Simultaneous estimation of arbitrary scale-parameters under arbitrary quadratic loss

by

Anirban DasGupta
Purdue University

Technical Report #84-32

Purdue University
West Lafayette, Indiana

August 1984
(Revised December 1985)

1. Introduction

Ever since the monumental work of Stein (1956) on the multivariate normal, much research has been done to show that Stein-effect is a very general phenomenon, having little to do with continuity of the underlying probability structure and often the exact form of the loss-function. Emphasis, for natural reasons, has however shifted now to more delicate questions of greater practical importance. Decision-theorists now often want to know how good and usable the alternative estimators are; for example, are they simple and do they allow for considerable frequentist risk improvement (at least in some parts of the parameter space)? Some of us are also concerned equally about the loss-specific nature of many of these Stein-type improved estimators and recognize the paramount need to build up such improved estimators which are consistent with prior beliefs about the parameters. Many of these problems have been successfully addressed in the multivariate normal situation; see the more recent works of Berger (1980a, 1982a, 1982b, 1983). Outside the normal distributions, however, these questions remain largely unanswered, mostly because of the difficulties of obtaining neat answers to most of these questions. The objective of this paper is to briefly touch on some of these questions in the general scale-parameter family; the flavor is still mostly frequentist and the relatively neater results obtained are also frequentist in nature.

It was shown in Brown (1966) that the best invariant estimate of a location vector is typically inadmissible for $p \geq 3$ under very general classes of loss-functions, not necessarily convex. Shinozaki (1984) obtained explicit James-Stein type improved estimators of a vector of location parameters in many important cases. On the other hand, the best equivariant estimate of p independent gamma scale-parameters was shown to be inadmissible in Berger (1980b) and relatively simpler improved estimators with come empirical Bayes interpretations have been recently found by DasGupta (1984). In the spirit of Shinozaki (1984) and DasGupta (1984), in the next section we first obtain explicit improved estimators of p independent arbitrary scale-parameters under sum of squared error losses $\sum_{i=1}^{p}(a_i-\theta_i)^2$. The results can be generalized to more general losses of the form $(a-\theta)'\,Q(\theta)(a-\theta)$. The best coordinatewise equivariant estimate $\delta_0(\underset{\sim}{x})$ is shown inadmissible for $p \geq 2$. A typical improved estimator is of the form $\delta_i(\underset{\sim}{x}) = \delta_{0,i}\,(\underset{\sim}{x}) + c\cdot(\prod_j x_j)^{1/p}$, where $c > 0$ is a constant depending on the underlying distributions only through a few moments; moreover, they also clearly satisfy the desirable need of loss-robustness to a certain extent. Possible generalizations have been indicated and the results are compared to similar results earlier obtained in other distributions. As in DasGupta (1984), exact analytical representations of the risk improvement are possible to obtain in the general scale parameter case. Consequently, Bayes risk calculations also turn out to be relatively simple and easy. Some results are obtained towards locating an improved estimator which minimizes the Bayes risk with respect to a given prior (the problem of minimizing the relative savings loss (RSL), a concept introduced by Efron and Morris, is equivalent to this). Thus, strictly speaking, in the general scale-parameter case, we have improved estimators which are attractively simple and moreover allow for incorporation of prior information in some form. Numerical calculations of Bayes risks have, however, convinced us that estimators with better Bayesian performance need to be found. The concluding section gives two real life applications of the new estimators and encouraging improvements in mean squared error are found to obtain in both examples.

2. Improved estimators for arbitrary scale-parameters

Let $X_1, X_2, \ldots, X_p$ ($p \geq 2$) be independently distributed and let X_i have density of the form $g_i(x_i/\theta_i) = \frac{1}{\theta_i} f_i(\frac{x_i}{\theta_i})$, $1 \leq i \leq p$. Thus θ_i (assumed positive) is a scale-parameter for the distribution of X_i. We will throughout denote expectation under $\underset{\sim}{\theta} = \underset{\sim}{1}$ as $E(\cdot)$. Moreover we will assume $E(X_i^2) < \infty$ for every i. The best equivariant estimate of $\underset{\sim}{\theta}$ is $\delta_0(\underset{\sim}{X}) = A\underset{\sim}{X}$, where $A = \text{diag}(a_1, a_2, \ldots, a_p)$ and $a_i = \frac{EX_i}{EX_i^2}$. In this section we first show that $A\underset{\sim}{X}$ is inadmissible under losses of the type mentioned in section 1 and suggest explicit estimators, in the spirit of Das Gupta (1984), with uniformly smaller frequentist risk. The proof essentially mimicks the proof of Theorem 2 in Das Gupta (1984) for the special gamma case.

