Estimation in the Multiparameter Exponential Family: Admissibility and inadmissibility results.

by

Anirban DasGupta and Bimal K. Sinha

Purdue University | University of Pittsburgh

Technical Report #84-8

Department of Statistics
Purdue University

March, 1984

(Revised July 1985)

Abstract

The problem of estimating $k(\geq 1)$ linear combinations of the component means in the multiparameter exponential family with possibly dependent coordinates is treated for squared-error loss. It is shown that if the natural parameter space is the whole of $\mathbb{R}^p$, the UMVUE is always admissible if $k \leq 2$. Also, sufficient conditions for admissibility of an arbitrary generalized Bayes estimate are obtained. If the natural parameter space is not the entire $\mathbb{R}^p$ (e.g. in the gamma case), inadmissibility results are proved to indicate that the admissibility pattern of the "natural" estimate may be different even when $k = 1$. Finally, the problem of estimating a vector of smooth parametric functions more general than the means is considered and admissible generalized Bayes estimates are obtained; in particular, a result of Ghosh and Meeden (1977) follows as a corollary.

1. Introduction

Let $\underset{\sim}{X} = (X_1, X_2, \ldots, X_p)$ be a random variable in $\mathbb{R}^p$ with possibly dependent coordinates and suppose $\underset{\sim}{X}$ has a density (with respect to some σ-finite measure μ on $\mathbb{R}^p$) given by

$$f_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = e^{\sum_{i=1}^{p} \theta_i x_i - \psi(\underset{\sim}{\theta})} \qquad (1.1).$$

Thus, the distribution of the random vector $\underset{\sim}{X}$ is in the Multiparameter Exponential family. Here $\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ stands for the vector of natural parameters, and $\theta \varepsilon (H) = \{\underset{\sim}{\theta}: \int e^{\Sigma \theta_i x_i} d\mu(x) < \infty\}$. It is well known that (H) is convex in $\mathbb{R}^p$ and in the interior of (H),

$$\mu_i(\underset{\sim}{\theta}) = E_\theta(X_i) = \frac{\partial}{\partial \theta_i} \psi(\underset{\sim}{\theta}), \; 1 \leq i \leq p. \qquad (1.2)$$

A great deal of research has been done on estimating the mean-vector $\nabla\psi(\underset{\sim}{\theta}) = (\mu_1(\underset{\sim}{\theta}), \ldots, \mu_p(\underset{\sim}{\theta}))$ under a sum of squared-error losses

$$L(\underset{\sim}{\mu}, \underset{\sim}{a}) = \sum_{i=1}^{p} (a_i - \mu_i(\underset{\sim}{\theta}))^2 \qquad (1.3)$$

or various weighted quadratic losses, both for particular distributions in the exponential family as also for the exponential family as a whole. Since the pioneering work of Stein (1956) in the multinormal, the Stein-effect has been shown to be present in a variety of probability structures (without any resemblance to the normal) with an infinite sample space (the fact that an infinite sample space is crucial for the Stein-effect was shown in Guttman (1982)). Apart from various inadmissibility results, the literature now contains useful techniques to construct explicit improvements on the inadmissible "standard estimates and elegant results on methods to meaningfully incorporate available prior information in constructing improved estimators; see, for example, Stein (1973), Hudson (1978), Berger (1980), Hwang (1982), Berger (1982a, 1982b, 1982c), Ghosh, Hwang and Tsui (1983), Berger and DasGupta (1985) etc.

This paper, in part, deals with the problem of estimating, under a sum of squared error losses, any k coordinates or in general, any k linear combinations of all the component means. The first serious attempt to handle a problem of this kind was made by Cohen (1965), who characterized all linear admissible estimates $\sum_{i=1}^{p} \phi_i X_i$ of a single linear combination $\sum_{i=1}^{p} \gamma_i \theta_i$ in the multivariate normal distribution (the result in Cohen (1965) can be somewhat extended to distributions for which the covariance structure is independent of the unknown parameter $\underset{\sim}{\theta}$). For p = 2, similar results were obtained in Makani (1972, 1977) for independent binomials and independent Poissons. Blyth (1974) proved that for p independent binomials, $\sum_{i=1}^{p} \frac{X_i}{n_i}$ is admissible for $\sum_{i=1}^{p} p_i$ for any $p \geq 1$. If the vector $\underset{\sim}{\theta}$ is a location parameter for the random vector X, this problem was treated in fairly complete generality by Berger (1976a, 1976b, 1976c), who showed that the best invariant estimator of a single linear combination of θ_i's is often admissible up to p = 3 and inadmissible for $p \geq 4$. It is important to keep in mind that the assumptions of Berger (1976b) do not hold for multivariate normal and hence the inadmissibility result for $p \geq 4$ excludes the normal distribution (as can also be seen from Cohen (1965)). Question arises if the best unbiased estimate of any linear combination of the means is admissible under a squared-error loss structure in the general exponential family, and if so, whether this result can be extended to estimating any two linear combinations of the means (that the standard estimate of three or more linear combinations cannot be admissible in general is well known). In Section 2, it has been shown that if $(H) = \mathbb{R}^p$, the UMVUE of any two linear combinations is admissible in the general exponential family

(1.1). Then we have also obtained other more general generalized Bayes estimates which are also admissible for the same parametric functions. The technique used is Blyth's (1951) and resembles a method of carrying out Blyth's technique as given in Brown and Hwang (1982). The results in Section 2 can be regarded as a generalization of a special case of Cohen (1965) and a two-fold generalization of Blyth (1974).

Next, it has been shown that if (H) is not the whole of $\mathbb{R}^p$, the admissibility status of the "natural" estimate of even a single linear combination of the means can be quite different; specifically, we have proved inadmissibility results for estimating one linear combination of the scale-parameters of several independent gamma distributions; a general inadmissibility theorem has been proved, using a technique found by Hwang (1982) and then used by DasGupta (1984), which shows the inadmissibility of many non-linear estimates arising naturally from Berger (1980); moreover, we have established, for all $p > 1$, the inadmissibility of many linear estimates including one which may be called a "natural" linear estimate. It turns out that the improved estimates have an empirical Bayesian interpretation and they also lead to new improved estimators of the vector of gamma scale (and natural) parameters under a wide variety of losses; these problems have been considered in Das Gupta (1984). Presumably, qualitatively similar inadmissibility results in the context of estimating one (or more) linear conbination of the means can also be obtained in other distributions with $(H) = R^p_+$, notably the Negative Binomial.

In Section 4, instead of estimating one or more linear functions of the means, we have treated the problem of estimating any p smooth scalar functions $\gamma_1(\theta),\ldots,\gamma_p(\theta)$ under a squared-error loss; the goal is to obtain a sufficient condition (like Brown and Hwang's (1982) for the mean-vector) for the admissibility of a generalized Bayes estimate. From the theorem proved in this section, a result of Ghosh and Meeden (1977) follows as a corollary; it's also a generalization of the main theorem in Brown and Hwang (1982) to parametric functions more general than the mean. Certain new results following as corollaries from this theorem roughly mean that if the parametric function $\underset{\sim}{\gamma}(\underset{\sim}{\theta})$ is "near the mean-vector $\nabla\psi(\theta)$", then the non-informative prior generalized Bayes estimate of $\underset{\approx}{\gamma}(\underset{\sim}{\theta})$ continues to be admissible for small enough p ($p \leq 2$ if $(H) = \mathbb{R}^p$, $p = 1$ if $(H) = R_+^p$); finally in closing, we have made some concluding remarks.

2. Admissibility for $(H) = \mathbb{R}^p$.

Consider the problem of estimating $L\nabla\psi(\underset{\sim}{\theta})$, where Rank $(L_{k\times p}) = k$, $1\leq k\leq p$, under the usual squared-error loss. The notations introduced below are much the same as in Section II of Brown and Hwang (1982).

For a given function h (perhaps vector-valued), define

$$I_x(h) = \int h(\underset{\sim}{\theta})\ f_{\underset{\approx}{\theta}}(\underset{\sim}{x})d\underset{\sim}{\theta}, \qquad (2.1)$$

whenever the integral exists. Let G be any prior distribution on (H); assume G has a density (Lebesgue) g which is almost everywhere (Lebesgue) differentiable. If $\delta_g(\underset{\sim}{X}) = (\delta_g^1(\underset{\sim}{X}),\ldots,\delta_g^k(\underset{\sim}{X}))$ denotes the generalized Bayes estimate of $L\nabla\psi(\underset{\sim}{\theta})$ with respect to G, then under mild conditions on g, (see Brown and Hwang (1982),

$$\delta_g(\underset{\sim}{x}) = L\underset{\approx}{X} + \frac{I_x(L\nabla g)}{I_x(g)} \qquad \text{(a.e. } d\mu) \qquad (2.2)$$

If G has a compact support (compact in (H)), the steps leading to (2.2) are easily verified. Note that (2.2) is frequently valid even when G does not have a compact support. A special case of interest is when $g\equiv 1$(i.e., G is Lebesgue measure), in which case δ_g reduces to $L\underset{\sim}{X}$, the UMVUE of $L\nabla\psi(\underset{\sim}{\theta})$. In what follows, we obtain sufficient conditions for the admissibility of $\delta_g(\underset{\sim}{X})$ for estimating $L\nabla\psi(\underset{\sim}{\theta})$. The analysis will be based on Stein's (1955) sufficient condition for admissibility (see also Blyth (1951)), stated below in the form given in Berger (1976a, page 345).

