ESTIMATION IN CONTINUOUS EXPONENTIAL FAMILIES:

BAYESIAN ESTIMATION SUBJECT TO

RISK RESTRICTIONS AND INADMISSIBILITY RESULTS

by

James Berger*
Purdue University

Mimeograph Series #81-27

Department of Statistics
Division of Mathematical Sciences

July 1981

*Research supported by the John Simon Guggenheim Memorial Foundation and by the National Science Foundation under Grants MCS78-02300A3 and MCS81-01670.

1. Introduction

Suppose we observe $\underset{\sim}{X} = (X_1,\ldots,X_p)$, where the X_i are independent and have positive densities

$$f_i(x_i|\theta_i) = \beta_i(\theta_i)t_i(x_i)e^{\theta_i x_i}$$

with respect to Lebesgue measure on $\mathcal{X}_i \subset R^1$. It is desired to estimate $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_p)$ where the loss in estimating $\underset{\sim}{\theta}$ by an estimate

$$\underset{\sim}{\delta}(\underset{\sim}{x}) = (\delta_1(\underset{\sim}{x}),\ldots,\delta_p(\underset{\sim}{x}))$$

is sum of squares error loss

$$L(\underset{\sim}{\theta},\underset{\sim}{\delta}) = \sum_{i=1}^{p} (\theta_i-\delta_i(\underset{\sim}{x}))^2.$$

The parameter space will be taken to be the natural parameter space, i.e.

$$\Theta = \{\underset{\sim}{\theta}=(\theta_1,\ldots,\theta_p): \int_{\mathcal{X}_i} t_i(x_i)d^{\theta_i x_i}dx_i<\infty \text{ for } i=1,\ldots,p\} .$$

An important feature of an estimator $\underset{\sim}{\delta}$ is its risk function (or expected loss)

$$R(\underset{\sim}{\theta},\underset{\sim}{\delta}) = E_{\underset{\sim}{\theta}}L(\underset{\sim}{\theta},\underset{\sim}{\delta}(\underset{\sim}{X})).$$

Two problems of major importance in this simultaneous estimation framework are as follows. Problem I involves classification of inadmissible and admissible estimators, and development of estimators offering significant improvement upon standard estimators that happen to be inadmissible. The most studied example of this is estimation of a multivariate normal mean $\underset{\sim}{\theta}$ in which the usual estimator $\underset{\sim}{\delta}^0(\underset{\sim}{x}) = \underset{\sim}{x}$ is inadmissible if $p \geq 3$ (Stein [18]).

Problem II is the problem of robust Bayesian estimation. In Bayesian estimation a prior distribution (possibly improper) $\pi(d\underset{\sim}{\theta})$ on Θ is determined, and nominally one would want to use the Bayes estimator $\underset{\sim}{\delta}^{\pi}$, defined as that estimator minimizing the Bayes risk

$$r(\pi,\underset{\sim}{\delta}) = E^{\pi}R(\underset{\sim}{\theta},\underset{\sim}{\delta}) = \int_{\Theta} R(\underset{\sim}{\theta},\underset{\sim}{\delta})\pi(d\underset{\sim}{\theta})$$

(or more generally minimizing the posterior expected loss). The determination of π is often very inexact, however, and hence it is important to consider the robustness (with respect to the specification of π)of the estimator selected. A general discussion of Bayesian robustness is given in Berger [3], in which it is argued that good measures of robustness can be obtained from $R(\underset{\sim}{\theta},\underset{\sim}{\delta})$. If, for example, a minimax estimator $\underset{\sim}{\delta}^{0}$ is the classical estimator for a problem, then if one restricts consideration to estimators satisfying

$$R(\underset{\sim}{\theta},\underset{\sim}{\delta}) \leq R(\underset{\sim}{\theta},\underset{\sim}{\delta}^{0}) + C, \tag{1.1}$$

it will be ensured that $r(\pi,\underset{\sim}{\delta}) \leq r(\pi,\underset{\sim}{\delta}^{0}) + C$ no matter how badly the prior distribution is misspecified. One can thus formally state as a Bayesian robustness problem

Problem II*: Select the estimator $\underset{\sim}{\delta}$ which minimizes $r(\pi,\underset{\sim}{\delta})$ subject to (1.1).

At first sight it may seem that the robustness requirement (1.1) is excessively harsh, but it will be seen that (1.1) can often be attained with surprisingly little sacrifice in Bayes risk (compared to the nominally optimal but often nonrobust Bayes estimator δ^{π}). Constraints other than (1.1) may sometimes be more natural. For example, in some problems it may be more reasonable to require that $R(\underset{\sim}{\theta},\underset{\sim}{\delta}) \leq R(\underset{\sim}{\theta},\underset{\sim}{\delta}^{0})(1+C)$.

Surprisingly, Problem II* is also frequently crucial in successful resolution of Problem I. This is because, if $\underset{\sim}{\delta}^{0}$ is a "standard" estimator which happens to be inadmissible, then the class of estimators better than $\underset{\sim}{\delta}^{0}$ is precisely the class of estimators satisfying (1.1) with C = 0. In selecting among these improved estimators it seems inescapable that prior information must be employed (see Berger [2] and [5]). Hence a reasonable solution would be to determine a prior distribution π and solve Problem II* when C = 0 in (1.1).

Problem II* has been considered for various situations in Hodges and Lehmann [13], Efron and Morris [11], Shapiro [15] and [16], Bickel [7] and [8], and Berger [4]. Exact mathematical solution is unfortunately very messy. For example, if $\underset{\sim}{X}$ has a normal distribution with identity covariance matrix, so that $\underset{\sim}{\delta}^{0}(\underset{\sim}{X}) = \underset{\sim}{X}$ is the usual estimator, the solution to Problem II* can be typically shown to be a (generalized) Bayes estimator with respect to a prior measure concentrated on a countably infinite number of shells. It is an extremely difficult numerical problem to determine the appropriate shells and their masses, and the resulting estimator is an abominable mess. For this reason we will consider a slightly modified version of Problem II*, one which is tractable mathematically and yields reasonably simple estimators.

The starting point for the investigation will be Stein's unbiased estimator of risk, discussed for this setting in Hudson [14] and Berger [1], which leads to the representation

$$R(\underset{\sim}{\theta},\underset{\sim}{\delta}) - R(\underset{\sim}{\theta},\underset{\sim}{\delta}^{0}) = E_{\underset{\sim}{\theta}}[\mathcal{D}\,\underset{\sim}{\delta}(\underset{\sim}{X})], \tag{1.2}$$

where $\mathcal{D}$ is usually a nonlinear differential operator. The condition (1.1) will clearly be satisfied if

$$\mathcal{D}\,\underset{\sim}{\delta}(\underset{\sim}{x}) \leq C, \tag{1.3}$$

and so we can formulate

Problem II**: Select the estimator which minimizes $r(\pi,\underset{\sim}{\delta})$ subject to (1.3).

The solutions to Problem II** seem to be very close to the solutions to Problem II*, and their comparative simplicity makes them considerably more attractive from a practical viewpoint.

Analysis in generality of any of the problems mentioned here is extremely difficult, and hence we will consider only various special cases. Section 2 will develop the needed form of the representation (1.2). Section 3 will present some results concerning Problem I, namely classification of inadmissible estimators. Section 4 will give an explicit solution to Problem II** when $\underset{\sim}{X}$ has a spherically symmetric normal distribution. The most important example of the theory in Section 4, namely the analysis when π is a conjugate prior, is presented in Section 5. The resulting estimators will be seen to have the rather startling property (for $p > 1$) of having nearly optimal Bayes risk even when C is very small (i.e., even when the estimators are constrained to have risks which never exceed the risk of the minimax estimator $\underset{\sim}{\delta}^0(\underset{\sim}{x}) = \underset{\sim}{x}$ by more than C). Section 5 can be understood (for the most part) without having read the previous sections.

2. The Unbiased Estimator of Risk

We begin by stating some conditions on the densities $f_i(x_i|\theta_i)$ and the estimators that will be considered. These are chosen for convenience of application, and can undoubtedly be generalized. The following notations will be used throughout the paper:

$$(2.1)\quad \mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_p, \quad t(\underset{\sim}{x}) = \prod_{i=1}^{p} t_i(x_i), \quad \beta(\underset{\sim}{\theta}) = \prod_{i=1}^{p} \beta_i(\theta_i),$$

$$f(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{i=1}^{p} f_i(x_i|\theta_i) = \beta(\underset{\sim}{\theta})t(\underset{\sim}{x})e^{\underset{\sim}{\theta}\underset{\sim}{x}^t},$$

$$h'(y) = \frac{d}{dy}h(y), \quad h^{(i)}(\underset{\sim}{x}) = \frac{\partial}{\partial x_i}h(\underset{\sim}{x}), \quad h^{(i,j)}(\underset{\sim}{x}) = \frac{\partial^2}{\partial x_i \partial x_j}h(\underset{\sim}{x}),$$

$$\nabla h(\underset{\sim}{x}) = (h^{(1)}(\underset{\sim}{x}),\ldots,h^{(p)}(\underset{\sim}{x})), \quad \nabla^2 h(\underset{\sim}{x}) = \sum_{i=1}^{p} h^{(i,i)}(\underset{\sim}{x}).$$

$$\underset{\sim}{\gamma}(\underset{\sim}{x}) = \underset{\sim}{\delta}(\underset{\sim}{x}) - \underset{\sim}{\delta}^{o}(\underset{\sim}{x}).$$

The estimator $\underset{\sim}{\delta}^{o}$ is to be thought of as the "standard" estimator or estimator under investigation, and $\underset{\sim}{\delta}$ as a competing estimator.

