $\rho = \sup_{x \in \Gamma} |x|$.) The region Ω will be said to be d-unbounded if $\sup_{x \in \Omega} d(x) = \infty$. For notational convenience, the following modification of "o" notation will be adopted for any functions r and h (possibly vector valued):

$$r(x) \text{ is } \bar{o}(h(x)) \text{ if } \lim_{T\to\infty} \sup_{\{x:\ d(x)>T\}} |r(x)|/|h(x)| = 0.$$

Needed assumptions on f, L, and G will now be given. These assumptions are not the most general possible, but should cover a wide variety of cases of interest, and despite their length are quite easy to verify.

Assumption A. (Conditions on the generalized prior G)

1. G is absolutely continuous with respect to Lebesque measure on R^p. (G will henceforth denote the density of the prior with respect to Lebesque measure.)
2. $0 \le G(\theta) \le B < \infty$.
3. There exists a d-unbounded region Ω such that if $\theta \in \Omega$, then the following conditions hold:
   - (a) $G(\theta) > 0$.
   - (b) G has continuous second order partial derivatives at θ.
   - (c) $|\nabla G(\theta)| = \bar{o}(G(\theta))$.
   - (d) There exists a positive increasing function w^*: $R^1 \to R^1$ for which the following hold:
     - (i) $[d - w^*(d)]$ is increasing in d and positive for some d.
     - (ii) There exists $q > 0$ such that $[w(\theta)]^{-q} = \bar{o}(|\nabla G(\theta)|^2/G(\theta))$, where $w(\theta) = w^*(d(\theta))$.

(iii) If $d(\theta) - w(\theta) \geq 0$, then for $1 \leq i, j \leq p$,

$$\sup_{\{\xi:\ |\theta-\xi| \leq w(\theta)\}} |G^{(i,j)}(\xi)| = \bar{o}(\nabla G(\theta)).$$

Assumption B. (Conditions on f and L)

1. $f(x,\theta) = f(x-\theta)$. (Thus θ is a location vector.)
2. L has all third order partial derivatives.
3. For any $D_2 < \infty$, there exist finite D_3 and D_4 such that if $|y| < D_2$, then $L^*(x+y) \leq D_3 + D_4 L^*(x)$, where L^* denotes L or any partial derivative of L through the third order.
4. For all $1 \leq i,j,k \leq p$, $E_0[|X|\,|L^{(i,j)}(X)|] < \infty$,

   $$E_0[|L^{(i,j,k)}(X)|] < \infty, \text{ and } E_0[|X|^{\max(2,q)}\{L(X) + |L^{(i)}(X)|\}] < \infty,$$

   where q is from assumption A3d(ii).
5. $h(c) = \int L(c+\theta)f(\theta)d\theta$ has a unique minimum at $c = c^o$.
6. The $(p \times p)$ matrix M with elements $m_{ij} = E_0[L^{(i)}(X+c^o)X_j]$ is nonsingular.
7. $L(x) \leq D_1 < \infty$ for all $x \in R^p$ (i.e. L is bounded), <u>or</u>

7'. $$\lim_{n\to\infty} \inf_{\{x:\ |x| > n\}} L(x) \geq h(c^o) + \varepsilon^*, \text{ for some } \varepsilon^* > 0.$$

Discussion of assumptions:

In Assumption A the key restriction is 3c. This places an allowable rate of decrease on G. If G goes to zero too quickly (as does $G(\theta) = \exp\{-|\theta|^2\}$) this condition will be violated, and indeed the approximation that will be developed is invalid. Such rapidly decreasing G are proper priors, however, and the resulting Bayes estimates are known to be admissible.

In condition A3d it usually suffices to choose $w(d) = cd$ (for some $0 < c < 1$), although choices such as $w(d) = d^{\alpha}$ (for some $0 < \alpha < 1$) may sometimes be necessary. Condition 3d(iii) somewhat surprisingly restricts G from being too flat. For example, $G(\theta) = 1 + \exp\{-|\theta|^2\}$ violates this condition. When $p \geq 3$, however, the result of Brown (1974) indicate that estimators arising from such G are inadmissible and can be improved upon by using G going to zero at an allowable rate.

