IMPROVING INADMISSIBLE ESTIMATORS UNDER QUADRATIC LOSS

by

Leon Jay Gleser
Purdue University

Technical Report #83-19

Department of Statistics
Purdue University

June 1983

This research was supported by the National Science Foundation under Grant No. MCS 81-21948.

IMPROVING INADMISSIBLE ESTIMATORS UNDER QUADRATIC LOSS

by

Leon Jay Gleser
Purdue University

1. INTRODUCTION

Let Z: mx1 be a random vector observation with distribution in a class $\{P_\theta: \theta \text{ in } \Theta\}$ indexed by a vector parameter θ: rx1. It is desired to estimate a vector-valued function $\mu = \mu(\theta)$: px1 of θ, $p \le r$, under a quadratic loss function

$$L(\theta,\delta(Z)) = (\delta(X)-\mu)'Q(\theta)(\delta(Z)-\mu), \qquad (1.1)$$

where $Q(\theta)$ is a pxp positive definite matrix ($Q(\theta) > 0$) for all θ. Let

$$R(\theta,\delta) = E_\theta[L(\theta,\delta(Z))] \qquad (1.2)$$

denote the risk function of $\delta = \delta(Z)$.

Suppose that $\delta_0(Z)$ is an estimator of μ having everywhere finite risk

$$R(\theta,\delta_0) < \infty, \quad \text{all } \theta \text{ in } \Theta, \qquad (1.3)$$

and that it is believed (or known) that δ_0 is inadmissible. Since the loss (1.1) is strictly convex in $\delta(Z)$, it follows that an estimator δ^* dominates δ_0 in risk if and only if $(1-\alpha)\delta_0 + \alpha\delta^*$ dominates δ_0 for all α, $0 < \alpha \le 1$. This fact suggests that if δ^* dominates δ_0, then among all estimators δ_c of the form $(1-c)\delta_0 + c\delta^*$, $-\infty < c < \infty$, the maximal subclass of estimators which dominate δ_0 has the form $\{(1-c)\delta_0 + c\delta^*: 0 < c \le \Delta\}$, where $\Delta \ge 1$. Even if δ^* does not dominate δ_0, it may be possible that $\delta_c = (1-c)\delta_0 + c\delta^*$ dominates δ_0 in risk for some $c > 0$. Theorem 1 below gives necessary and sufficient conditions for this to occur, and in this case also describes the maximal class of estimators of the form δ_c which dominate δ_0.

Theorem 1. Let

$$h(Z) = \delta_0(Z) - \delta^*(Z) \tag{1.4}$$

and define

$$B(\theta) = E_\theta[h'(Z)Q(\theta)(\delta_0(Z)-\mu)],\ A(\theta) = E_\theta[h'(Z)Q(\theta)h(Z)]. \tag{1.5}$$

In order that for some $c > 0$

$$\delta_c(\underline{Z}) = (1-c)\delta_0(Z) + c\delta^*(Z) = \delta_0(Z) - ch(Z) \tag{1.6}$$

dominates $\delta_0(Z)$ in risk, it is both necessary and sufficient that

$$\Delta = 2 \inf_\theta \frac{B(\theta)}{A(\theta)} > 0. \tag{1.7}$$

If (1.7) holds, and $(A(\theta))^{-1}B(\theta)$ is not constant for all θ, then δ_c dominates δ_0 in risk if and only if $0 < c \leq \Delta$. If $(A(\theta))^{-1}B(\theta) = K > 0$ all θ, then δ_0 and δ_{2K} have identical risk, and δ_c dominates δ_0 in risk if and only if $0 < c < \Delta = 2K$.

It should be noted that some of the members of the class $\{\delta_c: \ 0 < c \leq \Delta\}$ are themselves dominated in risk.

Theorem 2. If the condition (1.7) of Theorem 1 holds, then every estimator $\delta_c = \delta_0 - ch$, $0 < c < \frac{1}{2}\Delta$ is dominated in risk by $\delta_{\frac{1}{2}\Delta}$. If, further, $(A(\theta))^{-1}B(\theta) = K > 0$, all θ, then $\delta_{\frac{1}{2}\Delta}$ dominates all estimators δ_c, $c \neq \frac{1}{2}\Delta$, in risk.

Theorems 1 and 2 are proven in Section 2. In Section 3, these theorems are applied to the familiar problem where

$Z = X \sim MVN(\mu,\Sigma)$, μ: px1 unknown, $\Sigma > 0$ known.