Lemma 2.1 Suppose $g_n = gh_n^2$ is a sequence of finite priors such that

(i) $\int R(\underset{\sim}{\theta},\delta_{g_n})g_n(\underset{\sim}{\theta})d\underset{\sim}{\theta} < \infty$ for every $n \geq 1$

(ii) $h_n(\underset{\sim}{\theta}) \longrightarrow 1$ a.e. as $n \longrightarrow \infty$

(iii) $h_n(\underset{\sim}{\theta}) \geq \varepsilon$ for every $n \geq 1$ and for some $\varepsilon > 0$ on a set $S \subset (H)$ with $\int_S g(\underset{\sim}{\theta})d\underset{\sim}{\theta} > 0$

(iv) $\Delta_n = \int\{R(\underset{\sim}{\theta},\delta_{g_n}) - R(\underset{\sim}{\theta},\delta_g)\}\, g_n(\underset{\sim}{\theta})d\underset{\sim}{\theta} \longrightarrow 0$ as $n \longrightarrow \infty$.

Then $\delta_g(\underset{\sim}{X})$ is admissible.

Remark: If the sequence of functions h_n each vanishes outside a compact set, condition (i) is automatically satisfied. For given functions $\{h_n, n \geq 1\}$, each having a compact support and each almost everywhere differentiable, the generalized Bayes estimate of $L\nabla\psi(\underset{\sim}{\theta})$ against the prior density $g_n(\theta) = g(\theta)h_n^2(\theta)$ is given as

$$\delta_{g_n}(\underset{\sim}{X}) = L\underset{\sim}{X} + \frac{I_X(L\nabla g_n)}{I_X(g_n)} \tag{2.3}$$

Then, as in Brown and Hwang (1982),

$$\begin{aligned} \nabla_n &= \int ||\delta_g(\underset{\sim}{x}) - \delta_{g_n}(\underset{\sim}{x})||^2\, I_X(g_n)\, d\mu(x) \\ &\leq A_n + B_n \end{aligned} \tag{2.4}$$

$$\text{where } A_n = 2\int\left|\left|\frac{I_x(L\nabla g)}{I_x(g)} - \frac{I_x(h_n^2 L\nabla g)}{I_x g_n}\right|\right|^2 I_x(g_n)\, d\mu(x) \tag{2.5}$$

$$\text{and } B_n = 2\int\left|\left|\frac{I_x(gL\nabla h_n^2)}{I_x g_n}\right|\right|^2 I_x(g_n) d\mu(x) \tag{2.6}$$

If A_n, $B_n \longrightarrow 0$ for suitable choice of h_n's, it will then follow $\nabla_n \longrightarrow 0$ as $n \longrightarrow \infty$.

Now note that

$$B_n = 8\int\left|\left|\frac{I_x(gh_n L\nabla h_n)}{I_x(g_n)}\right|\right|^2 I_x(g_n)\, d\mu(x)$$

(writing $gh_n L\nabla h_n = \sqrt{g}h_n \times \sqrt{g}\, L\nabla h_n$ and then applying Schwartz's inequality)

$$\leq 8\int I_x(g||L\nabla h_n||)^2\, d\mu(x)$$

$$= 8\int g(\theta)||L\nabla h_n(\theta)||^2\, d\theta \quad \text{(by Fubini's Theorem)} \tag{2.7}$$

Hence, $B_n \longrightarrow 0$ if (2.7) goes to zero as $n \longrightarrow \infty$. Next, in order to show that $A_n \longrightarrow 0$, note that the integrand in A_n converges to 0 almost everywhere if h_n converges a.e. to 1. Moreover, proceeding as in Brown and Hwang (1982), it is uniformly (in n) bounded above by $I_x \frac{(||L\nabla g||^2)}{g}$. Hence, $A_n \longrightarrow 0$ as $n \longrightarrow \infty$ by the Bounded Convergence Theorem provided $\int I_x\left(\frac{||L\nabla g||^2}{g}\right) d\mu(x) < \infty$. Again, by Fubini's Theorem,

$$\int I_x\left(\frac{||L\nabla g||^2}{g}\right) d\mu(x) = \int \frac{||L\nabla g(\theta)||^2}{g(\theta)}\, d\theta \quad . \tag{2.8}$$

Now, in view of Lemma 2.1, (2.7), and (2.8), $\delta_g(\underset{\sim}{X})$ will be admissible if for suitable functions h_n of compact support satisfying (ii) and (iii) of Lemma 2.1,

(1) $\int g(\theta)||L\nabla h_n(\theta)||^2\, d\theta \longrightarrow 0$ as $n \longrightarrow \infty$

and (2) g is such that $\int \frac{||L\nabla g(\theta)||^2}{g(\theta)} d\theta < \infty$. (2.9)

For $L = I_{pxp}$, (1) and (2) reduce to the conditions in Brown and Hwang (1982). At this point we note that the UMVUE $L\underset{\sim}{X}$ corresponds to $g \equiv 1$ in which case (2) above becomes redundant. Also, if a sequence of functions $\{h_n\}$ can be found such that it satisfies conditions (i), (ii) and (iii) of Lemma 2.1 and moreover (1) and (2) above are satisfied with $L = I_{pxp}$ (so that by Brown and Hwang (1982) the generalized Bayes estimate δ of $\nabla\psi(\underset{\sim}{\theta})$ against the prior G is admissible), then (1) and (2) are satisfied for any L_{kxp}, since

$$\int g(\theta)||L\nabla h_n(\theta)||^2 \, d\theta \le \lambda_{max} \int g(\theta)||\nabla h_n(\theta)||^2 \, d\theta$$

$$\text{and } \int \frac{||L\nabla g(\theta)||^2}{g(\theta)} d\theta \le \lambda_{max} \int \frac{||\nabla g(\theta)||^2}{g(\theta)} d\theta ,$$

where λ_{max} denotes the maximum eigen-value of $L'L$. In view of the preceding analysis it will then follow that $\delta_g = L\delta$ is admissible for $L\nabla\psi(\theta)$. This of course corresponds to the fact that the admissibility of $L\delta$ for $L\nabla\psi(\theta)$ under squared error loss is equivalent to the admissibility of δ for $\nabla\Psi(\theta)$ under the loss $(a - \nabla\psi(\theta))' L'L(a - \nabla\psi(\theta))$, which by Shinozaki (1975) in turn is equivalent to the admissibility of δ for $\nabla\psi(\theta)$ under usual squared-error loss so long as $L'L$ is positive-definite. Specializing to the case $g \equiv 1$, it will mean that so long as $\underset{\sim}{X}$ is admissible for $\nabla\psi(\underset{\sim}{\theta})$, $L\underset{\sim}{X}$ will also be admissible for $L\nabla\psi(\theta)$; however, as is well known typically $\underset{\sim}{X}$ is not admissible for $\nabla\psi(\theta)$ beyond dimension $p = 2$, so that admissibility of $L\underset{\sim}{X}$ for $L\nabla\psi(\underset{\sim}{\theta})$ will not follow in a straightforward manner for $p > 2$. This also indicates that for (1) above to be satisfied for every $p \ge 1$ (and not just $p \le 2$), new choices of $\{h_n\}$ different from the sequence of Brown and Hwang (1982) need to be found. In what follows we show that for suitable h_n's condition (1) above holds for any $p \ge 1$ so long as $k \le 2$. It will be clear from the proof that if $k \geq 3$ (i.e., if one wants to estimate 3 or more linear functions of the means), then (1) immediately fails (as it should be). Before we get into the actual analysis, we remark that by making a non-singular linear transformation on the sample space (which will preserve the exponential structure) we may assume that the rows of L are unit vectors. In terms of our estimation problem it means that we are trying to estimate any k coordinates of the mean vector. We also remark that the admissibility question in the problem of estimating the coordinates is not interesting unless there is a dependent structure in the sample space, because the admissibility problem otherwise essentially reduces to a one-dimensional problem because of the independence of the coordinates X_i; see, for example, Lehmann (1983). In the following theorem we have taken $g \equiv 1$; later we shall indicate other g's which may be handled too.

<u>Theorem 1</u> If the distribution of $\underset{\sim}{X}_{p\times 1}$ is in the multiparameter exponential family (1.1) with $(H) = \mathbb{R}^p$, $L\underset{\sim}{X}$ is admissible for $L\nabla\psi(\theta)$ for any $L_{2\times p}$. (R(L) in the proof means rank of L).

Proof: Assume without loss of generality $R(L) = 2$ since if $R(L) = 0$ or 1, either $L\underset{\sim}{X}$ becomes proper Bayes or L can be essentially taken as a single row vector in which case admissibility follows from the case $R(L) = 2$ to be treated below.

If $R(L) = 2$, we may take the vectors of L as the first two unit vectors as mentioned before. In view of Lemma 2.1 and the discussion thereafter, we have to construct sequence of functions $\{h_n\}$ with compact support satisfying (ii) and (iii) of Lemma 2.1 and such that $\int(\frac{\partial}{\partial\theta_1} h_n(\theta))^2 \, d\theta + \int(\frac{\partial}{\partial\theta_2} h_n(\theta))^2 \, d\theta \longrightarrow 0$ as $n \longrightarrow \infty$.

For $n \geq 1$, define $||\theta||_n = |\theta_1| + |\theta_2| + \sum_{i\geq 3} |\theta_i|^{\alpha_n}$, where $\alpha_n \geq 1$ is a sequence of reals to be chosen later. Now define

$$h_n(\theta) = 1 \qquad \text{if } ||\theta||_n \leq 1$$
$$= 1 - \frac{\log ||\theta||_n}{\log n} \qquad \text{if } 1 < ||\theta||_n < n$$
$$= 0 \qquad \text{if } ||\theta||_n \geq n \tag{2.10}$$

Since $||\theta||_n$ is everywhere continuous in $\underset{\sim}{\theta}$ and differentiable outside the set $\{\underset{\sim}{\theta}: \theta_i = 0 \text{ for some } i\}$, and since $\{\underset{\sim}{\theta}: ||\theta||_n = \text{a constant}\}$ has Lebesgue measure zero, h_n is almost everywhere differentiable (in fact, by throwing out a countable union of null sets, we can find $(H_0) \subset (H)$ such that every h_n is differentiable on $(H)_0$ and the Lebesgue measure of $(H) - (H)_0$ is 0).

