A General Theorem on Decision Theory for Nonnegative Functionals: With Applications

by

Anirban DasGupta
Purdue University

Technical Report #86-25

Department of Statistics
Purdue University

July 1986
Revised April 1988
Revised August 1988

AMS 1980 subject classifications. Primary 62C15; Secondary 62F10, 62H12.

Key words: Inadmissibility, eigenvalues, scale-parameters, Wishart, elliptically symmetric, equivariant, Bayes, loss functions.

Research supported by NSF Grant DMS 8702620.

A General Theorem on Decision Theory for
Nonnegative Functionals: with Applications

by

Anirban DasGupta
Purdue University

ABSTRACT

A distribution-free inadmissibility theorem is proved for estimating under quadratic loss non-negative functionals that are allowed to depend on the unknown c.d.f. as well as the data. It follows from the theorem that, *subject to finiteness of the risk*, some natural estimators of the eigenvalues of many commonly occurring matrices in multivariate problems are inadmissible, typically from dimension 2, when samples are drawn from any multivariate elliptically symmetric distribution. As an example, if $X_1, \ldots, X_{k+1}$ are i.i.d. observations from such a distribution with dispersion matrix Σ, then the eigenvalues of $\frac{S}{k}$ are inadmissible for the eigenvalues of Σ if a certain condition holds where S is the sample sum of squares and products matrix. It also follows from the theorem that in the general scale-parameter family, the best equivariant estimator of the scale-parameters is inadmissible for $p \geq 2$, and some natural estimators of the losses of the best equivariant estimators are inadmissible, usually for $p \geq 2$. The theorem also has certain applications to the Ferguson family of distributions, the multivariate F distribution, and for unbiased estimators in some families of distributions.

1. Introduction.

Stein (1955) showed that the usual estimator of the mean of a p-variate normal distribution with identity covariance matrix is inadmissible under squared-error loss if $p \geq 3$. Since this monumental work, researchers in this general area have proved that the same inadmissibility phenomenon holds in an incredibly large number of problems. Brown (1966) proved the existence of the "Stein-effect" in general location problems under very general losses; Berger (1980) gave a general theory for the continuous exponential family, with applications in the gamma case, and Hwang (1982a) treated the same problem in the Discrete exponential family. The problem of simultaneously estimating multiple gamma scale-parameters was readdressed in DasGupta (1986a) and the following result was proved: for ordinary squared-error loss, the best equivariant estimate, say $\delta_0(X)$, of the scale-parameters is inadmissible for every $p \geq 2$, and a dominating estimator is $\delta_0(X) + c \cdot \left(\prod_j X_j\right)^{\frac{1}{p}} \cdot 1$, where $1' = (1, 1, \ldots, 1)$ and $c > 0$ is a suitable constant. Only later it was discovered that this *same* result holds for every $p \geq 2$ in the arbitrary scale-parameter family; see DasGupta (1984b). Interestingly, in all of these cases in which the idea of shifting by the geometric mean works very well for proving inadmissibility results, the parametric functions being estimated are always non-negative. The purpose of this paper is to prove a very simple theorem, motivated by Brown's (1979) heuristics, in the context of simultaneously estimating p non-negative functionals (which can depend on the unknown parameters *as well as* the random variables) under quadratic losses; roughly speaking, the theorem says that if $X = (X_1, \ldots, X_p)$ has a distribution $F \varepsilon \mathcal{F}$, where $\mathcal{F}$ is some family of distributions on $\mathbb{R}^p$, and if one wants to estimate non-negative functionals $\mu_1(X,F)$, $\mu_2(X,F), \ldots, \mu_p(X,F)$, then under a suitable technical condition (see Theorem 1 in the next section), an estimate $\delta(x) = \big(\delta_1(x), \delta_2(x), \ldots, \delta_p(x)\big)$ is inadmissible and a better estimator is $\delta(x) + c \cdot \lambda(x) \cdot 1$, where c is a suitable real number, and $\lambda(x)$ is a suitable size statistic, *not* necessarily the geometric mean of the δ_j''s. The examples in the next section show that a variety of size statistics, like $\big(\prod_j \delta_j(x)\big)^{\frac{1}{p}}$, $\sum_j \delta_j(x)$, $\min\big(\delta_1(x), \ldots, \delta_p(x)\big)$ etc., work in different problems. The critical dimension of inadmissibility is almost always 2, but sometimes is 3 (see examples in the next section). Besides the theoretically interesting facts that the theorem is nonparametric in nature and that it can handle estimation of functionals which depend on parameters as well as the data, the wide variety of problems to which the result applies is surprising. Some important problems in which the inadmissibility of natural estimators can be proved by directly using this theorem include the following: (i) estimation of the means in arbitrary scale-parameters families, (ii) estimation of the eigenvalues of the dispersion matrix Σ and the precision matrix Σ^{-1} in a general elliptically symmetric distribution, (iii) estimation of the eigenvalues of $\Sigma_1^{-1}\Sigma_2$ where Σ_i, $i = 1, 2$, are the dispersion matrices of two elliptically symmetric distributions, and independent random samples are available from the two populations, (iv) estimation of the eigenvalues of the matrix $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, which arises in multivariate regression, (v) estimation of loss-functions of the best equivariant estimate in the general scale-parameter family, (vi) estimation of the means in some members of the Ferguson family of distributions (see Ferguson (1962)), and (vii) estimation of the means in the multivariate F distribution. The theorem also gives some interesting inadmissibility results for unbiased estimates in certain cases. Note that problem (ii) arises in principal components analysis, and problem (iii) in MANOVA. Problem (ii) has received the attention of several authors, including Stein (1975). Problems (i), (v), (vi), and (vii) are interesting from the viewpoint of general decision theory. We believe that this theorem probably has potential applications in other problems involving estimation of nonnegative functionals not considered by us in this article.

In section 2, we give the main theorem and its applications and relate the main result to Brown's (1979) technique. In section 3 some concluding remarks about the amount of risk-improvement and the scope for choice of a dominating estimator are made. Before closing this section, we would like to point out that the results in this paper are valid *subject to finiteness of risk.*

2. The Main Result and Applications.

Theorem 1. Let $X = (X_1, X_2, \ldots, X_p)$ have a joint distribution F, where F belongs to some class $\mathcal{F}$ of distributions on $\mathbb{R}^p$. Let $\mu_i\ (X,F)$, $1 \le i \le p$, be nonnegative functionals and let $0 \le \delta_i(X)$, $1 \le i \le p$, be any estimators of $\mu_i\ (X,F)$. Let $\lambda(X_1, \ldots X_p)$ be any non-negative statistic. Define $A(F) = \dfrac{E\left\{\lambda \cdot \sum\limits_{i=1}^{p} \delta_i(x)\right\}}{E\left\{\lambda \cdot \sum\limits_{i=1}^{p} \mu_i(x,F)\right\}}$.

