IMPROVING UPON SOME STANDARD ESTIMATORS IN CONTINUOUS EXPONENTIAL FAMILIES UNDER ARBITRARY QUADRATIC LOSS

by

Shun-Yu Chen*
Purdue University

Technical Report #83-2

Department of Statistics
Division of Mathematical Sciences
Technical Report #83-2

February 1983

*Research supported by the National Science Foundation under Grant MCS-8101670 and David Ross Grant XR 0671.

IMPROVING UPON SOME STANDARD ESTIMATORS IN CONTINUOUS EXPONENTIAL FAMILIES UNDER ARBITRARY QUADRATIC LOSS

by

Shun-Yu Chen*
Purdue University

Summary

Let $X = (X_1,\ldots,X_p)^t$ have density $f(x|\theta) = \beta(\theta,\Sigma)t(x)e^{\theta^t\Sigma^{-1}x}$, where t satisfies some conditions given in Section 2. It is desired to estimate $\theta = (\theta_1,\ldots,\theta_p)^t$ under the quadratic loss $L(\theta,\delta) = (\theta-\delta)^tQ(\theta-\delta)$, where Σ, Q are positive definite matrices. First, we obtain an identity and using this, we show that the M.L.E. and unbiased estimator, $\delta_0(x) = -\Sigma\nabla \log t(x)$, is inadmissible ($p \geq 3$) and we obtain a class of better estimators. This result is applied to the multivariate normal situation, where Σ is known or partially unknown ($\Sigma = \sigma^2\Sigma_0$ with Σ_0 known and σ^2 unknown). A broad class of minimax estimators for θ is developed.

*Research supported by the National Science Foundation under Grant MCS-8101670 and David Ross Grant XR 0671.

1. Introduction.

Let $X = (X_1,\ldots,X_p)^t$ represent a vector valued random variable taking values x in a sample space $\mathcal{X} = \mathcal{X}_1 \times\ldots\times \mathcal{X}_p$, where the $\mathcal{X}_i$ are (possible infinite) intervals (a_i,b_i). Assume X has density

$$(1.1) \qquad f(x|\theta) = \beta(\theta,\ddagger)t(x)e^{\theta^t \ddagger^{-1} x}, \quad x \in \mathcal{X}$$

where $\ddagger$ is a given positive definite matrix, and let Θ represent the natural parameter space, i.e.

$$\Theta = \{\theta = (\theta_1,\ldots,\theta_p)^t\colon \int_{\mathcal{X}} t(x)e^{\theta^t \ddagger^{-1} x}dx < \infty\}.$$

It is desired to estimate θ using an estimator $\delta\colon \mathcal{X} \to \Theta$ under quadratic loss

$$(1.2) \qquad L(\theta,\delta) = (\delta-\theta)^t Q(\delta-\theta),$$

where Q is a given positive definite matrix. An estimator will be evaluated by its risk function

$$R(\theta,\delta) = E_\theta^X[L(\theta,\delta(x))] = \int_{\mathcal{X}} L(\theta,\delta(x))f(x|\theta)dx.$$

Here E stands for expectation, with subscripts denoting parameter value at which the expectation is to be taken and superscripts denoting random variable over which the expectation is to be taken.

An estimator δ_1 is defined to be as good as δ_2 if

$$(1.3) \qquad R(\theta,\delta_1) \leq R(\theta,\delta_2)$$

for all $\theta \in \Theta$. The estimator δ_1 is said to dominate δ_2 (or is better than δ_2) if, in addition to (1.3)

$$R(\theta,\delta_1) < R(\theta,\delta_2)$$

for some $\theta \in \Theta$. The estimator δ is admissible if there exists no better estimator, and is inadmissible otherwise.