2. PROOFS OF THEOREMS 1 AND 2

2.1 Proof of Theorem 1.

In what follows, it can be assumed that

$$A(\theta) = E_\theta[h'(Z)Q(\theta)h(Z)] < \infty, \text{ all } \theta, \tag{2.1}$$

as shown by the following lemma.

Lemma 1. *If $A(\theta_0) = \infty$ for any θ_0 in Θ, then there does not exist $\delta_c = \delta_0 - ch$, $c \neq 0$, for which δ_c dominates δ_0 in risk. Further, in this case,* $\Delta = 2 \inf_\theta (A(\theta))^{-1}|B(\theta)| = 0$.

Proof. Let $1_k(Z)$ be the indicator function of

$$\{Z: \; h'(Z)Q(\theta_0)h(Z) \leq k\}.$$

Define $\mu_0 = \mu(\theta_0)$,

$$A_k(\theta_0) = E_{\theta_0}[1_k(Z)h'(Z)Q(\theta_0)h(Z)],$$

$$B_k(\theta_0) = E_{\theta_0}[1_k(Z)h'(Z)Q(\theta_0)(\delta_0(Z)-\mu_0)], \tag{2.2}$$

$$R_k(\theta_0,\delta_c) = E_{\theta_0}[1_k(Z)L(\theta_0,\delta_c(Z))].$$

By the Cauchy-Schwartz inequality,

$$B_k(\theta_0) \leq [A_k(\theta_0)R_k(\theta_0,\delta_0)]^{\frac{1}{2}},$$

so that from (1.1), (1.2), (1.4) through (1.6), and (2.3),

$$\begin{aligned} R_k(\theta_0,\delta_c) &= c^2A_k(\theta_0) - 2cB_k(\theta_0) + R_k(\theta_0,\delta_0) \\ &\geq (cA_k^{\frac{1}{2}}(\theta_0) - R_k^{\frac{1}{2}}(\theta_0,\delta_0))^2, \text{ all } c. \end{aligned} \tag{2.3}$$

Since $R(\theta_0,\delta_0) < \infty$ by (1.3) and $A(\theta_0) = \infty$, by the given, taking $k \to \infty$ in (2.3) yields

$$R(\theta_0,\delta_c) = \lim_{k\to\infty} R_k(\theta_0,\delta_c) = \infty > R(\theta_0,\delta_0), \text{ all } c \neq 0.$$

Consequently, no δ_c, $c \neq 0$, can dominate δ_0 in risk. Finally,

$$0 \leq \Delta \leq \frac{2|B(\theta_0)|}{A(\theta_0)} = \lim_{k\to\infty} \frac{2|B_k(\theta_0)|}{A_k(\theta_0)} \leq \lim_{k\to\infty}\left[\frac{R_k(\theta_0,\delta_0)}{A_k(\theta_0)}\right]^{\frac{1}{2}} = 0,$$

proving that $\Delta = 0$. □

Lemma 1 verifies Theorm 1 for the case where $A(\theta_0) = \infty$, some θ_0. Thus, assume that (2.1) holds. Note that

$$|B(\theta)| \leq [A(\theta)R(\theta,\delta_0)]^{\frac{1}{2}} \tag{2.4}$$

by the Cauchy-Schwartz inequality. Thus, it follows from (1.3) and (2.1) that $|B(\theta)| < \infty$. Hence, for fixed θ, fixed $c > 0$,

$$R(\theta,\delta_c) = c^2A(\theta) - 2cB(\theta) + R(\theta,\delta_0), \tag{2.5}$$

and consequently

$$\begin{aligned} R(\theta,\delta_c) \leq R(\theta,\delta_0) &\text{ if and only if } c(cA(\theta) - 2B(\theta)) \leq 0 \\ &\text{ if and only if } c \leq \frac{2B(\theta)}{A(\theta)}. \end{aligned} \tag{2.6}$$

Further,

$$R(\theta,\delta_c) < R(\theta,\delta_0) \text{ if and only if } c < \frac{2B(\theta)}{A(\theta)}. \tag{2.7}$$

Therefore, for $c > 0$,

$$\delta_c \text{ dominates } \delta_0 \text{ in risk if and only if} \begin{cases} 0 < c \leq \dfrac{2B(\theta)}{A(\theta)}, \text{ all } \theta, \\ 0 < c < \dfrac{2B(\theta_0)}{A(\theta_0)}, \text{ some } \theta_0. \end{cases} \tag{2.8}$$