Theorem 2.1

Consider the problem of estimating $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_p)$ under the loss $L(\underset{\sim}{\theta}, a) = \sum_{i=1}^{p} (a_i - \theta_i)^2$. Let $\delta_0(\underset{\sim}{X}) = A\underset{\sim}{X}$ be the best equivariant estimate of $\underset{\sim}{\theta}$ and let $\delta(\underset{\sim}{X}) = A\underset{\sim}{X} + c\cdot(\prod_{i=1}^{p} X_i^{1/p}) \cdot \underset{\sim}{1}$, where $c > 0$ is as in (2.5). For $p \geq 2$, $R(\underset{\sim}{\theta}, \delta) < R(\underset{\sim}{\theta}, \delta_0)$ for every $\underset{\sim}{\theta}$ and hence $\delta_0(\underset{\sim}{X})$ is inadmissible.

Proof:

Let $a_{jp} = EX_j^{\frac{1}{p}}$

$$b_{jp} = EX_j^{\frac{2}{p}}$$

$$c_{jp} = EX_j^{1+\frac{1}{p}}$$

$$a_j = \frac{EX_j}{EX_j^2}, \quad 1 \leq j \leq p \qquad (2.1)$$

Also, let $t = \prod_{i=1}^{p} x_i^{1/p}$.

Then,

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

$$= E_\theta\,[A\underset{\sim}{X} + c t\underset{\sim}{1} - \underset{\sim}{\theta}]'[A\underset{\sim}{X} + c t\underset{\sim}{1} - \underset{\sim}{\theta}] - E_{\underset{\sim}{\theta}}\,[A\underset{\sim}{X} - \underset{\sim}{\theta}]'[A\underset{\sim}{X} - \underset{\sim}{\theta}] \qquad (2.2)$$

$$= E_\theta[pc^2t^2 + 2ct\underset{\sim}{1}'(A\underset{\sim}{X} - \underset{\sim}{\theta})]$$

$$= pc^2 \prod_j b_{jp}\theta_j^{2/p} + 2c \sum_j a_j \frac{c_{jp}}{a_{jp}}\theta_j \cdot(\prod_j a_{jp}\,\theta_j^{1/p}) - 2c \sum_j \theta_j(\prod_j a_{jp}\theta_j^{1/p}) \qquad (2.3)$$

Now, by Liapunov's inequality, $\varepsilon = \min_j\,(1 - \frac{a_j c_{jp}}{a_{jp}}) > 0$. Hence, from (2.3),

$$\Delta(\underset{\sim}{\theta}) \leq pc^2(\prod_j b_{jp}\,\theta_j^{2/p}) - 2c\varepsilon \sum_j \theta_j \cdot(\prod_j a_{jp}\theta_j^{1/p}) \qquad (2.4)$$

Setting $b_p = \max_j b_{jp}$ and $\underline{a}_p = \min_j a_{jp}$, from (2.4).

$$\Delta(\underset{\sim}{\theta}) \leq pc^2 \cdot b_p^p(\prod_j \theta_j^{2/p}) - 2c\varepsilon\, \underline{a}_p^p\,(\sum_j \theta_j)(\prod_j \theta_j^{1/p})$$

$$\leq (\prod_j \theta_j^{2/p})\,[pc^2 b_p^p - 2pc\varepsilon\underline{a}_p^p) \qquad (2.5)$$

< 0 for every $\underset{\sim}{\theta}$ if

$$0 < c < 2\varepsilon(\frac{\underline{a}_p}{b_p})^p$$

<u>Remarks, generalizations, and discussion of the assumptions</u>

1. Theorem 2.1 implies that in the <u>general scale-parameter family</u> uniform mean squared error improvement can be obtained by shifting by multiples of the geometric

mean whenever $p > 1$. Shifting by the geometric mean in a scale-parameter problem is much like shifting by the arithmetic mean of the coordinates in a location problem because it is well known that if x_i's have a scale parameter family of distributions, then $\log x_i$'s have a location parameter family of distributions. It is also interesting that Theorem 2.1 holds even if the coordinate distributions do not come from the same parameteric family, because the functional form of the density of X_i was allowed to depend on i.