The following notations will be used throughout the paper. First, let $h: R^p \to R$ be a differentiable function. Then

$$h^{(i)}(x) = \frac{\partial}{\partial x_i} h(x), \quad h^{(i,j)}(x) = \frac{\partial^2}{\partial x_i \partial x_j} h(x),$$

$$\nabla h(x) = (h^{(1)}(x),\ldots,h^{(p)}(x))^t, \; \nabla^2 h(x) = \sum_{i=1}^{p} h^{(i,i)}(x),$$

and h is superharmonic if $\nabla^2 h(x) \leq 0$ for all $x \in R^p$. Second, if $\gamma(x) = (\gamma_1(x),\ldots,\gamma_p(x))^t$ then div $\gamma(x) = \sum_{i=1}^{p} \gamma_i^{(i)}(x)$. Finally, if A is a pxp positive definite matrix then trA = trace of A, $ch_{max}A$ = the maximum characteristic root of A, and $A^{\frac{1}{2}}$ denotes the unique positive definite square root of A.

For the multivariate normal case, Stein (1955) proved the surprising result (for $\Sigma = Q = I_p$) that the standard estimator $\delta_0(x) = x$ is inadmissible when $p \geq 3$. A better estimator, δ^{J-S}, was found in James and Stein (1960), which has the form

$$\delta^{J-S}(x) = (1 - \frac{p-2}{\sum_{i=1}^{p} x_i^2})x.$$

Since then, a considerable amount of work by a number of authors has gone into finding significant improvements upon the usual 'standard' estimators in more general settings. Berger (1980a) also obtained Stein type results for the gamma distribution.

In section 2, we use integration by parts (first so used by Stein - see Stein (1981)) to establish a powerful identity for continuous exponential families, which will be useful for establishing the subsequent general dominance results.

In section 3, generalizing Hudson (1978), we consider a class of estimators of the form

$$(1.4) \qquad \delta(x) = -\ddagger\nabla \log t(x) + 2\ddagger Q^{*-\frac{1}{2}} \frac{\nabla h(y)}{h(y)},$$

where $Q^* = \ddagger Q \ddagger$ and $y = Q^{*-\frac{1}{2}}x$. If h is superharmonic and satisfies some regularity conditions, then δ dominates the unbiased estimator $\delta_0(x) = -\ddagger\nabla\log t(x)$. Specifically, when $p \geq 3$ the estimator

$$(1.5) \qquad \delta(x) = -\ddagger\nabla\log t(x) - \frac{2\lambda(p-2)}{r} Q^{-1}\ddagger^{-1}x,$$

where $r = x^t Q^{*-1} x$ and $0 < \lambda \leq 1$, dominates δ_0.

In section 4, we consider the nonsymmetric multivariate normal case. Estimators having uniformly smaller risk than $\delta_0(x) = x$ for this situation have been found by Berger (1976a, 1976b, 1979, 1980b, 1982b), Bhattacharya (1966), Bock (1975), Efron and Morris (1976), Hudson (1974), Lin and Tsai (1973), Strawderman (1978), and many others. Here we consider $X \sim N_p(\theta, \sigma^2\ddagger)$, $\ddagger$ is a given positive definite matrix, and σ^2 is unknown but a variable S^2 (independent of X) is observed with S^2/σ^2 having a chi-square distribution with n degree of freedom. A class of more simple form of minimax estimators is given by

$$(1.6) \qquad \delta(x) = x + \frac{2S^2}{n+2} \ddagger Q^{*-\frac{1}{2}} \frac{\nabla h(y)}{h(y)},$$

where $y = Q^{*-\frac{1}{2}}x$ and h is a superharmonic function satisfying the reqularity conditions.

In section 5, it is shown that if $X_1 \sim N(\theta_1, \sigma^2)$ and X_2 is independent of X_1 with a gamma (α, θ_2) distribution where $\alpha > 2$, then the standard estimator

$$(1.7) \qquad \delta_0(x) = (x_1, \frac{\alpha-2}{x_2})^t$$

is inadmissible under the sum of squared error loss function. We develop a class of improved estimators which exhibits the same phenomenon as that in Berger (1980a) (Berger considered a simultaneous estimation of a two -- instead of one -- dimensional normal mean and a gamma scale parameter), that the improved estimator treats the coordinates quite differently; one is shrunk towards zero, while the other is shifted towards infinity.

