Shrinkage Estimators: Pseudo-Bayes Rules for Normal Mean Vectors

by

M. E. Bock*
Purdue University

Purdue University

Technical Report #87-18

Department of Statistics
Purdue University

May 1987

* This author's research is supported by NSF Grant No. DMS-8702620.

o

Shrinkage Estimators: Pseudo-Bayes Rules for Normal Mean Vectors

M. E. Bock
Department of Statistics
Purdue University
West Lafayette, Indiana, U.S.A.

1. INTRODUCTION

We examine the problem of estimating the mean vector $\underset{\sim}{\theta}$ of a normal random vector X with covariance matrix I_p. Stein [9] has given an unbiased estimator

$$p + D(\delta, X)$$

for the risk

$$R(\theta, \delta) = E_\theta ||\delta(X) - \theta||^2$$

of an estimator δ of θ. A standard method for showing that one estimator $\delta^{(1)}$ dominates another $\delta^{(2)}$ is to show that

$$D(\delta^{(1)}, X) \leq D(\delta^{(2)}, X)$$

for all X (with strict inequality on a set of positive measure). Brown [4] shows that such a technique will always fail for certain inadmissible estimators such as the James-Stein positive part estimator

$$\delta_c(X) = (1 - \frac{c}{||X||^2}) X I_{(c,\infty)}(||X||^2)$$

For $p-2 \leq c < 2(p-2)$, Brown shows that there is no estimator δ satisfying the inequality

$$D(\delta, X) \leq D(\delta_c, X)$$

for all X (with strict inequality on a set of positive measure). One might suspect that there is an admissible estimator which satisfies the inequality not necessarily for all X but at least for all X with $||X||^2 \geq c$. Theorem 3 shows that this too is impossible for $p - 2 < c \leq 2(p-2)$.

In general we examine estimators δ which are required to have small Bayes risk $r(\pi, \delta)$ with respect to a prior density π, yet which are not necessarily Bayes rules. For instance, we consider the robust Bayes estimator of Berger [1] of the form

$$\delta_B(X) = I_{(0,A)}(||X-\mu||^2)\delta_\pi(X) + I_{(A,\infty)}(||X-\mu||^2)(1 - \frac{2(p-2)}{||X-\mu||^2})X$$

where $A = 2(p-2)(1+\tau^2)$ and where δ_π is the Bayes rule for θ based on the normal prior π with mean μ and covariance matrix $\tau^2 I_p$. Berger shows this estimator cannot be dominated by a generalized Bayes rule δ with $r(\pi, \delta) < r(\pi, \delta_B)$ satisfying

$$(*) \qquad D(\delta, X) \leq 0$$

for all X (with strict inequality on a set of positive measure). Theorem 3 shows that even if we require the inequality of (*) to hold for just those X with $||X-\mu||^2 \geq A$ (with strict inequality on a set of positive measure), there is still no generalized Bayes rule δ with $r(\pi,\delta) < r(\pi,\delta_B)$.

The James-Stein positive part estimator and Berger's robust rule are special cases of "pseudo-Bayes" rules. They have the form

$$\delta(X) = X + \nabla ln\ m(X)$$

where m is a positive function called a pseudo-marginal density. (If δ is actually a Bayes rule, then it can always be written in this form where m is the Bayes marginal density for X.) Properties and characteristics of "pseudo-Bayes" rules are described in Section 2.

For each pseudo-Bayes rule δ^m, Section 3 describes a class of rules C_m^* which contains all the admissible and generalized Bayes rules dominating δ^m. Section 4 describes properties of certain pseudo-Bayes rules which are formed from a given Bayes rule in the following fashion: The pseudo-marginal density is defined to be equal to the given Bayes marginal density $m_\pi(X)$ when $m_\pi(X)$ is large. When $m_\pi(X)$ is small, the pseudo-marginal density is defined to be a larger function of X.

2. PSEUDO-BAYES RULES

Let $X \sim N(\theta, I_p)$ and estimate θ by $\hat{\theta}$ with loss function

$$L(\theta,\hat{\theta}) = ||\hat{\theta}-\theta||^2,$$

where θ is in $\mathbb{R}^p$. Define the conditional risk of an estimator $\hat{\theta}$ to be

$$R(\theta,\hat{\theta}) = E_\theta[||\hat{\theta}(X)-\theta||^2].$$

An unbiased estimate of the risk function was given by Stein [1]:

Define

$$D(\delta,X) = ||\delta(X)-X||^2 + 2\nabla\cdot(\delta(X)-X).$$

Then

$$E_\theta[p + D(\delta,X)] = R(\theta,\delta)$$

for all θ when δ is "a.e. differentiable". (See the appendix for a definition.) It is assumed that

$$E_\theta[|\frac{\partial}{\partial X_i}[\delta_i(X)-X_i]|] < \infty, i = 1,\ldots,p$$

and

$$E_\theta||\delta(X)-X||^2 < \infty.$$

The function D is useful for comparing the risks of two estimators $\delta^{(1)}$ and $\delta^{(2)}$. For instance, assume that for all X,

$$D(\delta^{(1)},X) \leq D(\delta^{(2)},X)$$

with strict inequality on a set of positive measure for X. Then $\delta^{(1)}$ dominates $\delta^{(2)}$ and for all θ,

$$R(\theta, \delta^{(1)}) < R(\theta, \delta^{(2)}).$$

Furthermore, under the density π for $\theta, \delta^{(1)}$ is preferable to $\delta^{(2)}$, i.e.

$$r(\pi, \delta^{(1)}) < r(\pi, \delta^{(2)})$$

where

$$r(\pi, \delta^{(i)}) = E_\pi[R(\theta, \delta^{(i)})].$$

If m_π is the marginal density of X, we define

$$r^*(m_\pi, \delta) = \int_{\mathbb{R}^p} \{D(\delta, X) + p\} m_\pi(X) dX$$

where

$$m_\pi(X) = \int_{\mathbb{R}^p} \pi(\theta)(2\pi)^{-p/2} e^{-\frac{1}{2}||X-\theta||^2} d\theta.$$

Clearly, $r^*(m_\pi, \delta)$ equals $r(\pi, \delta)$. The Bayes rule δ^π for the prior π is closely related to the marginal density m_π since it has the form

$$\delta^\pi(X) = X + \nabla ln\ m_\pi.$$

For a proper prior density $\pi(\theta)$, there does not exist δ such that

$$D(\delta, X) \le D(\delta^\pi, X)$$

for all X with strict inequality on a set of positive measure under m_π. (That would imply that

$$r^*(m_\pi, \delta) < r^*(m_\pi, \delta^\pi),$$

a contradiction to the fact that

$$r^*(m_\pi, \delta^\pi) \le r^*(m_\pi, \delta)$$

for all δ.)

We will examine other estimators δ^m which are not Bayes (or generalized Bayes) but which have similar properties for the function D. We call δ^m a "pseudo-Bayes" rule and define

$$\delta^m(X) = X + \nabla ln\ m(X)$$

where m is a positive real-valued function such that

$$r^*(m, \delta^m) = \int_{\mathbb{R}^p} \{D(\delta^m, X) + p\} m(X) dX$$

is finite. We refer to the function m as a "pseudo-marginal density". We call it a "strict pseudo-marginal density" if it *cannot* be written as

$$m(X) = \int_{\mathbb{R}^p} p(\theta) e^{-\frac{1}{2}||X-\theta||^2} d\theta$$

for a positive real-valued function p. (If it could, then δ^m would be a Bayes or a generalized Bayes estimator.) Define C_m^* to be the class of a.e. differentiable estimators δ such that

$$r^*(m, \delta^m) \leq r^*(m, \delta).$$

Clearly there is no δ in C_m^* satisfying

$$D(\delta, X) \leq D(\delta^m, X)$$

for all X (with strict inequality on a set of positive measure). Brown [4] shows that there are no estimators δ with that property for many reasonable δ^m. In Section 3 we will examine what estimators are in C_m^* and see that in many cases it includes all the admissible or generalized Bayes estimators that dominate δ^m.

3. DESCRIPTION OF C_m^*

In this section we examine C_m^* a collection of estimators which contains the admissible and generalized Bayes estimators dominating the pseudo-marginal estimator. The remark that follows Lemma 2 shows that in many cases the class contains all generalized Bayes rules with bounded risk. We conclude with an example of a common form for the pseudo-marginal density m when X is large.

It is convenient to examine a subset of C_m^* which we call C_m. The rules δ in C_m are defined to satisfy

$$\int_{\mathbb{R}^p} \nabla \cdot (m(\delta - \delta^m)) dX = 0 \qquad .$$

and are a.e. differentiable. It can be shown that C_m is contained in C_m^* because Theorem 1 shows that $r^*(m, \delta) \geq r^*(m, \delta^m)$ for all δ in C_m. Corollary 1 shows that no δ in C_m has smaller unbiased estimate of risk function than δ^m has. (This has been shown by Brown [4].)