2. Several generalizations of Theorem 2.1 are fairly easily obtained. First, the vector $\underset{\sim}{1}$ in the improved estimator can be generalized to an arbitrary positive vector $\underset{\sim}{\alpha}$. This will enable us to give unequal shifts in different coordinates. This scope of choice in the vector $\underset{\sim}{\alpha}$ also leads to the natural question of selecting an $\underset{\sim}{\alpha}$ that best suits the available prior knowledge on $\underset{\sim}{\theta}$ so that good risk-improvements would be obtained in that part of the parameter space where $\underset{\sim}{\theta}$ is likely to lie.

Next, the statement of, Theorem 2.1 holds for more general losses of the form $(\underset{\sim}{a}-\underset{\sim}{\theta})'Q(\underset{\sim}{\theta})(\underset{\sim}{a}-\underset{\sim}{\theta})$ where $Q(\theta)$ is any positive definite matrix such that $\inf_{\theta} \lambda_{min}(Q) > 0$ and $\sup_{\theta} \lambda_{max}(Q) < \infty$ where λ_{min}, λ_{max} are the minimum and the maximum eigenvalues of Q. Also, as in DasGupta (1984), improved estimators in the entire scale-parameter family can also be obtained for all losses of the form $\sum_{i=1}^{p} c_i \theta_i^{m_i}(a_i-\theta_i)^2$, where $m_i \neq 0$. We have, however, no results for the invariant loss.

3. The estimators in Theorem 2.1 are somewhat loss-robust, but not immendsely so (for a beautiful treatment of the problem of constructing loss robust estimators in the normal (and some others) distribution, see Hwang (1983); however, robustness with respect to that large a class of losses may often be unachievable and is perhaps also a conservative formulation of the problem). Consider first the case of $Q = I$; the allowed range of c in (2.5) is

$$0 < c < 2\varepsilon\cdot\left(\frac{a_p}{b_p}\right)^p .$$

Taking $c = c_0 = \varepsilon\cdot\left(\frac{a_p}{b_p}\right)^p$, one will have uniformly smaller frequentist risk for the estimate $A\underset{\sim}{X} + c_0 t\underset{\sim}{1}$ under squared-error loss. It is natural to ask if one can still achieve uniform domination with this <u>same</u> estimator when the loss is $(\underset{\sim}{a} - \underset{\sim}{\theta})'Q(\underset{\sim}{a} - \underset{\sim}{\theta})$ where Q is not necessarily the identity matrix. We have been able to prove that uniform domination can be obtained by using the <u>same</u> estimator described above for all Q such that

$$\lambda_{min} \geq \frac{trQ}{2p} \tag{2.6}$$

Comparable loss-robustness was achieved in Berger (1976) for estimating a multi-normal mean. It was proved in Berger (1976) that if $\underset{\sim}{X} \sim N(\underset{\sim}{\theta},I)$, then for estimating $\underset{\sim}{\theta}$ under loss $L_Q(\underset{\sim}{\theta},\underset{\sim}{a}) = (\underset{\sim}{a}-\underset{\sim}{\theta})'Q(a-\underset{\sim}{\theta})$, the usual James-Stein estimate continues to be minimax for $p \geq 3$, if

$$\lambda_{max}(Q) \leq \frac{2trQ}{p+2} \tag{2.7}$$

Both (2.6) and (2.7) essentially mean that the eigenvalues of Q should not be very scattered. Thus a moderate amount of loss-robustness can be achieved with our estimators.

