SIMULTANEOUS ESTIMATION IN THE MULTIPARAMETER GAMMA DISTRIBUTION UNDER WEIGHTED QUADRATIC LOSSES

by

Anirban Das Gupta

Technical Report #84-6

Department of Statistics
Purdue University

(Revised July 1985)

AMS 1970 Subject classifications: Primary 62C15, 62F10; Secondary 62H99

Key words and phrases. Loss functions, Inadmissibility, Gamma scale parameters, Empirical Bayes, Linear estimates, Risk-improvement.

Abstract

A new class of solutions to a general differential inequality often encountered in multiparameter estimation problems is obtained. Using these solutions as guidelines, inproved estimators for the scale-parameters as well as the natural parameters of independent Gamma distributions are obtained for a large class of weighted quadratic losses. The improved estimators have an empirical Bayes interpretation. They also permit an exact analytical representation of the risk-improvement. For the ordinary squared-error loss, a larger class of improved estimates is obtained which may allow for incorporation of prior information in choosing an alternative estimate. Numerical results are given which indicate the extent of risk-improvement in certain situations.

1. Introduction

Since the pioneering work of Stein (1956), a great amount of research has been done on exhibiting the presence of the Stein-effect in various probability structures with an infinite number of points in the sample space. Scattered inadmissibility results were gradually unified after the powerful technique of improving upon an inadmissible estimator via integration by parts was found by Stein (1973). The technique of explicitly constructing improved estimators by solving differential (difference) inequalities on the sample space has since been very productively used by many authors, notably, Hudson (1978), Berger (1980), Hwang (1982), and Ghosh, Hwang, and Tsui (1983). The beauty of the method lies in the facts that it often allows for consideration of a large number of losses of general quadratic form since the solutions to the differential inequalities often follow a general pattern and that the technique also offers one a choice from a big class of improved estimators. (For an indication of how this scope for choice leads to highly interesting selection problems, see Berger (1982)). However, one should perhaps mention in the same breath that the improved estimators thus obtained may be most extremely loss-specific and also moderately to highly unwiedly; in fact some of the recent skepticism about inadmissibility results has a lot to do with these undesirable features.

This paper deals primarily with simultaneous estimation of parameters in independent gamma distributions, although some of the results in the next section extend to the problem of estimating the vector of natural parameters in the general continuous exponential family. In Berger (1980), the problem of estimating the vector of scale-parameters $\theta_1^{-1}, \theta_2^{-2}, \ldots, \theta_p^{-1}$ of p independent Gamma distributions was considered for four different losses $\sum_{i=1}^{p} \theta_i^m(\theta_i\theta_i-1)^2$, where $m = -2, -1, 0,$ or 1, and explicit improved estimators were obtained. Unfortunately, the improved estimators looked completely different for the four different losses and their functional form did not allow for study of actual risk improvement except through possible simulation.(This of course is the typical picture in multiparameter estimation problems.) In the next section we shall treat the problem of estimating gamma scale-parameters under general weighted quadratic losses $\sum_{i=1}^{p} c_i\theta_i^{m_i}(\delta_i\theta_i-1)^2$ where c_i and m_i ($\neq 0$) are any constants. Losses of this kind will obviously include three of the four losses studied in Berger (1980) as special cases The excluded case $m = 0$ refers to the invariant quadratic loss and more of why the invariant loss has to be left out will be said later.) Note also that similar losses were considered by Hwang (1982) in the multiparameter Poisson problem and analogs of these losses in the normal case were considered in Brown (1980), although with a somewhat different purpose. Improved estimators are obtained for all these losses for $p \geq 2$ and it will be seen that the improved estimates are functionally similar and look alike for the different losses. Improved estimates are also obtained for a variety of losses for the natural parameters and again they are functionally similar for different losses.

It was shown by Berger (1980) how certain terms in his differential inequalities played the dominant role in obtaining improved estimators. It will be seen that solutions to the dominant inequality suggest improved estimators in all the problems we consider. These suggested estimators can then be shown to be actually dominating by exactly calculating the risk-difference. Towards this end, in the next section we obtain a new class of solutions to the general differential inequality

$$\Delta(x) = \psi(x) \sum_{i=1}^{p} v_i(x_i) \phi_i^{i(1)} + \sum_{i=1}^{p} w_i(x) \phi_i^2(x) < 0 \tag{1.1}$$

first studied in Berger (1980). Solutions were first found by him and then his class of solutions was extended by Ghosh and Parsian (1980) in the spirit of Efron and Morris (1976). Our solutions are new and these are then used to form possibly improved estimators in the gamma problems. We then calculate the risk-difference analytically and show that the estimators heuristically obtained are indeed improvements in terms of risk for all the losses mentioned above. The scale-parameters as well as the natural parameters are considered. Next, for the ordinary squared-error loss (m = -2), the inadmissibility results have been extended to give a broader class of improved estimators. For the ordinary squared-error loss, we also show that our class of improved estimators has an empirical Bayesian justification. The question of the actual amount of risk improvement is of great interest in practice. Since our improved estimators permit exact analytical representation of the risk improvement, we have studied this aspect analytically to some extent and then actually calculated the percentage risk-improvements in some situations. For the ordinary squared error loss, there is considerable improvement in terms of risk.

2. Construction of improved estimators

In this section we first obtain a new class of solutions to the differential inequality (1.1). The importance of this general differential inequality in multiparameter estimation is now very well known. See Berger (1980) and Ghosh and Parsian (1980) for an extensive discussion.