Next note that $\{\underset{\sim}{\theta} = |\theta_i| > n \text{ for some } i \geq 1\}$

$$\subseteq \{\underset{\sim}{\theta} = ||\underset{\sim}{\theta}||_n > n\} \qquad (\text{as } \alpha_n \geq 1 \text{ for all } n)$$
$$\subseteq \{\underset{\sim}{\theta} = h_n(\underset{\sim}{\theta}) = 0\}$$

$\Rightarrow \{\underset{\sim}{\theta} = h_n(\underset{\sim}{\theta})\} > 0\} \subseteq \{\theta = |\theta_i| \leq n \text{ for all } i \geq 1\}$,

which is compact in $(H) = \mathbb{R}^P$. Since h_n's have compact support and $0 \leq h_n \leq 1$, (i) of Lemma 2.1 follows.

To prove (ii), first note that if $\frac{\alpha_n}{\log n} \longrightarrow 0$ as $n \longrightarrow \infty$. then for every fixed $\underset{\sim}{\theta}$, $||\theta||_n < n$ for all large enough n. This is because $|\theta_i| < \frac{n}{3}$ for $i = 1,2$ for large enough n, and for $i \geq 3$, $|\theta_i|^{\alpha_n} < \frac{n}{3(p-2)}$ for large enough n since $\frac{\alpha_n}{\log n} \longrightarrow 0$ as $n \longrightarrow \infty$.

Also, $\frac{\log ||\theta||_n}{\log n} \leq \frac{\log (|\theta_1| + |\theta_2| + \{\sum_{i=3}^{p} |\theta_i|\}^{\alpha_n})}{\log n} \longrightarrow 0$ as $n \longrightarrow \infty$

for every fixed $\underset{\sim}{\theta}$. It follows from the definition of h_n that $h_n(\theta) \longrightarrow 1$ as $n \longrightarrow \infty$.

In order to verify (iii), define

$$S = \{\underset{\sim}{\theta} : \sum_{i=1}^{p} |\theta_i| \leq 1\}.$$

Clearly, $S \subseteq \{\underset{\sim}{\theta}: |\theta_i| \leq 1 \text{ for all } i \geq 1\}$

$$\subseteq \{\underset{\sim}{\theta}: |\theta_i| \leq 1,\ |\theta_2| \leq 1,\ |\theta_i|^{\alpha_n} \leq |\theta_i| \text{ for all } i \leq 3\}.$$

Hence, for $\underset{\sim}{\theta} \varepsilon S$, $||\theta||_n = |\theta_1| + |\theta_2| + \sum_{i=3}^{p} |\theta_i|^{\alpha_n} \leq \sum_{i=1}^{p} |\theta_i| \leq 1$

$\Rightarrow S \subseteq \{\underset{\sim}{\theta}: h_n(\theta) = 1\}$.

Since $\int_S d\theta > 0$, (iii) holds. We now need to show that

$$\int(\frac{d}{d\theta_1} h_n(\theta))^2 d\theta + \int(\frac{d}{d\theta_2} h_n(\theta))^2 d\theta \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Towards this end, first observe that $\{\underset{\sim}{\theta}: 1 < ||\theta||_n < n\}$ is an open set in $(H) = \mathbf{R}^P$. Hence,

$$\frac{d}{d\theta_i} h_n(\theta) = - \frac{1}{||\theta||_n \log n} \frac{d}{d\theta_i} ||\theta||_n I_{1 < ||\theta||_n < n} \quad \text{a.e. (Lebesgue)} \tag{2.11}$$

Now, $\frac{d}{d\theta_1} ||\theta|| = \operatorname{sgn} \theta_1$ and $\frac{d}{d\theta_2} ||\theta||_n = \operatorname{sgn} \theta_2$ (a.e.), so that from (2.11), $(\frac{d}{d\theta_1} h_n(\theta))^2 = (\frac{d}{d\theta_2} h_n(\theta))^2 = \frac{1}{||\theta||_n^2 (\log n)^2} I_{1 < ||\theta||_n < n}$ a.e.

Hence, it suffices to show $I_n = \frac{1}{(\log n)^2} \int \frac{1}{||\theta||_n^2} I_{1 < ||\theta||_n < n} \longrightarrow 0$ as $n \longrightarrow \infty$.

Note that the integrand in I_n is a function of $|\theta_i|$'s; hence

$$I_n = \frac{2^p}{(\log n)^2} \int_{\substack{\theta:\ \theta_i > 0\ \forall i \\ 1<||\theta||_n<n}} \frac{1}{||\theta||_n^2}\, d\theta = \frac{2^p}{(\log n)^2}\, I_n^0 (\text{say}) \tag{2.12}$$

On the first quadrant, namely, $\{\underset{\sim}{\theta}: \theta_i > 0 \text{ for every } i\}$, transform first to variables $y_1 = \theta_1$, $y_2 = \theta_2$, $y_i = \theta_i^{\alpha_n}$ for $i \geq 3$. Hence,

$$I_n^0 = \int_{\substack{1<y_i<n \\ y_i>0,\ \forall i}} \frac{1}{\left(\sum_{i=1}^{p} y_i\right)^2} \frac{1}{\alpha_n^{p-2}} \prod_{i=3}^{p} y_i^{\frac{1}{\alpha_n}-1}\, d\underset{\sim}{y} \tag{2.13}$$

Next transform to variables $t_1, t_2 \ldots t_p$ with $t_i = y_i + \ldots + y_p$, $1 \leq i \leq p$. Note $t_i + 1 \leq t_i$ for every i. Thus, from (2.13),

$$I_n^0 = \frac{1}{\alpha_n^{p-2}} \int_{\substack{1< t_1<n \\ 0<t_i+1\leq t_i\ \forall i}} \frac{1}{t_1^2} \prod_{i=3}^{p} (t_i - t_{i+1})^{\frac{1}{\alpha_n}-1}\, d\underset{\sim}{t}$$

$$= \frac{1}{\alpha_n^{p-2}} \frac{\alpha_n^2}{(p-2)^2} \frac{\prod_{i=1}^{p-3} B\left(\frac{1}{\alpha_n}, \frac{i}{\alpha_n}\right)\left(n^{\frac{p-2}{\alpha_n}} - 1\right)}{\left(1 + \frac{p-2}{\alpha_n}\right)} \tag{2.14}$$

Now for any fixed r,

$$B\left(\frac{1}{\alpha_n}, \frac{r}{\alpha_n}\right) = \int_0^{\varepsilon} x^{\frac{1}{\alpha_n}-1} (1-x)^{\frac{r}{\alpha_n}-1} dx + \int_{\varepsilon}^{1-\varepsilon} x^{\frac{1}{\alpha_n}-1} (1-x)^{\frac{r}{\alpha_n}-1} dx +$$

$$\int_{1-\varepsilon}^{1} x^{\frac{1}{\alpha_n}-1} (1-x)^{\frac{r}{\alpha_n}-1} dx = O(\alpha_n) \text{ if } \alpha_n \longrightarrow \infty,$$ since the middle term

behaves like a constant for a fixed $\varepsilon > 0$ and the first and the second terms are $O(\alpha_n)$.

Hence, by (2.12) and (2.14),

$$I_n = O\left[\frac{\alpha_n}{(\log n)^2} \quad (n^{\frac{p-2}{\alpha_n}} - 1)\right] \tag{2.15}$$

Now choosing $\alpha_n = \frac{(p-2)\log n}{\log\log n}$,

$$\frac{\alpha_n \cdot n^{\frac{p-2}{\alpha_n}}}{(\log n)^2} = \frac{\alpha_n}{\log n} \longrightarrow 0 \text{ and } \frac{\alpha_n}{(\log n)^2} \longrightarrow 0 \text{ as } n \longrightarrow \infty .$$

Hence $I_n \longrightarrow 0$ as $n \longrightarrow \infty$, verifying (iv) of Lemma 2.1.

This proves Theorem 1.

Remarks

1. The choice of α_n is clearly not unique. Other choices of α_n can also work.

2. If L had k rows, then defining $||\theta||_n = |\theta_1|+|\theta_2|+\ldots+|\theta_k|+\sum_{i=k+1}^{p}|\theta_i|^{\alpha_n}$

   and h_n as before, it is seen as in (2.15) that

   $$I_n = O\left[\frac{n^{k-2+\frac{p-k}{\alpha_n}} - 1}{(\log n)^2(k-2+\frac{p-k}{\alpha_n})}\right].$$ Clearly, if $k \geq 3$, then for no positive sequence α_n, the quantity in brackets can go to zero. On the other hand, if we allow α_n to be negative, clearly $||\theta||_n$ can be small even when $\sum_{i=1}^{p}\theta_i^2$ is large and thus h_n's cease to have compact supports. Thus the above analysis works only up to k = 2 as it should because for $k \geq 3$ admissibility usually is not true.

3. For the case k=1, one can choose $\alpha_n \equiv p-1$ and it can be seen that with this choice of α_n, $I_n = 0(\frac{1}{\log n}) \longrightarrow 0$ as $n \longrightarrow \infty$. However, for k=2, it is necessary to take $\alpha_n \longrightarrow \infty$. Heuristically speaking, this means as the dimensionality of the problem increases, flatter priors are needed to prove admissibility.