3. <u>Selecting an</u> $\underset{\sim}{\alpha}$

For the case Q = I, uniform domination can be achieved by choosing any positive vector $\underset{\sim}{\alpha}$. It is natural to ask if $\underset{\sim}{\alpha}$ can be chosen so as to maximize the risk-improvement in some desirable parts of the parameter space or perhaps to minimize the Bayes risk against a certain prior. We will briefly touch on both the problems so that the reader gets a flavor of how the present estimates allow for incorporation of prior information to some extent.

Proposition 3.1. Assume X_i's are iid when all $\theta_i = 1$, and that $Q = I$. Then on the set $\{\underset{\sim}{\theta}:\theta_i = \theta\ \forall_i \geq 1\}$, $-\Delta(\underset{\sim}{\theta}) = R(\underset{\sim}{\theta},\delta_0) - R(\underset{\sim}{\theta},\delta)$ is maximized by choosing $\underset{\sim}{\alpha}$ to be proportional to $\underset{\sim}{1}$.

Proof: $-\Delta(\underset{\sim}{\theta}) = E_\theta[A\underset{\sim}{X} - \underset{\sim}{\theta}]'[A\underset{\sim}{X} - \underset{\sim}{\theta}] - E_\theta[A\underset{\sim}{X} + ct\underset{\sim}{\alpha} - \underset{\sim}{\theta}]'[A\underset{\sim}{X} + ct\underset{\sim}{\alpha} - \underset{\sim}{\theta}]$

$$\text{where } 0 < c < \frac{2p}{\underset{\sim}{\alpha}'\underset{\sim}{\alpha}}\left(1 - \frac{ac_p}{a_p}\right)\left(\min_i \alpha_i\right)\left(\frac{a_p}{b_p}\right)^p , \tag{3.1}$$

Choosing c to be the mid-range,

$$-\Delta(\underset{\sim}{\theta}) = p.\ \frac{a_p^{2p}}{b_p^p}\ \left(1 - \frac{ac_p}{a_p}\right)^2 .\ \frac{\left[2\underset{\sim}{\alpha}'\underset{\sim}{\theta}(\Pi_j\theta_j)^{\frac{1}{p}}(\min_i \alpha_i) - p\cdot(\Pi_j\theta_j)^{\frac{2}{p}}\cdot(\min_i \alpha_i)^2\right]}{\underset{\sim}{\alpha}'\underset{\sim}{\alpha}} \tag{3.2}$$

Hence, on the set $\{\theta_i = \theta\ \forall_i \geq 1\}$

$$-\Delta(\theta) = \text{constant x } \left[\frac{2\Sigma\alpha_i(\min_i \alpha_i) - p(\min_i \alpha_i)^2}{\Sigma\alpha_i^2}\right]\theta^2 \tag{3.3}$$

Since $-\Delta(\theta)$ is scale-invariant in $\underset{\sim}{\alpha}$, one may assume without loss of generality that $\underset{\sim}{\alpha}'\underset{\sim}{\alpha} = p$. The objective is to show (3.3) is maximized when $\underset{\sim}{\alpha}$ is proportional to $\underset{\sim}{1}$.

Hence, it suffices to show that

$$\frac{2\Sigma\alpha_i(\min_i \alpha_i) - p(\min_i \alpha_i)^2}{\Sigma\alpha_i^2} \leq 1$$

$$\Leftrightarrow \quad \Sigma(\alpha_i - \min_i \alpha_i)^2 \geq 0 \tag{3.4}$$

Hence proved.

Remarks 1. Proposition 3.1 means that if all θ_i's are equal, then every coordinate should receive equal shift in order to maximize the risk-improvement. This is what is intuitively expected.

2. Proposition 3.1 immediately suggests that if θ_i are not necessarily equal but instead, are iid, then to minimize the Bayes risk one should choose $\underset{\sim}{\alpha}$ as proportional to $\underset{\sim}{1}$. This is indeed true. However, it would be more interesting to consider more general non-iid priors. This is what we do below. Let θ_i's be independently distributed with distributions Π_i; the problem is to select an $\underset{\sim}{\alpha}$ that minimizes the Bayes risk against the product prior $\Pi = \Pi_1 \otimes \Pi_2 \otimes \cdots \otimes \Pi_p$. Assume $Q = I$ and as usual also assume the coordinate distributions are identical if each θ_i is set equal to 1.