Theorem 1. Consider the differential inequality (1.1), and assume $\psi(x) > 0$. Define $g_i(x_i)$ as $g_i'(x_i) = \frac{1}{v_i(x_i)}$. Suppose for some $\alpha \neq \frac{1}{p+1}$, for some $k > 0$,

$$\frac{1}{\psi(x)} \sum_{i=1}^{p} w_i(x)|g_i(x_i)|^{2\alpha} \leq k \qquad \text{for all } \underset{\sim}{x} \tag{2.1}$$

Then $\underset{\sim}{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)$ with

$$\phi_i(x) = c \cdot \operatorname{sgn}\ g_i(x_i) \cdot \prod_{i=1}^{p} |g_i(x_i)|^{\frac{\alpha-1}{p}} \cdot |g_i(x_i)|^{\alpha} \tag{2.2}$$

solves $\Delta(x) < 0$, whenever $c\{\alpha(p+1)-1\} + kc^2 < 0$.

Proof: Clearly, for almost all $\underset{\sim}{x}$,

$$\phi_i^{i(1)}(x) = \frac{\partial}{\partial x_i}\phi_i(x) = c\alpha(\operatorname{sgn} g_i(x_i))^2|g_i(x_i)|^{\alpha-1}\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}} g_i'(x_i)$$

$$+ c\frac{(\alpha-1)}{p}(\operatorname{sgn} g_i(x_i))^2|g_i(x_i)|^{\alpha-1}\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}} g_i'(x_i)$$

$$= \frac{c[\alpha(p+1)-1]}{p}|g_i(x_i)|^{\alpha-1}\cdot\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}}\cdot g_i'(x_i) \qquad (2.3)$$

Hence,

$$\Delta(\underset{\sim}{x}) = \psi(x)\cdot\frac{c[\alpha(p+1)-1]}{p}\cdot\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}}\cdot\sum_{i=1}^{p}|g_i(x_i)|^{\alpha-1}$$

$$+ c^2\prod_{i=1}^{p}|g_i(x_i)|^{\frac{2(\alpha-1)}{p}}\cdot\sum_{i=1}^{p}w_i(x)\,|g_i(x_i)|^{2\alpha}$$

$$= \psi(x)\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}}\left[\frac{c[\alpha(p+1)-1]}{p}\sum_{i=1}^{p}|g_i(x_i)|^{\alpha-1}\right.$$

$$\left. + \frac{c^2}{\psi(x)}\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}}\cdot\sum_{i=1}^{p}w_i(x)|g_i(x_i)|^{2\alpha}\right]$$

$$\leq \psi(x)\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}}\left[\frac{c[\alpha(p+1)-1]}{p}\sum_{i=1}^{p}|g_i(x_i)|^{\alpha-1} + c^2k\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}}\right]$$

$$\leq \psi(x)\prod_{i=1}^{p}|g_i(x_i)|^{\frac{\alpha-1}{p}}\left[\frac{c[\alpha(p+1)-1]}{p}\sum_{i=1}^{p}|g_i(x_i)|^{\alpha-1} + \frac{c^2k}{p}\sum_{i=1}^{p}|g_i(x_i)|^{\alpha-1}\right]$$

$$\leq 0, \text{ if } c\{\alpha(p+1)-1\} + kc^2 < 0 \qquad (2.4)$$

Remarks 1. Under appropriate conditions similar to condition (2.1) above, a class of solutions to $\Delta(x) < 0$ was found in Berger (1980). The solutions there were essentially of the form

$$\phi_i(\underset{\sim}{x}) = \frac{-c\; g_i(x_i)}{b + \sum_{i=1}^{p} |g_i(x_i)|^{\beta}} \tag{2.5}$$

where b, c, β are suitable positive numbers. The constant c can be generalized to an appropriate function $\tau(\underset{\sim}{x})$ (See Ghosh and Parsian (1980)).

2. The new class of solutions in (2.2) above can be used to anticipate improved estimates of the scale as well as the natural parameters of independent gamma variables under a wide class of weighted quadratic losses. At this stage, we merely mention that if the loss is $\sum_{i=1}^{p} \eta_i^{m_i}(\delta_i\eta_i - 1)^2$ (where $\eta_i = \theta_i$ or $\frac{1}{\theta_i}$), one may heuristically arrive at $\frac{x_i}{\alpha+1}\,[1 + c_i x_i^{\frac{m_i}{2}}(\prod_{i=1}^{p} x_i^{\frac{-m_i}{2p}})]$ and $\frac{\alpha_i - 2}{x_i}\,[1 + c_i x_i^{\frac{-m_i}{2}}(\prod_{i=1}^{p} x_i^{\frac{m_i}{2p}})]$ as possibly improved estimators of the scale and natural parameters respectively by considering certain dominant parts of the relevant differential inequalities (see Berger (1980)).

We now actually prove that the estimators heuristically obtained above dominate the standard estimators in terms of risk.