Assumption B1 is made to keep the theory relatively simple. (See Brown (1974) for discussion of other cases.) Note that many scale parameter problems can be transformed into this setting by the usual log transform.

Assumptions B2, B3, and B4 are technical assumptions which could undoubtedly be weakened. Note that one should choose q as small as possible, in order to minimize the number of moments needed in B4.

Assumptions B7 or B7' should cover virtually all cases of interest. Note that any loss unbounded in all directions (such as a strictly convex loss) satisfies B7'.

Assumption B5 essentially says that the best invariant estimator of θ must be unique. (The best invariant estimator is thus $\delta^{o}(x) = x + c^{o}$.) When $h(c)$ has a nonunique minimum, the problem of approximating a generalized Bayes estimator becomes considerably more complex. If L is strictly convex, it is easy to check that $h(c)$ has a unique minimum. The following lemmas give other useful conditions for this to be true.

LEMMA 4.1. If $L(\theta)$ and $f(\theta)$ are symmetric functions in each coordinate θ_i, with $L(\theta)$ increasing in $|\theta_i|$ and $f(\theta)$ decreasing in $|\theta_i|$ $(1 \leq i \leq p)$, then $h(c)$ has a unique minimum at $c = 0$.

PROOF. Straightforward.||

LEMMA 4.2. If $L(\theta) = L^*(\theta^t Q\theta)$ and $f(\theta) = f^*(\theta^t \Sigma^{-1}\theta)$ where Q and Σ are $(p\times p)$ positive definite matrices, L^* is an increasing function, and f^* is a decreasing function, then $h(c)$ has a unique minimum at $c = 0$.

PROOF. Performing the linear change of variables, in the integral for $h(c)$, which simultaneously diagonalizes Q and Σ gives the desired result by Lemma 4.1.||

Assumption B5 also implies the following needed result.

LEMMA 4.3. If assumption B holds, then

(a) $E_0[\nabla L(X+c^o)] = 0$, and

(b) the $(p\times p)$ matrix $\mathcal{L}$ with elements $l_{ij} = E_0[L^{(i,j)}(X+c^o)]$ is positive definite.

PROOF. Under Assumption B it is easy to check that $h(c)$ can be differentiated twice under the integral sign. The lemma thus follows from the standard necessary conditions for c^o to be a unique minimum of $h(c)$.||

Assumption B6 essentially ensures that it is desired to estimate the full location vector. (See Brown (1966).) Two cases of considerable interest are given in the following lemmas.

LEMMA 4.4. If f is a p-variate normal density with known covariance matrix Σ, then $M = \mathcal{L}\Sigma$.

PROOF. Given in Berger (1976a).||

LEMMA 4.5.

(a.) Assume L in assumption B is of the form $L(x) = L^*(x^t Qx)$, where Q is positive definite and L^* is strictly increasing. Assume f is nondegenerate (i.e. not concentrated on a lower dimensional set with probability one). Then M is nonsingular.

(b.) If $L(x) = x^tQx$, then $M = 2Q\Sigma$, where $\Sigma = E_0\{[X-(E_0X)][X-(E_0X)]^t\}$ is the covariance matrix of f.

PROOF. (a) Letting L' denote the derivative of L*, M^t can be written

$$(4.1) \qquad M^t = E_0\{X[\nabla L(X+c^o)]^t\}$$

$$= E_0\{(X+c^o)[\nabla L(X+c^o)]^t\} \quad \text{(by Lemma 4.3(a))}$$

$$= 2E_0\{L'([X+c^o]^tQ[X+c^o])(X+c^o)(X+c^o)^t\}Q.$$

Note that for any $z \in R^p-\{0\}$,

$$z^tE_0\{L'([X+c^o]^tQ[X+c^o])(X+c^o)(X+c^o)^t\}z$$

$$= E_0\{L'([X+c^o]^tQ[X+c^o])([X+c^o]^tz)^2\} > 0$$

(since $L' > 0$ and $[X+c^o]$ cannot be perpindicular to z with probability one). Hence

$$E_0\{L'([X+c^o]^tQ[X+c^o])(X+c^o)(X+c^o)^t\}$$

is positive definite, and M is nonsingular.