If either

(a) $\inf\limits_{F \varepsilon \mathcal{F}} A(F) = r > 1$,

or

(b) $\sup\limits_{F \varepsilon \mathcal{F}} A(F) = R < 1$ holds,

and

(c) $E\big(\lambda^2(\underset{\sim}{x})\big) \leq A_0 \cdot E\{\lambda(\underset{\sim}{x}) \cdot \Sigma\delta_i\} \; \forall \; F$ and some $A_0 > 0$,

then $\delta(X)$ is inadmissible for estimating $\mu(X, F)$ under sum of squared-error losses $\sum\limits_{i=1}^{p}(\delta_i - \mu_i)^2$.

Moreover, a dominating estimator is given as $\delta_1(X) = \delta(X) + c\lambda \cdot 1$, where c is a suitable negative number in case (a), and a suitable positive number in case (b) (see (2.4) below).

Remark 1: In a large number of examples to be considered in this section, there is enough homogeneity that $A(F)$ is actually a constant for all F. In such cases verification of the sufficient condition(s) of Theorem 1 will be somewhat simplified.

It is interesting that problems (i) through (vii) listed in the introduction all have a "pseudo scale structure" in them. (We will not attempt to formally define a pseudo-scale structure, but roughly speaking we mean that $\dfrac{\sum_i E\big(\delta_i(\underset{\sim}{x})\big)}{\sum_i E\mu_i(\underset{\sim}{x},F)}$ and $\dfrac{\sum_i E\big(\delta_i^2(\underset{\sim}{x})\big)}{\sum_i E\big(\mu_i^2(\underset{\sim}{x},F)\big)}$ behave like constants uniformly over $F \in \mathcal{F}$). Suppose now we consider the problem of estimating, under sum of squared error losses, p positive parameters $\theta_1, \theta_2, \ldots, \theta_p$ by estimates $aX_1, aX_2, \ldots, aX_p$ respectively, where a is a real constant. Let $\mu_i(\theta_i) = E(X_i)$, $V_i(\theta_i) = E(X_i - \theta_i)^2$, $r(\underset{\sim}{x}) = (a-1)\underset{\sim}{x}$. Let's assume that $\sum \mu_i(\theta_i) = \sum \theta_i$. Brown's (1979) heuristic analysis then suggests that we should expect $\underset{\sim}{x} + r(\underset{\sim}{x})$ to be inadmissible if we can find a function $\lambda : \mathbb{R}^p \to \mathbb{R}^1$ satisfying the non-linear differential inequality

$$-2(a-1)\lambda(\underset{\sim}{\theta}) \sum \theta_i - 2\sum V_i(\theta_i)\frac{\partial}{\partial\theta_i}\lambda(\underset{\sim}{\theta}) - p\lambda^2(\underset{\sim}{\theta}) \geq 0 \; \forall \; \underset{\sim}{\theta}$$

$$\Leftrightarrow p\lambda^2(\underset{\sim}{\theta}) + 2(a-1)\lambda(\underset{\sim}{\theta}) \sum \theta_i + 2\sum V_i(\theta_i)\frac{\partial}{\partial\theta_i}\lambda(\underset{\sim}{\theta}) \leq 0 \; \forall \; \underset{\sim}{\theta}. \tag{2.1}$$

(this is just inequality (6) on page 964 in Brown (1979).) Suppose now $\mu_1 \le \frac{V_i(\theta_i)}{\theta_i^2} \le \mu_2 < \infty$ (in some problems it is easier to check that $\mu_1 \le \frac{\sum_i V_i(\theta_i)}{\sum_i \theta_i^2} \le \mu_2 < \infty$); if one now lets $\lambda(\underset{\sim}{\theta}) = c\big(\prod_j \theta_j\big)^{\frac{1}{p}}$, then (2.1) reduces to

$$pc^2(\prod_j \theta_j)^{\frac{2}{p}} + 2c(a-1)(\prod_j \theta_j)^{\frac{1}{p}} \sum_i \theta_i + 2\frac{c}{p}(\prod_j \theta_j)^{\frac{1}{p}} \sum_i \frac{V_i(\theta_i)}{\theta_i} \le 0 \tag{2.2}$$

Using the facts that $\frac{V_i(\theta_i)}{\theta_i^2} \le \mu_2$ and $\prod_j (\theta_j)^{\frac{1}{p}} \le \frac{\sum \theta_j}{p}$, it easily follows that (2.2) will hold if there exists a $c > 0$ such that $c + 2(a-1) + 2\frac{\mu_2}{p} < 0$; if $a < 1$, then clearly such a $c > 0$ exists for large p because $\frac{\mu_2}{p} \to 0$ as $p \to \infty$. Brown's (1979) analysis then strongly indicates that in problems with a pseudo-scale structure, $a\underset{\sim}{X}$ should be uniformly dominated by $a\underset{\sim}{X} + c(\prod_j X_j)^{\frac{1}{p}}$ for large enough p, where $c > 0$ is a suitable real number. Note that in order to formally carry out Brown's (1979) analysis, one would have to first make a variance stabilizing transformation. Theorem 1, however, uses arguments different from Brown's, although it is motivated by Brown (1979). Brown's (1979) heuristic analysis can be used to generate similar provable inadmissibility results in many estimation problems; following this theorem, we have given a large number of examples in which the function λ is different from the geometric mean and in all of these examples, Brown's (1979) heuristics do suggest that such other λ's can be used to obtain uniform domination. In choosing an appropriate function λ for constructing dominating estimators, the following heuristic prescription also seems to work very well: find a function $\lambda(\underset{\sim}{X})$ such that $\frac{\Sigma X_i}{\lambda(\underset{\sim}{X})}$ is independent of either $\lambda(\underset{\sim}{X})$ or ΣX_i when all θ_i's are equal. Then $\delta(\underset{\sim}{X}) + c\lambda(\underset{\sim}{X})$ is a plausible estimator dominating $\delta(\underset{\sim}{X})$. For example, let $X_1, \ldots, X_{k+1}$ be iid $N_p(\underset{\sim}{\mu}, \Sigma)$, where $\underset{\sim}{\mu}, \Sigma$ are unknown, and let $\ell_1 \ge \ldots \ge \ell_p > 0$ be the ordered eigenvalues of $S = \Sigma(X_j - \overline{X})(X_j - \overline{X})$; consider the problem of estimating $(\lambda_1, \ldots, \lambda_p)$, the ordered eigenvalues of Σ. It is well known that $\frac{\sum \ell_i}{(\prod_j \ell_j)^{\frac{1}{p}}} = \frac{\mathrm{tr} S}{|S|^{\frac{1}{p}}}$ is independent of tr S when all λ_i's are equal to some λ (i.e., $\Sigma = \lambda I$); this follows from Basu's (1955) theorem because tr$S$ is complete sufficient for λ and tr $S / \mid S \mid^{\frac{1}{p}}$ is an ancillary statistic. Thus, in the eigenvalue problem described above, $\mid S \mid^{\frac{1}{p}}$ is a plausible choice for $\lambda(\underset{\sim}{\ell})$.