If $\Delta = 2 \inf_\theta A^{-1}(\theta)B(\theta) > 0$, then it is easily seen from (2.8) that δ_c dominates δ_0 in risk for all c, $0 < c < \Delta$. Unless $2(A(\theta))^{-1}B(\theta) = \Delta > 0$, all θ, $c = \Delta$ also satisfies the right side of (2.8), proving that δ_Δ

dominates δ_0. If $2(A(\theta))^{-1}B(\theta) = \Delta > 0$, all θ, then it follows from (2.5) that $R(\theta,\delta_\Delta) = R(\theta,\delta_0)$, all θ, so that δ_Δ and δ_0 are risk equivalent.

On the other hand, if δ_{c*} dominates δ_0 in risk for some $c* > 0$, then since $A(\theta) \geq 0$, all θ, it follows from (2.8) that $\Delta = 2 \inf_\theta (A(\theta))^{-1}B(\theta) \geq c* > 0$. Consequently, condition (1.7) of Theorem 1 is satisfied. This completes the proof of Theorem 1. □

2.2 Proof of Theorem 2. Because of Lemma 1 and (2.4), it can be assumed that $A(\theta)$ and $B(\theta)$ are finite for all θ. From (2.5),

$$R(\theta,\delta_{\frac{1}{2}\Delta}) - R(\theta,\delta_c) = A(\theta)(\tfrac{1}{2}\Delta+c-2\,\frac{B(\theta)}{A(\theta)})(\tfrac{1}{2}\Delta-c). \qquad (2.9)$$

Since $A(\theta) > 0$, all θ, it follows from (2.9) that if $0 < c < \frac{1}{2}\Delta$,

$$R(\theta,\delta_{\frac{1}{2}\Delta}) - R(\theta,\delta_c) \leq A(\theta)(\tfrac{1}{2}\Delta + c-\Delta)(\tfrac{1}{2}\Delta-c) < 0$$

for all θ, proving that $\delta_{\frac{1}{2}\Delta}$ dominates δ_c. Now suppose that $2(A(\theta))^{-1}B(\theta) = \Delta$, all θ. Let $\frac{1}{2}\Delta = K$. Then, from (2.9),

$$R(\theta,\delta_K) - R(\theta,\delta_c) = A(\theta)(K+c-2K)(K-c) = -A(\theta)(K-c)^2 < 0,$$

all $c \neq K$, all θ. This completes the proof of Theorem 2. □

2.3 Remarks.

A. In an entirely analagous fashion to the proof of Theorem 1, it can be shown that δ_c dominates δ_0 for some $c < 0$ if and only if

$$(1.7') \quad -\Delta = 2 \sup_\mu \frac{B(\theta)}{A(\theta)} < 0.$$

If (1.7') holds, and $(A(\theta))^{-1}B(\theta)$ is not everywhere constant, then δ_c dominates δ_0 if and only if $-\Delta \leq c < 0$, but $\delta_{-\frac{1}{2}\Delta}$ dominates δ_c in risk for all $-\frac{1}{2}\Delta < c < 0$. If $(A(\theta))^{-1}B(\theta) = -K$, $K > 0$, all θ, then δ_{-K} dominates δ_c for all $c \neq -K$.

B. It follows from Theorem 1 that for δ^* to dominate δ_0 in risk, it is necessary that:

$$B(\theta) = E_\theta\{(\delta_0(Z)-\delta^*(Z))'Q(\theta_0)(\delta_0(Z)-\mu)\} \geq 0, \qquad \text{all } \theta. \tag{2.10}$$

It is recommended that this requirement be used to weed out bad candidates δ^*. In many cases, simply taking limits of $B(\theta)$ as θ goes to various extreme values in Θ will show that (2.10) cannot hold.

It is often the case that $\delta_0(Z)$ is an unbiased estimator of μ. In this case $B(\theta)$ is a weighted sum of the covariances

$$\text{cov}(\delta_{0i}(Z)-\delta_i^*(Z),\delta_{0j}(Z)),$$

where $\delta_0(Z) = (\delta_{01}(Z),\ldots,\delta_{0p}(Z))'$, $\delta^*(Z) = (\delta_1^*(Z),\ldots,\delta_p^*(Z))'$. The requirement (2.10) states that this weighted sum of covariances must be nonnegative for all θ.