(b) This problem follows from (4.1) (since $L'(\cdot) \equiv 1$) and an easy calculation which shows that $c^o = -(E_0X)$.||

The desired approximation to δ^G is given in the following theorem. Note that this is essentially the result developed heuristically in Brown (1974).

THEOREM 4.6. If Assumptions A and B are satisfied and $x \in \Omega$, then

$$\delta^G(x) = c^o + x + \mathcal{L}^{-1} M[\nabla \log G(x)] + \tilde{o}(\nabla \log G(x)). \tag{4.2}$$

The proof of this theorem will be given at the end of the section. The following corollary gives several examples of application.

COROLLARY 4.7. (a) Assume $G(\theta) = a/[b+(\theta^t C\theta)^\alpha]$, where a, b, and α are positive constants and C is a positive definite $(p\times p)$ matrix. Assume $L(x) = (x^t Qx)^\beta$, where $\beta = 1$ or $\beta \geq 3/2$. Finally, assume f is a nondegenerate location density with $2(\alpha+\beta) + 3$ finite moments. Then

$$\delta^G(x) = c^o + x - 2\alpha \mathcal{L}^{-1} MCx/(x^t Cx) + \tilde{o}(x).$$

(Recall $\tilde{o}(x)$ denotes a vector whose norm goes to zero as $|x| \to \infty$).

(b) If, in addition, f is a normal density, then

$$\delta^G(x) = x - 2\alpha \Sigma\, Cx/(x^t Cx) + \tilde{o}(x).$$

(c) If $\beta = 1$ in part (a), then

$$\delta^G(x) = -(E_0 X) + x - 2\alpha \Sigma Cx/(x^t Cx) + \tilde{o}(x).$$

PROOF. Assumptions A and B must first be verified. For Assumption A, choose $\Omega = \{\theta: |\theta| > 1\}$, $d(\theta) = |\theta| - 1$, $w^*(d) = d/3$, and $q = 2\alpha + 3$. It is straightforward to check that Assumption A then holds for the given G.

The first four conditions of Assumption B are easy to check, as is B7'. Condition B5 is satisfied because L is strictly convex, and condition B6 follows from Lemma 4.5 (a.).

A calculation gives

$$[\nabla \log G(x)] = -2\alpha(x^tCx)^{(\alpha-1)}Cx/[b+(x^tCx)^{\alpha}]$$
$$= -2\alpha Cx/(x^tCx) + \tilde{o}(x).$$

Clearly $d(x) \to \infty$ is equivalent to $|x| \to \infty$, so the result follows from Theorem 4.6.

(b.) This follows from part (a.), Lemma 4.4, and the easily verified fact that $c^o = 0$.

(c.) This follows from part (a.) and Lemma 4.5 (b.), noting that $c^o = -(E_0X)$. ||

Note that the above corollary implies, for normal densities, that estimators of the form (1.3) with $h(y) = y^{-1}$ and $B = k \ddagger C$ (for some $k > 0$) are directionally consistent. This complements the necessity of the condition $B = k \ddagger C$ established in Section 3.

From Theorem 4.6 follows the following generalization of Theorem 3.1.

THEOREM 4.8. Assume Assumption B holds for all $q > 0$ (see B4). Then an estimator

$$\delta(x) = c^o + x - h(x^tCx)Bx + \tilde{o}(x)|x|h(x^tCx),$$

with h, B, and C as in Theorem 3.1, is directionally inconsistent ($\mathcal{G}_1$) unless $B = kM^{-1}\mathcal{L}C$ for some constant k. (Here $\mathcal{G}_1$ is the class of all G satisfying Assumption A with $\Omega = R^p - \Gamma$, Γ a compact set.)