Condition 1:

(i) The $\mathcal{X}_i$ are (possibly infinite) intervals (a_i,b_i);

(ii) the functions t_i are differentiable and $E_{\underset{\sim}{\theta}}|\nabla \log t(\underset{\sim}{x})|^2 < \infty$;

(iii) $E_\theta|\underset{\sim}{\delta}^{o}(\underset{\sim}{X})|^2 < \infty$.

Condition 2: For $i=1,\ldots,p$ and all $\underset{\sim}{\theta} \in \Theta$, γ_i satisfies

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

$$\lim_{x_i\to a_i} \{\gamma_i(\underset{\sim}{x})t_i(x_i)e^{\theta_i x_i}\} = \lim_{x_i\to b_i} \{\gamma_i(\underset{\sim}{x})t_i(x_i)e^{\theta_i x_i}\} = 0;$$

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

Condition 3: For some positive differentiable functions m_o and g,

(i) $\underset{\sim}{\delta}^{o}(\underset{\sim}{x}) = \nabla \log m_o(\underset{\sim}{x}) - \nabla \log t(\underset{\sim}{x})$;

(ii) $\underset{\sim}{\gamma}(\underset{\sim}{x}) = 2\nabla \log g(\underset{\sim}{x})$.

Comment: If $\pi(d\underset{\sim}{\theta})$ is a (generalized) prior distribution on $\overline{\Theta}$ (the closure of Θ) and the marginal density of $\underset{\sim}{X}$, given by

$$m(\underset{\sim}{x}) = t(\underset{\sim}{x}) \int_{\Theta} e^{\underset{\sim}{\theta}\underset{\sim}{x}^t} \beta(\underset{\sim}{\theta})\pi(d\underset{\sim}{\theta}), \tag{2.2}$$

is finite, then it is well known that the Bayes estimator of $\underset{\sim}{\theta}$ is

(2.3) $$\underset{\sim}{\delta}^{\pi}(\underset{\sim}{x}) = \nabla \log m(\underset{\sim}{x}) - \nabla \log t(\underset{\sim}{x}).$$

Furthermore, Berger and Srinivasan [6] show that any admissible estimator must be of this form. Hence the restrictions in Condition 3 are natural.

Theorem 1. If Conditions 1, 2, and 3(i) hold, then

(2.4) $$R(\underset{\sim}{\theta},\underset{\sim}{\delta}) - R(\underset{\sim}{\theta},\underset{\sim}{\delta}^{o})$$
$$= E_{\theta}[2 \sum_{i=1}^{p} \gamma_i^{(i)}(\underset{\sim}{X}) + 2 \sum_{i=1}^{p} \gamma_i(\underset{\sim}{X}) \frac{\partial}{\partial X_i} \log m_o(\underset{\sim}{X}) + \sum_{i=1}^{p} \gamma_i^2(\underset{\sim}{X})].$$

If, furthermore, Condition 3(ii) holds, then

(2.5) $$R(\underset{\sim}{\theta},\underset{\sim}{\delta}) - R(\underset{\sim}{\theta},\underset{\sim}{\delta}^{o}) = E_{\theta}[\frac{4}{g(\underset{\sim}{X})} \mathcal{D}\, g(\underset{\sim}{X})],$$

where

(2.6) $$\mathcal{D}\, g(\underset{\sim}{X}) = \nabla^2 g(\underset{\sim}{X}) + \nabla g(\underset{\sim}{X}) \cdot \nabla \log m_o(\underset{\sim}{X}).$$

Proof. Although versions of this theorem are given in Hudson [14] and Berger [1], we sketch the proof under this set of assumptions. Clearly

(2.7) $$R(\underset{\sim}{\theta},\underset{\sim}{\delta}) - R(\underset{\sim}{\theta},\underset{\sim}{\delta}^{o}) = \sum_{i=1}^{p} E_{\theta}[2\gamma_i(\underset{\sim}{X})(\delta_i^o(\underset{\sim}{X}) - \theta_i) + \gamma_i^2(\underset{\sim}{X})].$$

Now

(2.8) $$E_{\underset{\sim}{\theta}}[\gamma_i(\underset{\sim}{X})\theta_i]$$
$$= \int_{\prod_{j\neq i} \mathcal{X}_j} \beta(\underset{\sim}{\theta}) \prod_{j\neq i} [t_j(x_j) e^{\theta_j x_j}] \int_{a_i}^{b_i} \gamma_i(\underset{\sim}{x}) t_i(x_i) \theta_i e^{\theta_i x_i} dx_i \prod_{j\neq i} dx_j.$$

Observe, using Conditions 1 and 2 and the Cauchy-Schwartz inequality, that

$$(2.9) \qquad E_{\underset{\sim}{\theta}}\left[\frac{1}{t_i(X_i)} \left| \frac{\partial}{\partial X_i} \{\gamma_i(\underset{\sim}{X}) t_i(X_i)\} \right| \right]$$

$$= E_{\underset{\sim}{\theta}} |\gamma_i^{(i)}(\underset{\sim}{X}) + \gamma_i(\underset{\sim}{X}) \frac{d}{dX_i} \log t_i(X_i)|$$

$$\leq E_{\underset{\sim}{\theta}} |\gamma_i^{(i)}(\underset{\sim}{X})| + \{[E_{\underset{\sim}{\theta}} \gamma_i^2(\underset{\sim}{X})][E_{\underset{\sim}{\theta}}(\frac{d}{dX_i} \log t_i(X_i))^2]\}^{1/2}$$

$$< \infty .$$

It further follows from Condition 2 that, with probability one, $\gamma_i(\underset{\sim}{x})t_i(x_i)$ is absolutely continuous on $[c,d]$ for $a_i < c < d < b_i$, which together with (2.9) establishes the validity (with probability one) of the integration by parts

$$\int_c^d \gamma_i(\underset{\sim}{x})t_i(x_i)\theta_i e^{\theta_i x_i} dx_i = \gamma_i(\underset{\sim}{x})t_i(x_i)e^{\theta_i x_i}\Big|_{x_i=c}^{x_i=d}$$

$$- \int_c^d [\gamma_i^{(i)}(\underset{\sim}{x})t_i(x_i) + \gamma_i(\underset{\sim}{x})t_i'(x_i)]e^{\theta_i x_i} dx_i .$$

Letting $c \to a_i$, $d \to b_i$, applying Condition 2(i) and inserting the result in (2.8) gives the equality

$$E_{\underset{\sim}{\theta}}[\gamma_i(\underset{\sim}{X})\theta_i] = -E_{\underset{\sim}{\theta}}[\gamma_i^{(i)}(\underset{\sim}{X})] - E_{\underset{\sim}{\theta}}[\gamma_i(\underset{\sim}{X}) \frac{d}{dX_i} \log t_i(X_i)].$$

Applying this in (2.7) and using Condition 3(i) yields (2.4). Equation (2.5) follows by a direct calculation. ||

3. Inadmissibility

Proofs of inadmissibility of various estimators $\underset{\sim}{\delta}^0$ using theorems analogous to Theorem 1 have been carried out in Hudson [14] and Berger [1] in a rather haphazard manner. A systematic approach to the problem would be to attempt to solve

$$(3.1) \qquad \mathcal{D}\, g(\underset{\sim}{x}) = 0$$

(see (2.6)) for g, and observe that if a solution $g > 0$ is found then

$$\mathcal{D}[g(\underset{\sim}{x})]^{\alpha} = \alpha[g(\underset{\sim}{x})]^{\alpha-1}\mathcal{D}g(\underset{\sim}{x}) + \alpha(\alpha-1)[g(\underset{\sim}{x})]^{\alpha-2}\sum_{i=1}^{p}[g^{(i)}(\underset{\sim}{x})]^2$$

$$= \alpha(\alpha-1)[g(\underset{\sim}{x})]^{\alpha-2}\sum_{i=1}^{p}[g^{(i)}(\underset{\sim}{x})]^2 < 0$$

for $0 < \alpha < 1$. From (2.5) it would follow that the estimator

$$\underset{\sim}{\delta}(\underset{\sim}{x}) = \underset{\sim}{\delta}^{o}(\underset{\sim}{x}) + \underset{\sim}{\gamma}(\underset{\sim}{x}) = \underset{\sim}{\delta}^{o}(\underset{\sim}{x}) + 2\alpha\nabla \log g(\underset{\sim}{x})$$

has smaller risk than $\underset{\sim}{\delta}^{o}$ for all $\underset{\sim}{\theta}$.

Unfortunately, closed form solution of (3.1) is possible only in certain special cases, such as when $p = 1$, when $\underset{\sim}{X}$ has a spherically symmetric normal distribution and $\underset{\sim}{\delta}^{o}$ is a spherically symmetric estimator, and when the X_i are from Gamma distributions with equal degrees of freedom and

$$\nabla \log m_o(\underset{\sim}{x}) = (x_1^{-1},\ldots,x_p^{-1}) + \nabla \log \phi(\sum_{i=1}^{p} x_i^2).$$

We will consider the situation when $p = 1$ as an example.