Proof of Theorem 1: We will prove the theorem when case (a) holds. The other case is exactly similar. By definition,

$$\Delta(F) = E\sum_{i=1}^{p} \{\delta_i(X) + c\lambda - \mu_i(X,F)\}^2 - E\sum_{i=1}^{p} \{\delta_i(X) - \mu_i(X,F)\}^2$$

$$= E\left\{pc^2\lambda^2 + 2c\lambda \sum_{i=1}^{p} \big(\delta_i(x) - \mu_i(X,F)\big)\right\}$$

$$\leq E\left\{(pc^2A_0+2c)\lambda\cdot\sum_{i=1}^{p}\delta_i(X)-2c\lambda\sum_{i=1}^{p}\mu_i(X,F)\right\} \tag{2.3}$$

$$= E\left[\lambda\cdot\sum_{i=1}^{p}\mu_i(X,F)\right]\cdot\{(pc^2A_0+2c)\cdot A(F)-2c\}$$

$$< 0,\ \text{if } 0 > c > \frac{2(1-r)}{prA_0}. \tag{2.4}$$

Hence proved.

The permissible values of c in case (b) are $0 < c < \frac{2(1-R)}{pRA_0}$. With slight modifications, the assertion of the above theorem holds when the loss is $L_Q = (\delta-\mu)'\,Q(\delta-\mu)$ where Q is a p.d. matrix of known constants and when the unit vector 1 is replaced by an $\alpha > 0$ (i.e., each component is positive) in the dominating estimator. Note that while it follows from Shinozaki (1975) that if $\underset{\sim}{\delta}(\underset{\sim}{x})$ is inadmissible under squared error loss, then it is also inadmissible under L_Q, it does not follow from Shinozaki (1975) that essentially the same type of estimator dominates $\underset{\sim}{\delta}(\underset{\sim}{x})$ under both losses.

Remark 2: In some problems, it is easier to show that $C(F) = -E\{\lambda\cdot\Sigma(\delta_i(\underset{\sim}{x})-\mu_i(\underset{\sim}{x},F))\} > 0$ and that $\frac{E(\lambda^2(\underset{\sim}{x}))}{C(F)}$ is uniformly bounded over $F \in \mathcal{F}$. The line preceding (2.3) shows that in such a case one can find a small enough $c > 0$ such that $\Delta(F) < 0$ $\forall\, F \in \mathcal{F}$. See example 3 for an illustration in the Pareto distribution.

Remark 3: If a family of statistics $\{\lambda_\alpha, \alpha\epsilon T\}$ satisfy conditions (a) and (c) (or (b) and (c)) of Theorem 1, then very often a convex mixture $\lambda = \int_T \lambda_\alpha dP(\alpha)$ also satisfies conditions (a) and (c) ((b) and (c)) of Theorem 1. This fact is useful in generating new improved estimators by taking convex combinations of different estimators. Of course, that such convex combinations are also uniformly dominating estimators follows from the convexity of squared-error loss. What is interesting is that such convex combinations also satisfy the hypothesis of Theorem 1. See DasGupta (1986b) for details.

We now give applications of this theorem to various estimation problems. At this point, we like to remind the reader that in almost all the examples given below, the functional $A(F)$ defined in the theorem is, in fact, a constant independent of F. For a related result in this direction, see Gleser (1986).

Example 1: Estimation of arbitrary scale-parameters.

Let $X_i \overset{\text{indep.}}{\sim} \frac{1}{\theta_i}f_i\left(\frac{X_i}{\theta_i}\right)$, $\quad X_i,\ \theta_i > 0, \quad 1 \leq i \leq p$. The standard estimate of θ_i is the best equivariant estimate $\delta_i(X_i) = a_iX_i$, where $a_i = \frac{E_{\theta_i=1}(X_i)}{E_{\theta_i=1}(X_i^2)}$. Take $\lambda(\underset{\sim}{X}) = (\prod_j X_j)^{\frac{1}{p}}$ in Theorem 1.

In this case,

$$A(\theta) = \frac{E_\theta\left[\sum\limits_{i=1}^{p} a_i X_i (\prod\limits_{j=1}^{p} X_j)^{\frac{1}{p}}\right]}{\sum\limits_{i=1}^{p} \theta_i \cdot E_\theta (\prod\limits_{j=1}^{p} X_j)^{\frac{1}{p}}}$$

$$= \frac{E_\theta\left[\sum\limits_{i=1}^{p} a_i X_i{}^{1+\frac{1}{p}} (\prod\limits_{j\neq i} X_j)^{\frac{1}{p}}\right]}{(\sum_{i=1}^{p} \theta_i \cdot E_\theta (\prod_{j=1}^{p} X_j)^{\frac{1}{p}}} \tag{2.5}$$

Now observe,

$$a_i E_{\theta_i}(X_i{}^{1+\frac{1}{p}}) = \frac{E_{\theta_i=1}(X_i)}{E_{\theta_i=1}(X_i^2)} \cdot E_{\theta_i=1}(X_i{}^{1+\frac{1}{p}}) \cdot \theta_i{}^{1+\frac{1}{p}}$$

$$< E_{\theta_i=1}(X_i^{\frac{1}{p}})\theta_i{}^{1+\frac{1}{p}}$$

$$= E_{\theta_i}(X_i^{\frac{1}{p}})\ \theta_i \tag{2.6}$$

(the inequality follows form Liapunov's inequality which implies that $E(X^{1+\frac{1}{p}})\ E(X) < E(X^2)\ E(X^{\frac{1}{p}})$ for any positive random variable X; see DasGupta (1984b)).

The mutual independence of the X_i's, (2.5), and (2.6) now imply that $\sup\limits_{\theta}\ A(\theta) < 1$ and hence the best equivariant estimate is inadmissible for every $p \geq 2$ (Note that $A(\theta)$ is actually a constant free of θ). This example thus establishes inadmissibility of the standard estimates of the means (and in fact, *all* moments) in the Gamma, Pareto, lognormal, Rectangular, F, Weibull, and the Half-normal distributions with unknown scale parameters. The class of improved estimators in DasGupta (1986a) are thus proved to be robust in the sense that in the entire scale parameter family, they dominate the best scale equivariant estimator of the scale parameter $\underset{\sim}{\theta}$. Recently, Shinozaki (1984) has proved that the James-Stein estimators are also fairly robust and give uniform domination for many location parameter distributions.

Example 2. Inadmissibility of unbiased estimates.

The result of Theorem 1 can also be used to prove inadmissibility of unbiased estimates in certain families of distributions. Let $0 \leq X_i$, $1 \leq i \leq p$, be independently distributed as $F_1, F_2 \ldots, F_p$ respectively. Suppose we want to estimate $\mu_i(F_i) = E_{F_i}(X_i)$, $1 \leq i \leq p$. Identifying $\delta_i(X)$ as X_i in Theorem 1, we will show that $\inf_F A(F) > 1$ if $\sqrt{X_i}$ has a coefficient of variation uniformly bounded away from zero (over F_i), and hence X is inadmissible for its expectation. We take $\lambda(\underset{\sim}{x}) = (\prod\limits_{j} x_j)^{\frac{1}{p}}$.