THEOREM 1. *Let m be a positive real-valued function on $\mathbb{R}^p$ such that*

$$\nabla \ln\, m \text{ is a.e. differentiable}$$

and

$$\delta^m(X) = X + \nabla \ln\, m.$$

Define C_m to be the class of estimators δ such that δ is a.e. differentiable and

$$\int_{\mathbb{R}^p} \nabla \cdot [m(\delta - \delta^m)] dX = 0.$$

Then

$$r^*(m,\delta) - r^*(m,\delta^m)$$
$$= \int_{\mathbb{R}^p} ||\delta - \delta^m||^2 m(X)dX$$

if δ *is in* C_m.

Proof: Recall that

$$D(\delta, X) = ||\delta - X||^2 + 2\nabla \cdot (\delta - X).$$

Thus

$$\begin{aligned} D(\delta,X) - D(\delta^m,X) &= ||\delta - X||^2 - ||\delta^m - X||^2 + 2\nabla \cdot (\delta - \delta^m) \\ &= ||\delta - \delta^m||^2 + 2m^{-1}\{\nabla \cdot [m(\delta - \delta^m)]\} \\ &\text{(since } \nabla m = m^{-1}(\delta^m - X)\text{)}. \end{aligned}$$

This implies

$$\begin{aligned} &r^*(m,\delta) - r^*(m,\delta^m) \\ &= \int_{\mathbb{R}^p} \{D(\delta,X) - D(\delta^m,X)\}m(X)dX \\ &= \int_{\mathbb{R}^p} (||\delta - \delta^m||^2 m + 2\nabla \cdot [m(\delta - \delta^m)])dX \\ &= \int_{\mathbb{R}^p} ||\delta - \delta^m||^2 m \, dX + 2\int_{\mathbb{R}^p} \nabla \cdot [m(\delta - \delta^m)]dX \\ &= \int_{\mathbb{R}^p} ||\delta - \delta^m||^2 m \, dX \end{aligned}$$

(because δ is in C_m). □

The next corollary follows immediately from the theorem and shows that there is no estimator δ in C_m whose unbiased estimate of risk is always smaller than that of δ^m.

COROLLARY 1. Under the conditions of Theorem 1 there does not exist δ in C_m with

$$D(\delta, X) \leq D(\delta^m, X)$$

for all X with strict inequality on a set of positive (m) measure.

The results that follow further describe the classes of rules C_m^* and C_m. Also, the particular case that the pseudo-marginal density m is a function of a quadratic form in X is examined more closely.

LEMMA 1. Fix the vector μ and the positive definite matrix B. Suppose that for all X with $(X - \mu)^t B(X - \mu)$ sufficiently large, we have

$$(X - \mu)^t B(\delta - \delta^m) \geq 0.$$

Then if δ is a.e. differentiable it is in C_m^*.

Proof: Because (as in the proof of Theorem 1),

$$D(\delta, X) - D(\delta^m, X) = ||\delta - \delta^m||^2 + 2m^{-1}\{\nabla \cdot [m(\delta - \delta^m)]\},$$

we have

$$\begin{aligned} &r^*(m, \delta) - r^*(m, \delta^m) \\ &\quad = \int ||\delta - \delta^m||^2 m + 2\int \nabla \cdot [m(\delta - \delta^m)] \end{aligned}$$

and it suffices to show

$$\int \nabla \cdot [m(\delta - \delta^m) \geq 0$$

in order to show $[r^*(m, \delta) - r^*(m, \delta^m)]$ is nonnegative.

Gauss' Divergence Theorem implies that

$$\begin{aligned} &\int \nabla \cdot [m(\delta - \delta^m)]dX \\ &\quad = \lim_{a \to \infty} \int_{(X-\mu)^t B(X-\mu) \leq a^2} \nabla \cdot [m(\delta - \delta^m)]dX \\ &\quad = \lim_{a \to \infty} \int_{(X-\mu)^t B(X-\mu) = a^2} \frac{m(X)(X-\mu)^t B(\delta - \delta^m)}{\{(X-\mu)^t B(X-\mu)\}^{\frac{1}{2}}} dX. \end{aligned}$$

Since $m \geq 0$, this is clearly nonnegative if for some a sufficiently large,

$$(X - \mu)^t B(\delta - \delta^m) \geq 0$$

for all X with $(X - \mu)^t B(X - \mu) \geq a^2$. □

A slightly weaker condition on δ combined with a condition on m insures that δ is in C_m in Lemma 2.