Theorem 2. Let $p = 1$ and suppose that Condition 1 holds and that δ^{o} is as in Condition 3(i). Define

$$\psi_1(x) = \int_a^x \frac{1}{m_o(y)}\, dy \text{ and } \psi_2(x) = \int_x^b \frac{1}{m_o(y)}\, dy,$$

and suppose for all $x \in \mathcal{X} = (a,b)$ that either $\psi_1(x) < \infty$ or $\psi_2(x) < \infty$. Letting ψ_i (i=1 or 2) be the finite function chosen, assume further that

(i) $E_\theta \left| \frac{d}{dX} \log \psi_i(X)\right|^2 < \infty$, and

(ii) $\lim_{x\to a}\{t(x)e^{\theta x} \frac{d}{dx} \log \psi_i(x)\} = \lim_{x\to b}\{t(x)e^{\theta x} \frac{d}{dx} \log \psi_i(x)\} = 0.$

Then δ^{o} is inadmissible and a better estimator is given by

(3.2) $$\delta(x) = \delta^{o}(x) + \gamma(x),$$

where

$$\gamma(x) = 2\alpha \frac{d}{dx} \log \psi_i(x)$$

and $0 < \alpha < 1$.

Proof. We begin by verifying Condition 2 of Section 2. Only the condition $E_\theta|\gamma'(X)| < \infty$ is not immediately obvious. But

$$\gamma'(x) = -\frac{1}{2\alpha}\gamma^2(x) - \gamma(x)\frac{d}{dx}\log m_o(x),$$

so

$$E_\theta|\gamma'(X)| \le \frac{1}{2\alpha} E_\theta[\gamma^2(x)] + \{[E_\theta\gamma^2(X)][E_\theta(\frac{d}{dx}\log m_o(X))^2]\}^{1/2}.$$

All terms are finite by assumption (the finiteness of $E_\theta(\frac{d}{dX}\log m_o(X))^2$ following from Condition 1 (ii) and (iii)).

Theorem 1 thus applies, and it is easy to check that $\mathcal{D}\psi_i(x) = 0$. The discussion preceeding this theorem completes the proof. ||

Comment: Because m_o is continuous and positive, ψ_i will be finite for all $a < x < b$ if it is finite for any x. Hence only the behavior of $m_o(x)$ at the boundaries of $\mathcal{X}$ is relevant to the admissibility problem.

Example. Suppose X has a Gamma distribution, i.e. has density on $\mathcal{X} = (0,\infty)$

$$f(x|\theta) = (-\theta)^\alpha x^{\alpha-1} e^{\theta x}/\Gamma(\alpha).$$

(The natural parameter space here is $\Theta = (-\infty,0)$. It is easy to transpose the results below to the more common parameterization in which $\Theta = (0,\infty)$, however.) Clearly $t(x) = x^{\alpha-1}$, so that $E_\theta|\frac{d}{dX}\log t(X)|^2 < \infty$ if and only if $\alpha = 1$ or $\alpha > 2$. Henceforth assume that $\alpha > 2$.

Suppose, now, that

$$(3.3)\qquad m_0(y) = \begin{cases} k_1 y^r(1+o(1)) & \text{as} \quad y \to 0 \\ k_2 y^s(1+o(1)) & \text{as} \quad y \to \infty \ . \end{cases}$$

If $r < 1$ then ψ_1 is finite, while if $s > 1$ then ψ_2 is finite. In either case it is easy to verify the remaining conditions needed to apply Theorem 2, and hence (3.2) gives a better estimator.

Observe that the possible inadmissibility of δ^0 due to the behavior of $m_0(y)$ at ∞ is of little practical concern since for reasonable (generalized) priors (see 2.2) $m(y)$ will not be blowing up at ∞. Inadmissibility due to the behavior of $m_0(y)$ near zero is of concern, however, as can be seen by considering the prior distribution

$$\pi(d\theta) = \frac{2}{\pi(1+\theta^2)}\, d\theta \ .$$

A simple calculation gives

$$m(y) = \frac{2y^{\alpha-1}}{\pi\Gamma(\alpha)} \int_{-\infty}^{0} (-\theta)^\alpha e^{y\theta} \frac{d\theta}{(1+\theta^2)}$$

$$= \frac{2}{\pi\Gamma(\alpha)} \int_0^\infty \frac{\eta^\alpha e^{-\eta}}{y^2+\eta^2}\, d\eta \ .$$

Clearly

$$\lim_{y\to 0} m(y) = \frac{2\Gamma(\alpha-1)}{\pi\Gamma(\alpha)} = \frac{2}{(\alpha-1)\pi} \ ,$$

and

$$m(y) = \frac{2\alpha}{\pi y^2}(1+o(1)) \quad \text{as } y \to \infty \ .$$

Hence (3.3) is satisfied with $r = 0 < 1$, so the Bayes estimator with respect to π, given by $\delta^\pi(x) = \nabla \log m(x) - \nabla \log t(x)$, is inadmissible. (Although π is proper, it can be checked that the Bayes risk for the problem is infinite. Hence δ^{π} is being defined as the estimator minimizing the posterior expected loss.)

It can be shown that if π has α moments, then $m(y) = O(y^{\alpha-1})$ as $y \to 0$, which (since $\alpha > 2$) means that r will be greater than 1 and the inadmissibility theorem will not apply.

Comment: Once δ^{0} has been determined to be inadmissible, the problem of selecting a good improvement still remains. As mentioned in the introduction, one possible method of tackling this problem is to solve Problem II** with C = 0. A second method which might have some potential is to exploit the relationship between (2.6) and diffusion processes (first observed by Brown [9]). The operator $\mathcal{D}$ happens to be the infinitesimal generator of the diffusion process on $\mathcal{X}$ which has local mean $\mu(\underset{\sim}{x}) = \nabla \log m_0(\underset{\sim}{x})$ and local covariance matrix 2I. It will typically be the case that the diffusion is transient if and only if $\mathcal{D}\, g(\underset{\sim}{x}) = 0$ has a suitable positive solution, i.e. if and only if the estimator $\underset{\sim}{\delta}^{0}$ is inadmissible. Furthermore, if X_t denotes a (random) sample path of the (transient) diffusion and $E_{\underset{\sim}{x}}$ stands for expectation when the process starts at $\underset{\sim}{x}$ at time t = 0, then for appropriate positive functions h the function

$$g_h(\underset{\sim}{x}) = E_{\underset{\sim}{x}} \int_0^{\infty} h(X_t)dt$$

will be finite and satisfy $\mathcal{D}\, g_h(\underset{\sim}{x}) < 0$. Thus a large class of improved estimators

$$\underset{\sim}{\delta}(\underset{\sim}{x}) = \underset{\sim}{\delta}^{0}(\underset{\sim}{x}) + 2\nabla \log g_h(\underset{\sim}{x})$$

could be produced, and perhaps h could be chosen to accommodate the available prior information. Formidable difficulties are unfortunately also encountered in this approach to the problem.

4. Ristricted Risk Bayes Rules

Using (2.4), (2.5) and (2.6), and providing Conditions 1 and 3(i) hold, we can formally state Problem II** as that of minimizing $r(\pi,\underset{\sim}{\delta})$ among all estimators satisfying Conditions 2 and 3(ii) and also satisfying

(4.1) $$\mathcal{D}_C g(\underset{\sim}{x}) \equiv \nabla^2 g(\underset{\sim}{x}) + \nabla g(\underset{\sim}{x})\cdot\nabla \log m_0(\underset{\sim}{x}) - \frac{C}{4}\, g(\underset{\sim}{x}) \leq 0$$

or, equivalently,

(4.2) $$\mathcal{D}^*_C \underset{\sim}{\gamma}(\underset{\sim}{x}) \equiv 2\sum_{i=1}^{p} \gamma_i^{(i)}(\underset{\sim}{x}) + 2\sum_{i=1}^{p} \gamma_i(\underset{\sim}{x}) \frac{\partial}{\partial x_i} \log m_0(\underset{\sim}{x}) + \sum_{i=1}^{p} \gamma_i^2(\underset{\sim}{x}) - C \leq 0.$$

This is basically a calculus of variations minimization problem with side constraints, and the answer will typically be that the solution, $\underset{\sim}{\delta}^C$, must be a smooth blending of the unconstrained minimizing estimator $\underset{\sim}{\delta}^\pi$ and estimators arising from solutions to $\mathcal{D}_C\, g(\underset{\sim}{x}) = 0$ (or $\mathcal{D}^*_C \underset{\sim}{\gamma}(\underset{\sim}{x}) = 0$).

The major problem in determining $\underset{\sim}{\delta}^C$ is that of solving the elliptic partial differential equation $\mathcal{D}_C g(\underset{\sim}{x}) = 0$. Indeed, as discussed at the beginning of Section 3, this can only be solved in closed form for certain special cases. In this section we will analyze the spherically symmetric normal situation.