Towards this end, note that

$$E\left\{\sum_i X_i \cdot (\prod_j X_j^{\frac{1}{p}})\right\}$$

$$= \sum_i E\left\{X_i^{1+\frac{1}{p}} \cdot \left(\prod_{j \neq i} X_j^{\frac{1}{p}}\right)\right\}$$

$$\geq \left\{(EX_i)^{1+\frac{1}{p}} \cdot E(\prod_{j \neq 1} X_j^{\frac{1}{p}})\right\}$$

(because of independence and Holder's inequality)

$$= \Sigma_i \left\{\mu_i(F_i) \cdot (EX_i)^{\frac{1}{p}} \cdot E(\prod_{j \neq i} X_j^{\frac{1}{p}})\right\}. \tag{2.7}$$

Now by assumption

$$\sup_{F_i} \frac{EX_i^{\frac{1}{2}}}{(EX_i)^{\frac{1}{2}}} < 1$$

$$\Rightarrow \sup_{F_i} \frac{EX_i^r}{(EX_i)^r} < 1 \quad \text{if } 0 < r < 1. \tag{2.8}$$

(as long as $0 < r < 1$, the value of r does not matter). Taking $r = \frac{1}{p}$ in (2.8), it now follows from (2.7) that

$$\inf_F \frac{E\{\Sigma_i X_i \cdot (\prod_j X_j \cdot \frac{1}{p}\}}{E\{\prod_j X_j^{\frac{1}{p}}\} \cdot \Sigma \mu_i(F_i)} > 1.$$

Distributions for which $\sqrt{X_i}$ has a uniformly bounded (below) coefficient of variation have, in some sense, pseudo-scale structures in them. It can be quite easily proved that scale mixtures of distributions uniformly satisfying (2.8) also satisfy (2.8). See DasGupta (1986b) for details.

One very simple example of this sort is $h(x) = \varepsilon \frac{1}{\theta_1} f_1(\frac{x}{\theta_1}) + (1-\varepsilon)\frac{1}{\theta_2} f_2(\frac{x}{\theta_2})$, where $1 > \varepsilon > 0$ is known. Note that this is *not* a scale-parameter density.

Example 3. Ferguson family of distributions.

The heuristic prescription given in the paragraph preceding the proof of Theorem 1 for finding an appropriate function $\lambda(\underset{\sim}{x})$ in order to construct the improved estimator seems to work well for some well known members of the Ferguson family of distributions (for the definition of the Ferguson family of distributions, see Ferguson (1962)). An important feature of the Ferguson family of distributions is that if $X_i \overset{\text{indep.}}{\sim} L(\theta_i, \sigma_i^2, \gamma_i)$ have distributions belonging to the Ferguson family, then one can find appropriate size functions $G(\underset{\sim}{x})$ (a statistic G is called a size function if $G(a\underset{\sim}{x}) = aG(\underset{\sim}{x})$ for all $a > 0$) such that $\frac{\underset{\sim}{X}}{G(\underset{\sim}{X})}$ and $G(\underset{\sim}{X})$ are independently distributed. Note that $G(\underset{\sim}{x})$ can actually depend on the parameters also: for example, if $X_i \overset{\text{indep.}}{\sim}$ *Pareto* (θ_i) where θ_i's are the scale-parameters, then $G(\underset{\sim}{X}) = \min(\frac{X_1}{\theta_1}, \ldots, \frac{X_p}{\theta_p})$. For the purpose of the following discussion, it will be understood that $G(\underset{\sim}{X})$ is such that $\frac{\underset{\sim}{X}}{G(\underset{\sim}{X})}$ and $G(\underset{\sim}{X})$ are independent under $\underset{\sim}{\theta} = \underset{\sim}{1}$.

It turns out that in two important members of the Ferguson family of distributions, namely, the Pareto, and the lognormal, the natural estimators of the means (which are usually constant multiples of the X_j's) can be uniformly dominated in risk by expanding them by a suitable positive multiple of $G(\underset{\sim}{x})$.

For the lognormal case, the size $G(\underset{\sim}{X})$ is $(\prod_j X_j)^{\frac{1}{p}}$ and thus, rather coincidentally, the result of example 1 can be invoked to prove that uniform domination obtains. In the Pareto case, we will present the actual proof only for $p = 2$; the calculations get somewhat complicated for larger p. Before we proceed further, we would like to point out that the appropriate size function $G(\underset{\sim}{x})$ such that $\frac{\underset{\sim}{x}}{G(\underset{\sim}{x})}$ and $G(\underset{\sim}{x})$ are independent is usually found by equating $G(\underset{\sim}{x})$ to a sufficient statistic when all θ_i's are equal to some θ (see Lemma 4.1 in James (1979)). Let now $X_i \overset{\text{indep.}}{\sim}$ Pareto (a, θ_i), where $a > 2$ is the common shape-parameter; i.e.,

$$f(x_i/\theta_i) = \frac{a}{\theta_i}\left(\frac{x_i}{\theta_i}\right)^{-a-1}, \quad x_i > \theta_i, \quad a > 2. \tag{2.9}$$

The best equivariant estimate of θ_i is $\frac{a-2}{a-1}x_i$. Recall that $G(\underset{\sim}{x}) = \min(x_1, \ldots, x_p)$ in this case. Now

$$\begin{aligned}\sum_i E_{\underset{\sim}{\theta}}\Big[\frac{a-2}{a-1}x_i + cG(\underset{\sim}{x}) - \theta_i\Big]^2 &- \sum_i E_{\underset{\sim}{\theta}}\Big[\frac{a-2}{a-1}x_i - \theta_i\Big]^2 \\ &= E_{\underset{\sim}{\theta}}\Big[pc^2\, G^2(\underset{\sim}{x}) + 2c\cdot\frac{a-2}{a-1}\cdot G(\underset{\sim}{x})\sum_i X_{(i)} - 2c\sum_i \theta_i\cdot G(\underset{\sim}{x})\Big],\end{aligned} \tag{2.10}$$

where $X_{(1)} \leq \ldots \leq X_{(p)}$ are the sample order statistics. Direct calculation (in the case $p = 2$) yields that for $\theta_1 \leq \theta_2$, the risk-difference is equal to

$$\Big[pc^2\cdot\frac{a}{a-2}f_1(r) - \frac{2ca}{a-1}f_2(r)\Big]\theta_2^2, \tag{2.11}$$

where

$$r = \frac{\theta_1}{\theta_2} \leq 1,$$

$$f_1(r) = r^2 - \frac{r^a}{a-1},$$

and
$$f_2(r) = \frac{ar^a}{(a-1)(2a-1)} - \frac{r^{a+1}}{2a-1} + \frac{r}{(a-1)^2}. \tag{2.12}$$

A similar identity holds for $\theta_2 \leq \theta_1$.