4. The sequence of priors has formal similarity to the sequence used in Berger (1976a); however, Berger's (1976a) sub-additive norm has to be suitably modified to treat a changed expression for the Bayes risk difference Δ_n. It is important to remember that the fastest convergence to zero of I_n will usually be achieved by taking h_n as the minimizing solution of an appropriate variational problem. See the discussion on page 979 of Brown (1979).

5. In the above analysis g was taken as 1. It is natural to ask for what other choices of g, the same sequence of h_n's may work. For k = 1, if

$$g(\theta) \leq C\, t_1^{\frac{\alpha-(p-1)}{\alpha}} \text{ for some } \alpha > 0, \text{ where } t_1 = |\theta_1| + \sum_{i=2}^{p} |\theta_i|^{\alpha},$$

it is still true that $I_n = 0(\frac{1}{\log n})$ and thus converges to zero. Clearly, for priors with bounded densities $\alpha = p-1$ will work. However, for admissibility one also needs $\int \frac{(\frac{d}{d\theta_1} g(\theta))^2}{g(\theta)} d\theta < \infty$. This will be true if in addition to $g(\theta) \leq c \cdot t_1^{1-\frac{p-1}{\alpha}}$, $\frac{\frac{d}{d\theta_1} g(\theta)}{g(\theta)} = 0(t_1^{-1-\varepsilon})$ for some $\varepsilon > 0$. Of course, both these need to be true only for large $\underset{\sim}{\theta}$.

6. Continuing with other choices of g which can be accommodated also when $k = 2$, a special class of priors which is of interest in many problems (and especially the normal problem) is the class of spherically symmetric priors. If, for example, $g(\theta) = \phi(||\theta||)$ where ϕ is bounded and a.e. differentiable, condition (1) in (2.9) holds with h_n's as in Theorem 2 up to $k = 2$. Also, by straightforward calculations,

$$\left(\frac{d}{d\theta_i} g(\theta)\right)^2 \leq \text{constant} \times \phi'^2(||\theta||) \text{ for } i \geq 1.$$

Hence, condition (2) in (2.9) holds whenever $\int_0^\infty \frac{\phi'^2(x)x^{p-1}}{\phi(x)} dx < \infty$; as an example, take $\phi(x) = \frac{1}{1+x^{2r}}$ where $r > 0$ (for $r = 0$, one has the uniform prior; these priors were also considered in Berger (1976a)). It is easy to check that with this choice of ϕ, $\int_0^\infty \frac{\phi'^2(x)x^{p-1}}{\phi(x)} dx < \infty$ whenever $2-4r < p < 2r+2$, and hence the corresponding generalized Bayes estimate $\delta_g(x)$ is admissible for (μ_1, μ_2). Actually if $p < 2r$, g becomes a finite prior so that admissibility of δ_g needs this proof only for $2r \leq p < 2r+2$ (note that if $p \geq 2$, $p > 2-4r$ is anyway true for every $r > 0$).

7. In the normal case, the priors $g(\theta) = C \cdot ||\theta||^{2-p-\varepsilon}$, $\varepsilon > 0$, also have some interest. It is easy to check and is also mentioned in Brown and Hwang (1982) that for such g's, with h_n as theirs, $\int ||\nabla h_n(\theta)||^2 g(\theta)\, d\theta \longrightarrow 0$ and $\int \frac{||\nabla g(\theta)||^2}{g(\theta)} d\theta < \infty$ for every $p \geq 1$. Consequently, such priors generate admissible estimates of $L\nabla\psi(\theta)$ for any L (recall the discussion preceding the statement of Theorem 1).

8. More general weighted quadratic losses can perhaps be tackled for the case $k = 2$. But it seemed to us that in this problem of estimating linear combinations of the means, probably the only meaningful weights are constants. Of course, in view of Shinozaki (1975), a separate admissibility proof is not needed for constant weights.
9. For the special case $g \equiv 1$, an alternative admissibility proof for $k = 1$ and 2 can be written down using the multiparameter information inequality.

3. Inadmissibility results in several independent Gamma distributions. If the restriction $(H) = \mathbb{R}^p$ is relaxed to, say, $(H) = (0,\infty)^p$ (as in the gamma or the negative Binomial case), clearly h_n's need to be redefined because h_n's of the previous section are 1 for $\underset{\sim}{\theta}$ near 0, and thus such h_n's cannot have compact support if $(H) = (0,\infty)^p$. Keeping the gamma problem in mind, one might want to define h_n as in section 1, replacing θ_i by $\log \theta_i$. Usually such an approach works well because the gamma problem is similar to the normal problem via the logarithm (see page 576 in Brown (1980)). A crucial requirement for this similarity to hold is the special scale-structure of the gamma distribution. However, if one wants to estimate a linear combination $\Sigma\gamma_i\theta_i^{-1}$ by a linear estimate $\Sigma c_i x_i$, this structure is lost unless only one γ_i and the same c_i is non-zero. One will then anticipate that the superficial similarity between the gamma and the normal problems may be absent in this problem. In what follows, we shall show this indeed is the case. By using the much-too-well-known technique of constructing an improved estimator by solving an appropriate differential inequality on the sample space, we will now show the inadmissibility of many linear estimates even when we are estimating just a single linear combination of the gamma scale-parameters.

Theorem 2 Let $X_i \overset{ind.}{\sim} f(x_i|\theta_i) = \dfrac{e^{-x_i\theta_i} x_i^{\alpha_i - 1} \theta_i^{\alpha_i}}{\Gamma\alpha_i}$, $x_i > 0$, $\alpha_i > 0$.

Let $\gamma(\underset{\sim}{\theta}) = \sum_{i=1}^{p} \ell_i \theta_i^{-1}$, $\ell_i > 0$ for each i. Let $\delta_0(\underset{\sim}{x}) = \sum_{j=1}^{p} a_j x_j$ be a linear estimate such that $a_j < (>) \dfrac{\ell_j \sum_{j=1}^{p} \alpha_j}{\alpha_j (1 + \sum_{j=1}^{p} \alpha_j)}$, $j=1,2....,p$. Then $\delta_0(\underset{\sim}{X})$ is inadmissible for $\gamma(\underset{\sim}{\theta})$ under squared-error loss.

Proof: Let $\delta(\underset{\sim}{X}) = \delta_0(\underset{\sim}{X}) + h(\underset{\sim}{X})$ be a competing strategy. Then,

$$\Delta(\underset{\sim}{\theta}) = R(\theta,\delta) - R(\underset{\sim}{\theta},\delta_0)$$

$$= E_\theta\left[\delta_0(\underset{\sim}{X}) + h(\underset{\sim}{X}) - \sum_{i=1}^{p} \ell_i \theta_i^{-1}\right]^2 - E_\theta\left[\delta_0(\underset{\sim}{X}) - \sum_{i=1}^{p} \ell_i \theta_i^{-1}\right]^2$$

$$= E_\theta\left[h^2(\underset{\sim}{X}) + 2h(\underset{\sim}{X}) \sum_{j=1}^{p} a_j x_j - 2\sum_{i=1}^{p} \ell_i \theta_i^{-1} h(\underset{\sim}{X})\right] \qquad (3.1)$$

Under appropriate tail and integrability conditions on h, $E_\theta[\theta_i^{-1} h(\underset{\sim}{X})]$ $= E_\theta\left[X_i^{1-\alpha_i} g_i'(\underset{\sim}{X})\right]$, where $g_i(\underset{\sim}{x})$ is such that $x_i^{1-\alpha_i} \frac{\partial}{\partial x_i} g_i(x) = h(x)$

(see Berger (1980), page 549) (3.2)

Hence, from (3.1),

$$\Delta(\underset{\sim}{\theta}) = E_\theta\left[h^2(\underset{\sim}{X}) + 2h(\underset{\sim}{X}) \sum_{j=1}^{p} a_j x_j - 2\sum_{j=1}^{p} \ell_j X_j^{1-\alpha_j} g_j(\underset{\sim}{X})\right] = E_\theta[D(\underset{\sim}{X})] \text{ (Say) } (3.3).$$

It's clear that if function g_i and h can be chosen such that subject to (3.2) and the tail-conditions of Berger (1980), $D \leq 0$ for almost all $\underset{\sim}{X}$, then δ_0 will be proved inadmissible. It is also clear from (3.2) that the ultimate solutions g_i must be such that $x_i^{1-\alpha_i} \frac{\partial}{\partial x_i} g_i(x)$ is independent of i; we will call such g_i's "coordinate-consistent". We now provide coordinate-consistent solutions g_i to $D(X) \leq 0$.

$$\text{Let } g_i(\underset{\sim}{x}) = \frac{C}{\alpha_i} \, x_i^{\alpha_i} \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} , \quad i = 1,2,\ldots p \tag{3.4}$$

$$\text{Then } h(\underset{\sim}{x}) = x_i^{1-\alpha_i} \frac{\partial}{\partial x_i} g_i(\underset{\sim}{x})$$

$$= \frac{C(1+\Sigma\alpha_j)}{\Sigma\alpha_j} \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} \tag{3.5}$$

By hypothesis, $a_j < \dfrac{\ell_j \quad \Sigma\alpha_j}{\alpha_j(1+\Sigma\alpha_j)}$ for every j.

$$\text{Let } 0 < \varepsilon = \min_j \left\{ \frac{\ell_j}{\alpha_j} - \frac{a_j(1+\Sigma\alpha_j)}{\Sigma\alpha_j} \right\} \tag{3.6}$$