2. An Identity.

Let X have density $f(x|\theta)$, given by (1.1), in this paper we will assume that $t(x)$ satisfies the following conditions:

(i) The function t is differentiable and $E_\theta^X||\nabla \log t(x)||^2 < \infty$;

(ii) for $i = 1,\ldots,p$ and all η, where $\eta = \ddagger^{-1}\theta$,

$$\lim_{x_i \to a_i} t(x)e^{\eta_i x_i} = \lim_{x_i \to b_i} t(x)e^{\eta_i x_i} = 0.$$

Let

$$\delta_0(x) = \ddagger\nabla \log m_0(x) - \ddagger\nabla \log t(x), \tag{2.1}$$

where m_0 is a positive differentiable function, be thought of as the "standard" estimator or estimator under investigation and δ as a competing estimator. Write δ in the form

$$\delta(x) = \delta_0(x) - \ddagger\gamma(x) \tag{2.2}$$

where γ_i satisfies the following regularity conditions for $i = 1,\ldots,p$:

(i) except possibly for $(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_p)$ in a set of probability zero, $\gamma_i(x)$ is a continuous piecewise differentiable function of x_i and

$$\lim_{x_i \to a_i} \gamma_i(x)t(x)e^{\eta_i x_i} = \lim_{x_i \to b_i} \gamma_i(x)t(x)e^{\eta_i x_i} = 0$$

for all η, where $\eta = \ddagger^{-1}\theta$.

(ii) $E_\theta[\gamma_i^2(x)] < \infty$ and $E_\theta|\gamma_i^{(i)}(x)| < \infty$.

Under the loss function $L(\theta,\delta)$ defined in (1.2), the following identity (which is a generalization of identities in Hudson (1978), Berger (1982), and Stein (1981)) holds.

Theorem 1. Let δ_0 and δ be defined as in (2.1) and (2.2), respectively. If γ satisfies the regularity conditions, then

(2.3) $$R(\theta,\delta) - R(\theta,\delta_0) = E_\theta[\mathcal{D}\,\gamma(x)],$$

where

(2.4) $$\mathcal{D}[\gamma(x)] = \gamma^t(x)Q^*\gamma(x) - 2\gamma^t(x)Q^*\nabla \log m_0(x) - 2\mathrm{tr}(VQ^*)$$

and $Q^* = \ddagger Q \ddagger$, $V = (v_{ij})$ with $v_{ij} = \frac{\partial \gamma_i}{\partial x_j}$.

Proof. Let $\eta = \ddagger^{-1}\theta$ and for any given pxp matrix A, let $\lambda(x) = A\gamma(x)$. Clearly, λ also satisfies the reqularity conditions. Using the regularity conditions on λ and the Cauchy-Schwartz inequality, it is easy to show that

$$E_\theta\left[\frac{1}{t(x)}\left|\frac{\partial}{\partial x_i}\{\lambda_i(x)t(x)\}\right|\right] < \infty.$$

Then, using integration by parts and regularity condition (i), we have

$$\int_{a_i}^{b_i} \lambda_i^{(i)}(x)t(x)e^{\eta_i x_i}dx_i = -\int_{a_i}^{b_i} [\eta_i\lambda_i(x) + \lambda_i(x)\frac{\partial \log t(x)}{\partial x_i}]t(x)e^{\eta_i x_i}dx_i.$$

Thus

$$E_\eta[\lambda_i^{(i)}(x)] = \int_{\pi_{j\neq i}\mathcal{X}_j} \beta(\theta,\ddagger)\, \pi_{j\neq i}\, e^{\sum_{j\neq i}^{p} \eta_j x_j} \left[\int_{a_i}^{b_i} \lambda_i^{(i)}(x)t(x)e^{\eta_i x_i}dx_i\right] \pi_{j\neq i}\, dx_j$$

$$= -E_\eta[\eta_i\lambda_i(x) + \lambda_i(x)\frac{\partial \log t(x)}{\partial x_i}].$$