It's clear that $f_2(r) > 0 \; \forall \; r \varepsilon (0,1]$; furthermore, $\frac{f_1(r)}{f_2(r)} \to 0$ as $r \to 0$, and $\frac{f_1(r)}{f_2(r)}$ has no singularities on (0,1]. Consequently, $\frac{f_1(r)}{f_2(r)}$ is uniformly bounded on (0,1]. Hence one can choose a small enough $c > 0$ such that the risk-difference in (2.11) is uniformly negative. This proves that for two independent Pareto random variables, the best equivariant estimate can be uniformly dominated in risk by expanding it by a multiple of $\min(X_1, X_2)$. We believe that these calculations can be carried out for a general p; the algebra, of necessity, gets more involved.

Example 4. Estimation of the means in the multivariate F distribution.

Let $X_0, X_1, \ldots, X_p$ be independently distributed as $\Gamma(\alpha_i, \theta_i)$, $\alpha_i > 0$, $\theta_i > 0$. The joint distribution of $Z_1, Z_2, \ldots, Z_p$, where $Z_i = \frac{X_i}{X_0}$, is called the multivariate F distribution.

In multifactor ANOVA problems, X_0 plays the role of the error sum of squares, and $X_1, X_2, \ldots, X_p$ are the sums of squares due to the treatment effect, the block effect, etc. The joint distribution of the ratios $Z_i = \frac{X_i}{X_0}$ is important for *simultaneously* testing that there is no treatment effect, no block effect, etc.

It's easy to verify that $E(Z_i) = \frac{\alpha_i}{\alpha_0 - 1} \cdot \frac{\theta_0}{\theta_i}$; also, the best scale equivariant estimate of $\frac{\theta_0}{\theta_i}$ is $\frac{\alpha_0 - 2}{\alpha_i + 1} Z_i$. Note that $\{Z_i\}$ are not independent. We will prove that for $p \geq 2$, the best scale-equivariant estimate is inadmissible and dominated uniformly in risk by $\frac{\alpha_0 - 2}{\alpha_i + 1} Z_i + c(\prod_j Z_j)^{1/p}$, for a suitable $c > 0$. In view of Theorem 1, we merely need to show that

$$\sup_{\underset{\sim}{\theta}} \frac{E_\theta \left[\sum_i \frac{\alpha_0 - 2}{\alpha_i + 1} Z_i \cdot (\prod_j Z_j)^{1/p} \right]}{\theta_0 \cdot \sum_i \theta_i^{-1} \cdot E_\theta (\prod_j Z_j)^{1/p}} < 1 \tag{2.13}$$

Using the facts that $X_0, X_1, \ldots, X_p$ are independent, a direct calculation shows that the numerator in (2.13) equals $\frac{1}{\alpha_0 - 1} (\sum_i \frac{\alpha_i + \frac{1}{p}}{\alpha_i + 1}) (\prod_j \frac{\Gamma(\alpha_j + \frac{1}{p})}{\Gamma \alpha_j})$ and the denominator equals $\frac{p}{\alpha_0 - 1} (\prod_j \frac{\Gamma(\alpha_j + \frac{1}{p})}{\Gamma \alpha_j})$ (the ratio is independent of $\theta_0, \theta_1, \ldots, \theta_p$, and thus can as well be calculated under $\underset{\sim}{\theta} = \underset{\sim}{1}$). Thus (2.13) equals $\frac{\sum_i \frac{\alpha_i + \frac{1}{p}}{\alpha_i + 1}}{p}$, which clearly is less than 1 for every $p \geq 2$. This proves the inadmissibility result.

Example 5. Estimation of the eigenvalues of the dispersion matrix in multivariate normal distributions.

Let $X_1, X_2, \ldots, X_{k+1}$ be i. i. d. observations from $N_p(\mu, \Sigma), \mu, \Sigma$ (p.d.) both unknown. Let $S = \sum_{j=1}^{k+1} (X_j - \overline{X})\ (X_j - \overline{X})'$; it is well-known that $S \sim W_p(k, \Sigma)$ and that S can be written as $S = \sum_{j=1}^{k} U_j\ U_j'$, where $U_j \sim N_p(0, \Sigma)$. Moreover, the unbiased estimate of Σ is $\frac{S}{k}$ and the *mle* is $\frac{S}{k+1}$. Consequently, if $\ell_1 \geq \ell_2 \geq \ldots \ell_p > 0$ are the eigenvalues of S, then $\frac{\ell_i}{k}$ as well as $\frac{\ell_i}{k+1}$ are in some sense "natural" estimates of the eigenvalues of Σ. It is known that $\frac{\ell_i}{k+1}$ are the *mle*s of the ordered eigenvalues of Σ, but $\frac{\ell_i}{k}$ are not unbiased for them except in the case $p = 1$. Note that in this problem there is no best choice of the multiple of the $\{\ell_i\text{'s}\}$ and motivated by the case $p = 1$, one may even consider the estimators $\frac{\ell_i}{k+2}$.

In general, if one sets $\delta_i(X) = a\ell_i$, then,

$$A(\Sigma) = \frac{a \cdot E_\Sigma \left\{ \cdot \text{ tr } S \cdot \mid S \mid^{\frac{1}{p}} \right\}}{\text{tr } \Sigma \cdot E_\Sigma \mid S \mid^{\frac{1}{p}}} (\text{we take } \lambda(\underset{\sim}{\ell}) = |S|^{\frac{1}{p}}). \tag{2.14}$$

$A(\Sigma)$ is easily seen to be constant in Σ. Thus, if $a = \frac{1}{k}$, then inadmissibility will follow if we can prove that

$$A(I) = \frac{E_I\{\text{tr } S \cdot |S|^{1/p}\}}{E_I(\text{tr } S) \cdot E_I(|S|^{1/p})} > 1, \tag{2.15}$$

$$\Longleftrightarrow \text{Cov}_I(\text{tr } S, |S|^{1/p}) > 0. \tag{2.16}$$

Direct calculation yields that COV_I (tr S, $|S|^{1/p}) = \frac{k+\frac{2}{p}}{k}$ so that $\left\{\frac{\ell_i}{k}\right\}$ are inadmissible for $p \geq 2$.

