ON MINIMAX ESTIMATION OF A NORMAL MEAN VECTOR FOR GENERAL QUADRATIC LOSS

William E. Strawderman
Rutgers University

Let $X \sim N_p(\theta, \Sigma)$ (Σ known) and consider the problem of estimating the mean vector when loss is general quadratic loss $(\delta - \theta)'Q(\delta - \theta)$. Many results are known for the case $\Sigma = Q = I$. There is also a relatively large literature for the case of general Σ and Q but it is relatively less well developed. The purpose of this paper is to unify many of the results in the general case by relating them to the simpler case $\Sigma = Q = I$. We give a reduction of the general case to a canonical form ($\Sigma = I$, $Q =$ Diagonal) and show that a natural correspondence between priors, marginals, and estimators in the two versions of the problem preserves risk, admissibility, minimaxity and Bayesianity. This allows many results on minimaxity and admissibility in the case $\Sigma = Q = I$ to be extended to the general case and allows an expansion of the classes of known minimax estimators in the general case. It also seems to make the general case somewhat more comprehensible.

1. Introduction

Let $X \sim N_p(\theta, \Sigma)$ and consider the problem of estimating the mean vector θ with loss $L(\theta, d) = (d - \theta)'Q(d - \theta)$.

A great deal is known about this problem when $\Sigma = Q = I$ (and more generally when Σ and Q are known multiples of I). Relatively less is known when the covariance matrix, Σ, and the matrix Q are general positive definite matrices. The purpose of this paper is to close, to a degree, the gap between the case $Q = \Sigma = I$ and the general case.

In Section 2, we briefly present a snapshot of results for the case where Σ and Q are known multiples of I. In Section 3, we extend these results to the case where Σ and Q are diagonal and in Section 4, to the case of general positive definite Σ and Q.

The spirit of the development herein is to derive procedures in the general case corresponding to procedures in the $\Sigma = Q = I$ case which are Bayes (proper, generalized, or pseudo) minimax and/or admissible and which preserve these properties in the general case. There are a number of results along these lines in the literature. This paper unifies and generalizes many of these results and gives a comprehensive and, it is hoped, comprehensible picture of the general case.

We will use the notation $\nabla m(X)$, $\nabla \cdot m(X)$ and $\nabla^2 m(X)$ for the gradient, divergence, and laplacian of a function $m(X)$. Recall that $\nabla^2 m(X) = \sum_1^p \partial^2 m(X)/\partial X_i^2$ and that a function $m(X)$ is superharmonic if and only if $\nabla^2 m(X) \leq 0\ \forall X$.

AMS subject classifications: 62C20, 62C15, 62C10.
Keywords and phrases: Stein estimation; minimaxity; Bayes, quadratic loss.

See Strawderman (1971), Alam (1973), Bock (1975, 1988), Berger (1976), Faith (1978), Shinozaki (1980), Stein (1981), George (1986), Berger and Robert (1990), Berger and Strawderman (1996), and Fourdrinier, Strawderman, and Wells (1998) for background and results related to this paper.

The main technical contributions of the paper are part (c) of Theorem 3.2 and part (b) of Example 4.1 which may be viewed as extensions of Berger (1976) along the lines Fourdrinier, Strawderman and Wells (1998).

2. The case $\Sigma = \sigma^2 I$, $Q = I$

The basis of much of the modern development in Stein estimation is Stein's 1981 paper. The following is a summary of that development. We assume throughout that $g(X)$ is weakly differentiable and $E\|g(X)\|^2 < \infty$.

Theorem 2.1 (Stein, 1981). *Let $X \sim N_p(\theta, \sigma^2 I)$ and suppose the loss function is $\|d - \theta\|^2$ (i.e., $\Sigma = \sigma^2 I$, $Q = I$).*

(a) $E[(X-\theta)' g(X)] = \sigma^2 E\,\nabla \cdot g(X)$.

(b) *If* $\delta(X) = X + \sigma^2 g(X)$, $E\|\delta - \theta\|^2 = R(\theta, \delta) = p\sigma^2 + E[\sigma^4(\|g(X)\|^2 + 2\nabla \cdot g)]$.