LEMMA 2. If δ is a.e. differentiable and if for all X with $||X||$ sufficiently large we have

$$||\delta - \delta^m|| \leq d_0 < \infty,$$

then δ is in C_m provided

$$\lim_{c \to \infty} \int_{||X||=c} m(X)dX = 0.$$

Proof: As in the proof of Lemma 1,

$$\begin{aligned}
&|\int_{\mathbb{R}^p} \nabla \cdot [m(\delta - \delta^m)]dX| \\
&\quad \leq \lim_{c\to\infty} \int_{||X||=c} |X^t(\delta - \delta^m)| \frac{m(X)dX}{||X||} \\
&\quad \leq \lim_{c\to\infty} \int_{||X||=c} ||\delta - \delta^m|| m(X)dX \\
&\quad \leq d_0 \lim_{c\to\infty} \int_{||X||=c} m(X)dX \\
&\quad = 0. \qquad \square
\end{aligned}$$

REMARK: Suppose $||\delta^m - X||$ is bounded and

$$\lim_{c\to\infty} \int_{||X||=c} m(X)dX = 0.$$

Then any generalized Bayes rule δ with bounded risk (i.e. $\sup_\theta R(\theta, \delta) < \infty$) is in C_m.

Proof: Brown [3] shows that $||\delta - X||$ is bounded if and only if $\sup_\theta R(\theta, \delta) < \infty$ (in his Corollary 3.3.2). Thus $||\delta^m - \delta||$ is bounded since $||\delta^m - X||$ and $||\delta - X||$ are. The conditions of Lemma 2 are now satisfied. $\square$

Note: (See Brown [3]) $\sup_\theta R(\theta, \delta) < a$ only if the closed convex hull of the generalized prior for δ is $\mathbb{R}^p$.

LEMMA 3.

$$\begin{aligned}
r^*(m, \delta) - r^*(m, \delta^m) - \int ||\delta - \delta^m||^2 m dX \\
= 2\int \nabla \cdot [m(\delta - X]dX - 2\int \nabla^2 m dX.
\end{aligned}$$

Proof: By definition,

$$\begin{aligned}
&r^*(m, \delta) - r^*(m, \delta^m) \\
&\quad = \int \{D(\delta, X) - D(\delta^m, X\} m(X)dX.
\end{aligned}$$

As in the proof of Theorem 1,

$$D(\delta, X) - D(\delta^m, X) = ||\delta - \delta^m||^2 + 2m^{-1}\{\nabla \cdot [m(\delta - \delta^m)]\}.$$

Thus,

$$(*) = r^*(m,\delta) - r^*(m,\delta^m) - \int ||\delta - \delta^m||^2 m(X)dX$$
$$= 2\int \nabla\cdot[m(\delta - X)]dX - 2\int \nabla\cdot[m(\delta^m - X)]dX.$$

Because

$$\int \nabla\cdot[m(\delta^m - X)]dX = \int \nabla^2 m dX,$$

we have the result. □

It is clear that we can show δ is in C_m using Lemma 3 if we show

$$\int \nabla^2 m dX = 0 \text{ and } \int \nabla\cdot[m(\delta - X)]dX = 0.$$

The next lemma considers the case where δ depends on a quadratic form.

LEMMA 4. If an a.e. differentiable estimator δ has the form

$$\delta(X) = X + h((X-\mu)^t B(x-\mu))B(X-\mu)$$

for some real-valued function h, then

$$\int \nabla\cdot[m(\delta - X)]dX = 0$$

if and only if

$$\lim_{a\to\infty} ah(a^2)\int_{D_a} m(X)dX = 0$$

where

$$D_a = \{X:\ (X-\mu)^t B(X-\mu) = a^2\}.$$

Proof. We may write

$$(*) = \int \nabla\cdot[m(\delta - X)]dX$$
$$= \lim_{a\to\infty}\int_{(X-\mu)^t B(X-\mu)\le a^2} \nabla\cdot[m(\delta - X)]dX$$
$$= \lim_{a\to\infty}\int_{D_a} \frac{(X\mu)^t B(\delta - X)m(X)dX}{\{(X-\mu)^t B^2(X-\mu)\}^{\frac{1}{2}}}$$

by Gauss' Divergence Theorem. Thus,

$$(*) = \lim_{a\to\infty} \cdot h(a^2)\cdot\int_{D_a}\{(X-\mu)^t B^2(X-\mu)\}^{\frac{1}{2}} m(X)dX.$$

The integral is bounded above and below by

$$ab_p \int_{D_a} m(X)dX \text{ and } ab_1 \int_{D_a} m(X)dX$$

where b_p and b_1 are the largest and smallest eigenvalues of B, respectively. Thus (*) is zero if and only if

$$\lim_{a\to\infty} ah(a^2) \int_{D_a} m(X)dX = 0. \qquad \square$$

Theorem 2 and its corollaries which follow consider the case when the pseudo-marginal density m is a function of a quadratic form.