Then $2h(\underset{\sim}{x}) \, \Sigma a_j x_j - 2\Sigma \ell_j x_j^{1-\alpha_j} \, g_j(\underset{\sim}{x})$

$$= 2C \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} \cdot \left[\frac{1+\Sigma\alpha_j}{\Sigma\alpha_j} \sum_j a_j x_j - \sum_j \frac{\ell_j}{\alpha_j} x_j \right].$$

$$\leq -2c\varepsilon \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} \sum_j x_j \tag{3.7}$$

Now, from (3.3),

$$D(\underset{\sim}{x}) \leq \frac{C^2(1+\Sigma\alpha_j)^2}{(\Sigma\alpha_j)^2} \prod_j x_j^{\frac{2\alpha_j}{\Sigma\alpha_j}} - 2c\varepsilon \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} \cdot \sum_j x_j$$

$$= C\prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} \left[\frac{C(1+\Sigma\alpha_j)^2}{(\Sigma\alpha_j)^2} \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} - 2\varepsilon \sum_j x_j \right]$$

$$\leq C \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} \left[\frac{cp_0(1+\Sigma\alpha_j)^2}{(\Sigma\alpha_j)^2} \sum_j x_j - 2\varepsilon \sum_j x_j \right] \quad \left(p_0 = \max_j \frac{\alpha_j}{\Sigma\alpha_j}\right)$$

$$\leq 0, \text{ if } 0 < C < 2\varepsilon p_0^{-1}\left(\frac{\Sigma\alpha_j}{1+\Sigma\alpha_j}\right)^2 . \qquad (3.8)$$

If $a_j > \frac{\ell_j}{\alpha_j} \frac{\Sigma\alpha_j}{(1+\Sigma\alpha_j)}$ for every j, $D(\underset{\sim}{x}) \leq 0$ for every $\underset{\sim}{X}$ by choosing C negative and sufficiently close to zero. We now note that if g_i,h as defined in (3.4) and (3.5) above satisfy the tail-conditions in Berger (1980), then $\delta_0(\underset{\sim}{X}) = \sum_j a_j x_j$ will be proved inadmissible. We merely remark that it is easy to check that they do. This proves the theorem.

Some remarks are in order.

Remarks 1. If $\gamma(\underset{\sim}{\theta}) = \sum_j E_\theta(X_j)$, then $\ell_j = \alpha_j$. The best componentwise linear estimator for $E_\theta(X_j)$ (with respect to squared-error loss) is $a_j X_j$, where $a_j = \frac{\alpha_j}{1+\alpha_j}$; it is easy to verify $a_j < \frac{\Sigma\alpha_j}{1+\Sigma\alpha_j}$ for every j whenever $p > 1$. Hence $\sum_j \frac{\alpha_j X_j}{1+\alpha_j}$ is inadmissible for $\sum_j \alpha_j \theta_j^{-1}$ whenever $p > 1$.

2. Some controversy remains regarding whether $\sum_j \frac{\alpha_j x_j}{1+\alpha_j}$ is after all a "natural" linear estimate for $\sum_j \alpha_j \theta_j^{-1}$ or not. Let us, for simplicity, treat the case $\alpha_j \equiv \alpha$ for all j. It is easy to check there is no "best" linear estimate of $\sum_j \theta_j^{-1}$; this is because the risk-function of linear estimates involves both $\sum_j \theta_j^{-2}$ and $(\sum_j \theta_j^{-1})^2$. However, some linear estimates can be shown to be trivially inadmissible in that one can dominate them by other linear estimates. For example, if each $a_j = k$, it follows quite easily that $k \cdot \sum_j x_j$ is inadmissible for $\sum_j \theta_j^{-1}$ if $k < \frac{1}{\alpha+1}$ or $k > \frac{1}{\alpha+\frac{1}{p}}$. Roughly speaking, in the first case $k \cdot \sum_j x_j$ is too much of an under-estimate and by giving it a positive linear shift one can improve

upon it. So the only k's which pose non-obvious admissibility problems are $\frac{1}{\alpha+1} \leq k \leq \frac{1}{\alpha+\frac{1}{p}}$. From Theorem 2 above it follows if $k < \frac{1}{\alpha+\frac{1}{p}}$, again one has inadmissibility. However, the theorem does not say anything about $k = \frac{1}{\alpha+\frac{1}{p}}$ and it seems plausible it is an admissible choice. More generally, $\frac{\Sigma\alpha_j}{1+\Sigma\alpha_j} \cdot \sum_j x_j$ may be an admissible estimator of $\sum_j \alpha_j \theta_j^{-1}$. In view of this, one may rather call this a natural estimate of $\sum_j \alpha_j \theta_j^{-1}$ instead of $\sum_j \frac{\alpha_j}{1+\alpha_j} x_j$. Unfortunately, our attempts to prove its admissibility have not been successful.

3. If $\alpha_j = \alpha$ for every j, the improved estimate shifts by a multiple of the geometric mean. It turns out the improved estimate is Empirical Bayes; one can show this assuming a conjugate gamma prior for the θ_i's with an unknown scale-parameter r, and incorporating a data-based estimate of r. A more detailed discussion on this aspect in some related gamma problems appears in Das Gupta (1984).

4. Because of the form of the improved estimate, an exact analytical form of the risk-improvement $\Delta(\underset{\sim}{\theta})$ is easy to find. This is because, under quadratic loss, one requires only the first and the second moments of the estimates to find $\Delta(\underset{\sim}{\theta})$ and those of course are easily found. Thus, risk-simulations will not be necessary if one wants to evaluate $\Delta(\underset{\sim}{\theta})$.

5. In section 2 it was seen that if $(H) = \mathbb{R}^p$ the natural estimate of any two linear combinations remains admissible, while for $(H) = R_+^p$, inadmissibility seems to prevail even for one linear combination. A similar phenomenon was found to be true for the problem of estimating the whole mean-vector of independent gamma variables in an article of Berger (1980); it was shown that typically, in the

gamma problems, the critical dimension of inadmissibility is one less than that in problems with (H) = $\mathbb{R}^p$. Whether this similar inadmissibility behavior in the two problems can be related may be of some theoretical interest.

6. In Berger (1980), improved estimates of the vector of scale-parameters were obtained under losses $L_m = \sum_{i=1}^{p} \theta_i^m (\delta_i\theta_i - 1)^2$, for m = 0,±1, and 2. Letting $\underset{\sim}{\delta}^B_m$ denote the improved estimate under loss-function L_m, it is of interest to know whether $\sum_{i=1}^{p} \delta^B_{i,m}(\underset{\sim}{X})$ provides us with an admissible estimate of $\Sigma\theta_j^{-1}$. The estimates δ^B_m are, however, highly non-linear and it is difficult to find their risk or try to dominate them by solving differential inequalities. However, using a method first used by Hwang (1982), one can often conclude $\sum_{i=1}^{p} \delta^B{}_{i,m}(\underset{\sim}{X})$ is actually inadmissible. The basic lemma to be used in the proof is stated below; see Hwang (1982) or Das Gupta (1984) for details.

Lemma 3.1 Let $X = X_1,\ldots,X_p$ have an arbitrary multiparameter probability distribution. Let $\gamma(\underset{\sim}{\theta})$ be any parametric function and let $\underset{\sim}{\delta}_1(\underset{\sim}{X})$ and $\underset{\sim}{\delta}_2(\underset{\sim}{X})$ be two estimates of $\underset{\sim}{\gamma}(\underset{\sim}{\theta})$ such that $\underset{\sim}{\delta}_2$ dominates δ_1 under squared-error loss. If $d = \delta_2 - \delta_1$, then any other estimator $\underset{\sim}{\delta}$ of $\underset{\sim}{\gamma}$ is inadmissible if it satisfies $\underset{\sim}{d}\cdot\underset{\sim}{\delta} \leq \underset{\sim}{d}\cdot\underset{\sim}{\delta}_1$ for all $\underset{\sim}{x}$. For a scalar parametric function γ, this lemma roughly means if δ_1 is an under-estimate (over-estimate) of γ and is inadmissible, then a further under-estimate (over-estimate) is also inadmissible. This simple-looking lemma proved by Hwang (1982) is surprisingly powerful in establishing inadmissibility of a whole class of estimators. The notations used below are as in Das Gupta (1984).

Let $\gamma(\underset{\sim}{\theta}) = \sum_{j=1}^{p} \ell_j\theta_j^{-1}$ be a linear combination of the scale-parameters of p independent gamma variables $X_1, X_2, \ldots X_p$. Let A be a p-dimensional rectangle in R_+^p; let

$$\lambda^*(x) = \sum_j a_j x_j [1+\phi^*(x) I_A(x)]$$

$$\lambda(\underset{\sim}{x}) = \sum_j a_j x_j [1+\phi(x) I_A(x)] \qquad (3.9)$$

be two estimators of $\gamma(\underset{\sim}{\theta})$. The idea is to choose ϕ, ϕ^*, and A such that λ^* dominates λ. Lemma 3.1 then will imply inadmissibility of any δ satisfying $(\lambda^*-\lambda)\delta \leq (\lambda^*-\lambda)\lambda$; since $\lambda^*-\lambda = 0$ outside A, if inequality holds for $\underset{\sim}{x}$ in A, δ will be inadmissible. The main task is to actually show λ^* dominates λ; in order to do this, we break up $R(\underset{\sim}{\theta},\lambda^*) - R(\underset{\sim}{\theta},\lambda)$ as $E_\theta[\Delta(\phi^*(\underset{\sim}{x})) - \Delta(\phi(\underset{\sim}{x})]$ plus a negative quantity where Δ is a differential operator, and then get hold of ϕ,ϕ^* such that $\Delta\phi^*(\underset{\sim}{x}) < \Delta\phi(x)$. As is well known, the right place to search for ϕ,ϕ^* is in an appropriate class of estimators $\delta_0[1+\phi]$ which dominate $\delta_0(x) = \sum_{j=1}^{p} a_j x_j$.