(c) *If* $\|g(X)\|^2 + 2\nabla \cdot g(X) \le 0\ \forall X$, *then* $\delta(X) = X + \sigma^2 g(X)$ *is minimax.*

(d) *If* $\theta \sim \pi(\theta)$, *the Bayes estimator* $\delta_\pi(X) = X + \sigma^2 \nabla m(X)/m(X)$, *where* $m(X) = (\sqrt{2\pi}\sigma)^{-2} \int \exp[-(2\sigma^2)^{-1}\|X - \theta\|^2]\pi(\theta)\, d\theta$ *is the marginal distribution of* X.

(e) *If* $\theta \sim \pi(\theta)$,

$$R(\theta, \delta_\pi) = p\sigma^2 + E\left[\sigma^4\left(2\frac{m(X)\nabla^2 m(X) - \|\nabla m(X)\|^2}{m^2(X)}\right)\right]$$
$$= p\sigma^2 + 4\sigma^4 E\left(\frac{\nabla^2\sqrt{m(X)}}{\sqrt{m(X)}}\right).$$

Hence δ_π is minimax provided $\sqrt{m(X)}$ is superharmonic.

Note that superharmonicity of $\pi(\theta)$ implies superharmonicity of $m(X)$ which in turn implies superharmonicity of $\sqrt{m(X)}$. Hence superharmonicity of $\pi(\theta)$, $m(X)$, or $\sqrt{m(X)}$ implies minimaxity of $\delta_\pi(X)$.

It is also convenient to introduce the notion of a pseudo-Bayes estimate (see Bock, 1988). We say $\delta_m(X)$ is pseudo-Bayes if $\delta_m(X) = X + \sigma^2 \frac{\nabla m(X)}{m(X)}$, where $m(X)$ is any function for which $\nabla m(X)$ exists (we assume further that $\nabla^2 m(X)$ exists and $E\|\nabla m\|^2/m^2 < \infty$). Hence a pseudo-Bayes estimate has the form of a Bayes estimate but $m(X)$ may not be a true marginal distribution resulting from a (generalized) prior $\pi(\theta)$. For example, if $m(X) = \left(1/\|X\|^2\right)^b$, the resulting pseudo-Bayes estimator is $\delta_m(X) = X - (2b\sigma^2/\|X\|^2)X$, a James-Stein type estimator. The next corollary follows essentially directly from Theorem 2.1(d) and (e).

Corollary 2.1. *A pseudo-Bayes estimator $\delta_m(X)$ is minimax provided $\sqrt{m(X)}$ is superharmonic.*

Example 2.1. Suppose $m(X) = (1/\|X\|^2)^b$. Then

$$\nabla\sqrt{m(X)} = -\frac{bX}{\|X\|^{b+2}},$$

and

$$\begin{aligned}\nabla^2\sqrt{m(X)} &= -b\left[\frac{p\|X\|^{b+2} - (b+2)\|X\|^{b+2}}{\|X\|^{2b+4}}\right] \\ &= -\frac{b[(p-2)-b]}{\|X\|^{b+2}}.\end{aligned}$$

Hence $\nabla\sqrt{m(X)} \leq 0$ (m is superharmonic) provided $0 \leq b \leq (p-2)$. It follows that the James-Stein estimator $(1 - a\sigma^2/\|X\|^2)X$ is minimax for $0 \leq a \leq 2(p-2)$. Note that superharmonicity of $m(X)$ holds only for $0 \leq b \leq \frac{p-2}{2}$ corresponding to the range $0 \leq a \leq (p-2)$. The risk of $\delta_m(X)$ is equal to $p\sigma^2 - 4\sigma^4 E[b[(p-2)-b]/\|X\|^2]$ by Theorem 2.1(e).

Fourdrinier, Strawderman, and Wells (1998) (FSW) point out that the distinction between superharmonicity of $\sqrt{m(X)}$ vs $m(X)$ is important and that it is impossible for $m(X)$ to be simultaneously proper (corresponding to a proper (integrable) prior $\pi(\theta)$) and superharmonic. They show that propriety of $m(X)$ and superharmonicity of $\sqrt{m(X)}$ are indeed possible, if $p \geq 5$.

FSW also give a fairly general class of Bayes minimax estimators. A main result of that paper is the following result concerning hierarchical (generalized and proper) Bayes minimax estimators.