PROOF. The argument starts with (4.2), and proceeds exactly as does the proof of Theorem 3.1, with $\ddagger^{-1}$ replaced by $M^{-1}\mathcal{L}$. ||

Except for the unfortunate limitation to $\mathcal{G}_1$, the above theorem says that the directional inconsistency results hold for wide ranges of losses and densities. This, hopefully, alleviates the concern that directional inconsistency might be highly dependent on the loss used. Note, in particular, that if f is a normal density, then since $M^{-1}\mathcal{L} = \ddagger^{-1}$ (Lemma 4.4), directional inconsistency ($\mathcal{G}_0$) of an estimator for quadratic loss implies directional inconsistency ($\mathcal{G}_1$) of the estimator for all losses satisfying Assumption B. This lends considerable additional force to the criticism of directional inconsistency.

Before proceeding with the proof of Theorem 4.6, a final example is in order to demonstrate the scope of Theorem 4.6 (and to justify the apparent complexity of the Ω, d notation). Often, only partial prior information is available. Typically, for example, G may be a function only of θ_1. The following corollary considers such a situation.

COROLLARY 4.9. Assume $G(\theta) = a/(b+\theta_1^2)$, $a > 0$ and $b > 0$, $L(x) = x^tQx$, and f is a nondegenerate location density with seven finite moments. Then

$$\delta^G(x) = -(E_0X) + x - 2\ddagger(x_1^{-1},0,\ldots,0)^t + \bar{o}(x_1),$$

where $\bar{o}(x_1)$ is a vector whose norm goes to zero as $|x_1| \to \infty$.

PROOF. Choosing $\Omega = R^p$, $d(x) = |x_1|$, and $w^*(d) = d/3$, the verification of assumptions A and B and use of Theorem 4.6 is straightforward.||

The proof of Theorem 4.6 concludes the section.

PROOF of Theorem 4.6.

It is desired to find $\gamma^G(x) = \delta^G(x) - x - c^o$ which minimizes

$$I = \int L(\gamma(x) + x - \theta + c^o)f(x-\theta)G(\theta)d\theta. \tag{4.3}$$

Part1: Assume $\gamma(x) = \bar{o}(1)$. (It will be shown in part 2 that $\gamma^G(x) = \bar{o}(1)$.) By Assumption A3d, d(x) can be chosen large enough so that $d(x) - w(x) > 0$. Defining $V = \{\theta: |\theta-x| < w(x)\}$, it then follows from property (ii) of d that

$d(\theta) > d(x) - w(x) > 0$. Since $d(x) > 0$ also, the convexity of $\mathfrak{R}$ implies that $G(\theta)$ can be expanded in a Taylor expansion about x (up to third order terms) for $\theta \in V$. Line (4.3) can then be written

(4.4)
$$I = I_1 + I_2 + I_3 + I_4,$$
$$I_1 = G(x) \int_V L(\gamma(x) + x - \theta + c^o) f(x - \theta) d\theta,$$
$$I_2 = \sum_{i-1}^{p} G^{(i)}(x) \int_V (\theta_i - x_i) L(\gamma(x) + x - \theta + c^o) f(x - \theta) d\theta,$$
$$I_3 = \frac{1}{2} \sum_i \sum_j \int_V (\theta_i - x_i)(\theta_j - x_j) L(\gamma(x) + x - \theta + c^o) f(x - \theta) G^{(i,j)}(x^*) d\theta,$$
$$I_4 = \int_{V^c} L(\gamma(x) + x - \theta + c^o) f(x - \theta) G(\theta) d\theta,$$

where x^* is a point on the line segment between x and θ.

A simple Chebyshev argument using Assumption B3, B4, A2, and A3d(ii) shows that

(4.5)
$$I_4 \leq K[w(x)]^{-q} = \bar{o}(|\nabla G(x)|^2 / G(x)).$$

All future integrals over V^c are handled similarly and will simply be replaced by the appropriate $\bar{o}$ term with no further comment.