The mle's of the eigenvalues in this problem correspond to $a = \frac{1}{k+1}$. The ratio $A(\Sigma)$ in this case is $\frac{k}{k+1} \cdot \frac{k+\frac{2}{p}}{k} < 1$ for $p \geq 3$ and hence it follows from Theorem 1 that the mle is inadmissible for $p \geq 3$. The Associate Editor has pointed out, using a different argument, that the mle is inadmissible even for $p = 2$ and in fact for $p = 2$ is dominated by $\frac{2k+1}{2k(k+2)}(\ell_1, \ell_2)$. Note that for $a = \frac{1}{k}$ the constant c of the theorem is negative so that the improved estimators are shrinkage estimators, while for $a = \frac{1}{k+1}$ the constant c is positive so that the improved estimators are now expanders. Theorem 1 indeed shows that for each $a \neq (k + \frac{2}{p})^{-1}$, $\{a\ell_i\}$ are inadmissible, and $a = (k + \frac{2}{p})^{-1}$ is *not* a natural choice except when $p = 1$ or 2. We do not know if this choice of a implies admissibility of $a\ell_1, \ldots, a\ell_p$ or not.

Example 6. Estimation of the eigenvalues of Σ in elliptically symmetric distributions.

Let $X_1, X_2, \ldots, X_{k+1}$ be i.i.d. observations from a distribution with density $\mid \Sigma \mid^{-\frac{1}{2}} \cdot f\big((x-\mu)'\Sigma^{-1}(x-\mu)\big)$, where μ, $\Sigma(> 0)$ are unknown. Once this sample is obtained, we define the matrix S in the usual fashion. It's easy to show that $E(\frac{S}{k}) = \Sigma$. Consider the problem of simultaneously estimating the eigenvalues of Σ by constant multiples of the eigenvalues of S. It turns out that $\{\frac{\ell_i}{k}\}$ are inadmissible if f is of the form $f((x-\mu)'\Sigma^{-1}(x-\mu)) = \text{constant} \times \int \frac{e^{-\frac{1}{2\tau^2}(x-\mu)'\Sigma^{-1}(x-\mu)}}{\tau^p} dF(\tau)$ where F is a probability measure on $(0,\infty)$. This can be seen by reasoning conditionally on τ^2 and then integrating over τ^2. See DasGupta (1986b) for details. Important densities which are such scale mixtures of normals are the spherically symmetric Cauchy, t, and the Double exponential densities.

The estimation of the eigenvalues of Σ when $\underset{\sim}{X}_j$'s are non-normal is of importance in principal components analysis. As a dimension reduction technique, principal components analysis does not make any distributional assumptions whatsoever, and the ordered eigenvalues of Σ are the variances of the principal components whether or not the X_j's are normal.

Example 7. A larger class of improved eigenvalue estimators when $S \sim W_p(k,\Sigma)$.

Broader classes of improved estimators for the eigenvalues of Σ when $S \sim W_p(k,\Sigma)$ are obtainable using Theorem 1.

The identities of Haff (1980) in conjunction with Theorem 1 imply that $\frac{\ell_i}{k} + cr(t)$ dominates $\frac{\ell_i}{k}$ for suitable $c < 0$, whenever

$$\text{(i)}\ \ 0 < \frac{r(t)}{t} \le 1$$

$$\text{and}\quad \text{(ii)}\ \ t^{-\varepsilon} r(t) \ \text{ is nondecreasing in } t \text{ for some } \varepsilon > 0, \tag{2.17}$$

where $t = |S|^{\frac{1}{p}}$.
Note that condition (i) is needed for hypothesis (c) of Theorem 1 to go through.

The identities of Haff(1980) can be used to generate improved estimators for the eigenvalues of Σ by shifting by suitable functions of tr$S$ as well. We have been able to show that $\frac{\ell_i}{k} + cr(\text{tr } S)$ dominates $\frac{\ell_i}{k}$ for suitable $c < 0$, whenever

(i) there exists an $\varepsilon > 0$ such that $\varepsilon \le \frac{r(W)}{W} \le 1 \ \forall\ w$

(ii) $\frac{r(W)}{W}$ is non-increasing. (2.18)

A nice thing about these fairly wide classes of uniformly improved estimators is that one can now formally seek to select an improved estimator by using criteria like gamma-minimaxity or the restricted risk Bayes principle. For some recent works on such selection problems, see Marazzi (1985), DasGupta and Berger (1986c), DasGupta and Rubin (1987a) and DasGupta and Bose (1987b). Also see L. Brown's discussion on Berger (1983).

Remark 4: Results similar in spirit to those in Example 5 have been obtained for estimation of the eigenvalues of the precision matrix Σ^{-1}, of $\Sigma_1^{-1}\Sigma_2$ where $S_i \overset{\text{indep.}}{\sim} W_p(k,\Sigma_i), i =$

$1, 2$, and of the residual matrix $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ when Σ is partitioned as $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. These problems arise naturally in the context of testing problems in multivariate statistics and in multivariate regression. See DasGupta (1986b).

Example 8. Estimation of losses in the scale-parameter family.

We now give an application of Theorem 1 to a completely different type of problem. Let $X_i \overset{\text{indep.}}{\sim} \frac{1}{\theta_i} f_i(\frac{x_i}{\theta_i})$. As mentioned in example 1, the natural estimate of θ_i under squared-error loss is the best equivariant estimate $a_i X_i$, where $a_i = \frac{E\theta_i{=}1(X_i)}{E\theta_i{=}1(X_i^2)}$. The loss in estimating θ_i by $a_i X_i$ is $(a_i X_i - \theta_i)^2$. An important decision-theoretic problem is that of estimating the actual loss incurred in estimating an unknown parameter; the idea here is that a statistician should report not only an estimate of the unknown parameter, but also a measure of accuracy of this estimate. An estimated loss serves this purpose. The related problem of estimating the risk function is also of some importance. Note that the loss incurred is a function of both the parameter and the data; we will use squared-error as a criterion for choosing among different estimators of the loss. In the context of our scale-parameter problem, then, the problem under consideration is simultaneous estimation of the componentwise losses $(a_i X_i - \theta_i)^2$, $1 \le i \le p$, under a second stage sum of squared error loss. The problem of estimating the overall loss $\Sigma_{i=1}^p (a_i X_i - \theta_i)^2$ could also be attempted; see Johnstone (1987). For ease in presentation, we will work out this example in the gamma case, but *every* step in the following analysis goes through in the arbitrary scale- parameter family.

Let $X_i \overset{\text{indep.}}{\sim} \Gamma(\alpha, \theta_i)$, with $EX_i = \frac{\alpha}{\theta_i}$, $1 \le i \le p$. In this case, the best equivariant estimate of θ_i^{-1} is $a_i X_i$, where $a_i = \frac{1}{\alpha+1}$. For estimating the loss $(a_i X_i - \theta_i^{-1})^2$, natural one-dimensional estimates are $c_i X_i^2$; one reason for this is that the expected loss is proportional to θ_i^{-2}, and another reason is that unbiased estimators of the overall loss $\sum_{i=1}^p (a_i X_i - \theta_i^{-1})^2$ are of the form $\sum_{i=1}^p c_i X_i^2$ (i.e., for suitable constants c_i's, $E_\theta\{\sum_{i=1}^p c_i X_i^2 - \sum_{i=1}^p (a_i X_i - \theta_i^{-1})^2\} \equiv 0$).