Theorem 2.2 (FSW). *Suppose θ has a prior distribution with the following structure*

$$\theta|\lambda \sim N_p\left(0, \frac{(1-\lambda)\sigma^2}{\lambda} I\right), \quad \lambda \sim h(\lambda), \quad 0 < \lambda < 1.$$

(a) *The marginal distribution of X conditional on λ is*

$$X \mid \lambda \sim N_p\left(0, \frac{\sigma^2}{\lambda} I\right).$$

(b) *The marginal distribution of* X *is*

$$m(X) \propto \int \lambda^{p/2} \exp\left[-\frac{\lambda}{2\sigma^2}(X'X)\right] h(\lambda)\, d\lambda.$$

(c) $\sqrt{m(X)}$ *is superharmonic (and hence* $\delta_\pi(X)$ *is minimax) provided* $h(\lambda)$ *satisfies*

$$\frac{\lambda h'(\lambda)}{h(\lambda)} = \ell_1(\lambda) + \ell_2(\lambda), \tag{2.1}$$

where $\ell_1(\lambda) \leq A$, *is nonincreasing in* λ, *and* $0 \leq \ell_2(\lambda) \leq B$ *where* $\frac{1}{2}A + B \leq (p-6)/4$, *and*

$$\lim_{\lambda \to 0} \lambda^{p/2} h(\lambda) = 0, \quad \lim_{\lambda \to 1} h(\lambda) < \infty. \tag{2.2}$$

Note. FSW developed their results in terms of a hierarchical distribution of the form

$$\theta \mid V \sim N(0, V\sigma^2 I), \quad V \sim \lambda(V),\ 0 < V < \infty.$$

The above result re-expresses Theorem 1 of FSW in terms of the current parameterization.

Example 2.2. The prior distribution given by $h(\lambda) = (1-a)\lambda^{-a}$ given in Strawderman (1971) satisfies the conditions of the theorem for $-a \leq (p-6)/2$. These priors are proper provided that $-a > -1$ and lead to admissible estimators provided $-a \geq -2$ by Brown (1971). Noting that the interval $-1 < -a \leq (p-6)/2$ is nonempty for $p \geq 5$, and $-2 < -a \leq (p-6)/2$ is nonempty for $p \geq 3$, the above prior gives proper Bayes minimax estimators for $p \geq 5$ and admissible minimax estimators for $p \geq 3$.

See also Faith (1978), Alam (1973) for other mixing distributions $h(\lambda)$ leading to Bayes minimax estimators.

George (1986) studied multiple shrinkage estimators. Here one may have, for example, several possible points ν_i, $i = 1, \ldots, k$ towards which to shrink. It is desired to adaptively shrink toward one of the points so that the resulting procedure is minimax. The key observation is that if each of $\pi_i(\theta)$ is superharmonic, then so is $\sum \alpha_i \pi_i(\theta)$ (and more generally for an arbitrary mixture of superharmonic functions). Here is a version of George's result.

Theorem 2.3. *Suppose* $\pi_\alpha(\theta)$, $\alpha \in A$, *are a collection of superharmonic (generalized) priors and* $m_\alpha(X)$ *is the corresponding collection of marginal distributions. Let* $\lambda(\alpha)$ *be any finite mixing distribution on* $\alpha \in A$. *Then*

(a) $\int \pi_\alpha(\theta)h(\alpha)\,d\alpha$ *is superharmonic with corresponding superharmonic marginal*

$$\int m_\alpha(X)h(\alpha)\,d\alpha.$$

(b) *The resulting generalized (or pseudo, if $m_\alpha(X)$ are given) Bayes estimator is minimax.*

Example 2.3. Suppose $\nu_1, \dots, \nu_k$ are k given vectors in R^p, and $m_i(X) = (1/\|X - \nu_i\|^2)^b$ for $0 < b < (p-2)/2$. Let $m(X) = (1/k)\sum_{i=1}^k m_i(X)$. Then $m_i(X)$ and $m(X)$ are superharmonic and the resulting pseudo-Bayes estimator is minimax. This estimator is given by

$$\delta(X) = X - \sigma^2\left[\frac{\sum_{i=1}^k 2b(X-\nu_i)/(\|X-\nu_i\|^2)^{b+1}}{\sum_{i=1}^k 1/(\|X-\nu_i\|^2)^b}\right]$$

and "adaptively" shrinks X toward the "closest" ν_i or alternatively is a weighted combination of James-Stein like estimators shrinking toward the ν_i with greater weights $(1/\|X - \nu_i\|^{2b})$ put on ν_i closest to X.