Therefore

$$E_\eta[\eta^t \lambda(x)] = -E_\eta[\lambda^t(x) \nabla \log t(x) + \text{div } \lambda(x)]$$

and so

$$(2.4) \qquad E_\theta[\theta^t \Sigma^{-1} A\gamma(x)] = -E_\theta[\gamma^t(x) A^t \nabla \log t(x) + \text{div}(A\gamma(x))].$$

It is easy to show that

$$(2.5) \qquad R(\theta,\delta) - R(\theta,\delta_0) = E_\theta[\gamma^t(x) Q^*\gamma(x) - 2\delta_0^t(x) Q\Sigma\gamma(x)] + 2E_\theta[\theta^t(Q\Sigma\gamma(x))].$$

Now, let $A = Q^*$ in (2.4) we have

$$E_\theta[\theta^t(Q\Sigma\gamma(x))] = E_\theta[\theta^t \Sigma^{-1} Q^*\gamma(x)]$$
$$= -E_\theta[\gamma^t(x) Q^* \nabla \log t(x) + \text{div}(Q^*\gamma(x))].$$

Inserting the result in (2.5) and noting that

$$\text{div}(Q^*\gamma(x)) = \text{tr}(Q^* V) = \text{tr}(VQ^*),$$

we get the result.||

Remark 1. If $(a_i,b_i) = (-\infty,\infty)$ then $E_\theta|\gamma_i(x)| < \infty$ implies

$$\lim_{x_i \to a_i} \gamma_i(x)t(x)e^{\eta_i x_i} = \lim_{x_i \to b_i} \gamma_i(x)t(x)e^{\eta_i x_i} = 0.$$

Remark 2. Proofs of inadmissibility of various estimators δ_0 using versions of theorem 1 have been carried out in Hudson (1978) and Berger (1980a, 1982a). A systematic approach to the problem is to solve the nonlinear differential inequality

$$(2.6) \qquad \mathcal{D}\gamma(x) \leq 0$$

for some γ satisfying the regularity conditions, where $\mathcal{D}\gamma(x)$ is defined as in (2.4).

Remark 3. If γ satisfies (2.6), then $\lambda\gamma$ also satisfies the differential inequality for any constant λ with $0 < \lambda \leq 1$.

Remark 4. In solving the differential inequality (2.6), usually we will consider $\gamma(x) = -2\nabla \log g(x)$ for some positive differentiable function g and (2.6) then becomes

$$(2.7) \qquad \frac{4}{g(x)}\,[\nabla^t g(x) Q^* \nabla \log m_0(x) + \mathrm{tr}(GQ^*)] \le 0,$$

where $G = (g^{(i,j)}(x))$.

(2.7) is easier to handle than (2.6). Also, from the statistical view point, we know that any admissible estimator must be of the form

$$\ddagger \nabla \log g(x) - \ddagger \nabla \log t(x)$$

(see Berger and Srinivasan (1978)).

Remark 5. In (2.1), if we choose $\nabla \log m_0(x) = 0$, then $\delta_0(x) = -\ddagger \nabla \log t(x)$ is M.L.E. and unbiased estimator of θ.

3. Improvements upon the Unbiased Estimator.

In this section, we use the identity in Theorem 1 to prove that the 'standard' unbiased estimator

$$(3.1) \qquad \delta_0(x) = -\ddagger \nabla \log t(x)$$

is inadmissible under the quadratic loss function $L(\theta,\delta)$ defined in (1.2), and a class of better estimators is proposed. Let δ be defined as in (1.4), where $Q^* = \ddagger Q \ddagger$, and $y = Q^{*^{-\frac{1}{2}}} x$.

Theorem 2. If $\frac{\nabla h(y)}{h(y)}$ satisfies the regularity conditions, then

$$R(\theta,\delta) - R(\theta,\delta_0) = E_\theta^X\left[\frac{4}{h(Y)}\,\nabla^2 h(Y)\right].$$

Furthermore, if h is superharmonic then δ dominates δ_0.