THEOREM 2. *Suppose the pseudo-marginal density $m(X)$ has the form*

$$m(X) = \phi((X-\mu)^t B(X-\mu))$$

for a positive definite matrix B, a fixed vector μ and a positive real-valued function ϕ. Assume that the estimator $\delta(X)$ is a.e. differentiable and has the form

$$\delta(X) = X + h((X-\mu)^t B(X-\mu))B(X-\mu).$$

Then

$$\int_{\mathbb{R}^p} \nabla \cdot [m(\delta - X)]dX = 0$$

if and only if

$$\lim_{a\to\infty} \phi(a^2)h(a^2)a^p = 0.$$

Proof. Observe that

$$\begin{aligned}\int_{D_a} m(X)dX &= \phi(a^2)\int_{D_a} dX \\ &= \phi(a^2)K' a^{p-1}\end{aligned}$$

(where K' is independent of a). Thus

$$h(a^2)a\int_{D_a} m(X)dX = \phi(a^2)h(a^2)a^p K',$$

and this yields the result when we apply Lemma 4. $\square$

COROLLARY 2. If the pseudo-marginal density $m(X)$ has the form

$$m(X) = \phi((X-\mu)^t B(X-\mu))$$

for a positive definite matrix B, a fixed vector μ and a positive real-valued function ϕ, then

$$\int \nabla^2 m dX = 0$$

if and only if

$$\phi'(a^2)a^p \underset{a\to\infty}{\longrightarrow} 0.$$

Proof: Setting $\delta(X) = \delta^m(X)$ in Theorem 2 we have

$$h(a^2) = \phi'(a^2)/\phi(a^2)$$

so that

$$\phi(a^2)h(a^2)a^p = \phi'(a^2)a^p$$

and the result follows. □

The next corollary follows from the use of Theorem 2 and Corollary 3 in Lemma 3.

COROLLARY 3. Define

$$d^2 = (X-\mu)^t B(X-\mu)$$

for a fixed positive definite matrix B and vector μ. Let the pseudo-marginal density $m(X)$ have the form

$$m(X) = \phi(d^2),$$

with $\lim_{a\to\infty} \phi'(a^2)a^p = 0$, and assume the a.e. differentiable estimator δ has the form

$$\delta(X) = X + h(d^2)B(X-\mu).$$

Then

$$\begin{aligned} r^*(m,\delta) &- r^*(m,\delta^m) \\ &= \int ||\delta - \delta^m||^2 m(X)dX \end{aligned}$$

if and only if

$$\lim_{a\to\infty} \phi(a^2)h(a^2)a^p = 0,$$

(assuming the integral and the r^*'s are finite).

In the following corollary we consider a common form for the value of the pseudo-marginal density m when X is large.

COROLLARY 4. Consider the estimator

$$\delta(X) = X - \frac{r_0(X)}{d^2}B(X-\mu)$$

where r_0 is a real-valued function and μ is a fixed vector and B is a positive definite matrix and where

$$d^2 = (X-\mu)^t B (X-\mu).$$

For fixed vector η and the positive definite matrix C, define

$$f^2 = (X-\eta)^t C (X-\eta).$$

For *large* f^2, define

$$m(X) = \frac{K}{f^{2j}}.$$

(Note that we have not defined m for small values of f^2.)

(a) If r_0 is bounded, then δ is in C_m, if $j > \frac{(p-2)}{2}$.

(b) If $r_0(X) = r_0(d^2)$, then δ is in C_m

$$\text{if } \lim_{a\to\infty} r(a^2) a^{p-2-2j} = 0.$$

Proof. The result in (a) follows from the condition that $\lim_{a\to\infty} a^{p-2-2j}$ be zero. Now we show (b). Set $h(d^2)$ equal to $(r_0(d^2)/d^2)$. For

$$e^2 = (\eta-\mu)^t C (\eta-\mu),$$

we have

$$(g-e)^2 \le f^2 \le (g+e)^2$$

where

$$g^2 = (X-\mu)^t C (X-\mu).$$

Now assume d is sufficiently large so that

$$(dt_1^{\frac{1}{2}} - e)^2 \le (g-e)^2$$

and

$$(g+e)^2 \le (dt_p^{\frac{1}{2}} + e)^2,$$

since

$$t_1 d^2 \le g^2 \le t_p d^2.$$

Thus

$$\frac{K}{(dt_p^{\frac{1}{2}} + c)^{2j}} \le m(X) \le \frac{K}{(dt_1^{\frac{1}{2}} - e)^{2j}}.$$

Define $D_a = \{X : (X-\mu)^t B (X-\mu) = a^2\}$.