It seems worth noting that in Theorem 2.3 we required $\pi_i(\theta)$, and not $\sqrt{\pi_i(\theta)}$, to be superharmonic. This is significant in that this makes it impossible that the $\pi_i(\theta)$ (and hence $\pi(\theta)$) be proper priors. It need not be the case that the square root of mixtures of functions whose square roots are themselves superharmonic. Hence it is difficult to carry out George's development for mixtures of proper priors (or even proper pseudo-marginals as in Example 2.3).

3. Results for the Case Σ, Q Diagonal

In this section we study the case where $\Sigma = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$ and $Q = \operatorname{diag}(d_1, \dots, d_p) = D$. The extension of Theorem 2.1 to this case is the following result.

Theorem 3.1. *Suppose $X \sim N_p(\theta, \Sigma)$ and $L(\theta, d) = (d-\theta)'D(d-\theta)$, where Σ and D are as above. Then*

(a) *If $\delta(X) = X + \Sigma g(X)$, where $g(X)$ is weakly differentiable and $E\|g\|^2 < \infty$, then $R(\theta,\delta) = \operatorname{tr}\Sigma D + E[\sum_{i=1}^p \sigma_i^4 d_i(g_i^2(X) + 2\partial g_i(X)/\partial X_i)]$.*

(b) *If $\theta \sim \pi(\theta)$, the Bayes estimator of θ is $\delta_\pi(X) = X + \Sigma\nabla m(X)/m(X)$.*

(c) *The risk of a proper (generalized, pseudo-) Bayes estimator of the form* $\delta_m(X) = X + \Sigma\nabla m(X)/m(X)$ *is given by*

$$\begin{aligned} R(\theta, \delta_m(X)) &= \operatorname{tr}\Sigma D + E\left[\frac{2m(X)(\sum_{i=1}^p \sigma_i^4 d_i \partial^2 m(X)/\partial X_i^2)}{m^2(X)} - \frac{\sum_{i=1}^p \sigma_i^4 d_i (\partial m(X)/\partial X_i)^2}{m^2(X)}\right] \\ &= \operatorname{tr}\Sigma D + 4E\left[\frac{\sum_{i=1}^p \sigma_i^4 d_i \partial^2 \sqrt{m(X)}/\partial X_i^2}{\sqrt{m(X)}}\right]. \end{aligned}$$

(d) *If the term in brackets in the last line of (c) is non-positive, the proper (generalized, pseudo-) Bayes estimator* $\delta_m(X)$ *is minimax.*

Proof. The proof is basically the same as Theorem 2.1 and is essentially in Stein (1981). □

Here is the key observation that allows us to construct Bayes minimax procedures for the present case based on procedures for the case $\Sigma = Q = I$.

Lemma 3.1. *Suppose* $\eta(X)$ *is such that* $\nabla^2\eta(X) = \sum \partial^2\eta(X)/\partial X_i^2 \leq 0$. *Then* $\eta^*(X) = \eta(\Sigma^{-1}D^{-1/2}X)$ *is such that* $\sum \sigma_i^4 d_i \partial^2 \eta^*(X)/\partial X_i^2 \leq 0$.

Proof. The proof follows by straightforward calculation noting that

$$\frac{\partial^2}{\partial X_i^2}\eta^*(X) = \sigma_i^{-4} d_i^{-1} \frac{\partial^2}{\partial y_i^2}\eta(y)\big|_{y=\Sigma^{-1}D^{-1/2}X}.$$

Hence

$$\sum \sigma_i^4 d_i \frac{\partial^2}{\partial X_i^2}\eta^*(X) = \sum \frac{\partial^2}{\partial y_i^2}\eta(y)\big|_{y=\Sigma^{-1}D^{-1}X} \leq 0.$$ □

Note by the same reasoning that if $\eta(X)$ is superharmonic, then so is $\eta(aX)$ for any scalar a.

The following theorem is the main result of this section.