<u>Theorem 2</u> Let $X_i \overset{\text{indep.}}{\sim}$ Gamma (α_i,θ_i), where α_i's are considered known. Consider the problem of estimating $(\theta_1^{-1},\theta_2^{-1},\ldots,\theta_p^{-1})$ under a loss

$L(\underset{\sim}{\theta},\underset{\sim}{\delta}) = \sum_{i=1}^{p} c_i\theta_i^{m_i}(1-\delta_i\theta_i)^2$ where $c_i > 0$, $m_i \neq 0$, $i = 1,2,\ldots,p$. Consider the estimates $\delta_0(x)$ and $\delta(x)$ defined as

$$\delta_{o,i}(\underset{\sim}{x}) = \frac{x_i}{\alpha_i+1}$$

$$\delta_i(\underset{\sim}{x}) = \frac{x_i}{\alpha_i+1}(1+\phi_i(x))$$

where

$$\phi_i(x) = -c(\text{sgn}\,m_i)x_i^{\frac{m_i}{2}}\left(\prod_{i=1}^{p} x_i^{-\frac{m_i}{2p}}\right) \qquad (2.6),$$

$c > 0$ sufficiently small (see (2.10).

Then $R(\theta,\delta) < R(\theta,\delta_0)$ for every $\underset{\sim}{\theta}$ if $p \geq 2$ and hence $\delta_0(X)$ is inadmissible.

Proof: Letting $\Delta(\theta) = R(\theta,\delta) - R(\theta,\delta_0)$,

$$\Delta(\theta) = E\left[\sum_{i=1}^{p} c_i\theta_i^{m_i}\left\{\frac{x_i}{\alpha_i+1}\theta_i-1 + \frac{x_i\phi_i(x)}{\alpha_i+1}\theta_i\right\}^2 - \sum_{i=1}^{p} c_i\theta_i^{m_i}\left\{\frac{x_i}{\alpha_i+1}\theta_i-1\right\}^2\right]$$

$$= E_\theta\left[\sum_{i=1}^{p} c_i\theta_i^{m_i+2}\frac{x_i^2}{(\alpha_i+1)^2}\phi_i^2(x) + 2\sum_{i=1}^{p} c_i\theta_i^{m_i+2}\frac{x_i^2}{(\alpha_i+1)^2}\phi_i(x)\right.$$

$$\left. - 2\sum_{i=1}^{p} c_i\theta_i^{m_i+1}\frac{x_i}{\alpha_i+1}\phi_i(x)\right] \qquad (2.7)$$

Assuming that $\alpha_i > \frac{m_i}{p}$, $\alpha_i + 2 > \frac{m_i(1-p)}{p}$ and $\alpha_i + 1 > \frac{m_i(1-p)}{2p}$, by direct calculations,

$$E\, x_i^2\phi_i^2(x) = \frac{c^2\,\Gamma(\alpha_i+m_i+2-\frac{m_i}{p})}{\Gamma(\alpha_i-\frac{m_i}{p})}\cdot\theta_i^{-m_i-2}\prod_{j=1}^{p}\frac{\Gamma(\alpha_j-\frac{m_j}{p})}{\Gamma\alpha_j}\cdot\theta_j^{\frac{m_j}{p}}$$

$$E\, x_i^2\phi_i(x) = \frac{-c(\mathrm{sgn}\, m_i)\;\Gamma(\alpha_i+2+\frac{m_i}{2}-\frac{m_i}{2p})}{\Gamma(\alpha_i-\frac{m_i}{2p})}\,\theta_i^{-2-\frac{m_i}{2}}\prod_{j=1}^{p}\frac{\Gamma(\alpha_j-\frac{m_j}{2p})}{\Gamma\alpha_j}\,\theta_j^{\frac{m_j}{2p}}$$

and

$$E\, x_i\phi_i(x) = \frac{-c(\mathrm{sgn}\, m_i)\;\Gamma(\alpha_i+1+\frac{m_i}{2}-\frac{m_i}{2p})}{\Gamma(\alpha_i-\frac{m_i}{2p})}\cdot\theta_i^{-1-\frac{m_i}{2}}\prod_{j=1}^{p}\frac{\Gamma(\alpha_j-\frac{m_j}{2p})}{\Gamma\alpha_j}\,\theta_j^{\frac{m_j}{2p}} \qquad (2.8)$$

Setting

$$\frac{\Gamma(\alpha_i+m_i+2-\frac{m_i}{p})}{\Gamma(\alpha_i-\frac{m_i}{p})} = a_i \qquad \frac{\Gamma(\alpha_i-\frac{m_i}{p})}{\Gamma\alpha_i} = b_i$$

$$\frac{\Gamma(\alpha_i+1+\frac{m_i}{2}-\frac{m_i}{2p})}{\Gamma(\alpha_i-\frac{m_i}{2p})} = d_i \qquad \text{and} \qquad \frac{\Gamma(\alpha_i-\frac{m_i}{2p})}{\Gamma\alpha_i} = e_i,$$

(2.7) and (2.8) give

$$\Delta(\theta) = c^2 \sum_{i=1}^{p} \frac{c_i a_i}{(\alpha_i+1)^2} \cdot \left(\prod_{i=1}^{p} b_i \theta_i^{\frac{m_i}{p}}\right) - 2c \sum_{i=1}^{p} \frac{c_i(\operatorname{sgn} m_i)}{(\alpha_i+1)^2}\left(\alpha_i+1+\frac{m_i}{2}-\frac{m_i}{2p}\right) d_i \theta_i^{\frac{m_i}{2}} \left(\prod_{i=1}^{p} e_i \theta_i^{\frac{m_i}{2p}}\right)$$

$$+ 2c \sum_{i=1}^{p} \frac{c_i(\operatorname{sgn} m_i)}{(\alpha_i+1)} d_i \theta_i^{\frac{m_i}{2}} \left(\prod_{i=1}^{p} e_i \theta_i^{\frac{m_i}{2p}}\right)$$

$$= c^2 k \prod_{i=1}^{p} b_i \theta_i^{\frac{m_i}{p}} - 2c \sum_{i=1}^{p} \frac{c_i}{(\alpha_i+1)^2} \frac{|m_i|(p-1)}{2p} d_i \theta_i^{\frac{m_i}{2}} \left(\prod_{i=1}^{p} e_i \theta_i^{\frac{m_i}{2p}}\right), \qquad (2.9)$$

where $k = \sum_{i=1}^{p} \frac{c_i a_i}{(\alpha_i+1)^2}$.