First observe minimizing (with respect to $\underset{\sim}{\alpha}$) $r(\Pi,\delta) = \int R(\theta,\delta)d\Pi(\theta)$ is equivalent to maximizing $r(\Pi,\delta_0) - r(\Pi,\delta)$. Let

$$E\theta_j^{\frac{1}{p}} = \gamma_j$$

$$E\theta_j^{\frac{2}{p}} = \varepsilon_j$$

$$\text{and } E\theta_j^{1+\frac{1}{p}} = \delta_j \tag{3.5}$$

From (3.2),

$$r(\Pi,\delta_0) - r(\Pi,\delta) = \frac{\text{constant}}{\underset{\sim}{\alpha}'\underset{\sim}{\alpha}} \cdot [2(\min \alpha_i)(\prod_j \gamma_j)(\Sigma\alpha_i \frac{\delta_i}{\gamma_i}) - p(\min \alpha_i)^2(\prod_j \varepsilon_j)]$$

$$= \frac{\text{constant}}{\underset{\sim}{\alpha}'\underset{\sim}{\alpha}} \quad [2(\min\alpha_i)\Sigma\alpha_i V_i - cp(\min \alpha_i)^2] \tag{3.6}$$

where $V_i = \frac{\delta_i}{\gamma_i}$, $c = \prod_j \varepsilon_j \Big/ \prod_j \gamma_j$.

Again, because of scale-invariance, one may assume $\underset{\sim}{\alpha}'\underset{\sim}{\alpha} = p$. It is easy to show the maximizing $\underset{\sim}{\alpha}$ must be such that the ordering of α_i's is the same as the ordering of V_i's. Assume without loss $V_1 \leq V_2 \leq \ldots \leq V_p$. Suppose in the maximizing $\underset{\sim}{\alpha}$, the first k α_i's are equal to $\alpha_1 = \min_i \alpha_i$. We have to thus maximize

$2\alpha_1^2 \sum_{i=1}^{k} V_i + 2\alpha_1 \sum_{i=k+1}^{p} \alpha_i V_i - cp\ \alpha_1^2$ subject to $k\alpha_1^2 + \sum_{i=k+1}^{p} \alpha_i^2 = p$ (note k is <u>not</u> known and has to be found out).

A standard argument using a Lagrangian multiplier λ shows that the maximizing $\underset{\sim}{\alpha}$ satisfies

$$4\alpha_1 \sum_{i=1}^{k} V_i + 2 \sum_{i=1+k}^{p} \alpha_i V_i - 2cp\ \alpha_1 + 2\lambda k\alpha_1 = 0 \tag{3.7}$$

$$\alpha_i = -\frac{\alpha_1}{\lambda} V_i \ , \ i \geq k+1 \tag{3.8}$$

Using (3.8), (3.7) reduces to

$$S(k) = 2\sum_{i=1}^{k} V_i - \frac{\sum_{i=k+1}^{p} V_i^2}{\lambda} - cp + \lambda k = 0 \tag{3.9}$$

Since $\alpha_i > \alpha_1$ for $i \geq k+1$, $-\lambda < V_i \ \forall\ i \geq k+1$. Hence,

$$S(k+1) - S(k) = 2V_{k+1} + \frac{V_{k+1}^2}{\lambda} + \lambda$$

$$= \frac{1}{\lambda}\,[\lambda + V_{k+1}]^2 < 0$$ (as $\lambda < 0$ for α_i to be positive for $i \geq k+1$)

Consequently, there is exactly one k satisfying (3.9) subject to the ordering restriction $\alpha_i > \alpha_1$ for $i \geq k+1$, or equivalently $-\lambda < V_{k+1}$.