In I_1, I_2, and I_3, $L(\gamma(x) + x - \theta + c^o)$ will be expanded in Taylor expansions about $(x - \theta + c^o)$ up to fourth, third, and second order terms respectively. The integrals I_1^1, I_2^1, and I_3^1 resulting from the $L(x - \theta + c^o)$ term of the Taylor expansions are all independent of $\gamma(x)$. The dominant term is

$$I_1^1 = G(x) \int_V L(x - \theta + c^o) f(x - \theta) d\theta$$
$$= G(x)\{ \int L(x - \theta + c^o) f(x - \theta) d\theta - \int_{V^c} L(x - \theta + c^o) f(x - \theta) d\theta\}$$
$$= G(x)\{h(c^o) + \bar{o}(1)\}.$$

Using Assumptions A3c and A3d(iii), it is easy to see that $I_2^1 + I_3^1 = \bar{o}(G(x))$. Thus

(4.6) $$I_1^1 + I_2^1 + I_3^1 = G(x)\{h(c^o) + \bar{o}^*(1)\},$$

where "*" indicates the term is independent of $\gamma(x)$.

Using the assumptions and Lemma 4.3, the second, third, and fourth terms of the expansion of I_1 are

(4.7) $$I_1^2 = G(x) \sum_{i=1}^{p} \gamma_i(x)\{ \int L^{(i)}(x - \theta + c^o) f(x - \theta) d\theta \quad - \int_{V^c} [\quad] d\theta\}$$

$$= G(x) \sum_{i=1}^{p} \gamma_i(x)\{0 - \bar{o}(\nabla G)\} = \bar{o}(|\gamma(x)| \nabla G(x)),$$

$$I_1^3 = \frac{1}{2} G(x) \; \gamma^t(x) \mathscr{L} \gamma(x)(1 + \bar{o}(1)),$$

$$I_1^4 = \bar{o}(|\gamma(x)|^2 G(x)).$$

Similarly, the remaining terms of I_2 are

(4.8) $$I_2^2 = \sum_{i=1}^{p} \; G^{(i)}(x) \sum_{j=1}^{p} \gamma_j(x) \int_V L^{(j)}(x - \theta + c^o)(\theta_i - x_i) f(x - \theta) d\theta$$

$$= -[\nabla G(x)]^t M^t \gamma(x) + \bar{o}(|\gamma(x)| \nabla G(x)),$$

$$I_2^3 = \bar{o}(|\gamma(x)| \nabla G(x)).$$

Again using Assumption A3d(iii), the remaining term of I_3 is

(4.9) $$I_3^2 = \bar{o}(|\gamma(x)| \nabla G(x)).$$

Combining (4.4), (4.5), (4.6), (4.7), (4.8), and (4.9) gives

$$(4.10) \qquad I = G(x)\{h(c^o) + \bar{o}^*(1)\} + \tfrac{1}{2}G(x)\gamma^t(x)\mathscr{L}\gamma(x)$$

$$- [\nabla G(x)]^t M^t \gamma(x) + \bar{o}(G(x)|\gamma(x)|^2)$$

$$+ \bar{o}(|\gamma(x)|\nabla G(x)) + \bar{o}(|\nabla G(x)|^2/G(x)).$$

It is easy to check (using Assumption B6 and Lemma 4.3) that $\phi(\gamma)$, defined by

$$\phi(\gamma) = \tfrac{1}{2}G(x)\gamma^t(x)\mathscr{L}\gamma(x) - [\nabla G(x)]^t M^t \gamma(x),$$

is minimized at

$$\gamma^o(x) = \mathscr{L}^{-1}M[\nabla \log G(x)],$$

attaining the minimum value

$$\phi(\gamma^o) = -[\nabla G(x)]^t M^t \mathscr{L}^{-1}M[\nabla G(x)]/(2G(x)) < 0.$$

(Note by Assumption A3c that $\gamma^o(x) = \bar{o}(1)$ as was assumed in the above derivation.) Since the first term on the right hand side of (4.10) is independent of $\gamma(x)$, it follows that the γ which minimizes I (among those γ which are $\bar{o}(1)$) is of the form

$$\gamma(x) = \mathscr{L}^{-1}M[\nabla \log G(x)] + \bar{o}(\nabla \log G(x)).$$

Part 2: $\gamma^G(x)$ must be $\bar{o}(1)$. To see this, note that if $\gamma^G(x)$ minimizes I, it must be true that

$$(4.11) \qquad I = \int L(\gamma^G(x) + x - \theta + c^o)f(x-\theta)G(\theta)d\theta$$

$$< \int L(\gamma^o(x) + x - \theta + c^o)f(x-\theta)G(\theta)d\theta$$

$$= G(x)\{h(c^o) + \bar{o}(1)\}.$$

If $\gamma^G(x)$ is not $\bar{o}(1)$, then there exists an $\varepsilon > 0$ and a sequence of points $\{x^i\}$ in Ω such that