If $\underset{\sim}{X}$ has a p-variate normal distribution with identity covariance matrix, its distribution is as in (2.1) with

$$t(\underset{\sim}{x}) = e^{-|\underset{\sim}{x}|^2/2} \text{ and } \beta(\underset{\sim}{\theta}) = (2\pi)^{-p/2} e^{-|\underset{\sim}{\theta}|^2/2}\ .$$

Suppose now that the prior distribution $\pi(d\underset{\sim}{\theta})$ is symmetric about a point $\underset{\sim}{\mu} = (\mu_1,\ldots,\mu_p)$, so that the marginal density of $\underset{\sim}{X}$ will be of the form

(4.3) $$m(\underset{\sim}{x}) = h(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2).$$

The Bayes estimator for this problem can be written (see (2.3))

(4.4) $$\underset{\sim}{\delta}^{\pi}(\underset{\sim}{x}) = \underset{\sim}{x} + \frac{2h'(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)}{h(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)}(\underset{\sim}{x}-\underset{\sim}{\mu}).$$

In trying to solve Problem II**, it is natural to restrict attention to estimators which are spherically symmetric about $\underset{\sim}{\mu}$, i.e. to estimators of the form

(4.5) $$\delta(\underset{\sim}{x}) = \underset{\sim}{x} - \rho(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)(\underset{\sim}{x}-\underset{\sim}{\mu}).$$

To put this in the general framework of section 2, we can define

(4.6) $$\underset{\sim}{\delta}^{0}(\underset{\sim}{x}) = \underset{\sim}{x}, \quad g(\underset{\sim}{x}) = \phi(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2),$$

$$\rho(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2) = -\frac{4\phi'(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)}{\phi(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)}, \quad \underset{\sim}{\gamma}(\underset{\sim}{x}) = -\rho(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)(\underset{\sim}{x}-\underset{\sim}{\mu}),$$

so that an estimator of the form (4.5) can be written

(4.7) $$\underset{\sim}{\delta}(\underset{\sim}{x}) = \underset{\sim}{\delta}^{0}(\underset{\sim}{x}) + \underset{\sim}{\gamma}(\underset{\sim}{x}) = \underset{\sim}{\delta}^{0}(\underset{\sim}{x}) + 2\nabla \log g(\underset{\sim}{x}),$$

which is the form assumed in (2.1) and Condition 3(ii). For convenience, we will denote the corresponding quantities for $\underset{\sim}{\delta}^{\pi}$ by $g^{\pi}(\underset{\sim}{x})$, $\rho^{\pi}(\underset{\sim}{x})$, and $\underset{\sim}{\gamma}^{\pi}(\underset{\sim}{x})$, and note that

(4.8) $$\rho^{\pi}(\underset{\sim}{x}) = -\frac{2h'(|\underset{\sim}{x}-\underset{\sim}{\mu}|)^2}{h(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)}.$$

A simple calculation using (4.6) shows that $\mathcal{D}_c g(\underset{\sim}{x})$ in (4.1) can be written (letting $r = |\underset{\sim}{x}-\underset{\sim}{\mu}|^2$)

(4.9) $$\bar{\mathcal{D}}_c \phi(r) \equiv \mathcal{D}_c g(\underset{\sim}{x}) = 2p\,\phi'(r) + 4r\,\phi''(r) - \frac{c}{4}\,\phi(r)$$

$$= \tfrac{1}{4}\phi(r)[-2p\,\rho(r)+r\rho^2(r)-4r\,\rho'(r)-c].$$

The following lemmas present the solutions to the differential equation

(4.10) $$\bar{\mathcal{D}}_c\ \phi(r) = 0.$$

Lemma 1.

(i) If $C = 0$, positive solutions to (4.10) exist for all $r > 0$ only if $p > 2$, and are given (up to a multiplicative constant) by

$$(4.11)\qquad \phi_{C,\lambda}(r) = \begin{cases} \lambda + r^{(2-p)/2} & \text{for} \quad \lambda \geq 0 \\ 1 & \text{for} \quad \lambda = \infty \end{cases} .$$

(We will use λ to index the solutions.)

(ii) If $C > 0$, the positive solutions to (4.10) are given (up to a multiplicative constant) by

$$(4.12)\qquad \phi_{C,\lambda}(r) = \begin{cases} r^{(2-p)/4}[\lambda I_\nu(\tfrac{1}{2}\sqrt{Cr}) + K_\nu(\tfrac{1}{2}\sqrt{Cr})] & \text{for} \quad \lambda \geq 0 \\ r^{(2-p)/4} I_\nu(\tfrac{1}{2}\sqrt{Cr}) & \text{for} \quad \lambda = \infty \end{cases} ,$$

where $\nu = |p-2|/2$ and I_ν and K_ν are the modified Bessel functions determined by

$$I_\nu(r) = e^{-i\pi\nu/2} J_\nu(re^{i\pi/2}),$$

$$K_\nu(r) = \frac{\pi}{2} e^{i\pi\nu/2}[e^{i\pi(\nu+1)/2} I_\nu(r) - Y_\nu(re^{i\pi/2})],$$

where J_ν and Y_ν are the Bessel functions of the first and second kind respectively and of order ν.

Proof. When $C = 0$, (4.10) can be solved explicitely, yielding (4.11) as solutions. For $C > 0$, making the transformation $w(r) = r^{p/4}\phi(r)$ in (4.10) results in the equivalent differential equation

$$w''(r) + [\frac{-C}{16r} - \frac{p}{4}(\frac{p}{4} - 1)\frac{1}{r^2}]w(r) = 0.$$

The positive solutions to this equation are known to be of the form $w_\lambda(r) = \sqrt{r}[\lambda I_\nu(\tfrac{1}{2}\sqrt{Cr}) + K_\nu(\tfrac{1}{2}\sqrt{Cr})]$ for $\lambda \geq 0$ and $w_\infty(r) = \sqrt{r}\, I_\nu(\tfrac{1}{2}\sqrt{Cr})$.

Transforming back gives the desired result. $||$

<u>Lemma 2.</u> The functions ρ (as defined in (4.6)) corresponding to the $\phi_{C,\lambda}$ are

(i) when $C = 0$ and $p \geq 2$, given by

$$\rho_{C,\lambda}(r) = \begin{cases} \dfrac{2(p-2)}{\lambda r^{p/2}+r} & \text{for} \quad \lambda \geq 0 \\ 0 & \text{for} \quad \lambda = \infty \, ; \end{cases} \tag{4.13}$$

(ii) when $C > 0$ and $p > 1$, given by

$$\rho_{C,\lambda}(r) = \begin{cases} -\dfrac{\sqrt{C}}{\sqrt{r}} \; \dfrac{[\lambda I_{\nu+1}(\frac{1}{2}\sqrt{Cr})-K_{\nu+1}(\frac{1}{2}\sqrt{Cr})]}{[\lambda I_{\nu}(\frac{1}{2}\sqrt{Cr})+K_{\nu}(\frac{1}{2}\sqrt{Cr})]} & \text{for} \quad \lambda \geq 0 \\ -\dfrac{\sqrt{C}}{\sqrt{r}} \; \dfrac{I_{\nu+1}(\frac{1}{2}\sqrt{Cr})}{I_{\nu}(\frac{1}{2}\sqrt{Cr})} & \text{for} \quad \lambda = \infty \, ; \end{cases} \tag{4.14}$$

(iii) when $C > 0$ and $p = 1$, given by the expressions in (4.14) minus $2/r$.

<u>Proof.</u> The results follow from straightforward calculation and the fact that $\lambda I_\nu'(y) + K_\nu'(y) = \lambda I_{\nu+1}(y) - K_{\nu+1}(y) + \frac{\nu}{y}[\lambda I_\nu(y)+K_\nu(y)]$. ||

Some knowledge of the behavior of the functions $\rho_{C,\lambda}$ will be needed and is given in the following lemma.

<u>Lemma 3.</u>

(i) $\rho_{C,\lambda}(r)$ is decreasing and continuous in λ and hence

$$\rho_{C,\lambda}(r) \leq \rho_{C,\lambda}(r) \leq \rho_{C,0}(r).$$

Furthermore

$$\rho_{C,0}(r) = \begin{cases} 0 & \text{if} \quad C = 0 \text{ and } p \leq 2 \\ 2(p-2)/r & \text{if} \quad C = 0 \text{ and } p > 2 \\ \sqrt{C} \, / \, \sqrt{r} & \text{if} \quad C > 0 \text{ and } p = 1 \\ \dfrac{\sqrt{C}}{\sqrt{r}} \; \dfrac{K_{\nu+1}(\frac{1}{2}\sqrt{Cr})}{K_{\nu}(\frac{1}{2}\sqrt{Cr})} & \text{if} \quad C > 0 \text{ and } p > 1 \quad , \end{cases}$$

$$\rho_{C,\infty}(r) = \begin{cases} 0 & \text{if} \quad C = 0 \\ -\dfrac{\sqrt{C}}{\sqrt{r}} \coth(\tfrac{1}{2}\sqrt{Cr}) & \text{if} \quad C > 0,\ p = 1 \\ -\dfrac{\sqrt{C}}{\sqrt{r}} \dfrac{I_{\nu+1}(\tfrac{1}{2}\sqrt{Cr})}{I_{\nu}(\tfrac{1}{2}\sqrt{Cr})} & \text{if} \quad C > 0,\ p > 1\ . \end{cases}$$

(ii) As a function of r, $\rho_{C,\lambda}(r)$ has bounded derivatives on compact sets in $(0,\infty)$.