Let λ and k denote the actual solutions to (3.9) Now the restriction $\underset{\sim}{\alpha}'\underset{\sim}{\alpha} = p$ will give us

$$k\alpha_1^2 + \frac{\alpha_1^2}{\lambda^2} \sum_{i=k+1}^{p} V_i^2 = p$$

$$\Leftrightarrow \alpha_1^2\left(k + \frac{\sum_{i=k+1}^{p} V_i^2}{\lambda^2}\right) = p$$

$$\Leftrightarrow \alpha_1^2 = \frac{\lambda^2 p}{k\lambda^2 + \sum_{i=k+1}^{p} V_i^2} \tag{3.10}$$

(Note if $k = p$, the problem becomes trivial).

Substituting (3.10) into (3.8) now gives $\underset{\sim}{\alpha}$.

3. Note k has to be determined sequentially using (3.9). Once k and λ are found, the rest is fairly straightforward. It seems to us for small p, the algorithm can pretty well be carried out using only a calculator. We will shortly give an example.

4. The special case when θ_i's are iid is of some interest. In this case, it is possible to come up with an attractive closed form solution of the problem. In fact, it is not difficult to show that in this case the optimizing $\underset{\sim}{\alpha}$ is the vector $\underset{\sim}{1}$. We omit the proof.

Below we give an example to show how the sequential algorithm described in remark 2 above is fairly easy to implement for even reasonably large values of p. The example also shows that moderately good incorporation of prior information is possible using the optimizing $\underset{\sim}{\alpha}$ of remark 2.

<u>Example</u> Assume $X_i \overset{\text{indep.}}{\sim}$ Gamma (α,θ_i) and let $\alpha = 1$; take $p = 8$. Assume $\theta_i \overset{\text{indep.}}{\sim}$ Inverse Gamma $(2,r_i)$ with density $\Pi_i(\theta_i) = e^{-\frac{r_i}{\theta_i}} r_i^2 \theta_i^{-3} I_{\theta_i > 0}$.

If the priors are of these forms, their $\Rightarrow$

$$V_i = \frac{\delta_i}{\gamma_i} = \frac{\Gamma(1-\frac{1}{p})}{\Gamma(2-\frac{1}{p})} \cdot r_i$$

$$c = \prod_i \frac{\varepsilon_i}{\gamma_i} = \left(\frac{\Gamma(2-\frac{2}{p})}{\Gamma(2-\frac{1}{p})}\right)^p \cdot (\prod_i r_i)^{\frac{1}{p}} \tag{3.11}$$

Let $\underset{\sim}{r} = (.25,.5,1,1.5,2,8,40,50)$ (note $E\theta_i = r_i$; so one may, for example, suspect that a few normal variances are close to these numbers respectively and try the corresponding conjugate priors with these parameters). The λ obtained from the quadratic equation (3.9) does not meet the requirement $-\lambda < V_{k+1}$ if $k \leq 5$. For $k = 6$, one gets $-\lambda = 30.9432 < V_7 = 45.7143$. Consequently, the optimizing $\underset{\sim}{\alpha}$ will be as follows: using (3.10), $\alpha_1 = .8307$; now (3.8) gives $\alpha_7 = 1.2273$, and $\alpha_8 = 1.5340$. Clearly it is necessary to find out whether the best $\underset{\sim}{\alpha}$ actually does significantly better than, for example $\underset{\sim}{\alpha} = \underset{\sim}{1}$, in order to understand how well prior information is actually incorporated. In what follows, we shall let δ^* denote the improved estimate with $\underset{\sim}{\alpha}$ equal to the best choice, δ denote the improved estimate with $\underset{\sim}{\alpha} = \underset{\sim}{1}$, and δ_0 denote the best equivariant estimate. Both δ and δ^* uniformly dominate δ_0.

If no prior information is available, it may be natural to use δ as an alternative to δ_0, in which case one will get Bayes risk improvement equal to $r(\Pi,\delta_0) - r(\Pi,\delta)$. By using δ^*, the Bayes risk improvement will be $r(\Pi,\delta_0) - r(\Pi,\delta^*)$. Consequently, the increase in the Bayes risk improvement by using δ^* in preference to δ is $r(\Pi,\delta) - r(\Pi,\delta^*)$. If we express this as a fraction of the Bayes risk improvement that was already available by using δ itself, we get the quantity

$$I(\Pi) = \frac{r(\Pi,\delta) - r(\Pi,\delta^*)}{r(\Pi,\delta_0) - r(\Pi,\delta)} .$$

If $I(\Pi)$ is reasonably large, we may be satisfied that we have estimators which not only provide uniform frequentist domination but also allow for incorporation of prior information.