C. Suppose that (2.10) holds. In this case, for δ^* to dominate δ_0 in risk, it is necessary that

$$\Delta = 2 \inf_\theta \frac{B(\theta)}{A(\theta)} \geq 1. \tag{2.11}$$

The condition (2.11) is also sufficient unless $(A(\theta))^{-1}B(\theta)$ is constant for all θ (in which case, strict inequality in (2.11) is necessary and sufficient). Since (2.11) is not always easy to verify, it is helpful to note that (2.4) implies that

$$\Delta \leq 2 \inf_\theta\left[\frac{R(\theta,\delta_0)}{A(\theta)}\right]^{\frac{1}{2}}$$

and consequently,

$$A(\theta) \leq 4R(\theta,\delta_0), \text{ all } \theta, \tag{2.12}$$

is a necessary condition for δ^* to dominate δ_0 in risk. As in Remark B, limits as θ goes to extremes in (2.12) can frequently provide enough evidence to eliminate δ^* from consideration. Note that (2.12) places a bound on the expected length of $h(Z) = \delta_0(Z) - \delta^*(Z)$. If δ_0 is an equalizer rule ($R(\theta,\delta_0) = L$, all θ), and $Q(\theta) = Q > 0$, all θ, (2.12) implies that

$$E_\theta[h'(Z)h(Z)] \leq 4L(\lambda_{min}(Q))^{-1},$$

and is implied by

$$E_\theta[h'(Z)h(Z)] \leq 4L(\lambda_{max}(Q))^{-1},$$

where $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the smallest and largest eigenvalues, respectively, of a matrix A.

D. If one can show that δ^* dominates δ_0 in risk, then Theorem 1 states that $\{(1-\alpha)\delta_0 + \alpha\delta^*, 0 < \alpha \leq 1\}$ is a class of rules, each of which dominates δ_0 in risk. However, Theorem 2 shows that every member of the subclass $\{(1-\alpha)\delta_0 + \alpha\delta^*, 0 < \alpha < \frac{1}{2}\}$ is itself dominated in risk. Indeed, if $\Delta > 2$ in (ii) of Theorem 1, the entire class $(1-\alpha)\delta_0 + \alpha\delta^*$, $0 \leq \alpha \leq 1$, of rules is dominated in risk by $(1-\frac{1}{2}\Delta)\delta_0 + \frac{1}{2}\Delta\delta^*$.

3. NORMAL DISTRIBUTION: KNOWN COVARIANCE MATRIX

Consider

$X \sim MVN(\mu,\Sigma)$, μ: px1 unknown, $\Sigma > 0$ known.

Let $\delta_0(X) = X$, the UMVUE and MLE of μ. For the quadratic loss function

$$L(\theta,\delta(X)) = (trQ\Sigma)^{-1}(\delta(X)-\mu)'Q(\delta(X)-\mu), \quad (3.1)$$

where $Q > 0$ is a known matrix, $\delta_0(X)$ is minimax. An extensive literature [see Berger (1982)] considers classes of estimators of the form

$$\delta_c(X) = \delta_0(X) - ch(X) \tag{3.2}$$

which dominate δ_0 in risk when $p \geq 3$. On the other hand, δ_0 is known to be admissible when $p = 1,2$. This last fact, plus Theorem 1 and Remark B of Section 2, yields the following interesting result.

Lemma 2. *When $p = 1,2$, no measurable $p\times 1$ function $\ell(X)$ can exist, $\ell(X) = (\ell_1(X),\ldots,\ell_p(X))'$, for which*

$$\inf_\mu \frac{\sum_{i=1}^p \mathrm{cov}_\mu(x_i,\ell_i(X))}{\sum_{i=1}^p E_\mu(\ell_i^2(X))} > 0.$$

Proof. Let $h(X) = Q^{-1}\ell(X)$ in Theorem 1, and use Remark B of Section 2, plus the fact that $\ell'Q^{-1}\ell \leq \ell'\ell\ \lambda_{max}(Q^{-1})$. □

From now on, assume that $p \geq 3$. For any vector-valued function $h(X)$: $p\times 1$ which is absolutely continuous with respect to p-dimensional Lebesgue measure, let

$$\nabla h(X) = \left(\left(\frac{\partial h_j(X)}{\partial x_i}\right)\right):\quad p\times p \tag{3.3}$$

be the gradient of $h(X)$ at X. Using Stein's integration-by-parts method,

$$B(\mu) = B(\theta) = \frac{E_\mu[h'(X)Q(X-\mu)]}{tr(Q\Sigma)} = \frac{E_\mu[tr(Q\Sigma\nabla h(X))]}{tr(Q\Sigma)}, \tag{3.4}$$

whenever $E_\mu[h'(X)h(X)] < \infty$.