By familiar computations,

$$R(\theta,\lambda^*) - R(\theta,\delta_0)$$

$$= E[\phi^{*2}(x)\ (\sum_j a_j x_j)^2\ I_A(x) + 2\phi^*(x)\ (\sum_j a_j x_j)^2\ I_A(x) - 2\sum_j \ell_j \theta_j^{-1*}\phi(x)\ (\sum_j a_j x_j) I_A(x)] \qquad (3.10)$$

By Lemma 2.1 and (2.2) of Das Gupta (1984), for a given function $h(\underset{\sim}{x})$,

$$E[\theta_i^{-1} h(x) I_A(x)] = E[g_i(x) x_i^{1-\alpha_i}\ I_A(x)] - E[\theta_i^{-1}\ g_i(x) x_i^{1-\alpha_i}\ \frac{\partial}{\partial x_i} I_A(x)],$$

where $\frac{\partial}{\partial x_i}\ g_i(x) = h(x) x_i^{\alpha_i - 1}$. (3.11)

For a function $g(x)$ of a single variable x, $a < x < b$, such that $\lim_{x\to b} g(x) f_\theta(x) = 0$, $E_\theta[g(x)\ \frac{d}{dx}\ I_{[a,M]}(x)]$ is defined as $-g(M) f_\theta(M)$, where $a < M \leq b$; in the multivariate situation when A is a product rectangle as above, partial derivatives of products of step-functions is defined as

$\frac{\partial}{\partial x_i} \prod_j g_j(x_j) = \frac{d}{dx_i} g_i(x_i). \prod_{j\neq i} g_j(x_j)$; for details see either Hwang (1982) or DasGupta (1984). Using (3.10), an analogous expression for $R(\theta,\lambda) - R(\theta,\delta_0)$, and (3.11), $R(\theta,\lambda^*) - R(\theta,\lambda)$

$$= \{R(\theta,\lambda^*) - R(\theta,\delta_0)\} - \{R(\theta,\lambda) - R(\theta,\delta_0)\}$$

$$= E\left[\{\Delta\phi^*(x) - \Delta\phi(x)\} I_A(x) + 2\sum_i \ell_i\theta_i^{-1} (g_i^*(x) - g_i(x))x_i^{1-\alpha_i} \frac{\partial}{\partial x_i} I_A(x)\right] \quad (3.12)$$

where $\frac{\partial}{\partial x_i} g_i^*(x) = \phi^*(x) (\sum_j a_jx_j) x_i^{\alpha_i-1}$

$$\frac{\partial}{\partial x_i} g_i(x) = \phi(x) (\sum_j a_jx_j) x_i^{\alpha_i-1}, \quad (3.12)$$

$$\Delta\phi^*(x) = \phi^{2*}(x) (\sum_j a_jx_j)^2 + 2\phi^*(x) (\sum_j a_jx_j)^2 - 2\sum_i \ell_i g_i^*(x)x_i^{1-\alpha_i},$$

and $\Delta\phi(x) = \phi^2(x) (\sum_j a_jx_j)^2 + 2\phi(x) (\sum_j a_jx_j)^2 - 2\sum_i \ell_i g_i(x)x_i^{1-\alpha_i}$ (3.14)

We now go into the question of choosing $\phi(\underset{\sim}{x})$, $\phi^*(\underset{\sim}{x})$ such that $R(\theta,\lambda^*) - R(\theta,\lambda)$ is negative for all θ. By Theorem 2 it is known that whenever $a_j < (>)$ $\frac{\ell_j \sum_j \alpha_j}{\alpha_j(1+\sum_j\alpha_j)}$ for every j, $\sum_j a_jx_j + C\cdot\prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}}$ dominates $\sum_j a_jx_j$ for suitable C. This suggests choosing

$$\phi(\underset{\sim}{x}) \sum_j a_jx_j = C \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} \quad (3.15)$$

If $g_i(x) = \frac{C \cdot \sum_j\alpha_j}{\alpha_i(1+\sum_j\alpha_j)} x_i^{\alpha_i} \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}}$, (3.13) is easily checked. Assume $\{a_j\}$ satisfy the inequalities mentioned above.

Letting $\varepsilon = \min_j\{\frac{\ell_j}{\alpha_j} \frac{\Sigma\alpha_j}{1+\Sigma\alpha_j} - a_j\} > 0$, and $p_j = \frac{\alpha_j}{\Sigma\alpha_j}$, by (3.14),

$$\Delta\phi(x) = C^2 \prod_j x_j^{p_j}\left[\prod_j x_j^{p_j} - \sum_j p_jx_j\right] + \prod_j x_j^{p_j} \sum_j x_j (C^2p_j-2C\varepsilon)$$

$$+ 2C\left[\prod_j x_j^{p_j} \sum_j \left(\varepsilon - \frac{\Sigma\alpha_j}{1+\Sigma\alpha_j} \cdot \frac{\ell_j}{\alpha_j} + a_j\right) x_j\right]$$

$$= C^2 \prod_j x_j^{p_j}\left[\prod_j x_j - \sum_j p_j x_j\right] + \prod_j x_j^{p_j} \sum_j x_j (C^2 p_j - C^2 p_0) + \prod_j x_j^{p_j} \sum_j x_j (C^2 p_0 - 2C\varepsilon)$$

$$+ 2C \prod_j x_j^{p_j} \sum_j \left(\varepsilon - \frac{\Sigma\alpha_j}{1+\Sigma\alpha_j} \frac{\ell_j}{\alpha_j} + a_j\right) x_j$$

$$= A+B+C+D \quad \text{(say)} \quad (\text{here } p_0 = \max_j p_j) \tag{3.16}$$

Clearly, A,C,D are ≤ 0 for $C \geq 0$ and B is minimized for every $\underset{\sim}{x}$ at $C = \frac{\varepsilon}{p_0}$.

Hence, defining ϕ^* as in (3.15) with $C = \frac{\varepsilon}{p_0}$, it follows $\Delta\phi^*(x) < \Delta\phi(x)$ whenever $0 < C < \frac{\varepsilon}{p_0}$. Going back to (3.12), the second term in the risk-difference,

$$E\left[\sum_i \ell_i \theta_i^{-1} (g_i^*(x) - g_i(x)) x_i^{1-\alpha_i} \frac{\partial}{\partial x_i} I_A(x)\right]$$

$$= E \sum_i \ell_i \theta_i^{-1} \frac{(\varepsilon p_0^{-1} - C)\, \Sigma\alpha_j}{\alpha_i(1+\Sigma\alpha_j)} x_i \left(\prod_j x_j^{p_j}\right) \frac{\alpha}{\alpha x_i} I_A(x) \quad .$$

≤ 0 if $\ell_i \geq 0$ and if A is of the form $(0,M)^p$ for some $M > 0$.

Consequently, (3.12) implies λ^* dominates λ. Lemma 3.1 now enables us to state the following Theorem.

<u>Theorem 3</u> For $\ell_i \geq 0$, let $\gamma(\theta) = \Sigma\ell_i\theta_i^{-1}$ be estimated under squared-error loss. Let $\delta_0(X) = \sum_j a_j x_j$ be any linear estimate such that $a_j < \frac{\ell_j \; \Sigma\alpha_j}{\alpha_j(1+\Sigma\alpha_j)}$ for every $j \geq 1$. Let $\varepsilon = \min_j \left\{\frac{\ell_j \; \Sigma\alpha_j}{\alpha_j(1+\Sigma\alpha_j)} - a_j\right\} > 0$. If $\delta(X)$ is any other estimate such that for some $M > 0$ and some $0 < C \leq \frac{\varepsilon}{p_0}$,

$$\delta(x) \leq \sum_j a_j x_j + C \cdot \prod_j x_j^{\frac{\alpha_j}{\Sigma\alpha_j}} \qquad \text{for } x\varepsilon(0,M)^p,$$

then δ must be inadmissible.

Remarks: 1. A similar result will be true if the hypothesis of the above theorem holds for $\underset{\sim}{x}$ near ∞ for some $\{a_j\}$ with $a_j > \frac{\ell_j \Sigma\alpha_j}{\alpha_j(1+\Sigma\alpha_j)}$ for every j and some $0 > C > \frac{-\varepsilon}{p_0}$, where $\varepsilon = \min_j \left\{ a_j - \frac{\ell_j \ \Sigma\alpha_j}{\alpha_j(1+\Sigma\alpha_j)} \right\}$.

2. Theorem 3 shows the inadmissibility of some of the alternative estimates proposed in Theorem 2. There does not exist any C which minimizes $D(\underset{\sim}{x})$ for every $\underset{\sim}{x}$ in Theorem 2. However, with $\ell_j = \alpha_j$, the bound on $D(\underset{\sim}{x})$ in (3.8) is minimized at $C = \frac{p_0^{-1} \ \Sigma\alpha_j}{1+\Sigma\alpha_j} \min_j \left\{ \frac{\Sigma\alpha_j}{1+\Sigma\alpha_j} - a_j \right\}$.

Thus, in some sense, $\Sigma a_j x_j + p_0^{-1} \min_j \left\{ \frac{\Sigma\alpha_j}{1+\Sigma\alpha_j} - a_j \right\} \cdot \ \Pi x_j^{\frac{\alpha_j}{\Sigma\alpha_j}}$ may be regarded as a natural alternative to $\Sigma a_j x_j$. Theorem 3 does not show this estimator to be inadmissible as the Theorem is true only for C strictly smaller than $\frac{\varepsilon}{p_0}$ there. It may actually be an admissible estimator.