Let $b = \max b_i$, $e = \min e_i$ and $d = \min \frac{c_i |m_i| d_i}{(\alpha_i+1)^2}$. Then (2.9) yields,

$$\Delta(\theta) \le c^2 k\, b^p \prod_{i=1}^{p} \theta_i^{\frac{m_i}{p}} - \frac{cde^p(p-1)}{p} \prod_{i=1}^{p} \theta_i^{\frac{m_i}{2p}} \sum_{i=1}^{p} \theta_i^{\frac{m_i}{2}}$$

$$\le c \prod_{i=1}^{p} \theta_i^{\frac{m_i}{p}} \{ckb^p - de^p(p-1)\}$$

$$< 0, \quad \text{if } 0 < c < \frac{de^p(p-1)}{kb^p}. \qquad (2.10)$$

This completes the proof of Theorem 2.

Remarks 1. The bounds following (2.9) have been rather crude; for the special case $\alpha_i \equiv \alpha$, $c_i \equiv 1$ and $m_i \equiv m$, (2.9) gives

$$\Delta(\theta) = \frac{c^2 p}{(\alpha+1)^2} \frac{\Gamma(\alpha+m+2-\frac{m}{p})}{\Gamma(\alpha-\frac{m}{p})} \cdot \left(\frac{\Gamma(\alpha-\frac{m}{p})}{\Gamma\alpha}\right)^p \cdot \prod_{i=1}^{k} \theta_i^{\frac{m}{p}} - \frac{c|m|(p-1)}{(\alpha+1)^2 p}$$

$$\cdot \frac{\Gamma(\alpha+1+\frac{m}{2}-\frac{m}{2p})}{\Gamma(\alpha-\frac{m}{2p})} \sum_{i=1}^{p} \theta_i^{\frac{m}{2}} \cdot \left(\frac{\Gamma(\alpha-\frac{m}{2p})}{\Gamma\alpha}\right)^p \prod_{i=1}^{p} \theta_i^{\frac{m}{2p}} \qquad (2.11)$$

< 0

if $0 < c < \frac{|m|(p-1)}{p} \cdot \frac{\Gamma(\alpha+1+\frac{m}{2}-\frac{m}{2p})(\Gamma(\alpha-\frac{m}{2p}))^{p-1}}{\Gamma(\alpha+m+2-\frac{m}{p})\ (\Gamma(\alpha-\frac{m}{p}))^{p-1}}$.

Further specializing to the case m = -2 (ordinary squared-error loss), the range of c is $0 < c < \frac{2(p-1)}{p} \cdot \left(\frac{\Gamma(\alpha+\frac{1}{p})}{\Gamma(\alpha+\frac{2}{p})}\right)^p$. It is clear that there is no unique value of c which maximizes the improvement in risk. However, from (2.11) it is clear that the upper bound on $\Delta(\theta)$ is minimized at $c = \frac{(p-1)}{p} \left(\frac{\Gamma(\alpha+\frac{1}{p})}{\Gamma(\alpha+\frac{2}{p})}\right)^p$. Thus there is no optimal choice of c but if one has to choose one, the mid-point of the allowable values may be a natural choice. Recent results in Berger and Das Gupta (1985), however, indicate that the upper bound of c may be the most appropriate choice in many restricted risk Bayes problems.

2. For the squared-error loss ($m_i \equiv -2$), the improved estimate shifts by a multiple of the geometric mean. In a recent paper, Das Gupta and Sinha (1984) have shown that for estimating $\sum_{i=1}^{p} \ell_i \theta_i^{-1}$ under squared-error loss, $\sum_{i=1}^{p} \frac{\ell_i X_i}{\alpha_i+1}$ is inadmissible and the improved estimate also shifts by a multiple of the geometric mean.

3. The shrinkage behavior of our improved estimators follows the same pattern as the improved estimators in Berger (1980); thus if $m_i < 0$, the improved estimate expands the natural estimate of that coordinate and if $m_i > 0$, the improved estimate shifts by a negative quantity.

4. For the invariant quadratic loss $\sum_{i=1}^{p} (1-\delta_i\theta_i)^2$, the estimate $\delta(\underset{\sim}{x})$ with $\delta_i(x) = \frac{x_i}{\alpha+1} + \frac{c}{\alpha+1} (\text{sgnlog } x_i)\, x_i \prod_{i=1}^{p} |\log x_i|^{-\frac{1}{p}}$ is suggested as the alternative estimator by the solutions to $\Delta(x) < 0$ in Theorem 1. For such an estimate, an analytical representation of the risk is difficult to obtain; hence the natural way to prove that it dominates the standard estimate would be by using the technique of solving an exact differential inequality (not just the dominant terms); as is well known, certain tail and integrability restrictions must be imposed on the solutions for inadmissibility to be proved. These integrability conditions are not met by the estimators described above and

hence the invariant loss has to be left out. We remark that for the invariant loss the improved estimates obtained in Berger (1980) are probably the most natural because on making a log transform, they resemble the James-Stein estimators.