(iii) As $r \to 0$,

(a) when $p = 1$

$$\rho_{C,\lambda}(r) = \begin{cases} (1 - \dfrac{4\lambda}{\pi})\dfrac{\sqrt{C}}{\sqrt{r}}(1+o(1)) & \text{if} \quad \lambda \neq \dfrac{\pi}{4} \text{ and } \lambda < \infty \\ -\dfrac{C}{6}(1+o(1)) & \text{if} \quad \lambda = \dfrac{\pi}{4} \\ -\dfrac{2}{r}(1+o(1)) & \text{if} \quad \lambda = \infty \quad ; \end{cases}$$

(b) when $p = 2$

$$\rho_{C,\lambda}(r) = \begin{cases} \dfrac{-4}{r \log r}(1+o(1)) & \text{if} \quad \lambda < \infty \\ -\dfrac{C}{4}(1+o(1)) & \text{if} \quad \lambda = \infty \quad ; \end{cases}$$

(c) when $p \geq 3$

$$\rho_{C,\lambda}(r) = \begin{cases} \dfrac{2(p-2)}{r}(1+o(1)) & \text{if} \quad \lambda < \infty \\ -\dfrac{C}{2p}(1+o(1)) & \text{if} \quad \lambda = \infty \quad . \end{cases}$$

Proof. The fact that $\rho_{C,\lambda}(r)$ is decreasing in λ follows from simply differentiating with respect to λ in (4.13) and (4.14) and observing that the derivative is negative. The remainder of the lemma follows from well known properties and asymptotic expansions of the modified Bessel functions. ||

In general, I_ν and K_ν are expressible in closed form for half integer ν (corresponding to odd dimensions p). In all cases, tables of I_ν and K_ν exist for small and moderate integer and half integer values of ν (i.e. all p of moderate size), so one need not resort to numerical work to evaluate the $\rho_{C,\lambda}$.

The function $\rho_{C,0}(r)$, being the largest solution to (4.10), will be of particular interest. The following lemma gives some indication of its behavior for $C > 0$ and $p > 1$. (The $C = 0$ and $p = 1$ cases were dealt with in Lemma 3(i).)

Lemma 4. If $C > 0$

(i) and $p = 3$, then

$$\rho_{C,0}(r) = \frac{\sqrt{C}}{\sqrt{r}} + \frac{2}{r} ;$$

(ii) and $p = 5$, then

$$\rho_{C,0}(r) = \frac{\sqrt{C}}{\sqrt{r}} + \frac{4}{r} + \frac{4}{r(2+\sqrt{Cr})} ;$$

(iii) and $p = 2$, then

$$\rho_{C,0}(r) = \frac{2}{r} \frac{\int_0^\infty \frac{\cos(t\sqrt{Cr}/2)}{(t^2+1)^{3/2}}\,dt}{\int_0^\infty \frac{\cos(t\sqrt{Cr}/2)}{(t^2+1)^{1/2}}\,dt} ;$$

(iv) then as $r \to \infty$,

$$\rho_{C,0}(r) = \frac{\sqrt{C}}{\sqrt{r}} + \frac{(p-1)}{r} + \frac{(p-3)(p-1)}{2r\sqrt{Cr}} + o(r^{-2}).$$

Proof. Simple calculation from known formulas and expansions for the modified Bessel functions. ||

At this point, the estimators (for which Conditions 2 and 3 hold) which satisfy (4.1) (or $\bar{\mathscr{D}}_C\, \phi(r) \leq 0$) can be described. They are the estimators corresponding to functions $\rho(r)$ which

(i) are continuous and piecewise differentiable;

(ii) satisfy $\rho_{C,\infty}(r) \leq \rho(r) \leq \rho_{C,0}(r)$;

(iii) satisfy $\sqrt{r}\,\rho(r) \to 0$ as $r \to 0$ when $p = 1$, and satisfy

$$\int_0^{\varepsilon} r\rho'(r)dr < \infty \quad \text{when } p = 2;$$

(iv) have the property that for any given point r_0, corresponding to which is the λ_0 such that $\rho_{C,\lambda_0}(r_0) = \rho(r_0)$, $\rho(r)$ must be greater than or equal to $\rho_{C,\lambda_0}(r)$ for all $r \geq r_0$. The graph of $\rho(r)$ can thus follow any curve $\rho_{C,\lambda}(r)$, but if it departs from such a curve it must go up and to the right.

The properties (i), (ii), and (iii) above are conditions which ensure that the estimator (4.5) satisfies Conditions 2 and 3(ii) of Section 2. (Property (ii) above is also, of course, needed to ensure that $\bar{\mathscr{D}}_C\phi(r) \leq 0$.) When $p = 1$, the estimator will violate Condition 2(i) unless $\sqrt{r}\,\rho(r) \to 0$ (ensuring continuity of the estimator as $r = |x-\mu|^2 \to 0$). For $p > 1$, discontinuity at $r = 0$ is allowed by Condition 2(i). The moment requirements in Condition 2 can be shown (using Lemma 3(iii)) to be satisfied for the estimators corresponding to $\rho_{C,\infty}(r)$ and $\rho_{C,0}(r)$ when $p \geq 3$, and hence by $\rho(r)$ satisfying properties (i), (ii), and (iv) above. Property (iii) above ensures satisfaction of the moment requirements in Condition 2 for $p = 1$ and $p = 2$. (It is possible to show using Lemma 3(iii) that, when $p = 1$, only $\rho_{C,\pi/4}$ satisfies Condition 2, while when $p = 2$ only $\rho_{C,\infty}$ satisfies Condition 2.)

We now proceed with the theorem formalizing the nature of the solution to Problem II**. Let $\underset{\sim}{\delta}^C$ denote the "optimal" estimator, i.e., the estimator which minimizes $r(\pi,\underset{\sim}{\delta})$ among all spherically symmetric (about $\underset{\sim}{\mu}$) estimators satisfying Conditions 2 and 3 of Section 2 and for which $\bar{\mathcal{D}}_C\ (r) \le 0$ (which implies that $R(\underset{\sim}{\theta},\underset{\sim}{\delta}) \le R(\underset{\sim}{\theta},\underset{\sim}{\delta}^o) + C = p + C$). Also, let ρ_C and ϕ_C be defined, as usual, by

(4.15) $$\rho_C(r) = -4\phi_C'(r)/\phi_C(r),$$

$$\underset{\sim}{\delta}^o(\underset{\sim}{x}) = \underset{\sim}{x} - \rho_C(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)(\underset{\sim}{x}-\underset{\sim}{\mu}).$$

<u>Theorem 2.</u> If $\rho_C(r) \ne \rho^\pi(r)$ for all $r\in[a,b]$ ($a > 0$, $b < \infty$), then it must be true that $\rho_C(r) = \rho_{C,\lambda}(r)$ for some λ and all $a \le r \le b$.

<u>Proof.</u> We will consider the case $\rho_C(r) < \rho^\pi(r)$ for $a \le r \le b$. The other cases are dealt with by similar arguments.To argue by contradiction, suppose there does not exist a λ such that $\rho_C(r) = \rho_{C,\lambda}(r)$ for all $a \le r \le b$.

Let λ^* be such that $\rho_{C,\lambda^*}(b) = \rho_C(b)$. (Such a λ^* must exist by Lemma 3(i), since it can be shown that $\rho_C(r)$ must be between $\rho_{C,\infty}(r)$ and $\rho_{C,0}(r)$ to satisfy $\bar{\mathcal{D}}_C\phi_C(r) \le 0$.) Define

$$d = \sup_{a\le r\le b} \{r: \rho_{C,\lambda^*}(r) \ne \rho_C(r)\}\ .$$

By continuity, $a < d \le b$. Next, choose $\varepsilon > 0$ so that $d - \varepsilon > a$ and $\rho_C(r) < \rho_{C,\lambda^*}(r) < \rho^\pi(r)$ for $d - \varepsilon < r < d$. (It can be shown that if $\rho_C(r_1) \ge \rho_{C,\lambda^*}(r_1)$ for some $d - \varepsilon < r_1 < d$, then it cannot be true that $\bar{\mathcal{D}}_C\phi_C(r) \le 0$ for all $r_1 < r < d$.) Without loss of generality, it can be assumed that b and ε were chosen so that $|\rho_C'(r)| < k_1 < \infty$ for $d - \varepsilon < r \le d$. Also, let

$$k_2 = |\sup_{a\le r\le b} \rho_{C,\lambda^*}'(r)|$$

(which is finite by Lemma 3(ii)). Observe that, for any $k \geq 2(k_1+k_0)$ and any point $r_0 \in (d-\frac{\varepsilon}{2}, d)$, the function $\psi(r) = \rho_{C,\lambda^*}(r_0) + k(r-r_0)$ must intersect $\rho_C(r)$ at some point $d - \varepsilon < r_1 < r_0$. (Choose r_1 to be the first point of intersection if several exist.) Finally, define

$$\tilde{\rho}(r) = \begin{cases} \rho_C(r) & \text{for} \quad r \leq r_1 \text{ and } r \geq d \\ \psi(r) & \text{for} \quad r_1 \leq r \leq r_0 \\ \rho_{C,\lambda^*}(r) & \text{for} \quad r_0 \leq r \leq d. \end{cases}$$