3. For the loss $L_{-1} = \sum_{i=1}^{p} \theta_i^{-1} (1-\delta_i\theta_i)^2$, the improved estimate of $(\theta_1^{-1},\theta_2^{-1},\ldots,\theta_p^{-1})$ in Berger (1980) was $\delta_i(x) = \frac{x_i}{\alpha_i+1} + \frac{C(\alpha_i+1)}{b+\sum_j(\alpha_j+1)^3 x_j^{-1}}$, $0 < C < 2(p-1)$, $b > 0$.

The natural choices of b and c are 0 and p-1 respectively. For the case $\alpha_i \equiv \alpha$, $\sum_{i=1}^{p} \delta_i(x) = \frac{1}{\alpha+1}\Sigma x_i + \frac{p(p-1)}{(\alpha+1)^2 \Sigma x_j^{-1}} \leq \frac{1}{\alpha+1} \Sigma x_i + \frac{p-1}{(\alpha+1)^2} \prod_j x_j^{\frac{1}{p}}$. If all α_j's are equal, then $\varepsilon = \frac{p-1}{(p\alpha+1)(\alpha+1)}$ and $p_0 = \frac{1}{p}$. The hypothesis of Theorem 3 holds with $a_j \equiv \frac{1}{\alpha+1}$, $\ell_j \equiv 1$ for every $p > 1$ and every $\alpha > 0$. Consequently, $\sum_{j=1}^{p} \delta_j(x)$ is an inadmissible estimator of $\sum_{j=1}^{p} \theta_j^{-1}$.

The componentwise sum of Berger's (1980) improved estimates corresponding to the loss $L_1 = \sum_{i=1}^{p} \theta_i(1-\delta_i\theta_i)^2$ is also easily seen to be inadmissible in our context. This is because the improved estimates now shrink the natural estimate rather than expand it (see page 559 in Berger (1980)), and hence the hypothesis of our Theorem 3 is trivially satisfied. A similar inadmissibility result can also be proved, with some effort, by taking the improved estimates corresponding to his case 4 (usual squared-error loss). We remark that unlike in the previous two cases, now inadmissibility can be proved by using Theorem 3 above only for some range of values of the α_j's. Finally, Theorem 3 cannot be used to establish inadmissibility of the sum of the coordinate-wise estimators that correspond to the invariant quadratic loss (case 2 in Berger (1980)). However, it is apparent that Theorem 3 can be used to handle quite a few non-linear estimates which arise naturally from the work of Berger (1980).

4. Generalization of a theorem of Brown and Hwang   Brown and Hwang (1982) gave a unified proof of admissibility of many generalized Bayes estimators of the vector of means in the exponential family. The most important thing in the main theorem in that paper was that the same sequence of multipliers $h_n(\theta)$ was shown to work in a Blyth-type admissibility proof in a wide variety of problems with different generalized priors g. The importance of such sequence of h_n's in admissibility problems was earlier emphasized in another context in Berger (1976a). We will show in this section that the technique of Brown-Hwang can

be used to prove admissibility of generalized Bayes estimators of parametric functions which are, in some sense, in a close vicinity of the mean-vector. The technique used is Brown-Hwang's; also, with some obvious modifications, the proof for the one-parameter case naturally extends to the multiparameter case. Because of these reasons, we will give only a short sketch of the proof in one-dimension and then mention some applications. The notations are as in section 2.

Let $\gamma(\theta)$ be any smooth parametric function. Fix α, c and a non-negative function $g(\theta)$. Define $\Pi(\theta) = g(\theta).\ e^{\psi(\theta) - c\int^{\theta}\gamma(t)dt + \alpha\theta}$, where $\int^{\theta}\gamma(t)\,dt$ is to be interpreted as a primitive of γ. It will be implicitly assumed in the following theorem that $\lim_{\theta\to\underline{\theta},\overline{\theta}} f_\theta(x)\pi(\theta) = 0$ and $I_x(\frac{|g'|}{g}\cdot\pi)$ and $I_x(\pi) < \infty$ for every x.

We have the following admissibility theorem (a version of Theorem 4 below has been independently obtained by Ghosh and Meeden (1983)).

Theorem 4 The generalized Bayes estimate $\delta_\Pi(x)$ of $\gamma(\theta)$ against the prior $\Pi(\theta)$ is admissible under squared-error loss if

(i) There exist functions h_n satisfying (i), (ii), and (iii) of Lemma (2.1) such that $\int\{h_n'(\theta)\}^2\ \Pi(\theta)\,d\theta \longrightarrow 0$ as $n \longrightarrow \infty$, and

(ii) $\int_{\underline{\theta}}^{\overline{\theta}}\left(\frac{g'(\theta)}{g(\theta)}\right)^2\ \Pi(\theta)\,d\theta < \infty$.

Proof: Since the loss is squared-error, by doing an integration by parts,

$$\delta_\Pi(x) = \frac{\int e^{\theta(x+\alpha)}g(\theta)\ \gamma(\theta)e^{-c\int^{\theta}\gamma(t)dt}d\theta}{\int e^{\theta x - \psi(\theta)}\ \Pi(\theta)d\theta}$$

$$= \frac{x+\alpha}{c} + \frac{1}{c}\ \frac{I_x\left[g'\ e^{\xi}\right]}{I_x\left[g\ e^{\xi}\right]}\ , \qquad (4.1)$$

where $\xi(\theta) = \psi(\theta) - c\int^{\theta}\gamma(t)dt + \alpha\theta$

Defining $g_n(\theta) = g(\theta)h_n^2(\theta)$ and $\pi_n(\theta) = h_n^2\,\pi(\theta) = g_n(\theta)\, e^{\xi(\theta)}$, one has,

$$\Delta_n = \int(\delta_\pi(x) - \delta_{\pi_n}(x))^2\, I_x(\pi_n)\, d\mu(x)$$

$$= \frac{1}{c^2}\int\left[\frac{I_x(g'e^{\xi})}{I_x(g\, e^{\xi})} - \frac{I_x(g_n'\, e^{\xi})}{I_x(g_n\, e^{\xi})}\right]^2 I_x(\pi_n)d\mu(x)$$

$$\leq \frac{2}{c^2}\int\left[\frac{I_x(g'\, e^{\xi})}{I_x(g\, e^{\xi})} - \frac{I_x(h_n^2\, g'\, e^{\xi})}{I_x(h_n^2\, g\, e^{\xi})}\right]^2 I_x(\pi_n)\, d\mu(x)$$

$$+ \frac{8}{c^2}\int\left[\frac{I_x(h_n\, h_n'\, g\, e^{\xi})}{I_x(h_n^2\, g\, e^{\xi})}\right]^2 I_x(\pi_n)\, d\mu(x)$$

$$= A_n + B_n \qquad \text{(say)} \tag{4.2}$$

As usual, the task rests in finding appropriate h_n's such that A_n, $B_n \to 0$.

Applying Cauchy-Schwartz's inequality on B_n,

$B_n \leq$ constant $\{\int h_n'(\theta)\}^2\, g(\theta)\, e^{\xi(\theta)}\, d\theta \to 0$ if hypothesis (i) in Theorem 4 holds.

In order to show $A_n \to 0$ under hypothesis (ii), as usual we apply the Dominated Convergence theorem. Since it is obvious that the integran in A_n converges to zero pointwise, it will suffice to show that integrands are uniformly (in n) bounded by the integrable (wrt $d\mu$) function $I_x\left\{\left(\frac{g'^2}{g}\right)\, \pi\right\}$.

Again, this follows, as in Brown-Hwang, by noting that the integrand is

$$\left(I_x\left\{g_n\, e^{\xi}\left[\frac{I_x(g'\, e^{\xi})}{I_x(g\, e^{\xi})} - \frac{g'}{g}\right]\right\}\right)^2 \quad I_x^{-1}\,(\pi_n)$$

$$\leq I_x\left\{g\, e^{\xi}\left[\frac{I_x(g'\, e^{\xi})}{I_x(g\, e^{\xi})} - \frac{g'}{g}\right]^2\right\}$$

$$\leq I_x\left((\tfrac{g'}{g})^2\, \pi\right) \tag{4.3}$$

This completes the proof of the theorem.

Remarks 1. The proof goes through without any problems in the multiparameter situation. Suppose $\nabla\gamma(\underset{\sim}{\theta})$ is to be estimated under a quadratic loss $||\underset{\sim}{a} - \nabla\gamma(\theta)||^2$. With $\Pi(\theta) = g(\theta)\, e^{\xi(\theta)}$, where $\xi(\theta) = \psi(\theta) - c\gamma(\theta) + \underset{\sim}{\alpha}'\underset{\sim}{\theta}$, the hypotheses of the theorem should be changed to

(i) there exist h_n's such that $\int ||\nabla h_n(\theta)||^2\ \Pi(\theta)\, d\theta \longrightarrow 0$, and

(ii) $\int_{(H)} ||\frac{\nabla g(\theta)}{g(\theta)}||^2\ \Pi(\theta)\, d\theta < \infty$.

We also mention that no problems arise in writing down appropriate sufficient conditions for more general weighted quadratic losses as in Brown-Hwang (1982).