5. Finally note that the improved estimates of Theorem 2 allow for smaller values of the shape-parameters α_i to be accommodated than could previously be done. For example, if all $m_i \equiv 1$, then Theorem 2 only requires $\alpha_i > \frac{1}{p}$ for every i, while Berger (1980) requires $\alpha_i \geq 4$ for every i if α_i's are possibly unequal. Note that for large p, $\alpha_i > \frac{1}{p}$ will probably be satisfied anyway.

We now state a general inadmissibility theorem for estimating the natural parameters. The proof will be omitted because of similarity to the proof of Theorem 2.

Theorem 3. Consider the problem of estimating the natural parameters $(\theta_1,\theta_2,\ldots,\theta_p)$ of p independent gamma distributions under a loss $L(\underset{\sim}{\theta},\underset{\sim}{\delta}) = \sum_{i=1}^{m} c_i\theta_i^{m_i}(\frac{\delta_i}{\theta_i} - 1)^2$ where $c_i > 0$, $m_i \neq 0$ are some constants; assume $\alpha_i > \frac{m_i}{p}$, $\alpha_i - 2 > \frac{m_i(1-p)}{p}$ and $\alpha_i - 1 > \frac{m_i(1-p)}{2p}$. Consider the estimates $\delta_0(\underset{\sim}{x})$ and $\delta(\underset{\sim}{x})$ defined as

$$\delta_{0,i}(\underset{\sim}{x}) = \frac{\alpha_i-2}{x_i}$$

$$\delta_i(\underset{\sim}{x}) = \frac{\alpha_i-2}{x_i}\,(1+c(\mathrm{sgn}\,m_i)\,x_i^{\frac{m_i}{2}}\,(\prod_{i=1}^{p} x_i^{\frac{-m_i}{2p}})), \tag{2.12}$$

where c is a sufficiently small positive number.

Then $R(\theta,\delta) < R(\theta,\delta_0)$ for every θ if $p \geq 2$ and hence $\delta_0(X)$ is inadmissible.

As in the estimation of the scale-parameters, the invariant loss cannot be handled. Note that the ordinary squared-error loss now corresponds to $m_i \equiv 2$ and the restrictions on α_i are automatically satisfied since any way $\alpha_i > 2$. Also, as before, the improved estimates for the different losses are functionally similar. For the special squared-error loss, each coordinate is shifted by the reciprocal of the geometric mean. This fact brings out a natural similarity between estimating the θ_i's and the θ_i^{-1}'s in the sense that the shift by the improved estimate in one problem is just the reciprocal of the shift in the other problem. In other words, the reciprocal transformation on the parametric function is exactly reflected also in the improved estimate. Finally note that so far inadmissibility of the usual estimate of the natural parameters was known only for the squared-error loss (see Berger (1980)); Theorem 3 establishes inadmissibility for a wide class of weighted quadratic losses.

We now get back to estimating the scale-parameters and generalize Theorem 2 in two directions for the ordinary squared-error loss. For notational simplicity, we have taken $\alpha_i \equiv \alpha$ but the proofs go through with arbitrary α_i's.

Theorem 4. Consider the situation in Theorem 2 with $m_i \equiv -2$. Let $\delta(X)$ be any estimate given as

$$\delta_i(x) = \frac{x_i}{\alpha+1} + \frac{1}{\alpha+1}\left[\alpha r(t) + \frac{tr'(t)}{p}\right], \tag{2.13}$$

where t stands for $\left(\prod_{i=1}^{p} x_i\right)^{\frac{1}{p}}$ and $r(\cdot)$ is such that

(i) $0 < \frac{r(t)}{t} \leq \frac{2(p-1)}{p(\alpha+1)^2}$

(ii) $r(t)$ is non-decreasing

and

(iii) $r(t)/t$ is non-increasing

Then $\delta(\underset{\sim}{X})$ dominates $\frac{X}{\alpha+1}$.

Proof: Define $h_i(x) = x_i^\alpha\, r(t)$. In view of Berger (1980), it is enough to show h_i solves (2.7) there and that with this choice of h_i, the improved estimate is as in the statement of Theorem 4.

First note $h_i^{i(1)}(x) = \alpha x_i^{\alpha-1}\, r(t) + \frac{1}{p}\, x_i^{\alpha-1}\, tr'(t)$ (2.14)

Hence, the ith coordinate of the improved estimate is

$$\delta_i(x) = \frac{x_i}{\alpha+1} + \frac{x_i^{1-\alpha_i} h_i^{i(1)}(x)}{\alpha+1} = \frac{x_i}{\alpha+1} + \frac{\alpha r(t)}{\alpha+1} + \frac{tr'(t)}{p(\alpha+1)} ,$$

which is of the form (2.12).