Theorem 3.2. *Let* $X \sim N(\theta, \Sigma)$ *and* $L(\theta, d) = (d-\theta)'D(d-\theta)$ *where* Σ *and* D *are diagonal as above.*

(a) *Suppose* $\sqrt{m(X)}$ *is superharmonic (a proper, generalized, or pseudo-marginal for the case* $\Sigma = Q = I$*). Then*

$$\delta_m(X) = X + \Sigma\frac{\nabla m(\Sigma^{-1}D^{-1/2}X)}{m(\Sigma^{-1}D^{-1/2}X)}$$

is a proper, generalized, or pseudo-Bayes minimax estimator.

(b) *If* $\sqrt{m(\|X\|^2)}$ *is spherically symmetric and superharmonic, then*

$$\delta_m(X) = X + \frac{2m'(X'\Sigma^{-1}D^{-1}\Sigma^{-1}X)D^{-1}\Sigma^{-1}X}{m(X'\Sigma^{-1}D^{-1}\Sigma^{-1}X)}$$

is minimax.

(c) *Suppose the prior distribution* $\pi(\theta)$ *for* θ *has the hierarchical structure*

$$\theta|\lambda \sim N_p(0, A_\lambda), \quad \lambda \sim h(\lambda),\ 0 < \lambda < 1,$$

where $A_\lambda = (c/\lambda)\Sigma D\Sigma - \Sigma$ *where* c *is such that* A_1 *is positive definite and* $h(\lambda)$ *satisfies the conditions of Theorem* 2.2(c). *Then*

$$\delta_\pi(X) = X + \Sigma\frac{\nabla_X m(X)}{m(X)}$$

is minimax. [Note that such a c *may always be found and that* $h(\lambda)$ *may be any mixing distribution for which* $\sqrt{m(X)}$ *is superharmonic such as Alam's, etc.]*

(d) *Suppose* $m_i(X)$, $i = 1, \ldots, k$ *are superharmonic. Then the multiple shrinkage estimator*

$$\delta_m(X) = X + \Sigma\left[\frac{\sum_{i=1}^k \nabla_X m_i(\Sigma^{-1}D^{-1/2}X)}{\sum_{i=1}^k m_i(\Sigma^{-1}D^{-1/2}X)}\right]$$

is a minimax multiple shrinkage estimator.

Proof. Part (a) follows directly from Theorem 3.1(c) and (d) and Lemma 3.1. Part (b) follows from part (a) and Theorem 3.1(b) on straightforward calculation.

To show part (c), note that $X - \theta|\lambda \sim N(0, \Sigma)$ and $\theta|\lambda \sim N(0, A_\lambda)$ and $X - \theta$ and θ are therefore conditionally independent given λ. Hence $X|\lambda \sim N(0, A_\lambda + \Sigma)$. It follows that

$$m(X) \propto \int_0^1 \lambda^{p/2} \exp\left[-\frac{\lambda}{c}X'\Sigma^{-1}D^{-1}\Sigma^{-1}X\right] h(\lambda|d\lambda)$$

but $m(X) = \eta(X'\Sigma^{-1}D^{-1}\Sigma^{-1}X/c)$, where $\sqrt{\eta(X'X)}$ is superharmonic by Theorem 2.2(c). Hence by part (b), $\delta_\pi(X)$ is minimax (and proper or generalized Bayes depending on whether $h(\lambda)$ is integrable or not).

Part (d) follows from part (a) noting that $\eta(X)$ superharmonic implies that $\sqrt{\eta(X)}$ is superharmonic.

Note again that we ask for superharmonicity of $m_i(X)$ and not of $\sqrt{m_i(X)}$ in part (d). □

Examples 3.1. (a) *Pseudo-Bayes minimax estimators.* We saw in Example 2.1 that James-Stein estimators result from pseudo-marginals of the form $m(X) = (1/\|X\|^2)^b$. It follows from part (b) of Theorem 3.2 and Example 2.1 that $m(X'\Sigma^{-1}D^{-1}\Sigma^{-1}X) = (1/X'\Sigma^{-1}D^{-1}\Sigma^{-1}X)^b$ has associated pseudo-Bayes estimator $\delta_m(X) = X - (2bD^{-1}\Sigma^{-1}X)/(X'\Sigma^{-1}D^{-1}\Sigma^{-1}X)$ and that this estimator is minimax for $0 < b \leq (p-2)$.