Proof. For simplicity, let

$$B = Q^{*-\frac{1}{2}} = (b_{ij}) \text{ and } \gamma(x) = -2B\,\frac{\nabla h(y)}{h(y)}.$$

Clearly γ satisfies the regularity conditions. From Theorem 1 we have

$$R(\theta,\delta) - R(\theta,\delta_0) = E_\theta[\gamma^t(x)Q^*\gamma(x) - 2\,\mathrm{tr}(VQ^*)],$$

where $V = (v_{ij})$ with $v_{ij} = \frac{\partial\gamma_i}{\partial x_j}$. Then

$$\begin{aligned} v_{ij} &= -2\,\frac{\partial}{\partial x_j}\left(\sum_{k=1}^{p} b_{ik}\,\frac{h^{(k)}(y)}{h(y)}\right) \\ &= -2\sum_{k=1}^{p} b_{ik}\sum_{\ell=1}^{p}\frac{\partial}{\partial y_\ell}\left(\frac{h^{(k)}(y)}{h(y)}\right)\frac{\partial y_\ell}{\partial x_j} \\ &= \frac{-2}{h(y)}\sum_{k=1}^{p}\sum_{\ell=1}^{p} b_{ik}h^{(k,\ell)}(y)b_{\ell j} + \frac{2}{h^2(y)}\sum_{k=1}^{p}\sum_{\ell=1}^{p} b_{ik}h^{(k)}(y)h^{(\ell)}(y)b_{\ell j}. \end{aligned}$$

So

$$V = \frac{-2}{h(y)}\,BHB + \frac{2}{h^2(h)}\,B\nabla h(y)\nabla^t h(y)B,$$

where

$$H = (h^{(i,j)}(y)).$$

Thus

$$\begin{aligned} \mathrm{tr}(VQ^*) &= \frac{-2}{h(y)}\,\mathrm{tr}(BHBQ^*) + \frac{2}{h^2(y)}\,\mathrm{tr}(B\nabla h(y)\nabla^t h(y)BQ^*) \\ &= \frac{-2}{h(y)}\,\mathrm{tr}\,H + \frac{2}{h^2(y)}\,\mathrm{tr}(\nabla h(y)\nabla^t h(y)). \end{aligned}$$

But

$$\gamma^t(x)Q^*\gamma(x) = 4\,\frac{\nabla^t h(y)\nabla h(y)}{h^2(y)},$$

and noting that $\mathrm{tr}\,H = \nabla^2 h(y)$, it follows that

$$R(\theta,\delta) - R(\theta,\delta_0) = 4E_\theta^X\left[\frac{\nabla^2 h(Y)}{h(Y)}\right].$$

If h is superharmonic, then $\nabla^2 h(y) \le 0$. Since $h(y) > 0$ it can be concluded that δ dominates δ_0.||

Corollary 1. Let C be a positive definite matrix and $r = x^t Cx$. Suppose ϕ is a positive twice differentiable real valued function and $\nabla \log \phi(r) (= \frac{2\phi'(r)}{\phi(r)} Cx)$ satisfies the regularity conditions. Let

$$\delta(x) = \delta_0(x) + \frac{4\phi'(r)}{\phi(r)} \ddagger Cx, \tag{3.2}$$

where δ_0 defined in (3.1). Then

$$R(\theta,\delta) - R(\theta,\delta_0) = E_\theta^X\{\frac{4}{\phi(r)} [4\phi''(r)X^t CQ^*CX + 2\phi'(r)tr(Q^*C)]\}. \tag{3.3}$$

In particular, when $C = Q^{*-1}$, then

$$R(\theta,\delta)-R(\theta,\delta_0) = E_\theta^X\{\frac{4}{\phi(r)} [4\phi''(r)r+2p\phi'(r)]\}. \tag{3.4}$$