Next, (2.7) in Berger (1980) is equivalent to

$$\Delta_{-2}(\underset{\sim}{x}) = -\frac{2r(t)}{\alpha+1} \sum_{i=1}^{p} x_i + \frac{2}{(\alpha+1)^2} \left(\alpha r(t) + \frac{tr'(t)}{p}\right) \sum_{i=1}^{p} x_i$$

$$+ \frac{p}{(\alpha+1)^2} \left(\alpha r(t) + \frac{tr'(t)}{p}\right)^2$$

$$= \frac{2}{(\alpha+1)^2} \sum_{i=1}^{p} x_i \cdot \left(\frac{tr'(t)}{p} - r(t)\right) + \frac{p}{(\alpha+1)^2} \left(\alpha r(t) + \frac{tr'(t)}{p}\right)^2$$

$$\le \frac{p(\alpha+\frac{1}{p})^2}{(\alpha+1)^2} r^2(t) - \frac{2(p-1)}{p(\alpha+1)^2} \sum_{i=1}^{p} x_i \quad r(t)$$

(since $0 \le \frac{tr'(t)}{p} \le \frac{r(t)}{p}$)

$$\le \frac{p(\alpha+\frac{1}{p})^2}{(\alpha+1)^2} r^2(t) - \frac{2(p-1)}{(\alpha+1)^2} \quad tr(t)$$

$$< 0.$$

Remarks. 1. The upper bound on $\frac{r(t)}{t}$ can be increased to $\frac{2(p-1)}{p(\alpha+\frac{1}{p})^2}$. However, for large p, this is likely to be immaterial.

2. With $r(t) = ct$ $(0 < c \le \frac{2(p-1)}{p(\alpha+1)^2})$, one gets the improved estimates of Theorem 2. One advantage of providing an extended class of improved estimators like in Theorem 4 is that there is more scope of incorporating prior information in choosing an alternative estimator (see Berger (1982)) and also that in this extended class one may find actually an alternative estimator which is admissible. In the normal problem, for example, the admissible minimax estimators were found from such an enlarged class of improved estimators (see Strawderman (1971) and Berger (1976)).

3. With $t = (\prod_{i=1}^{p} x_i)^{\frac{1}{2p}}$, a similar extended class of improved estimators is easy to find in the lines of Theorem 4 when $m_i \equiv -1$.

4. It is easy to show that for a broader class of linear estimates $\delta_0(\underset{\sim}{X}) = A\underset{\sim}{X}$ of the scale parameters, uniform risk domination can be achieved by

shifting by the geometric mean when the loss is squared-error. We have been able to prove that if the elements a_{ij} of A are such that $(\alpha_j + \frac{1}{p})\sum_{i=1}^{p} a_{ij} < 1(>1)$ for every $j \geq 1$, then $A\underset{\sim}{X}$ is inadmissible and $\delta(\underset{\sim}{X}) = A\underset{\sim}{X} + c(\prod_{i=1}^{p} X_i)^{\frac{1}{p}} \cdot \underset{\sim}{1}$ is a better estimator for suitable constants c. Many of such linear estimates $A\underset{\sim}{X}$, however, can be uniformly dominated in risk by other linear estimates and unfortunately, we have not been able to characterize all the admissible linear estimates in this case. Such characterizations were obtained in the normal case by Cohen (1966) and in the Poisson case by Brown and Farrell (1985).

3. Empirical Bayes interpretation

In the results presented so far, the emphasis has been on establishing inadmissibility. However, the story doesn't end in merely knowing that a particular estimator is inadmissible; in fact, if we may say so, the interesting problems arise exactly at this point. Perhaps the most important and most interesting question that needs to be answered is whether one can build up improved estimators which conform to one's prior beliefs about the unknown parameters. A lot of research has been done on these questions for problems involving the normal distribution and some results are known also in the case of Poisson distribution; in particular, in both problems, proper Bayes improved estimators are known for large enough p; also, the James-Stein (1960) and Clevenson-Zidek (1975) estimators have long been known to have an empirical Bayesian justification. In the Gamma problem, finding Bayes or generalized Bayes improved estimators seems difficult as the so-called "conditionally conjugate" two-stage priors are analytically intractable. In what follows, we show that the class of improved estimators obtained in Theorem 2 contain certain natural empirical Bayes estimators.

Assume $\theta_i \overset{\text{ind.}}{\sim}$ Gamma $(2,r)$, where $r^{-1} > 0$ is an unknown scale-parameter. For squared-error loss, the Bayes estimate (given r) of θ_i^{-1} is

$$\delta_i(\underset{\sim}{x}) = \frac{x_i+r}{\alpha+1} . \tag{3.1}$$

Marginally, the x_i's are iid with the joint p.d.f.

$$f(x_1,\ldots,x_p) \propto \frac{\prod_{i=1}^{p} x_i^{\alpha-1}}{\prod_{i=1}^{p} (x_i+r)^{\alpha+2}} .$$

In particular, the density of each x_i is

$$f(x_i) = \frac{1}{rB(\alpha,2)} \frac{(\frac{x_i}{r})^{\alpha-1}}{(1+\frac{x_i}{r})^{\alpha+2}} \tag{3.2}$$

Clearly, r is a scale-parameter for the distribution of $x_1, x_2,\ldots,x_p$. The best scale-invariant estimate of r in the restricted class of estimators which depends only on the geometric mean is

$$a_0(\underset{\sim}{x}) = c_0(\pi x_i)^{\frac{1}{p}},$$

$$\text{where} \quad c_0 = \frac{E_{r=1}(\prod_i x_i)^{\frac{1}{p}}}{E_{r=2}(\prod_i x_i)^{\frac{2}{p}}} = \frac{(\Gamma(\alpha+\frac{1}{p})/\Gamma\alpha)^p \cdot (\Gamma(2-\frac{1}{p})/\Gamma 2)^p}{(\Gamma(\alpha+\frac{2}{p})/\Gamma\alpha)^p \cdot (\Gamma(2-\frac{2}{p})/\Gamma 2)^p}$$

$$= \frac{(\Gamma(\alpha+\frac{1}{p})\,\Gamma(2-\frac{1}{p}))^p}{(\Gamma(\alpha+\frac{2}{p})\,\Gamma(2-\frac{2}{p}))^p} \tag{3.3}$$

The empirical Bayes estimate is what results when $a_0(x)$ is substituted for r in (3.1). It can be verified that $c_0 < \frac{2(p-1)}{p} \left(\frac{\Gamma(\alpha+\frac{1}{p})}{\Gamma(\alpha+\frac{2}{p})}\right)^p$ for $p \geq 2$. Hence this empirical Bayes estimate dominates the usual estimate under ordinary squared error loss. (See Remark 1 following Theorem 2.)