This implies

$$\frac{K}{(dt_p^{\frac{1}{2}}+c)^{2j}}\int_{D_a} dX \le \int_{D_a} m(X)dX \le \frac{K}{(dt_p^{\frac{1}{2}}+c)^{2j}}\int_{D_a} dX,$$

i.e.

$$\{\frac{a}{dt_p^{\frac{1}{2}}+e}\}^{2j}K'a^{p-1-2j} \le \int_{D_a} m(X)dX \le \{\frac{a}{dt_1^{\frac{1}{2}}-e}\}^{2j}K''a^{p-1-2j}.$$

Thus

$$\lim_{a\to\infty} ah(a^2)\int_{D_a} m(X)dX = 0$$

if and only if

$$\lim_{a\to\infty} h(a^2)a^{p-2j} = 0.$$

The conditions of Lemma 3 are satisfied if $\int_{\mathbb{R}^p}\nabla^2 m dX = 0$. By Corollary 3, for $j > (p-2)/2$, we have $\int_{\mathbb{R}^p}\nabla^2 m dX = 0$. □

4. ADJUSTING BAYES RULES

In this section a special kind of pseudo-Bayes rule is described. They arise from proper Bayes rules δ^π which are changed only for certain values of the observations. For instance, the form of the estimator might be changed when the proper Bayes marginal density $m_\pi(X)$ is very small. In particular, a new function $m(X)$ is substituted for $m_\pi(X)$ for certain values of X. The resulting rule is a pseudo-Bayes rule with the properties described in previous sections. Special cases of such rules are given by versions of the James-Stein positive part estimator and the approximate restricted Bayes rules of Berger [1] and Chen [5]. See also the work of George [7], Spruill [8] and DasGupta and Rubin [6].

Let π be a proper prior density for θ in $\mathbb{R}^p$ and let m_π be the corresponding proper marginal density of X. The Bayes rule is given by

$$\delta^\pi(X) = X + \nabla ln\ m_\pi(X).$$

Define A to be a closed convex set and consider the following adjustment to δ^π for X outside of A. (Perhaps $m_\pi(X)$ is very small for X outside of A.) For X outside of A, we replace m_π by a larger function m so that m and $\nabla ln\ m$ agree with m_π and $\nabla ln\ m_\pi$, respectively, on the boundary of A. Thus, for X outside of A, δ^π is changed to δ^m where

$$\delta^m(X) = X + \nabla ln\ m(X).$$

We extend the definition of m and δ^m to all of $\mathbb{R}^p$ by making them equal to m_π and δ^π, respectively, for X in A. The careful matching of the estimators at the boundary of A insures that the unbiased estimate of risk for δ^m exists and agrees with that of δ^π for X in A, i.e.

$$D(\delta^m, X) = D(\delta^\pi, X)$$

for all X in A.

We know from Corollary 1 that there is no δ in C_m with

$$D(\delta, X) \leq D(\delta^m, X)$$

for *all* X (with strict inequality on a set of positive measure in X). The next theorem considers the case of δ in C_m that satisfies the inequality above, not for all X, but just for all X outside of A. It is shown that $r(\pi, \delta) \geq r(\pi, \delta^m)$. (Recall that under the condition that δ^m has bounded risk and $\lim_{a \to \infty} \int_{||X||^2 = a} m(X)dX$ is zero, we can show that all the bounded risk generalized Bayes rules are in C_m, and these include all admissible estimators dominating δ^m.)

THEOREM 3. *Let $m_\pi(X)$ be a proper marginal density with respect to the prior density π on θ. For the closed convex set A in $\mathbb{R}^p$, let $\overline{A^c}$ be the closure of the complement of A. Define the function m such that*

(i) $m : \overline{A^c} \to \mathbb{R}^+$ and $\nabla \ln\, m$ is a.e. differentiable in $\overline{A^c}$, and

(ii) $m(X) \geq m_\pi(X)$ for all X in $\overline{A^c}$, and

(iii) for all X in the boundary of $A, m(X) = m_\pi(X)$ and $\nabla \ln\, m(X) = \nabla \ln\, m_\pi(X)$, where $\nabla \ln\, m$ is defined on the boundary of A by continuity.