(4.12) $$d(x^i) \to \infty \text{ and } |\gamma^G(x^i)| > \varepsilon > 0.$$

Due to the continuity of h and Assumption B5, there exists a $\tau > 0$ such that

(4.13) $$h(c) > h(c^o) + \tau \qquad \text{if } |c - c^o| > \varepsilon.$$

Define $\alpha = \min\{\tau/[4h(c^o) + 3\tau],\ \varepsilon^*/[12h(c^o) + \varepsilon^*]\}$, where ε^* is from Assumption B7'. (Choose $\varepsilon^* = 1$ if B7' doesn't apply.) Note also that a Taylor expansion of G verifies that for any fixed positive integer n,

$$\inf_{\{\theta:\ |x-\theta|<n\}} G(\theta) = G(x)(1+\bar{o}(1)).$$

Hence for each n, there exists an $x^{i(n)} \in \{x_i\}$ such that

(4.14) $$\inf_{\{\theta:\ |x^{i(n)}-\theta|<n\}} G(\theta) > G(x^{i(n)})(1-\alpha).$$

Clearly $\{x^{i(n)}\}$ can be chosen so that $d(x^{i(n)}) \to \infty$. Also, it is obvious that

(4.15) if $\{|\gamma^G(x^i)|\}$ is unbounded, then $\{x^{i(n)}\}$ can be chosen so that $|\gamma^G(x^{i(n)})| > 3n$.

Defining $\gamma^n = \gamma^G(x^{i(n)})$, it follows from (4.14) that

(4.16) $$I(n) = \int L(\gamma^n + x^{i(n)} - \theta + c^o)f(x^{i(n)} - \theta)G(\theta)d\theta$$

$$\geq \left[\inf_{\{\theta:\ |x^{i(n)}-\theta|<n\}} G(\theta)\right] \int_{|x^{i(n)} - \theta|< n} L(\gamma^n + x^{i(n)} - \theta + c^o)f(x^{i(n)}-\theta)d\theta$$

$$> G(x^{i(n)})(1 - \alpha) \int_{|\theta|<n} L(\gamma^n + \theta + c^o)f(\theta)d\theta.$$

Case 1: L is bounded (Assumption B7). Then for large enough n, say n > N,

(4.17) $$\int_{|\theta|>n} L(\gamma^n + \theta + c^o)f(\theta)d\theta \leq D_1 \int_{|\theta|>n} f(\theta)d\theta < \tau/4.$$

This, together with (4.12), (4.13), and (4.16) shows that for n > N,

$$I(n) > G(x^{i(n)})(1 - \alpha)\{ \int L(\gamma^n + \theta + c^o)f(\theta)d\theta - \int_{|\theta|>n} [\quad]d\theta\}$$

$$> G(x^{i(n)})(1 - \tau/[4h(c^o) + 3\tau])\{h(c^o) + \tau - \tau/4\}$$

$$= G(x^{i(n)})\{h(c^o) + \tau/2\},$$

which contradicts (4.11). Hence $\gamma^G(x)$ must be $\bar{o}(1)$.

Case 2: $\lim_{n\to\infty} \inf_{\{\theta: |\theta|> n\}} L(\theta) \geq h(c^o) + \varepsilon^*$ (Assumption B7').