Using (3.6),

$$r(\Pi,\delta_0) - r(\Pi,\delta) = \text{constant} \times 218.3116$$

and $r(\Pi,\delta_0) - r(\Pi,\delta^*) = \text{constant} \times 247.5395$

Hence, $I(\Pi)$ comes to a reasonable 13.39 per cent. If the r_i's are more spread out and p is a little larger, we will get still better values for $I(\Pi)$.

4. Risk-improvement

In this section, we will first calculate risk-functions numerically in a few scale parameter situations to get an idea of how much improvement is attainable in practice and then use our estimators on two actual data sets and check the amount of risk improvement actually obtained. Typically, the Stein-type improved estimators give the best risk improvement when θ_i's are nearly equal or similar in some sense. The following proposition can be regarded as a rough statement of the best possible risk performance of our estimators.

Proposition 4.1

Let $\delta(\underset{\sim}{X}) = A\underset{\sim}{X} + c_0 t\cdot\underset{\sim}{1}$, where

$$c_0 = \left(1 - \frac{ac_p}{a_p}\right) \cdot \left(\frac{a_p}{b_p}\right)^p. \qquad (4.1)$$

Then, on the set $D = \{\underset{\sim}{\theta}: \ \theta_i = \theta, \ 1 \leq i \leq p\}$,

$$\lim_{p\to\infty} \frac{R(\underset{\sim}{\theta},\delta_0) - R(\underset{\sim}{\theta},\delta)}{R(\underset{\sim}{\theta},\delta_0)} = 1 - \frac{(EX)^2}{EX^2} \quad .$$

Proof: Using (2.3), by direct algebra, for $\underset{\sim}{\theta}$ D,

$$\frac{R(\underset{\sim}{\theta},\delta_0) - R(\theta,\delta)}{R(\underset{\sim}{\theta},\delta_0)} = (1 - \frac{ac_p}{a_p})^2 \cdot \frac{a_p^{2p}}{b_p^p} \Bigg| \left(1 - \frac{(EX)^2}{EX^2}\right) \quad . \tag{4.2}$$

By the bounded convergence theorem, $\lim_{p\to\infty} \frac{c_p}{a_p} = 1$ if $P(x_i=0) = 0$. Also, $\lim_{p\to\infty} \frac{a_p^{2p}}{b_p^p} = 1$ if $E(x_i^{-s}) < \infty$ for some $s > 0$. Now (4.2) immediately gives the result. Below we briefly give a few examples on the above proposition.

<u>Examples</u>

(i) Let $X_i \stackrel{\text{index}}{\sim}$ Gamma (α,θ_i), where $\alpha > 0$ is known. Then $\lim_{p\to\infty} \frac{R(\underset{\sim}{\theta},\delta_0)-R(\underset{\sim}{\theta},\delta)}{R(\underset{\sim}{\theta},\delta_0)} = \frac{1}{\alpha+1}$ for $\underset{\sim}{\theta} \in D$.

Clearly, $\sup_{\alpha>0} \lim_{p\to\infty} \frac{R(\underset{\sim}{\theta},\delta_0)-R(\theta,\delta)}{R(\underset{\sim}{\theta},\delta_0)} = 1$, implying that excellent risk improvements are attainable in the gamma problem, approaching 100% as $\alpha \to 0$ and p gets large.

(ii) Let $X_i \stackrel{\text{index}}{\sim}$ Pareto (α,θ_i) with densities

$$f(x_i|\theta_i) = \frac{\alpha}{\theta_i} (\frac{\theta_i}{x_i})^{\alpha+1} \ I_{x_i>\theta_i} \ , \quad \alpha > 2.$$

Direct calculations yield that, for $\underset{\sim}{\theta} \in D$,

$$\lim_{p\to\infty} \frac{R(\underset{\sim}{\theta},\delta_0)-R(\underset{\sim}{\theta},\delta)}{R(\underset{\sim}{\theta},\delta_0)} = \frac{1}{(\alpha-1)^2} \to 1 \text{ as } \alpha \to 2.$$