Because of the scale-structure in the problem, there exists a c_i minimizing $E[c_i X_i^2 - (\frac{X_i}{\alpha+1} - \theta_i^{-1})^2]^2$; the optimum c_i satisfies

$$\left(c_i - \frac{1}{(\alpha+1)^2}\right) \cdot EX_i^4 + \frac{2}{\alpha+1} EX_i^3 - EX_i^2 = 0 \tag{2.19}$$

where $E(\cdot)$ denotes expectation under $\theta_i = 1$.

By direct calculations now, $c_i = \frac{1}{(\alpha+1)^2(\alpha+2)}$. Identifying $\delta_i(X)$ as $c_i X_i^2$ and $\mu_i(X, \theta)$

as $(\frac{X_i}{\alpha+1} - \theta_i^{-1})^2$ in Theorem 1, it suffices to show that

$$\frac{1}{(\alpha+1)^2(\alpha+2)} \sup_\theta \frac{E_\theta\left[\sum_{i=1}^p X_i^2 \cdot (\prod_{j=1}^p X_j)^{\frac{2}{p}}\right]}{E_\theta\left[(\prod_{j=1}^p X_j)^{\frac{2}{p}} \cdot \sum_{i=1}^p \left(\frac{X_i}{\alpha+1} - \theta_i^{-1}\right)^2\right]} < 1 \tag{2.20}$$

By straightforward calculations, the numerator of (2.20) equals

$$\frac{\Gamma(\alpha+2+\frac{2}{p})}{\Gamma(\alpha+\frac{2}{p})} \cdot \left(\frac{\Gamma(\alpha+\frac{2}{p})}{\Gamma\alpha}\right)^p \cdot \left(\prod_j \theta_j^{-\frac{2}{p}}\right) \cdot \sum_i \theta_i^{-2} \tag{2.21}$$

Also, the denominator of (2.20) equals

$$\begin{aligned} &\left\{\frac{(\alpha+1+\frac{2}{p})(\alpha+\frac{2}{p})}{(\alpha+1)^2} - \frac{2(\alpha+\frac{2}{p})}{\alpha+1} + 1\right\} \cdot \left(\frac{\Gamma(\alpha+\frac{2}{p})}{\Gamma\alpha}\right)^p \cdot \left(\prod_j \theta_j\right)^{-\frac{2}{p}} \Sigma_i \theta_i^{-2} \\ &= \frac{\frac{4}{p^2} - \frac{2}{p} + \alpha + 1}{(\alpha+1)^2} \cdot \left(\frac{\Gamma(\alpha+\frac{2}{p})}{\Gamma\alpha}\right)^p \cdot \left(\prod_j \theta_j^{-\frac{2}{p}}\right) (\Sigma_i \theta_i^{-2}) \end{aligned} \tag{2.22}$$

Hence, in the notation of Theorem 1,

$$A(\theta) \equiv \frac{1}{(\alpha+1)^2(\alpha+2)} \cdot \frac{(\alpha+1+\frac{2}{p})(\alpha+\frac{2}{p})(\alpha+1)^2}{\frac{4}{p^2} - \frac{2}{p} + \alpha + 1} \tag{2.23}$$

Interestingly enough, this ratio is exactly 1 for both $p = 1$ and 2 and thus the inadmissibility result of Theorem 1 does not apply. However, for $p \geq 3$, this ratio is < 1 and hence uniform domination can be achieved by using improved loss estimators $\frac{X_i^2}{(\alpha+1)^2(\alpha+2)} + c \cdot (\prod_j X_j)^{\frac{2}{p}}$, where $c > 0$ is a suitable number. Thus, although the best equivariant estimate of θ_i^{-1} is itself inadmissible for $p \geq 2$, a natural estimator of its loss can be shown to be inadmissible only for $p \geq 3$, and is probably admissible for $p = 1, 2$. This somewhat surprising phenomenon in the gamma case thus completely agrees with the recent results of Johnstone (1987) in the normal case.

In the general scale-parameter family, the expression analogous to (2.19) is given as

$$\begin{aligned} &(c_i - a_i^2)EX_i^4 + 2a_i E(X_i^3) - E(X_i^2) = 0 \\ \Leftrightarrow\ & c_i = \frac{EX_i^2 + a_i^2 EX_i^4 - 2a_i EX_i^3}{EX_i^4}, \end{aligned} \tag{2.24}$$

where $E(\cdot)$ again denotes expectation under $\theta_i = 1$. If the coordinate distributions under $\theta_i = 1$ are assumed to be identical, then $c_i \equiv c$ (say) and $a_i \equiv a$ (say) for every i. The constant ratio analogous to (2.23) now is given as

$$A(\underset{\sim}{\theta}) \equiv \frac{c \cdot m(2+\frac{2}{p})/m(\frac{2}{p})}{a^2 \cdot \frac{m(2+\frac{2}{p})}{m(\frac{2}{p})} - 2a\frac{m(1+\frac{2}{p})}{m(\frac{2}{p})} + 1}, \tag{2.25}$$

where $m(r) = EX_i^r$.

Because of the result we just proved for the gamma case, there is no hope of showing that this ratio is $\neq 1$ for every $p \geq 2$. Interestingly, however, in virtually *every other* standard scale-parameter family, this ratio is different from 1 for $p = 2$. Our theorem would then imply that the best scale-equivariant estimator of the loss is inadmissible in *all of these* distributions for $p = 2$, and *hence* for every $p \geq 2$.

We will now give a simple sufficient condition for the ratio in (2.25) to be different from 1 when $p = 2$ (in view of Theorem 1, that's all one needs for inadmissibility).

Simple algebra shows that the ratio in (2.25) is different from 1 for $p = 2$ iff

$$\frac{1}{2}\frac{m(4)}{m(3)} + \frac{1}{2}\frac{m(2)}{m(1)} \neq \frac{m(3)}{m(2)}. \tag{2.26}$$

Let now $\psi(r) = \frac{m(r+1)}{m(r)}$, where $r > 0$. It's clear that if $\psi(r)$ is strictly convex or strictly concave in the interval $1 \leq r \leq 3$, then (2.26) holds. For many standard scale-parameter distributions, showing that $\psi(r)$ is strictly convex (concave) is actually easier than verifying (2.26) directly. For the sake of completeness, we have given below a list of the appropriate ψ's in a few standard examples:

Distribution	**Density under $\theta = 1$**	$\psi(r)$
Gamma	$\frac{e^{-x}x^{\alpha-1}}{\Gamma\alpha}$	$\alpha + r$
Pareto	$ax^{-a-1}, x > 1, a > 4$	$\frac{a-r}{a-r-1}$
lognormal	$\frac{1}{\sqrt{2\pi}x}e^{-\frac{1}{2}(\log x)^2}$	$e^{r+\frac{1}{2}}$
Rectangular	$1(0 < x < 1)$	$\frac{r+1}{r+2}$
F	$\frac{1}{\sqrt{2x}}(\sqrt{n}B(\frac{1}{2},\frac{n}{2}))^{-1}(1+\frac{x}{n})^{-\frac{(n+1)}{2}}$ $(n > 8)$	$\frac{n(1+2r)}{n-2r-2}$
Weibull	$\beta x^{\beta-1}e^{-x^\beta}$	$\frac{\Gamma(\alpha+\frac{r+1}{\beta})}{\Gamma(\alpha+\frac{r}{\beta})}$