(b) *A hierarchical (proper) Bayes minimax estimator* (Berger, 1976). Suppose $h(\lambda) \propto \lambda^{-a}$, $0 < \lambda < 1$, for $(6-p)/2 \leq a < 1$ and that $A_\lambda = c\Sigma D\Sigma - \Sigma$, where $c > 1/\min(\sigma_i^2 d_i)$. The resulting proper Bayes estimator is minimax by Example 2.2 and Theorem 3.2. That it is a generalized Bayes admissible minimax estimator for $a \leq 2$ will follow immediately from Section 3.

(c) *A multiple shrinkage minimax estimator.* It follows from Example 2.3 and Theorem 3.2 that

$$m(X) = \sum_{i=1}^{k} \left[\frac{1}{(X-\nu_i)'\Sigma^{-1}D^{-1}\Sigma^{-1}(X-\nu_i)} \right]^b$$

satisfies the conditions of Theorem 3.2(d) for $0 < b \leq (p-2)/2$ and hence that

$$\begin{aligned} &\delta_m(X) \\ &\quad = X - 2b\frac{\sum_{i=1}^{k}[D^{-1}\Sigma^{-1}(X-\nu_i)]/[((X-\nu_i)'\Sigma^{-1}D^{-1}\Sigma^{-1}(X-\nu_i))^{b+1}]}{\sum_{i=1}^{k} 1/[((X-\nu_i)'\Sigma^{-1}D^{-1}\Sigma^{-1}(X-\nu_i))^{b}]} \end{aligned}$$

is a minimax multiple shrinkage (pseudo-Bayes) estimator.
It is worth pointing out that in this example and in Example 2.3, a generalized Bayes minimax estimator (as opposed to a pseudo-Bayes estimator) results from the generalized prior

$$\pi(\theta) = \sum_{i=1}^{k} \left(\frac{1}{(\theta-\nu_i)'\Sigma^{-1}D^{-1}\Sigma^{-1}(\theta-\nu_i)} \right)^b$$

for $0 < b \leq (p-2)/2$.

4. The Case of Generalized Σ and Q

In this section we show that the case of general Σ and Q (both positive definite) can be reduced to the canonical form $\Sigma = I$, $Q = \text{diag}(d_1, \ldots, d_p) = D$. We use the following well-known fact repeatedly:

Lemma 4.1. *For any pair of positive definite matrices, Σ and Q, there exists a non-singular matrix A such that $A\Sigma A' = I$ and $(A')^{-1}QA^{-1} = D$ where D is diagonal.*

This fact leads to the following canonical form of the estimation problem:

Theorem 4.1. *Suppose $X \sim N(\theta, \Sigma)$ and loss is $L_1(\theta, d) = (d-\theta)'Q(d-\theta)$. Let A and D be as in Lemma 4.1, $Y = AX \sim N(\nu, I)$, $\nu = A\theta$, and $L_2(\nu, d) = (d-\nu)'D(d-\nu)$.*

(a) *For every estimator $\delta_1(X)$ with risk function $R_1(\theta, \delta_1) = EL_1(\theta, \delta_1(X))$, the estimator $\delta_2(Y) = A\delta_1(A^{-1}Y)$ has risk function $R_2(\nu, \delta_2) = R_1(\theta, \delta_1) = EL_2(\nu, \delta_2(Y))$.*

(b) *$\delta_1(X)$ is proper or generalized Bayes with respect to the prior distribution $\pi_1(\theta)$ (or pseudo-Bayes with respect to pseudo-marginal $m_1(X)$) under loss L_1 if and only if $\delta_2(Y) = A\delta_1(A^{-1}Y)$ is proper or generalized Bayes with respect to $\pi_2(\nu) = \pi_1(A^{-1}\nu)$ (or pseudo-Bayes with respect to pseudo-marginal $m_2(Y) = m_1(A^{-1}Y)$).*

(c) *$\delta_1(X)$ is admissible (or minimax or dominates $\delta_1^*(X)$) under L_1 if and only if $\delta_2(Y) = A\delta_1(A^{-1}Y)$ is admissible (or minimax or dominates $\delta_2^*(Y) = A\delta_1^*(A^{-1}Y)$) under L_2.*

Proof. (a)

$$
\begin{aligned}
R_2(\nu, \delta_2) &= E[L_2(\nu, \delta_2(Y))] \\
&= E[(\delta_2(Y) - \nu)'D(\delta_2(Y) - \nu)] \\
&= E[(A\delta_1(A^{-1}Y) - A\theta)'D(A\delta_1(A^{-1}Y) - A\theta)] \\
&= E(\delta_1(X) - \theta)'A'DA(\delta_1(X) - \theta) \\
&= E(\delta_1(X) - \theta)'Q(\delta_1(X) - \theta) \\
&= R_1(\theta, \delta_1).
\end{aligned}
$$

This completes the proof of part (a).