Proof. Let $B = Q^{*-\frac{1}{2}}$ and $A = B^{-1}CB^{-1}$. Then

$$r = x^t Cx = x^t B(B^{-1}CB^{-1})Bx = y^t Ay,$$

where $y = Bx$. Let $h(y) = \phi(r)$. Clearly

$$\nabla h(y) = 2\phi'(r)Ay,$$

so that

$$\delta(x) = x + 2\ddagger B \frac{\nabla h(y)}{h(y)}.$$

But

$$\begin{aligned}\nabla^2 h(y) &= \sum_{i=1}^p h^{(i,i)}(y) \\ &= \phi''(r) \sum_{i=1}^p (\frac{\partial r}{\partial y_i})^2 + \phi'(r) \sum_{i=1}^p \frac{\partial^2 r}{\partial y_i^2} \\ &= 4\phi''(r)y^t A^t Ay + 2\phi'(r)tr\ A \\ &= 4\phi''(r)x^t CQ^*Cx + 2\phi'(r)tr(Q^*C).\end{aligned}$$

Hence, the result follows from Theorem 2.||

We can make any of the usual choices for ϕ which satisfy the regularity conditions and the differential inequality

$$\phi''(r)r + 2p\phi'(r) \le 0,$$

and thus obtain a uniform improvement over δ_0. In particular, if $\phi(r) = r^{-\frac{\lambda}{2}(p-2)}$ with $0 < \lambda \leq 1$ and $p \geq 3$, we have the following corollary which extends the result of Hudson (1974) (He dealt with $x_1, \ldots, x_p$ independent and $Q = I_p$).

Corollary 2. If $p \geq 3$, then the unbiased estimator δ_0 defined in (3.1) is inadmissible and is dominated by the estimator

(3.5) $$\delta(x) = \delta_0(x) - \frac{2\lambda(p-2)}{r} Q^{-1} \Sigma^{-1} x,$$

where $r = x^t Q^{*-1} x$ and $0 < \lambda \leq 1$.

Proof. The estimator δ in (3.5) corresponds to the estimator (3.2) with $C = Q^{*-1}$ and $\phi(r) = r^{-\frac{\lambda}{2}(p-2)}$. It is easy to show that

$$4\phi''(r)r + 2p\phi'(r) = -\lambda(1-\lambda)(p-2)^2 \frac{\phi(r)}{r} \leq 0.$$

Thus the only thing we need to do is show that $\frac{\phi'(r)}{\phi(r)} Q^{-1} \Sigma^{-1} x$ satisfies the regularity conditions. But this is equivalent to showing that $E\{r^{-1}\} < \infty$.

Let $z = \Sigma^{-1} x$. By transforming to polar coordinates and noting that $p \geq 3$, it is easy to show that $E(\frac{1}{z^t z}) < \infty$.

Since

$$\frac{1}{r} \leq (ch_{max} Q)(\frac{1}{z^t z}),$$

so we have $E\{r^{-1}\} < \infty$, and the conclusion follows.||

4. Multivariate Normal Distribution with Unknown Variance.

In applications, it is important to consider the situation in which the covariance matrix of X is partially unknown. Here, we will only consider the case where the covariance matrix is of the form $\sigma^2 \Sigma$, Σ known but σ^2 unknown. (This is a common situation in regression problems).

When σ^2 is unknown, assume a random variable S^2 is observable (independent of X), where S^2/σ^2 has a chi-square distribution with n degrees of freedom. A suitable estimator for σ^2 is $S^2/(n+2)$. Let $\delta_0(x) = x$. Then we have the following theorem:

Theorem 3. Let

$$\delta(x) = x + \frac{2S^2}{n+2} \ddagger Q^{*-\frac{1}{2}} \frac{\nabla h(y)}{h(y)},$$

where $Q^* = \ddagger Q \ddagger$ and $y = Q^{*-\frac{1}{2}}x$. If $\frac{\nabla h(y)}{h(y)}$ satisfies the regularity conditions, then

$$R(\theta,\delta) - R(\theta,\delta_0) = \frac{4n\sigma^4}{n+2} E_\theta^X\left[\frac{\nabla^2 h(Y)}{h(Y)}\right]. \tag{4.1}$$

Furthermore, if h is superharmonic, then δ dominates $\delta_0(x) = x$ (thus, δ is minimax).