In all of these cases, except the Weibull, it is very easy to check that $\psi(r)$ is indeed either strictly convex or strictly concave by simply looking at $\psi''(r)$. In the Weibull case with $\beta \neq 1$, denoting $\frac{\alpha+r}{\beta}$ as s and $\frac{1}{\beta}$ as t, the convexity (concavity) of $\psi(s) = \frac{\Gamma(s+t)}{\Gamma(s)}$ follows from the facts that $\frac{\Gamma(s+t)}{\Gamma(s)} = EX^t$ where $X \sim \text{Gamma}(s,1)$, that the Gamma$(s,1)$ densities are strict Polya Type ∞ and for any $t \neq 1$, x^t crosses any line at most twice and hence Karlin's sign change theorem (see Karlin (1957)) implies that $\psi(s)$ crosses any line at most twice. This argument was sketched by an anonymous referee to whom we are very grateful.

3. Final Remarks, Bayesian Motivations, Risk-Improvements, and Scope for Choice

As stated in section 2, in the Wishart eigenvalue problem, example 7 gives a wide class of improved estimators. This gives us a good scope for choice and past evidence suggests that the function $\lambda(S)$ can be chosen to incorporate prior information; for example, one can choose that λ which gives the smallest Bayes risk with respect to a given prior.

There are some important problems in which our theorem does *not* apply without non-trivial modifications. These include estimation of characteristic roots of the correlation matrix, estimation of the "proportions of variance explained" in Principal components analysis, i.e., $\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}$, where λ_i's are the eigenvalues of Σ, and estimation of canonical correlations. However, we hope that the techniques of this paper will provide mathematical insight into those problems.

One nice feature of the dominating estimators provided by our Theorem is that in almost all applications, the risk improvement is exactly analytically calculable. This not only makes subsequent Bayesian calculations analytically feasible, it also makes simulation unnecessary for the study of risk-improvement. Using these analytical expressions in the Wishart case, we found that by using $\lambda(S) = \text{tr } S$, and c as the mid-point of the admissible range, about 22.2% risk-improvement can be attained over the eigenvalues of $\frac{S}{k}$ when $p = 2$ and k is small. The improvement, however, starts decreasing as k increases.

Acknowledgment.

I am deeply indebted to the Associate Editor, and two referees for their extremely constructive suggestions and to Leon Jay Gleser for suggesting significant improvements over an earlier draft of this paper. My sincerest appreciation also goes to Herman Rubin for his non-trivial contributions to this paper and for graciously helping at every stage of this work, and to Jim Berger for some valuable discussion on the materials of section 2.

References

Anderson, T. W. (1965). Some optimum confidence bounds for roots of determinantal equations. *Ann. Statist.* **36**, 468–488.

Basu, D. (1955). On Statistics independent of a complete sufficient statistic. *Sankhyā* **15**, 377–380.

Berger, James (1975). Minimax estimation of location vectors for a wide class of densities.

Ann. Statist. **3**, 1318–1328.

Berger, James (1980). Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.* **8**,545–571.

Berger, James (1983). The robust Bayesian viewpoint. Robustness in Bayesian Statistics, J. Kadane (Ed.). North-Holland, Amsterdam.

Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37**, 1087–1135.

Brown, L. D. (1979). A heuristic method for determining admissibility of estimators — with applications. *Ann. Statist.* **7**, 960–994.

DasGupta, Anirban (1984a). Admissibility in the gamma distribution: two examples. *Sankhyā, Ser. A* **46**, 395–407.

DasGupta, Anirban (1984b). Simultaneous estimation of arbitrary scale parameters under arbitrary quadratic loss. Unpublished Technical report #84-32, Department of Statistics, Purdue University.

DasGupta, Anirban (1986a). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses. *Ann. Statist.* **14**, 206–219.

DasGupta, Anirban (1986b). A general theorem on decision theory for nonnegative functionals. Technical report #86-25, Department of Statistics, Purdue University.

DasGupta, Anirban and Berger, James (1986c). Estimation of multiple gamma scale-parameters: Bayes estimation subject to uniform domination. *Comm. in Statist., Theory and Methods,* Special issue on Stein-type multivariate estimation **15**, 2065–2086.

DasGupta, Anirban and Rubin, H. (1987a). Bayesian estimation subject to minimaxity of the mean of a multivariate normal distribution in the case of a common unknown variance. Statistical Decision Theory and Related Topics IV, S. S. Gupta and J. O. Berger eds., Volume 1, 325–345.

DasGupta, Anirban and Bose, A. (1987b). Γ-minimax and restricted risk Bayes estimation of multiple poisson means under ε-contaminations of the subjective prior. *To appear in Statistics and Decisions.*

Ferguson, T. S. (1962). Location and scale parameters on exponential families of distributions. *Ann. Math. Statist.* **33**, 986–1001.

Girshick, M. A. (1939). On the sampling theory of roots of determinantal equations. *Ann. Math. Statist.* **34**, 203–24.

Gleser, L. J. (1986). Improving inadmissible estimators under quadratic loss. *Stat. and Decisions* **4**, 37–43.

Haff, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8**, 586–597.

Hwang, J. T. (1982a). Improving upon standard estimators in discrete exponential families with applications to Poisson and Negative Binomial cases. *Ann. Statist.* **10**, 857–867.

Hwang, J. T. (1982b). Certain bounds on the class of admissible estimators in continuous exponential families. *Statistical Decision Theory and Related Topics III,* **2**, S. S. Gupta and J. Berger (eds.), Academic Press, New York, 15–29.

James, Ian R. (1979). Characterization of a family of distributions by the independence of size and shape variables. *Ann. Statist.* **7**, 869–881.

Johnstone, Iain (1987). On the inadmissibility of Stein's unbiased estimate of loss. Statistical Decision Theory and Related Topics IV, S. S. Gupta and J. O. Berger eds., Volume 1, 361–381.

Karlin, Samuel (1957). Pólya type distributions, II. *Ann. Math. Statist.* **28**, 281–308.

Marazzi, A. (1985). On constrained minimization of the Bayes risk for the linear model. *Stat. and Decisions* **3**, 277–296.

Roy, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220–238.

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

Shinozaki, N. (1984). Simultaneous estimation of location parameters under quadratic loss. *Ann. Statist.* **12**, 322–335.

Stein, C. (1975). Rietz lecture. 39*th* Annual meeting, IMS. Atlanta, Georgia.

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.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika* **24**, 471–494.