Extend the definition of m to all of $\mathbb{R}^p$ by setting

$$m(X) = m_\pi(X), \text{ for all } X \text{ in } A,$$

and define

$$\delta^m(X) = X + \nabla \ln\, m(X), \text{ for all } X \text{ in } \mathbb{R}^p.$$

Then there is no δ in C_m with

$$r(\pi, \delta) < r(\pi, \delta^m)$$

and

$$D(\delta, X) \leq D(\delta^m, X) \text{ for all } X \text{ in } A^c.$$

Proof. Suppose such a δ exists. Then

$$\begin{aligned}
&r(\pi, \delta) - r(\pi, \delta^m) \\
&= \int [D(\delta, X) - D(\delta^m, X)] m_\pi(X) dX \\
&= \int_A [D(\delta, X) - D(\delta^m, X)] m_\pi(X) dX \\
&+ \int_{A^c} [D(\delta, X) - D(\delta^m, X)] m(X) dX \\
&+ \int_{A^c} [D(\delta, X) - D(\delta^m, X)](m_\pi - m) dX.
\end{aligned}$$

Because $m(X)$ equals $m_\pi(X)$ for X in A,

$$r(\pi,\delta) - r(\pi,\delta^m) = r^*(m,\delta) - r^*(m,\delta^m) + \int_{A^c} [D(\delta,X) - D(\delta^m,X)](m_\pi - m)dX.$$

Since $(m_\pi - m)$ and $[D(\delta,X) - D(\delta^m,X)]$ are nonpositive on A^c, their product is nonnegative on A^c and

$$\int_{A^c} [D(\delta,X) - D(\delta^m,X)](m_\pi - m)dX \geq 0.$$

So

$$r(\pi,\delta) - r(\pi,\delta^m) \geq r^*(m,\delta) - r^*(m,\delta^m).$$

But Theorem 1 implies that

$$r^*(m,\delta) - r^*(m,\delta^m) \geq 0.$$

Thus $r(\pi,\delta) - r(\pi,\delta^m) \geq 0$, a contradiction. □

Note: The conditions of the theorem also imply that there is no δ in C_m with

$$r(\pi,\delta) \leq r(\pi,\delta^m)$$

and

$$D(\delta,X) \leq D(\delta^m,X) \text{ for all } X \text{ in } A^c$$

with strict inequality on a set of positive $m - m_\pi$ measure in A^c.

REMARK: To apply Theorem 3 we need to show that $m_\pi(X) \leq m(X)$ for X outside of A, or equivalently that

$$ln\ m_\pi(X) \leq ln\ m(X)$$

for X outside of A. When m, m_π and A are defined in terms of $||X||^2$, it suffices to show that

$$\frac{\partial}{\partial ||X||^2}(ln\ m_\pi) \leq \frac{\partial}{\partial ||X||^2}(ln\ m)$$

for X outside of A (since $ln\ m_\pi = ln\ m$ on the boundary of A). Because $(\delta^m - X) = \frac{\partial}{\partial ||X||^2}(ln\ m)2X$, we have

$$\frac{\partial}{\partial ||X||^2}(ln\ m_\pi) \leq \frac{\partial}{\partial ||X||^2}(ln\ m)$$

if and only if

$$(\delta^\pi - X)^t X \leq (\delta^m - X)^t X.$$

If $m(X) = k/||X||^{2j}$ for X outside of A and we write δ^π in the form

$$\delta^\pi(X) = \left(1 - \frac{r(||X||^2)}{||X||^2}\right)X,$$

the last inequality is equivalent to

$$r(||X||^2) \geq 2j$$

for X outside of A. A similar sort of argument holds when m_π, m and A are defined in terms of a general quadratic form in X.

In the next example we consider the James-Stein positive-part estimator.

EXAMPLE 1. A prior distribution that puts all mass on the zero vector yields an estimator which is the zero vector. Adjusting this estimator to make it minimax leads to the James-Stein positive part estimator

$$\delta_c(X) = (1 - \frac{c}{||X||^2}) I_{(c,\infty)}(||X||^2) X.$$

See Bickel [2] for $c = 2(p-2)$. Brown [4] has shown that for $p - 2 \leq c < 2(p-2)$ there is no estimator δ with

$$D(\delta, X) \leq D(\delta_c, X)$$

for all X. Theorem 3 and the remark that follows can be used to show that even if we look among estimators that satisfy the above inequality just for $||X||^2 \geq c$, we are still unable to find a generalized Bayes or admissible estimator dominating δ_c if $p - 1 < c \leq 2(p-2)$. (We use the fact that bounded risk generalized Bayes estimators δ have $||\delta - X||$ bounded to show that they are in C_m.) It can also be seen that no estimator of the form

$$\delta(X) = (1 - \frac{r(X)}{||X||^2}) X$$

(where r is bounded) will dominate δ_c for $p - 2 < c \leq 2(p-2)$ and satisfy

$$D(\delta, X) \leq D(\delta_c, X)$$

for $||X||^2 \geq c$.