Now it is clear that the estimator

$$\tilde{\delta}(x) = x - \tilde{\rho}(|x-\mu|^2)(x-\mu)$$

will satisfy Conditions 2 and 3 of Section 2 if δ^C does. To verify that $\bar{\mathscr{D}}_C\tilde{\phi}(r) \leq 0$ (where $\tilde{\rho}(r) = -4\tilde{\phi}'(r)/\tilde{\phi}(r)$), it is only necessary to check (see (4.9)) that

$$\xi(r) = -2p\psi(r) + r\psi^2(r) - 4r\psi'(r) - C \leq 0$$

for $r_1 \leq r \leq r_0$. (By assumption on ρ_C and definition of ρ_{C,λ^*}, $\bar{\mathscr{D}}_C\tilde{\phi}(r) \leq 0$ for $r \leq r_1$ and $r \geq r_0$.) From the definition of $\psi(r)$ it is clear that

$$\xi(r) = -2p[\rho_{C,\lambda^*}(r_0)+k(r-r_0)]$$

$$+ r[\rho^2_{C,\lambda^*}(r_0)+2k\rho_{C,\lambda^*}(r_0)(r-r_0)+k^2(r-r_0)^2] - 4rk - C.$$

Observe, however, from (4.9) and the fact that $\bar{\mathscr{D}}_C\phi_{C,\lambda^*}(r) = 0$, that

$$-2p\rho_{C,\lambda^*}(r_0) = -r_0\rho^2_{C,\lambda^*}(r_0) + 4r_0\rho'_{C,\lambda^*}(r_0),$$

and hence

$$\xi(r) = -2pk(r-r_0) + (r-r_0)\rho^2_{C,\lambda^*}(r_0) + 2k(r-r_0)\rho_{C,\lambda^*}(r_0)$$

$$+ rk^2(r-r_0)^2 + 4r_0\rho'_{C,\lambda^*}(r_0) - 4rk.$$

A moments reflection reveals that the k_1 and k_2 which work for a given ε

also work for all smaller ε . By choosing ε small enough we can ensure that $r \geq \frac{1}{2} r_0$ for $r_1 \leq r \leq r_0$, and hence that

$$
\begin{aligned}
4r_0\rho'_{C,\lambda^*}(r_0) - 4rk &\leq 4r_0\rho'_{C,\lambda^*}(r_0) - 2r_0k \\
&\leq 4r_0k_2 - 4r_0(k_1+k_2) \\
&= -4r_0k_1.
\end{aligned}
$$

Thus

$$\xi(r) \leq (r-r_0)[-2pk+\rho^2_{C,\lambda^*}(r_0)+2k\rho_{C,\lambda^*}(r_0)+rk^2(r-r_0)] - 4r_0k_1.$$

As $\varepsilon \to 0$, the expression in square brackets above stays bounded, but $(r-r_0) \to 0$ when $r_1 \leq r \leq r_0$. Hence $\xi(r) \leq 0$ for $r_1 \leq r \leq r_0$ and small enough ε , completing the argument that $\tilde{\rho}(r)$ indeed satisfies $\bar{\mathcal{D}}_C\tilde{\phi}(r) \leq 0$.

To complete the proof, we must show that $r(\pi,\tilde{\underset{\sim}{\delta}}) < r(\pi,\underset{\sim}{\delta}^C)$, contradicting the supposed optimality of $\underset{\sim}{\delta}^C$. But it is well known that, for any estimator $\underset{\sim}{\delta}$,

$$
\begin{aligned}
r(\pi,\underset{\sim}{\delta}) - r(\pi,\underset{\sim}{\delta}^\pi) &= E^m|\underset{\sim}{\gamma}(\underset{\sim}{X})-\underset{\sim}{\gamma}^\pi(\underset{\sim}{X})|^2 \\
&= E^m\{[\rho(|\underset{\sim}{X}-\underset{\sim}{\mu}|^2)-\rho^\pi(|\underset{\sim}{X}-\underset{\sim}{\mu}|^2)]^2|\underset{\sim}{X}-\underset{\sim}{\mu}|^2\},
\end{aligned}
$$

where m indicates that the expectation is with respect to the marginal distribution of $\underset{\sim}{X}$. From this and the fact that, by construction, $\tilde{\rho}(r)$ is closer to $\rho^\pi(r)$ than $\rho_C(r)$ is to $\rho^\pi(r)$ for $r_1 < r < d$, the desired conclusion follows. ||

It will typically be the case that

(4.16) $$\rho^\pi(r) = k_1 + k_2r + o(r)$$

as $r \to 0$, where $k_1 > 0$ and $k_2 \neq 0$. (This can be seen by considering (4.8) and expanding h(r) in a Taylors series, for typical π.) When this is true, it can be seen from Lemma 3(iii) that if $\rho_{C,\lambda}(r)$ is positive as $r \to 0$, it blows up at such a rate that $\rho_C(r)$ (the optimal solution) cannot equal $\rho_{C,\lambda}(r)$ for sufficiently small r. Hence, by Theorem 2, $\rho_C(r)$ must equal $\rho^{\pi}(r)$ on some interval $(0,b_1)$. (When this is the case and (4.16) holds, it is easy to verify that ρ_C will satisfy Conditions 2 and 3 - see the discussion after Lemma 4 - so no technical difficulties will be encountered.) Intervals in which $\rho_C(r)$ equals some $\rho_{C,\lambda}(r)$ and equals $\rho^{\pi}(r)$ will then alternate. The structure of $\rho_C(r)$ will thus usually be of the following form: for some numbers $0 = a_0 < b_1 \leq a_1 \leq b_2 \leq \dots$,

$$(4.17) \qquad \rho_C(r) = \begin{cases} \rho^{\pi}(r) & \text{for} \quad a_i \leq r \leq b_{i+1} \\ \rho_{C,\lambda_i}(r) & \text{for} \quad b_i \leq r \leq a_i \ , \end{cases}$$

where the λ_i are determined by the continuity constraints $\rho^{\pi}(b_i) = \rho_{C,\lambda_i}(b_i)$. The at first sight formidable task of finding the optimal sequences $\{a_i\}$ and $\{b_i\}$ is greatly simplified by the observation that, after b_i (and hence λ_i) have been selected, the subsequent a_i can only be a point for which $\rho_{C,\lambda_i}(a_i) = \rho^{\pi}(a_i)$. There will almost never be more than one or two points at which $\rho_{C,\lambda_i}(r)$ and $\rho^{\pi}(r)$ are equal for $r > b_i$, so the possibilities for the a_i are very limited. Furthermore, it will often happen that $\rho^{\pi}(r) > \rho_{C,0}(r)$ for $r \geq k$, in which case $\rho_C(r)$ must equal $\rho_{C,0}(r)$ for $r > k$. (This follows from Theorem 2 and Lemma 3(i).) Hence there will typically be very few b_i (and a_i) (i.e., very few switches between ρ^{π} and the $\rho_{C,\lambda}$), so that numerical minimization of the Bayes risk of estimators satisfying (4.17) over the b_i (and a_i) is quite feasible. This is particularly true because of the following relatively simple formula that can be used for the Bayes risk of $\underset{\sim}{\delta}^C$.

Lemma 5. If

$$\underset{\sim}{\delta}^C(\underset{\sim}{x}) = \underset{\sim}{x} - \rho_C(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)(\underset{\sim}{x}-\underset{\sim}{\mu}),$$

and ρ_C is as in (4.17), then

$$(4.18)\quad r(\pi,\underset{\sim}{\delta}^C) = p+\frac{S_p}{2}\sum_{i\geq 1}\left\{4\left[2b_i^{p/2}h'(b_i)-2a_{i-1}^{p/2}h'(a_{i-1}) - \int_{a_{i-1}}^{b_i}\frac{(h'(r))^2}{h(r)}r^{p/2}dr\right] + \int_{b_i}^{a_i} Ch(r)r^{(p-2)/2}dr\right\},$$

where $h(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)$ is the marginal density of $\underset{\sim}{X}$ and S_p is the surface area of the unit p-sphere given by $S_p = 2\pi^{p/2}/\Gamma(p/2)$.

Proof. From (2.5) and the observation that $R(\underset{\sim}{\theta},\underset{\sim}{\delta}^o) = p$, it follows that

$$r(\pi,\underset{\sim}{\delta}^C) = p + E^\pi E_{\underset{\sim}{\theta}}\left[\frac{4}{g_C(\underset{\sim}{X})}\mathscr{D}\, g_C(\underset{\sim}{X})\right]$$

$$= p + E^h\left[\frac{4}{g_C(\underset{\sim}{X})}\mathscr{D}\, g_C(\underset{\sim}{X})\right].$$

Making the transformation $r = |\underset{\sim}{x}-\underset{\sim}{\mu}|^2$, noting that $\mathscr{D}g(\underset{\sim}{x}) = \mathscr{D}_C g(\underset{\sim}{x}) + \frac{C}{4}g(\underset{\sim}{x})$, and using (4.9) gives

$$r(\pi,\underset{\sim}{\delta}^C) = p + \int_0^\infty \frac{4}{\phi_C(r)}[\bar{\mathscr{D}}_C\phi_C(r) + \frac{C}{4}\phi_C(r)]h(r)(\tfrac{1}{2}S_p r^{(p-2)/2}dr).$$

Now, from (4.17) and (4.8) it follows that

$$\phi_C(r) = \begin{cases} \phi_\pi(r) \equiv \sqrt{h(r)} & \text{for} \quad a_i \leq r \leq b_{i+1} \\ \phi_{C,\lambda_i}(r) & \text{for} \quad b_i \leq r \leq a_i \end{cases}.$$