2. The result of Ghosh and Meeden (1977) follows as a corollary. First note, in the context of Ghosh and Meeden (1977), $g(\theta) \equiv 1$ so that hypothesis (ii) causes no problem. Also, if $\int \Pi^{-1}(\theta)\, d\theta = \infty$ on both the tails, where $\Pi(\theta) = e^{\psi(\theta) - c\int\gamma(t)dt + \alpha\theta}$ (see Ghosh and Meeden (1977)), indeed appropriate h_n's exist so that hypothesis (i) also holds. To find such h_n's, one merely needs to write down the Euler equation for the variational problem of minimizing $\int \{u'(\theta)\}^2\ \Pi(\theta)\, d\theta$, subject to $u(\theta) = 0$ for $\theta \notin (a_n, b_n)$ and $u(\theta_0) = 1$, where $a_n \downarrow \underline{\theta}$, $b_n \uparrow \overline{\theta}$, and θ_0 is an interior-point. It is easily seen the solution, if called h_n, is given by

$$h_n(\underset{\sim}{\theta}) = \frac{\int_{\theta}^{b_n} \Pi^{-1}(t)\, dt}{\int_{\theta_0}^{b_n} \Pi^{-1}(t)\, dt} \qquad \text{if } \theta_0 \le \theta \le b_n$$

$$= \frac{\int_{a_n}^{\theta} \Pi^{-1}(t)\, dt}{\int_{a_n}^{\theta_0} \Pi^{-1}(t)\, dt} \qquad \text{if } a_n \leq \theta \leq \theta_0 \tag{4.4}$$

The h_n's clearly have compact support; since a_n, b_n are monotone sequences, condition (iii) of Lemma 2.1 causes no problem; also, as $\int \Pi^{-1}(t)\, dt \longrightarrow \infty$ in both the tails, quite clearly $h_n(\theta) \longrightarrow 1$ as $n \longrightarrow \infty$ (after a certain stage $\theta \in (a_n, b_n)$). Thus, under the hypothesis of Ghosh and Meeden (1977), $\delta_\Pi(X) = \frac{X+\alpha}{c}$ is admissible for $\gamma(\theta)$.

3. Let $(H) = \mathbb{R}^p$ and consider the uniform prior $\Pi \equiv 1$. Then with h_n as in Brown and Hwang (1982), hypothesis (i) in the multiparameter case holds up to p = 2. Hypothesis (ii) is equivalent to

$$\int || c\nabla\gamma - \nabla\psi - \underset{\sim}{\alpha} ||^2 \;\; d\theta < \infty \tag{4.5}$$

If $\nabla\gamma = a\nabla\psi + \underset{\sim}{b}$, clearly with appropriate c and $\underset{\sim}{\alpha}$, the integrand is zero and (4.5) holds, meaning that the uniform prior generalized Bayes estimate is admissible up to p = 2. Of course, since the loss is squared-error, this fact follows from Brown and Hwang (1982) and a separate proof is not necessary. But (4.5) says that if for suitable c, $\underset{\sim}{\alpha}$, the integral above is finite, then admissibility still holds. Thus, roughly speaking, if the parametric function is in a close vicinity of the mean-vector, the uniform prior Bayes rule remains admissible up to two dimension. It will be interesting to know if violation of a condition essentially like (4.5) will actually imply inadmissibility. In the special Normal distribution, examination of various parametric functions which are not in the long run essentially linear in θ in the sense of (4.5), has led us to believe this should be the case at least in that distribution.

4. Just as $\pi(\theta) \equiv 1$ is the non-informative prior in the normal problem, in the gamma problem with scale-parameters $\theta_1^{-1}, \theta_2^{-2}, \ldots, \theta_p^{-1}$, the prior $\Pi(\theta) = \prod_{i=1}^{p} \theta_i$ is the non-informative prior and is of some special interest. For p = 1, it is seen as in remark 3, that if $\int_0^\infty (c\gamma(\theta) - \frac{k}{\theta} - \alpha)^2 \, \theta d\theta < \infty$ for suitable c and α (k being the known shape parameter), then the Bayes solution against the non-informative prior remains admissible. We remark that it is easy to write down the analogous integrability condition for a general p and general weighted quadratic losses.

5. It is clear that if $\gamma(\theta)$ is linear in the natural parameter θ, then $\int_0^\infty (c\gamma(\theta) - \frac{k}{\theta} - \alpha)^2 \, \theta \, d\theta$ cannot be finite. This of course corresponds to the fact that the non-informative prior generalized Bayes estimate of the natural parameter is not admissible in the gamma problem. However, the prior which results in the admissible estimate $\frac{K-2}{X}$ cannot be handled by our Theorem 4, i.e., Theorem 4 fails to show $\frac{K-2}{X}$ is admissible for θ. Whether this means that Brown-Hwang's technique to carry out Blyth's theorem will work only for parametric functions "near E(X)" is not clear to us. However , it seems clear that built-in in Brown-Hwang's technique is the implicit assumption that the estimate $\delta_\Pi(x)$ is $\frac{x+\alpha}{c}$ plus a small-enough perturbation, and if $\gamma(\theta)$ is nowhere near the mean, there is no reason why the dominant term in $\delta_\Pi(x)$ should be linear in x; it is thus not surprising that such $\gamma(\theta)$'s cannot be handled by Theorem 4 of this section.

5. Final remarks. It will be interesting to obtain a characterization of linear admissible estimates $\Sigma\gamma_i x_i$ for a linear combination $\Sigma\phi_i\mu_i$ of the means. The sequence of priors in section 2 of this paper is best suited for the UMVUE and for other estimates, a different argument, possibly using Brown's (1979)

heuristics may be necessary. Also, it should be interesting to know if inadmissibility results similar to those of section 3 can be proved for the Negative Binomial distribution, another distribution with $(H) = R_+^p$. Presumably, the risk-identities in Hwang (1982) should prove useful. Also, estimating a linear combination of the means under quadratic loss is pretty much like estimating all the means under a linear combination of quadratic losses (both problems essentially combine all the means in an additive fashion); one expects similarity in admissibility patterns in the two problems. This is taken up in the multivariate gamma situation in Das Gupta (1984).

Acknowledgment. It is a pleasure to thank Malay Ghosh for carefully reading a large part of the manuscript and making very helpful suggestions. The first author is also grateful to James Berger and Leon J. Gleser for very helpful discussions on the materials of sections 2 and 3, and to Marty Meyer for her invaluable support and encouragement at every stage of this work.

References

Berger, James O. (1976a). Admissibility results for generalized Bayes estimators of coordinates of a location vector. Ann. Stat. 4 (334-356).

Berger, James O. (1976b). Inadmissibility results for generalized Bayes estimators of coordinates of a location vector. Ann Stat. 4 (302-333).

Berger, James O. (1976c). Inadmissibility results for the best invariant estimator of two coordinates of a location vector. Ann. Stat. 4 (1065-1076).

Berger, James O. (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of Gamma scale parameters. Ann. Stat. 8 (545-571).

Berger, James O. (1982a). Bayesian robustness and Stein effect. JASA 77 (358-368)

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

Berger, James (1982c). Estimation in continuous exponential families: Bayesian estimation subject to risk restrictions and inadmissibility results. Statistical Decision Theory and Related Topics III (vol. 2). Academic Press, New York.

Berger, James, and DasGupta, Anirban (1985). Estimation of multiple gamma scale-parameters: Bayes estimation subject to uniform domination. Technical Report #85-13, Department of Statistics, Purdue University (Submitted for publication).

Blyth, C. R. (1951). On minimax statistical decision procedures and their admissibility. Ann. Math. Statist. 22 (22-42).

Blyth, Colin R. (1974). Necessary and sufficient conditions for inequalities of Cramer-Rao type. Ann. Stat. 2 (464-473).

Brown, L. D. (1979). A heuristic method for determining admissibility of estimators - with applications. Ann. Stat. 7 (961-994).

Brown, L. D. (1980). Examples of Berger's phenomenon in the estimation of independent normal means. Ann. Stat. 8 (572-585).

Brown, Lawrence D. and Hwang, Jiunn Tzon (1982). A unified admissibility proof. Statistical Decision Theory and related topics III, vol. 1 (205-230).

Cohen, Arthur (1965). Estimates of linear conbinations of the parameters in the mean-vector of a multivariate distribution. Ann. Math. Statist. 36 (78-87).

Das Gupta, Anirban (1984). Admissibility in the Gamma distribution: two examples. Sankhyā Ser. A, 46 (395-407).

Das Gupta, Anirban (1984). Simultaneous estimation of Gamma scale and natural parameters under weighted quadratic losses. (To appear in Ann. Stat.)

Ghosh, M., Hwang, J. T., and Tsui, K. W. (1983). Construction of improved estimators in multiparameter estimation for continuous exponential families. Jour. Mult. Analysis. 14 (212-220).

Ghosh, Malay, and Meeden, Glen (1977). Admissibility of linear estimators in the one parameter exponential family. Ann. Stat. 5 (772-778).

Ghosh, M., and Meeden, G. (1983). Private communication.

Guttman, Sam (1982). Stein's paradox is impossible in problems with finite sample space. Ann. Stat. 10 (1017-1020).

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

Hwang, J. T. (1982). Improving upon standard estimators in discrete exponential families with applications to Poisson and negative binomial cases. Ann. Stat. 10 (857-867).

Hwang, J. T. (1982). Certain bounds on the class of admissible estimators in continuous exponential families. Statistical Decision Theory and related topics III, volume 2 (15-30).

Lehmann, E. L. (1983). Theory of point estimation. John Wiley and Sons.

Makani, S. M. (1972). Admissibility of linear functions for estimating sums and differences of exponential parameters. Ph.D. Thesis, University of California, Berkeley.

Makani, S. M. (1977). A paradox in admissibility. Ann. Statist. 5 (544-546).

Shinozaki, N. (1975). A study of generalized inverse of matrices and estimation with quadratic loss. Ph.D. Thesis, Keio University.

Stein, C. (1955). A necessary and sufficient condition for admissibility. Ann. Math. Stat. 26 (518-522).

Stein, C. (1956). 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.

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