Theorem 3. *In order that for some $c > 0$, an estimator of the form $\delta_c(X) = X - ch(X)$ dominates $\delta_0(X) = X$ in risk under the loss function (3.1), it is both necessary and sufficient that*

$$\Delta = 2 \inf_{\mu} \frac{E_\mu[tr(Q\Sigma\nabla h(X))]}{E_\mu[h'(X)Qh(X)]} > 0. \tag{3.5}$$

If (3.5) holds, $\delta_c(X)$ _dominates_ $\delta_0(X)$ _in risk if_ $0 < c < \Delta$, _and only if_ $0 < c \leq \Delta$.

Proof. Note that

$$\begin{aligned} &\lambda_{min}(Q)E_\mu(h'(X)h(X)) \\ &\quad \leq tr(Q\Sigma)A(\mu) = E_\mu[h'(X)Qh(X)] \\ &\quad \leq \lambda_{max}(Q)E_\mu(h'(X)h(X)). \end{aligned} \tag{3.6}$$

If $E_{\mu_0}[h'(X)h(X)] = \infty$ for any μ_0, then by (3.6), $A(\mu_0) = \infty$. In this case, Lemma 1 verifies Theorem 3. If $E_\mu[h'(X)h(X)] < \infty$, all μ, then Theorem 3 follows as a direct consequence of (3.4) and Theorem 1. □

Corollary 1. _Let_ $\delta^*(X)$ _be absolutely continuous with respect to p-dimensional Lebesgue measure_. _For_ $\delta^*(X)$ _to dominate_ $\delta_0(X) = X$ _in risk with respect to the loss function (3.1) it is necessary that_

$$\frac{2E_\mu[tr\{Q\Sigma\nabla(X-\delta^*(X))\}]}{E_\mu[(X-\delta^*(X))'Q(X-\delta^*(X))]} \geq 1, \text{ all } \mu. \tag{3.8}$$

The condition (3.8) is also sufficient for $\delta^*(X)$ _to dominate_ $\delta_0(X)$ _in risk, except that when_ $h(X) = X-\delta^*(X)$ _satisfies the partial differential equation_

$$2\, tr[Q\Sigma\nabla h(X)] = h'(X)Qh(X), \tag{3.9}$$

$\delta^*(X)$ _and_ $\delta_0(X)$ _have identical risk functions_.

Proof. The necessity of (3.8) follows directly from Theorem 3. Unless the inequality in (3.8) is an equality for all μ, Theorem 1 shows that (3.8) is also sufficient. However, if (3.8) is an equality for all μ, then

$$E_\mu[2tr(Q\Sigma\nabla h(X))] = E_\mu[h'(X)Qh(X)], \text{ all } \mu, \tag{3.10}$$

in which case Theorem 1 states that $\delta^*(X)$ and $\delta_0(X)$ are risk equivalent. Since the family $\{MVN(\mu,\Sigma), -\infty < \mu < \infty\}$ of distributions of X is complete, (3.10) can hold if and only if (3.9) holds almost everywhere. This completes the proof. □

It is worth noting that the class of nonzero solutions of (3.9) is nonempty, since

$$h(X) = \left(\frac{2(p-2)}{X'\Sigma^{-1}Q^{-1}\Sigma^{-1}X}\right)Q^{-1}\Sigma^{-1}X \tag{3.11}$$

satisfies (3.9). The class of all nonzero solutions to (3.9) is worth obtaining, since it is easily shown that every convex combination

$$h^*(X) = \sum_{i=1}^{a} \alpha_i h_i(X), \ \alpha_i \geq 0, \ \sum_{i=1}^{a} \alpha_i = 1,$$

of nonzero solutions $h_i(X)$ of (3.9) defines an estimator $\delta^*(X) = X - h^*(X)$ which dominates $\delta_0(X) = X$ in risk.

REFERENCES

Berger, James O. (1982). Selecting a minimax estimator of a multivariate normal mean. *Ann. Statist.* 10, 81-92.