Furthermore,

$$\bar{\mathscr{D}}_C\phi_\pi(r) + \frac{C}{4}\phi_\pi(r) = 2p\phi_\pi'(r) + 4r\phi_\pi''(r)$$

$$= \frac{ph'(r)}{\sqrt{h(r)}} + \frac{2rh''(r)}{\sqrt{h(r)}} - \frac{r[h'(r)]^2}{[h(r)]^{3/2}},$$

while, by definition, $\bar{\mathcal{D}}_C\, \phi_{C,\lambda_i}(r) = 0$. Hence

$$r(\pi, \underset{\sim}{\delta}^C) = p + \frac{S_p}{2} \sum_{i \geq 1} \left\{ 4 \int_{a_{i-1}}^{b_i} \left[\frac{ph'(r)}{h(r)} + \frac{2rh''(r)}{h(r)} - \frac{r(h'(r))^2}{(h(r))^2} \right] h(r) r^{(p-2)/2} dr \right.$$

$$\left. + \int_{b_i}^{a_i} Ch(r) r^{(p-2)/2} dr \right\} .$$

Integrating by parts gives

$$\int_{a_{i-1}}^{b_i} h''(r) r^{p/2} dr = h'(r) r^{p/2} \Big|_{a_{i-1}}^{b_i} - \int_{a_{i-1}}^{b_i} h'(r) \frac{p}{2} r^{(p-2)/2} dr ,$$

which when used above gives the desired result. ||

5. An Example

In this section we present perhaps the most important example of the theory of the preceeding section, namely the analysis for conjugate priors. (Although conjugate priors are usually not robust, that is of no concern here because of the risk restrictions employed.) Thus we assume $\pi(d\underset{\sim}{\theta})$ is a $\mathcal{N}_p(\underset{\sim}{\mu}, \tau^2 I)$ distribution. Since $\underset{\sim}{X}$ is $\mathcal{N}_p(\underset{\sim}{\theta}, I)$, it follows that the marginal distribution of $\underset{\sim}{X}$ is $\mathcal{N}_p(\underset{\sim}{\mu}, (1+\tau^2)I)$, i.e., the marginal density is

$$h(r) = [2\pi(1+\tau^2)]^{-p/2} e^{-r/[2(1+\tau^2)]},$$

where $r = |\underset{\sim}{x} - \underset{\sim}{\mu}|^2$. Hence (see (4.8))

$$\rho^{\pi}(r) = \frac{-2h'(r)}{h(r)} = \frac{1}{1+\tau^2} .$$

It is easy to see in this case (as indicated in the discussion after Theorem 2) that the optimal estimator is

$$(5.1) \qquad \underset{\sim}{\delta}^C(\underset{\sim}{x}) = \underset{\sim}{x} - \rho_C(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2)(\underset{\sim}{x}-\underset{\sim}{\mu}),$$

where

$$(5.2) \qquad \rho_C(r) = \begin{cases} \rho^\pi(r) = 1/(1+\tau^2) & \text{for} \quad 0 < r \leq b \\ \rho_{C,0}(r) & \text{for} \quad b \leq r \end{cases}$$

and b is defined by

$$(5.3) \qquad \rho^\pi(b) = (1+\tau^2)^{-1} = \rho_{C,0}(b).$$

(Lemma 3(i) and Lemma 4 describe $\rho_{C,0}(r)$.) Furthermore, using Lemma 5, a calculation (again using integration by parts) yields for the Bayes risk of $\underset{\sim}{\delta}^C$ the formula

$$(5.4) \qquad r(\pi,\underset{\sim}{\delta}^C) = p - \frac{p}{(1+\tau^2)}\psi_p\left(\frac{b}{1+\tau^2}\right) + C\left(1-\psi_p\left(\frac{b}{1+\tau^2}\right)\right) - \frac{2}{(1+\tau^2)\Gamma(p/2)}\left[\frac{b}{2(1+\tau^2)}\right]^{p/2} e^{-b/[2(1+\tau^2)]},$$

where $\psi_\nu(z)$ is the cumulative distribution function of the chi-square distribution with ν degrees of freedom.

To obtain some idea as to the effectiveness of the estimators $\underset{\sim}{\delta}^C$, we will present some tables of their risks. It is convenient to consider, instead of $r(\pi,\underset{\sim}{\delta}^C)$, the normalized relative savings risk of Efron and Morris [11] given by

$$(5.5) \qquad RSR(\pi,\underset{\sim}{\delta}) = \frac{r(\pi,\underset{\sim}{\delta})-r(\pi,\underset{\sim}{\delta}^\pi)}{r(\pi,\underset{\sim}{\delta}^0)-r(\pi,\underset{\sim}{\delta}^\pi)}.$$

This measures the proportion of the potential Bayesian improvement over $\underset{\sim}{\delta}^0(\underset{\sim}{x}) = \underset{\sim}{x}$ which is attained by the estimator $\underset{\sim}{\delta}$. The other side of the coin is the "robustness" of the estimator, which in this case is indicated by C, the amount by which the estimator could be worse than $\underset{\sim}{\delta}^0$. To put this on the same scale as RSR, we will formally consider the "relative risk robustness"

$$RRR(\pi,\underset{\sim}{\delta}) = \frac{\sup_{\underset{\sim}{\theta}}[R(\underset{\sim}{\theta},\underset{\sim}{\delta})-R(\underset{\sim}{\theta},\underset{\sim}{\delta}^0)]}{r(\pi,\underset{\sim}{\delta}^0)-r(\pi,\underset{\sim}{\delta}^\pi)} \quad .$$

(This measure is also realistic in the sense that one would be concerned about the possible harm in using $\underset{\sim}{\delta}$ instead of $\underset{\sim}{\delta}^0$ relative to the maximum potential gain available.) Thus small RSR indicates near optimality from a Bayesian viewpoint, while small RRR indicates near optimality from a classical or minimax or Bayesian robustness viewpoint.

For the remainder of the section, we will state results for the situation where $\underset{\sim}{X}$ is $\mathcal{N}_p(\underset{\sim}{\theta},\sigma^2 I)$, since this is the practical situation.

<u>Theorem 3.</u> If $\underset{\sim}{X}$ is $\mathcal{N}_p(\underset{\sim}{\theta},\sigma^2 I)$ and $\underset{\sim}{\theta}$ is $\mathcal{N}_p(\underset{\sim}{\mu},\tau^2 I)$, then

(5.6) $$\underset{\sim}{\delta}^C(\underset{\sim}{x}) = \underset{\sim}{x} - \rho_C^*(|\underset{\sim}{x}-\underset{\sim}{\mu}|^2/\sigma^2)(\underset{\sim}{x}-\underset{\sim}{\mu}),$$

where (letting $r = |\underset{\sim}{x}-\underset{\sim}{\mu}|^2/\sigma^2$)

(5.7) $$\rho_C^*(r) = \begin{cases} \sigma^2/(\sigma^2+\tau^2) & \text{for} \quad 0 < r \le b \\ \rho_{C/\sigma^2,0}(r) & \text{for} \quad b \le r \end{cases}$$

and b is defined by

(5.8) $$\sigma^2/(\sigma^2+\tau^2) = \rho_{C/\sigma^2,0}(b).$$

Furthermore,

(5.9) $$RRR(\pi,\underset{\sim}{\delta}^C) = \frac{C(\sigma^2+\tau^2)}{p\sigma^4} \quad ,$$

and

(5.10) $$RSR(\pi,\underset{\sim}{\delta}^C) = [1-\psi_p(y)][1+RRR(\pi,\underset{\sim}{\delta}^C)] - \frac{(y/2)^{p/2}e^{-y/2}}{\Gamma(1+p/2)} \quad ,$$

where $y = b\sigma^2/(\sigma^2+\tau^2)$ depends only on p and $RRR(\pi,\underset{\sim}{\delta}^C)$.

Proof. Formulas (5.6) through (5.10) follow from the preceeding analysis after dividing $\underset{\sim}{X}$, $\underset{\sim}{\theta}$ and $\sqrt{C}$ by σ, and observing that

$$R(\underset{\sim}{\theta},\underset{\sim}{\delta}^0) = r(\pi,\underset{\sim}{\delta}^0) = p\sigma^2 \text{ and } r(\pi,\underset{\sim}{\delta}^\pi) = \frac{p\sigma^2\tau^2}{\sigma^2+\tau^2} .$$

The last statement of the theorem follows trivially from (5.8) and Lemma 3(i) when C = 0 or p = 1, while for C > 0 and p > 1 (5.8) can be written

$$\frac{\sigma^2}{\sigma^2+\tau^2} = \frac{[C/\sigma^2]^{1/2}}{[y(\sigma^2+\tau^2)/\sigma^2]^{1/2}} \frac{K_{\nu+1}\left(\frac{1}{2}\left[\frac{C}{\sigma^2}\cdot\frac{y(\sigma^2+\tau^2)}{\sigma^2}\right]^{1/2}\right)}{K_\nu\left(\frac{1}{2}\left[\frac{C}{\sigma^2}\cdot\frac{y(\sigma^2+\tau^2)}{\sigma^2}\right]^{1/2}\right)}$$

or

$$1 = \left[\frac{p\,\mathrm{RRR}}{y}\right]^{1/2} \frac{K_{\nu+1}\left(\frac{1}{2}[py\,\mathrm{RRR}]^{1/2}\right)}{K_\nu\left(\frac{1}{2}[py\,\mathrm{RRR}]^{1/2}\right)} . \quad ||$$

The pleasant feature of using RRR and RSR, as indicated in Theorem 3, is that, for a given p, $\mathrm{RSR}(\pi,\underset{\sim}{\delta}^C)$ depends on C, σ^2, and τ^2 only through $\mathrm{RRR}(\pi,\underset{\sim}{\delta}^C)$. The following corollaries and tables present interesting special cases. The proofs are immediate from Theorem 3, Lemma 3(i), and Lemma 4.