If $\{|\gamma^n|\}$ is bounded, the argument goes exactly as in Case 1, with Assumptions B3 and B4 being used to verify the analogue of (4.17). If $\{|\gamma^n|\}$ is unbounded, then (4.15) can be assumed to hold. Choose n large enough, say $n > N > c^o$, so that

(4.18) $$\inf_{\{\theta: |\theta|> n\}} L(\theta) > h(c^o) + \varepsilon^*/2.$$

Since $|\gamma^n| > 3n$, it follows for $|\theta|< n$ that $|\gamma^n + \theta + c^o| > n$. Hence for n > N,

$$\int_{|\theta|<n} L(\gamma^n + \theta + c^o)f(\theta)d\theta > (h(c^o) + \varepsilon^*/2) \int_{|\theta|<n} f(\theta)d\theta.$$

Choosing n large enough and using (4.16) will again result in a contradiction of (4.11). Hence in all cases $\gamma^G(x)$ must be $\bar{o}(1)$ and the proof is complete.||

REFERENCES

[1] Berger, J. (1976 a). Tail minimaxity in location vector problems and its applications. Ann. Statist. 4, 33-50.

[2] Berger, J. (1976 b). Inadmissibility results for generalized Bayes estimators of coordinates of a location vector. Ann. Statist. 4, 302-333.

[3] Berger, J. (1976 c). Admissibility results for generalized Bayes estimators of coordinates of a location vector. Ann. Statist. 4, 334-356.

[4] Berger, J. (1976 d). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Multivariate Anal. 6, 256-264.

[5] Berger, J. (1976 e). A robust generalized Bayes estimator of a multivariate normal mean. Technical Report #480, Purdue University.

[6] Bock, M. E. (1975). Minimax estimators of the mean of a multivariate normal distribution. Ann. Statist. 3, 209-218.

[7] Brown, L. (1966). On the admissibility of estimators of one or more location parameters. Ann. Math. Statist. 37, 1087-1136.

[8] Brown, L. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. Ann. Math. Statist. 42, 855-904.

[9] Brown, L. (1974). An heuristic method for determining admissibility of estimators - with applications. Technical Report. Purdue University.

[10] Casella, G. (1977). Ph.D. Thesis, Purdue University.

[11] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. J. Amer. Statist. Assoc. 68, 117-130.

[12] Farrell, R. (1964). Estimators of a location parameter in the absolutely continuous case. Ann. Math. Statist. 35, 949-998.

[13] Farrell, R. H. (1966). Weak limits of sequences of Bayes procedures in estimation theory. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1, 83-111. Univ. of California Press.

[14] Farrell, R. H. (1968). On a necessary and sufficient condition for admissibility of estimators when strictly convex loss is used, Ann. Math. Statist. 38, 23-28.

[15] Hirschmann, I. I. and Widder, D. V. (1955). The Convolution Transform. Princeton Univ. Press.

[16] Hudson, M. (1974). Empirical Bayes estimation. Technical Report #58, Stanford University.

[17] Lawless, J. F. and Wang, P. (1976). A simulation study of ridge and other regression estimators. Commun. Statist. 5, 307-323.

[18] Rao, C. R. (1977). Simultaneous estimation of parameters - a compound decision problem. S. S. Gupta and D. S. Moore (Eds.) Statistical Decision Theory and Related Topics II, Academic Press.

[19] Rolph, J. E. (1976). Choosing shrinkage estimators for regression problems. Commun. Statist. A5(9), 789-802.

[20] Sacks, J. (1963). Generalized Bayes solutions in estimation problems. Ann. Math. Statist. 34, 751-768.

[21] Srinivasan, C. (1976). Admissible estimators and exterior dirichlet problem. Thesis. Indian Statistical Institute.

[22] Stein, C. (1955). A necessary and sufficient condition for admissibility. Ann. Math. Statist. 34, 751-768.

[23] Stein, C. (1965). Approximation of improper prior measures by probability measures. Bayes, Bernouilli, Laplace, Springer-Verlag, Berlin.

[24] Strawderman, W. E. and Cohen, A. (1971). Admissibility of estimators of the mean vector of a multivariate normal distribution with quadratic loss. Ann. Math. Statist. 42, 270-296.

[25] Thisted, R. A. (1976). Ridge regression, minimax estimation, and empirical Bayes methods. Technical Report No. 28, Division of Biostatistics, Stanford University.

[26] Widder, D. V. (1946). The Laplace Transform. Princeton Univ. Press.