Part (b) follows upon noting that the Bayes estimator for any quadratic loss is the posterior mean. Hence, since if $\theta \sim \pi_1(\theta)$, $\nu = A\theta \sim \pi_2(\nu) = \pi_1(A^{-1}\nu)$ (ignoring constants) $\delta_2(Y) = E(\nu|Y) = E(A\theta|Y) = E(A(\theta|AX)) = AE(\theta|X) = A\delta_1(X) = A\delta_1(A^{-1}Y)$.

Part (c) follows directly from part (a). □

Strawderman (1978) contains a result which is similar to parts (a) and (c) of the above theorem. It is worth noting that if $\Sigma^{1/2}$ is the positive definite square root of Σ, and $A = P\Sigma^{-1/2}$ where P is orthogonal, then $Y = AX \sim N(\nu, I)$ and the above argument gives equivalence of the two corresponding problems $X \sim N(\theta, \Sigma)$, $L_1(\theta, d) = (d-\theta)'Q(d-\theta)$ and $Y \sim N(\nu, I)$, $L_2(\nu, d) = (d-\nu)'P\Sigma^{1/2}Q\Sigma^{1/2}P'(d-\nu)$. Hence in Theorem 4.1, $A = P\Sigma^{-1/2}$ where P diagonalizes $\Sigma^{1/2}Q\Sigma^{1/2}$ will work.

Examples 4.1. (a) *A pseudo-Bayes minimax estimator.* It follows from Example 3.1(a) and Theorem 4.1 that

$$m(X'\Sigma^{-1}Q^{-1}\Sigma^{-1}X) = \left(\frac{1}{X'\Sigma^{-1}Q^{-1}\Sigma^{-1}X}\right)^b$$

for $0 < b \leq p-2$ results in the minimax James-Stein estimators

$$\delta_m(X) = X - \frac{2bQ^{-1}\Sigma^{-1}X}{X'\Sigma^{-1}Q^{-1}\Sigma^{-1}X}.$$

(b) *A hierarchical (proper) Bayes minimax estimator.* Similarly for Example 3.1(b) and the above theorem we have that if

$$\theta \mid \lambda \sim N(0, A_\lambda), \quad \lambda \sim \lambda^{-a},\ 0 < \lambda < 1$$

for $(6-p)/2 \leq a < 1$ and

$$A_\lambda = \frac{c}{\lambda}\Sigma Q \Sigma - \Sigma$$

for $c > \inf \nu_i$, where ν_i are the eigenvalues of $\Sigma^{1/2}Q\Sigma^{1/2}$, then the resulting Bayes estimator is proper Bayes and minimax. It is admissible minimax if $(6-p)/2 \leq a \leq 2$. In fact the same is true if $\lambda \sim h(\lambda)$, where $h(\lambda)$ is any mixing distribution for which the Bayes estimator is minimax when $\Sigma = Q = I$.

(c) *A multiple shrinkage pseudo-Bayes minimax estimator.* Applying the result of this section to Example 3.1(c) implies that

$$m(X) = \sum_{i=1}^{k}\left[\frac{1}{(X-\nu_i)'\Sigma^{-1}Q^{-1}\Sigma^{-1}(X-\nu_i)}\right]^b$$

for $0 < b \leq (p-2)/2$ leads to the minimax estimator

$$\begin{aligned}\delta_m(X) \\ = X - 2b\left[\frac{\sum_1^k Q^{-1}\Sigma^{-1}(X_i-\nu_i)/\left((X-\nu_i)'\Sigma^{-1}Q^{-1}\Sigma^{-1}(X-\nu_i)\right)^{b+1}}{\sum_1^k 1/\left((X-\nu_i)'\Sigma^{-1}Q^{-1}\Sigma^{-1}(X-\nu_i)\right)^b}\right].\end{aligned}$$