Corollary 1. If C > 0 and p = 1, then

$$\delta^C(x) = \begin{cases} x - \dfrac{\sigma^2}{\sigma^2+\tau^2}(x-\mu) & \text{if } |x-\mu|^2 \le C(\sigma^2+\tau^2)^2/\sigma^4 \\ x - \dfrac{\sqrt{C}}{|x-\mu|}(x-\mu) & \text{if } |x-\mu|^2 \ge C(\sigma^2+\tau^2)^2/\sigma^4 , \end{cases}$$

$$\mathrm{RRR}(\pi,\delta^C) = C(\sigma^2+\tau^2)/\sigma^4 \quad (\equiv \mathrm{RRR} \text{ for short}),$$

and

$$\mathrm{RSR}(\pi,\delta^C) = [1-\psi_1(\mathrm{RRR})][1+\mathrm{RRR}] - [2\mathrm{RRR}/\pi]^{1/2}e^{-\mathrm{RRR}/2} .$$

Table 1. $RRR(\pi,\delta^C)$ vs. $RSR(\pi,\delta^C)$ for $p = 1$, $C > 0$.

RRR	0	.002	.02	.10	.2	.4	.6	.8	1.0	1.4	4	5	∞
RSR	1	.93	.80	.58	.46	.32	.24	.18	.16	.10	.0115	.006	0

Corollary 2. If $C > 0$ and $p = 2$, then

$$\delta^C(x) = \begin{cases} x - \dfrac{\sigma^2}{\sigma^2+\tau^2}(x-\mu) & \text{if} \quad |x-\mu|^2 \le b\sigma^2 \\ x - \dfrac{\sqrt{C}}{|x-\mu|} \dfrac{K_1\left(\frac{1}{2}|x-\mu|\sqrt{C}/\sigma^2\right)}{K_0\left(\frac{1}{2}|x-\mu|\sqrt{C}/\sigma^2\right)} & \text{if} \quad |x-\mu|^2 \ge b\sigma^2 , \end{cases}$$

$$RRR(\pi,\delta^C) = C(\sigma^2+\tau^2)/[2\sigma^4],$$

and

$$RSR(\pi,\delta^C) = e^{-y/2}[1- \frac{y}{2} + RRR(\pi,\delta^C)] .$$

(See Theorem 3 for the definitions of b and y. Note that an integral representation for K_1/K_0 is given in Lemma 4.)

Table 2. $RRR(\pi,\delta^C)$ vs. $RSR(\pi,\delta^C)$ for $p = 2$, $C > 0$.

RRR	0	.024	.073	.135	.20	.28	.36	.44	.52
RSR	1	.41	.31	.24	.19	.16	.13	.11	.089

Corollary 3. If $C \ge 0$ and $p = 3$, then

$$\delta^C(x) = \begin{cases} x - \dfrac{\sigma^2}{\sigma^2+\tau^2}(x-\mu) & \text{if} \quad |x-\mu|^2 \le (\sigma^2+\tau^2)y \\ x - \left[\dfrac{\sqrt{C}}{|x-\mu|} + \dfrac{2\sigma^2}{|x-\mu|^2}\right](x-\mu) & \text{if} \quad |x-\mu|^2 \ge (\sigma^2+\tau^2)y, \end{cases}$$

$$RRR(\pi,\underset{\sim}{\delta}^C) = C(\sigma^2+\tau^2)/[3\sigma^4],$$

$$RSR(\pi,\underset{\sim}{\delta}^C) = [1-\psi_3(y)][1+RRR] - \frac{(y/2)^{3/2}e^{-y/2}}{3\sqrt{\pi}\,/4} \quad ,$$

and

$$y = \left\{\frac{3}{2}\,RRR + \frac{4}{3} + RRR[1+8/(3RRR)]^{1/2}\right\} .$$

Table 3. $RRR(\pi,\underset{\sim}{\delta}^C)$ vs. $RSR(\pi,\underset{\sim}{\delta}^C)$ for $p = 3$, $C \geq 0$.

RRR	0	.025	.075	.1	.135	.2	.4	.7	1.0	1.5
RSR	.296	.203	.151	.133	.116	.091	.052	.027	.014	.008

Corollary 4. If $C = 0$ and $p \geq 3$, then

$$\underset{\sim}{\delta}^C(\underset{\sim}{x}) = \begin{cases} \underset{\sim}{x} - \dfrac{\sigma^2}{\sigma^2+\tau^2}(\underset{\sim}{x}-\underset{\sim}{\mu}) & \text{if} \quad |\underset{\sim}{x}-\underset{\sim}{\mu}|^2 \leq 2(p-2)(\sigma^2+\tau^2) \\[2ex] \underset{\sim}{x} - \dfrac{2(p-2)\sigma^2}{|\underset{\sim}{x}-\underset{\sim}{\mu}|^2}(\underset{\sim}{x}-\underset{\sim}{\mu}) & \text{if} \quad |\underset{\sim}{x}-\underset{\sim}{\mu}|^2 \geq 2(p-2)(\sigma^2+\tau^2) \ , \end{cases}$$

$$RRR(\pi,\underset{\sim}{\delta}^C) = 0,$$

and

$$RSR(\pi,\underset{\sim}{\delta}^C) = [1-\psi_p(2(p-2))] - \frac{(p-2)^{p/2}e^{-(p-2)}}{\Gamma(1+p/2)} \quad .$$

Table 4. $RSR(\pi,\underset{\sim}{\delta}^C)$ for $p \geq 3$, $C = 0$.

p	3	4	5	6	7	8	9	10	15	20
RSR	.296	.135	.0727	.0427	.0267	.0174	.0117	.008	.0016	.0004

Tables 1 through 4 exhibit the almost startlingly impressive performance of the estimators $\underset{\sim}{\delta}^C$, especially for $p > 1$. When $p = 1$, a substantial sacrifice in Bayes risk improvement must be made if small RRR is desired. For $p = 2$ and $p = 3$, however, the situation is more promising. When $p = 3$, for example, one can guarantee that $\underset{\sim}{\delta}^C$ is no more than 10% worse than $\underset{\sim}{\delta}^0$ at a cost of only 13.3% of the potential Bayes risk improvement. (Note, in contrast, that the conjugate prior Bayes estimator $\underset{\sim}{\delta}^\pi$ has $RRR(\pi,\underset{\sim}{\delta}^\pi) = \infty$.) Table 4 is particularly startling since $C = 0$, i.e., $RRR(\pi,\underset{\sim}{\delta}^C) = 0$ so $\underset{\sim}{\delta}^C$ is minimax. When $p \geq 5$ one attains virtually all of the possible Bayesian gains at no cost (in terms of possible worsened performance compared to $\underset{\sim}{\delta}^0$). Of course, as discussed in Berger [4], this exceptional behavior is due to the Stein effect in simultaneous estimation. It is interesting that, even when $p = 2$, there is apparently considerable benefit derived from this effect.

Acknowledgments. The author would like to thank Burgess Davis and Herman Rubin for several valuable discussions of these problems.

References

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

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

[3] Berger, J. (1980). Statistical Decision Theory: Foundations, Concepts, and Methods. Springer-Verlag, New York.

[4] Berger, J. (1981). Bayesian robustness and the Stein effect. Technical Report #81-1, Purdue University.

[5] Berger, J. (1982). Selecting a minimax estimator of a multivariate normal mean. Ann. Statist. 10.

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

[7] Bickel, P. J. (1980). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. Technical Report, University of California at Berkeley.

[8] Bickel, P. J. (1980). Minimax estimation of the mean of a normal distribution subject to doing well at a point. Technical Report, University of California at Berkeley.

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

[10] Brown, L. (1981). The differential inequality of a statistical estimation problem. Technical Report, Cornell University.

[11] Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators - Part I: the Bayes case. J. Amer. Statist. Assoc. 66, 807-815.

[12] Ghosh, M. and Parsian, A. (1980). Admissible and minimax multiparameter estimation in exponential families. Technical Report, Iowa State University.

[13] Hodges, J. L., Jr. and Lehmann, E. L. (1952). The use of previous experience in reaching statistical decisions. Ann. Math. Statist. 23, 392-407.

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

[15] Shapiro, S. H. (1972). A compromise between Bayes and minimax approaches to estimation. Technical Report No. 31, Department of Statistics, Stanford University.

[16] Shapiro, S. H. (1975). Estimation of location and scale parameters - a compromise. Communications in Statist. 4(12), 1093-1108.

[17] Srinivasan, C. (1980). Admissible generalized Bayes estimators and exterior boundary value probem. Sankhya.

[18] 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, Berkeley.

[19] Stein, C. (1973). Estimation of the mean of a multivariate distribution. Proc. Prague Symp. Asymptotic Statist. 345-381.

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