Proof. Let

$$\gamma(x) = \frac{-2S^2}{n+2} Q^{*-\frac{1}{2}} \frac{\nabla h(y)}{h(y)}.$$

Then, as in Theorem 2, it can be shown that

$$V = \left(\frac{\partial\gamma_i}{\partial x_j}\right) = \frac{-4s^2}{n+2}\left[\frac{1}{h(y)} Q^{*-\frac{1}{2}} H Q^{*-\frac{1}{2}} - \frac{1}{h^2(y)} Q^{*-\frac{1}{2}} \nabla h(y) \nabla^t h(y) Q^{*-\frac{1}{2}}\right]$$

where $H = (h^{(i,j)}(y))$. Hence

$$\mathrm{tr}(VQ^*\sigma^2) = \frac{-4\sigma^2 S^2}{n+2}\left[\frac{\mathrm{tr}\, H}{h(y)} - \frac{\nabla^t h(y)\nabla h(y)}{h^2(y)}\right].$$

Since $\delta_0(x) = x$, $\nabla \log m_0(x) = 0$. Using the identity in Theorem 1, we have

$R(\theta,\delta)-R(\theta,\delta_0)$

$$= 4E^{S^2}_{\theta,\sigma^2} E^{X|S^2}[(\frac{S^4}{(n+2)^2} - \frac{\sigma^2 S^2}{n+2}) \frac{\nabla^t h(Y)\nabla h(Y)}{h^2(Y)} + \frac{\sigma^2 S^2}{n+2} \frac{\nabla^2 h(Y)}{h(Y)}]$$

$$= 4E^X_\theta E^{S^2|X}_{\sigma^2}[(\frac{S^4}{(n+2)^2} - \frac{\sigma^2 S^2}{n+2}) \frac{\nabla^t h(Y)\nabla h(Y)}{h^2(Y)} + \frac{\sigma^2 S^2}{n+2} \frac{\nabla^2 h(Y)}{h(Y)}].$$

Using the facts

$$E^{S^2}_{\sigma^2}(S^2) = n\sigma^2, \quad E^{S^2}_{\sigma^2}(S^4) = n(n+2)\sigma^4,$$

and noting that S^2, X are independent, we have

$$R(\theta,\delta) - R(\theta,\delta_0) = 4E^X_\theta E^{S^2|X}_{\sigma^2}[\frac{n\sigma^4}{n+2} \frac{\nabla^2 h(Y)}{h(Y)}]$$

$$= \frac{4n\sigma^4}{n+2} E^X_\theta[\frac{\nabla^2 h(Y)}{h(Y)}].$$

If, in addition, h is superharmonic, then

$$R(\theta,\delta) \leq R(\theta,\delta_0) \quad \text{for all } \theta.$$

So, δ dominates δ_0 and δ is also minimax.||

In particular, let

$$r = x^t Q^{*-1} x = y^t y$$

and

(4.2) $$h(y) = (y^t y)^{-\frac{\lambda}{2}(p-2)} = r^{-\frac{\lambda}{2}(p-2)}.$$

Then we have the following corollary:

<u>Corollary 3</u>. If $p \geq 3$, then the estimator

(4.3) $$\delta(x) = x - \frac{2\lambda(p-2)S^2}{(n+2)r} Q^{-1}\ddagger^{-1}x$$

dominates $\delta_0(x) = x$ when $0 < \lambda \leq 1$ and $r = x^t Q^{*-1} x$.

Proof. Clearly, the function h defined in (4.2) is superharmonic. The conclusion follows from Theorem 3.||

5. Improved Simultaneous Estimation of a Normal and a Gamma Parameter.

In this section, we consider an example in which two unrelated problems, one involving a one dimensional normal mean and one a gamma scale parameter, can be combined to obtain an estimator improving upon admissible estimators in each separate problem.

Let X_1 be $N(\theta_1,\sigma^2)$, where σ^2 known, and X_2 (independent of X_1) be a gamma (α,θ_2) distribution with $\alpha > 2$. (i.e. having density $g(x_2|\theta_2) = \theta_2^{\alpha} x_2^{\alpha-1} e^{-\theta_2 x_2}/\Gamma(\alpha)$, $x_2 > 0$).