Also a prior distribution of the form

$$\pi(\theta) = \sum_{i=1}^{k}\left(\frac{1}{(\theta-\nu_i)'\Sigma^{-1}Q^{-1}\Sigma^{-1}(\theta-\nu_i)}\right)^b$$

for $0 < b \leq (p-2)/2$ gives a minimax generalized Bayes estimator.
It follows from a result of Shinozaki (1980) that admissibility of a (generalized Bayes) estimator does not depend on the matrix Q in the loss function. Hence admissibility holds in the above case if $b = (p-2)/2$ from Brown (1971).

5. Summary and Conclusions

We have attempted in this paper to unify and extend much of the existing literature on minimax estimation of the mean of a multivariate normal distribution with arbitrary (known) covariance matrix and arbitrary quadratic loss.

The main results are

(a) If a (pseudo-) marginal $m(X)$ results in a minimax estimator for the case $\Sigma = Q = I$, then $m(\Sigma^{-1}D^{-1/2}X)$ results in a minimax estimator for the case Σ and D are diagonal, and

(b) If $m(X)$ (or $\pi(\theta)$) results in a minimax estimator for the case $\Sigma = I$ and $Q = D$ (diagonal), then $m(A^{-1}X)$ $(\pi(A^{-1}\theta))$ results in a minimax estimator for the general Σ, Q case where A is such that $A\Sigma A' = I$ and $(A')^{-1}QA^{-1} = D$.

(c) The general case can be reduced to the case $\Sigma = I$, $Q = D$ in such a way that risks, minimaxity, admissibility, and Bayesianity are all preserved through the correspondence $X \leftrightarrow AX$, $\theta \leftrightarrow A\theta$, $\delta(X) \leftrightarrow A\delta(A^{-1}X)$.

Most of the specific examples in the paper are not new but the point of view seems to be new and quite successful in unifying and extending known results.

Acknowledgements. The results in this paper were developed while preparing a series of lectures as Statistician in Residence at the University of British Columbia. The author thanks the Department of Statistics at UBC and especially Jim Zidek and Constance van Eeden for their hospitality during his stay. The author also thanks Dominique Fourdrinier and Marty Wells for many valuable discussions.

This research was supported by NSF grant DMS 97-04524.

REFERENCES

Alam, K. (1973). A family of admissible minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* 1, 517–525.

Berger, J.O. (1976). Admissible minimax estimators of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* 4, 223–226.

Berger, J.O. and Robert, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: On the frequentist interface. *Ann. Statist.* 18, 617–651.

Berger, J.O. and Strawderman, W.E. (1996). Choice of hierarchical prior: Admissibility in estimation of normal means. *Ann. Statist.* 24, 931–951.

Bock, M.E. (1975). Minimax estimation of the mean of a multivariate normal distribution. *Ann. Statist.* 3, 209–218.

Bock, M.E. (1988). Shrinkage estimators: Pseudo Bayes estimators for normal mean vectors. *Statistical Decision Theory IV* (S.S. Gupta and J.O. Berger, eds.), pp.281–298. Springer-Verlag, New York.

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

Faith, R.E. (1978). Minimax Bayes point estimators of a multivariate normal mean. *J. Mult. Anal.* 8, 372–379.

Fourdrinier, D., Strawderman, W.E., and Wells, M.T. (1998). On the construction of Bayes minimax estimators. *Ann. Statist.* 26, 660–671.

George, E.I. (1986). Minimax multiple shrinkage estimators. *Ann. Statist.* 14, 188–205.

Shinozski, N. (1980). Estimation of a multivariate normal mean with a class of quadratic loss functions. *J. Amer. Statist. Assoc.* 75, 973-976.

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

Strawderman, W.E. (1971). Proper Bayes minimax estimation of the multivariate normal mean. *Ann. Math. Statist.* 42, 385–388.

Strawderman, W.E. (1978). Minimax adaptive generalized ridge regression estimators. *J. Amer. Statist. Assoc.* 73, 623–627.

William E. Strawderman
Department of Statistics
561 Hill Center, Busch Campus
Rutgers University
Piscataway, NJ 08854-8019
USA
straw@stat.rutgers.edu