The "empirical Bayesian" interpretation of the improved estimate is somewhat ad hoc because of the restriction to the class of estimators for r which depend only on the geometric mean. The restriction, however, seems necessary because the usual Pitman estimate of r derived from the marginal distribution of all X_i's does not have an easy analytical form.

Note that for the four losses dealt in Berger (1980), no Bayesian interpretation of the improved estimators obtained there is known except that for $m_i \equiv -1$, an approximation to the generalized Bayes estimate against certain flat priors looks something like Berger's improved estimate (see Brown and Hwang (1982)). It will clearly be interesting to obtain improved estimates which can be actually linked with reasonable priors.

4. Risk-improvement

Whereas from a theoretical stand-point uniform domination is of interest, to the applied statistician is of utmost importance the question: how much risk improvement can be achieved in practice? For most problems, computer simulation is necessary to get an idea of the actual risk-improvement because the improved estimators are almost always such that any analytical representation of the risk difference is at least formidable, if not impossible. Fortunately, however, our improved estimators are such that such an exact analytical expression is possible to obtain; (2.9) and (2.10) do in fact give the risk-improvements $\Delta(\theta)$. Consequently, simulation of data will be quite unnecessary. Since ordinary risk-improvement can be unbounded, it is more meaningful to deal with the percentage risk-improvement.

Assume $\alpha_i \equiv \alpha$ and $m_i \equiv m(\neq 0)$. Using

$$c = \frac{|m|(p-1)\Gamma(\alpha+1+\frac{m}{2}-\frac{m}{2p})(\Gamma(\alpha-\frac{m}{2p}))^{p-1}}{2p\Gamma(\alpha+m+2-\frac{m}{p})(\Gamma(\alpha-\frac{m}{p}))^{p-1}}$$

(i.e., the mid-point of the allowed range for c), by direct calculations using (2.9),

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

$$= \frac{m^2(p-1)^2\ (\Gamma(\alpha+1+\frac{m}{2}-\frac{m}{2p}))^2(\Gamma(\alpha-\frac{m}{2p}))^{2p-2}}{4p(\alpha+1)\Gamma(\alpha+m+2-\frac{m}{p})(\Gamma(\alpha-\frac{m}{p}))^{p-1}\ (\Gamma\alpha)^p}\ \frac{\left[\frac{2}{p}\sum_{i=1}^{p}\theta_i^{\frac{m}{2}} - \prod_{i=1}^{p}\theta_i^{\frac{m}{2p}}\right]\prod_{i=1}^{p}\theta_i^{\frac{m}{2p}}}{\sum_{i=1}^{p}\theta_i^{m}} \qquad (4.1)$$

It may be interesting to find a measure of some kind of average percentage risk improvement in various parts of the parameter space. In particular, of some statistical interest is the limiting value (as $p\to\infty$) of the average (with respect to Lebesgue measure) percentage risk improvement when all θ_i belong to an interval [a,b]. Here, a could be thought of as the prior guess for $\min_{1\le i\le p} \theta_i$ and b as $\max_{1\le i\le p} \theta_i$.

Specializing to the ordinary squared-error loss, from (4.1),

$$\lim_{p\to\infty} \int \frac{R(\underset{\sim}{\theta},\delta_0) - R(\underset{\sim}{\theta},\delta)}{R(\underset{\sim}{\theta},\delta_0)}\, d\mu$$

$$= \lim_{p\to\infty} \frac{(p-1)^2\left(\Gamma(\alpha+\frac{1}{p})\right)^{2p}}{p^2(\alpha+1)\left(\Gamma(\alpha+\frac{2}{p})\right)^{p}(\Gamma 2)^p} \int \frac{\left(\frac{2}{p}\sum_{i=1}^{p}\theta_i^{-1} - \prod_{i=1}^{p}\theta_i^{-\frac{1}{p}}\right)\prod_{i=1}^{p}\theta_i^{-\frac{1}{p}}}{\frac{1}{p}\sum_{i=1}^{p}\theta_i^{-2}}\, d\mu \quad (4.2),$$

where μ stands for the normalized Lebesgue measure on $[a,b]^p$.