It is desired to estimate $\theta = (\theta_1,\theta_2)^t$ under the sum of squared error loss function $L(\theta,\delta) = (\theta_1-\delta_1)^2 + (\theta_2-\delta_2)^2$. We know that the M.L.E., unbiased estimator $\frac{\alpha-1}{x_2}$ of θ_2 is inadmissible and is dominated by the admissible estimator $\frac{\alpha-2}{x_2}$ under the squared error loss function. So, the standard estimator of θ consider here is

$$\delta_0(x) = (x_1, \frac{\alpha-2}{x_2})^t.$$

We will show that δ_0 is dominated by the estimator $\delta = (\delta_1,\delta_2)^t$ where

$$\delta_1(x) = x_1 - \frac{2\lambda}{x_1^2/\sigma^4+x_2^2} \frac{1}{\sigma^2} x_1$$

(5.1)

$$\delta_2(x) = \frac{\alpha-2}{x_2} + \frac{2\lambda}{x_1^2/\sigma^4+x_2^2} x_2,$$

for $0 < \lambda \leq 1$.

Corollary 4. δ dominates δ_0.

Proof. Let

$$\ddagger = \begin{pmatrix} \sigma^2 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } t(x) = x_2^{\alpha-1} e^{-\frac{x_1^2}{2\sigma^2}}$$

then $X = (x_1, x_2)^t$ has density of the form as in (1.1).

Let

$$\nabla \log m_0(x) = (0, \frac{1}{x_2})^t \text{ and } r = \frac{x_1^2}{\sigma^4} + x_2^2.$$

Then

$$\delta_0(x) = \ddagger \nabla \log m_0(x) - \ddagger \nabla \log t(x)$$

and

$$\delta(x) = \delta_0(x) - \frac{1}{r} \ddagger^{-1} x.$$

Noting that $1/r \leq \frac{1}{x_2^2}$, the regularity conditions hold. So, we can use Theorem 1 (for $Q = I_2$) to get

$$R(\theta,\delta) - R(\theta,\delta_0) = -4\lambda(1-\lambda)E\{r^{-1}\} \leq 0,$$

for $0 < \lambda \leq 1$. The conclusion follows.||

Acknowledgment. The author wishes to thank Professor James Berger for his helpful suggestions and patient corrections.

References

[1] Berger, J. (1976a). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. Ann. Statist. 4, 223-226.

[2] Berger, J. (1976b). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Multivariate Anal. 6, No. 2, 1976.

[3] Berger, J. (1979). Multivariate estimation with nonsymmetric loss functions. In optimizing Methods in Statistics. (J. S. Rustagi, ed.). Academic Press, New York.

[4] Berger, J. (1980a). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. Ann. Statist. 8, 545-571.

[5] Berger, J. (1980b). A robust generalized Bayes estimator and confidence region of a multivariate normal mean. Ann. Statist. 8, 716-761.

[6] Berger, J. (1982a). Estimation in continuous exponential families: Bayesian estimation subject to risk restrictions and inadmissibility results. Statistical decision theory and related topics III, eds. S. S. Gupta and J. Berger, New York: Academic Press.

[7] Berger, J. (1982b). Selecting a minimax estimator of a multivariate normal mean. Ann. Statist. 10, 81-92.

[8] Berger, J. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. Ann. Statist. 6, 783-801.

[9] Bhattacharya, P. K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. Ann. Math. Statist. 37, 1819-1824.

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

[11] Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. Ann. Statist. 4, 11-21.

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

[13] Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. Ann. Statist. 6, 473-484.

[14] James, W. and Stein, C. (1960). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Stat. Prob. 1, 361-379. University of California Press.

[15] Lin, Pi-Erh, and Tsai, Hui-Liang (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. Ann. Statist. 1, 142-145.

[16] Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. 1, 197-206. University of California Press.

[17] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135-1151.

[18] Strawderman, W. E. (1978). Minimax adaptive generalized ridge regression estimators. J. Amer. Statist. Assoc. 73, 623-627.