These results may be seen by noting that

$$r(\pi, \delta) = R(\theta = 0, \delta)$$

and noting that the pseudo-marginal $m(X)$ has the form $K/||X||^c$ for large $||X||$, as considered in Corollary 4.

EXAMPLE 2. Consider the normal prior π for θ with mean μ and covariance matrix $\tau^2 I_p$. The minimax robust Bayes estimator δ_B considered by Berger [1] has the form

$$\delta_B(X) = I_{(0,A)}(||X - \mu||^2)\delta^\pi(X) + I_{(A,\infty)}(||X - \mu||^2)(1 - \frac{2(p-2)}{||X||^2})X$$

where $A = 2(p-2)(1+\tau^2)$ and δ^π is the Bayes rule. Note that all the bounded risk or minimax generalized Bayes rules are in the class C_m. Theorem 3 implies that there is no generalized Bayes rule δ satisfying

$$D(\delta, X) \leq D(\delta_B, X)$$

for $||X-\mu||^2 \geq A$ and

$$r(\pi, \delta) < r(\pi, \delta_B).$$

(Note that $D(\delta_B, X) = 0$ for $||X-\mu||^2 \geq A$.)

EXAMPLE 3. This is a generalization of the previous example to the case of a normal prior π for θ with mean η and covariance matrix T. Then the marginal density $m_\pi(X)$ is a function of the quadratic form

$$f^2 = (X-\eta)^t C (X-\eta)$$

discussed in Corollary 4 with

$$C = [I_p + T]^{-1}.$$

A pseudo-Bayes rule of the form discussed in Corollary 4 is

$$\delta^m(X) = I_{(0,2j)}(f^2)\delta^\pi(X) + I_{[2j,\infty)}(f^2)[X - \frac{2j}{f^2}C(X-\eta)]$$

with $j > \frac{(p-2)}{2}$. Thus there is no generalized Bayes rule δ satisfying

$$D(\delta, X) \leq D(\delta^m, X)$$

for $f^2 \geq 2j$ and

$$r(\pi, \delta) < r(\pi, \delta^m).$$

Similar results might be developed for the case where we look at estimators satisfying

$$r(\pi, \delta) \leq r(\pi, \delta_B) + c.$$

APPENDIX

Definition: A function $h: \mathbb{R}^p \to \mathbb{R}$ is a.e. differentiable (Stein [9]) if there exists a function $\nabla h: \mathbb{R}^p \to \mathbb{R}^p$ such that for all Z in $\mathbb{R}^p$,

$$h(X+Z) - h(X) = \int_0^1 Z^t \nabla h(X+tZ)\,dt$$

for almost all X in $\mathbb{R}^p$. A function $\delta(X): \mathbb{R}^p \to \mathbb{R}^p$ is a.e. differentiable if each component $\delta_i(X)$ is a.e. differentiable.

REFERENCES

[1] Berger, James O. (1982). Estimation in continuous exponential families: Bayesian estimation subject to risk restrictions and inadmissibility results. *Statistical Decision Theory and Related Topics III*, Vol. 1, 109–141, Academic Press, New York.

[2] Bickel, P. J. (1984). Parametric robustness: small biases can be worthwhile. *The Annals of Statistics*, Vol. 12, No. 3, 864–879.

[3] Brown, Lawrence D. (1971). Admissible estimators, recurrent diffusions and insoluble boundary value problems. *Ann. Math. Statist.*, **42**, 855–903.

[4] Brown, Lawrence D. (1986). The differential inequality of a statistical estimation problem. *Proceedings of the Fourth Purdue Symposium on Statistical Decision Theory and Related Topics*, (S. S. Gupta and J. Berger, eds.), Springer-Verlag, New York.

[5] Chen, S. Y. (1983). Restricted risk Bayes estimation for the mean of the multivariate normal distribution. Technical Report No. 33, Department of Statistics, Purdue University.

[6] DasGupta, Anirban and Rubin, Herman (1987). Bayesian estimation subject to minimaxity of the mean of a multivariate normal distribution in the case of a common unknown variance: a case for Bayesian robustness. *Proceedings of the Fourth Purdue Symposium on Statistical Decision Theory and Related Topics*, (S. S. Gupta and J. Berger, eds.), Springer-Verlag, New York.

[7] George, Edward I. (1986). Minimax multiple shrinkage estimators. *Annals of Statistics*, Vol. 14, 188–205.

[8] Spruill, M. C. (1986). Some approximate restricted Bayes estimators of a normal mean. *Statistics and Decisions*, **4**, 337–351.

[9] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, Vol. 9, 1135–1151.