The first term in (4.2) monotonically increases to $\frac{1}{\alpha+1}$ as $p\to\infty$; also, since θ_i's are i.i.d. uniform, by the strong law of large numbers,

$$\lim_{p\to\infty} \int \frac{\left(\frac{2}{p}\sum_{i=1}^{p}\theta_i^{-1} - \prod_{i=1}^{p}\theta_i^{-\frac{1}{p}}\right)\prod_{i=1}^{p}\theta_i^{-\frac{1}{p}}}{\frac{1}{p}\sum_{i=1}^{p}\theta_i^{-2}}\, d\mu$$

$$= \lim_{p\to\infty} \frac{\int\left(\frac{2}{p}\sum_{i=1}^{p}\theta_i^{-1} - \prod_{i=1}^{p}\theta_i^{-\frac{1}{p}}\right)\prod_{i=1}^{p}\theta_i^{-\frac{1}{p}}\, d\mu}{\int \frac{1}{p}\sum_{i=1}^{p}\theta_i^{-2}\, d\mu}$$

$$= \lim_{p} \frac{2p\left(a^{-\frac{1}{p}} - b^{-\frac{1}{p}}\right)\cdot\left(\frac{p}{p-1}\right)^{p-1}\left(b^{1-\frac{1}{p}} - a^{1-\frac{1}{p}}\right)^{p-1} - \left(\frac{p}{p-2}\right)^{p}\left(b^{1-\frac{2}{p}} - a^{1-\frac{2}{p}}\right)^{p}}{\left(\frac{1}{a} - \frac{1}{b}\right)(b-a)^{p-1}}$$

$$= \frac{2ec\,\log c\, e^{-\frac{c\log c}{c-1}} - e^{2}c(c-1)e^{-\frac{2c\log c}{c-1}}}{c-1}, \qquad (4.3)$$

where $c = \frac{b}{a}$.

In particular, along the line in which all θ_i's are equal to some θ (i.e., c = 1), by L'Hospital's rule, the expression in (4.3) is equal to 1; hence, the percentage risk improvement along this ray can be considerable for small α, approaching 100% as $p \to \infty$, $\alpha \to 0$.

Interestingly, the limiting average risk improvement depends solely on the ratio $c = \frac{b}{a}$ and decreases monotonically as c increases. We will later provide actual values of this limit for various c.

Thus for m = -2, encouraging risk-improvements seem to be attainable. For other values of m it is difficult to obtain neat expressions for the iterated supremum as above since (4.1) is no longer so simple to handle. For m = 1, an iterated supremum as was calculated in Berger (1980b) for his improved estimator and it was found that up to approximately 10% risk-improvement is possible (as $\alpha \to 0$) along the same ray as we have considered. We do not have any corresponding results for m = 1 so that a direct comparison is not possible. However, certain numerical studies have led us to believe that for m = 1, Berger's (1980) estimators will usually give better percentage risk-improvement.

Since (4.3) gives an idea of attainable risk-improvement only for the squared-error loss, it is desirable to pursue this question in other situations as well. We have provided some numerical observations below. We repeat that no computer simulation was done because it is not necessary to do it. Percentage risk-improvements are shown for different ranges of θ_i's and different m. The improvements were calculated for a fixed set of random θ_i's uniformly distributed in the indicated range. For all the losses, percentage risk improvements are shown for corresponding to the constant c which is the mid-point of the allowed range. The values indicate that the percentage improvements are best for the squared-error loss; also, the improvements seem to be better for $m < 0$. Finally, the percentage improvements are larger for larger p. Throughout we have taken independent simple exponential distributions (i.e., $\alpha_i = 1$).

Table of % risk improvements

Range of θ's	m = -2		m = -1		m = 1	
	p=5	p=10	p=5	p=10	p=5	p=10
$(0,5]^p$	12.26	18.25	4.55	6.33	2.18	2.77
$[10,15]^p$	24.57	35.02	5.45	7.34	2.43	3.02
$(0,15]^p$	7.81	9.96	3.69	5.12	1.80	2.45

Table of limiting average percentage risk-improvements ($\alpha_i \equiv 1$)

c	Limiting risk-improvement
1	50
1.5	49.33
2	48.03
3	45.16
4	42.48
5	40.10
10	31.70
15	26.60
30	18.61
50	13.74
100	8.70

Acknowledgment.

I am deeply grateful to James Berger and an associate editor for their invaluable attention and constructive suggestions in the preparation of this paper. Sincere thanks are also due to Malay Ghosh and Herman Rubin for helpful suggestions, and to Marty Meyer for her support and encouragement during this work.

References

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

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

Berger, James O. (1980b). Risk improvement in Gamma problem (Private communication).

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

Berger, James O. and Das Gupta, Anirban (1985). Estimation of multiple gamma parameters: Bayes estimation subject to uniform domination. Technical Report #85-13, Department of Statistics, Purdue University.

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

Brown, L. D. and Farrell, R. H. (1985). All admissible linear estimates of a multivariate poisson mean. Ann. Statist. 13 (282-294).

Brown, L. D. and Hwang, J. T. (1982). A unified admissibility proof. Statistical Decision Theory and related topics III, volume 1 (205-230).

Clevenson, M. L. and Zidek, J. V. (1975). Simultaneous estimation of the mean of independent Poisson laws. J. Amer. Statist. Assoc. 70 (698-705).

Cohen, Arthur (1966). All admissible linear estimates of the mean vector. Ann. Math. Statist. 37 (458-463).

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

Das Gupta, Anirban and Sinha, Bimal K. (1984). Estimation in multiparameter Exponential family: Admissibility and inadmissibility results (To appear in Statistics and Decision).

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

Ghosh, M. Hwang, J. T. and Tsui, K. W. (1984.). Construction of improved estimators in multiparameter estimation for continuous exponential families. Jour. Mult. Analysis. 14 (212-220).

Ghosh, M. and Parisian, A. (1980). Admissible and minimax multiparameter estimation in exponential families. Jour. Mult. Analysis. 10 (551-564).

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

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

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

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).

Strawderman, W. E. (1971). Proper Bayes minimax estimators for the mean of a multivariate normal population. Ann. Math. Statist. 42 (385-388).