Again, very encouraging risk-improvements are possible in Pareto problems when p is large. Proposition (4.1) doesn't say anything about the extent of risk-improvement if θ_i's are unequal. Since our estimators permit an exact analytical representation of the risk-improvement for arbitrary $\underset{\sim}{\theta}$ (see (2.3)), it is easy to calculate percentage risk-improvements in different parts of the parameter space. Table 1 below gives the percentage improvements when $x_i \sim R(0,\theta_i)$ and $X_i \sim$ Pareto $(2.5, \theta_i)$ respectively. The improvements are calculated using a fixed set of θ_i's randomly generated from the indicated ranges.

Table 1. Percentage risk-improvements

Range of θ_i	Rectangular		Pareto	
	p=5	p=10	p=5	p=10
$(0,3)^p$	10.14	12.86	32.21	30.96
$(5,10)^p$	11.00	16.25	34.94	39.12
$(5,15)$	10.40	15.45	33.03	37.20

The numbers seem to indicate that satisfactory risk-improvements are possible even when θ_i's are unequal, especially in Pareto models. Finally, we now provide examples of two real-life situations where our estimators are used and check how much is gained.

Example 1. Proschan (1963) provides records giving the durations of time between successive failures of the air-conditioning systems of 13 different Boeing 720 jet airplanes. These durations are listed in Table 2 below. Thus plane number 7907 had the first failure after 194 hours of service, a second failure after another 15 hours, and so on. The time durations seemed to follow simple exponential

Table 2.

Table of the Durations of Time Between Successive Failures of the Air-Conditioning Systems of Each Member of a Fleet of 13 Boeing 720 Jet Airplanes (213 Observations in All)

Plane identification number												
7907	7908	7909	7910	7911	7912	7913	7914	7915	7916	7917	8044	8045
194	413	90	74	55	23	97	50	359	50	130	487	102
15	14	10	57	320	261	51	44	9	254	493	18	209
41	58	60	48	56	87	11	102	12	5		100	14
29	37	186	29	104	7	4	72	270	283		7	57
33	100	61	502	220	120	141	22	603	35		98	54
181	65	49	12	239	14	18	39	3	12		5	32
	9	14	70	47	62	142	3	104			85	67
	169	24	21	246	47	68	15	2			91	59
	447	56	29	176	225	77	197	438			43	134
	184	20	386	182	71	80	188				230	152
	36	79	59	33	246	1	79				3	27
	201	84	27	[a]	21	16	88				130	14
	118	44	[a]	15	42	106	46					230
	[a]	59	153	104	20	206	5					66
	34	29	26	35	5	82	5					61
	31	118	326		12	54	36					34
	18	25			120	31	22					
	18	156			11	216	139					
	67	310			3	46	210					
	57	76			14	111	97					
	62	26			71	39	30					
	7	44			11	63	23					
	22	23			14	18	13					
	34	62			11	191	14					
		[a]			16	18						
		130			90	163						
		208			1	24						
		70			16							
		101			52							
		208			95							

[a] Indicates major overhaul.

distributions (gamma with $\alpha = 1$) for all the 13 jets; for risk calculations, we take each θ_i to be its best unbiased estimate based on the sample of observations on the corresponding jet.

The estimator of Theorem 2.1 corresponding to $\underset{\sim}{\alpha} = \underset{\sim}{1}$ and $c_0 = (1-\frac{ac_p}{a_p}) \cdot (\frac{a_p}{b_p})^p$ was used to simultaneously estimate $\underset{\sim}{\theta}$ [$\underset{\sim}{\theta}^{-1}$ is taken as the vector (.012, .010, .012, .008, .008, .017, .013, .016, .005, .009, .003, .009, .012)]. Using (2.10) and (2.11),

$$R(\underset{\sim}{\theta},\delta_0) = 125521.9172, \text{ and } -\Delta(\underset{\sim}{\theta}) = R(\underset{\sim}{\theta},\delta_0) - R(\underset{\sim}{\theta},\delta) = 35031.31054,$$

implying that 27.91% risk-improvement is attained by using the new estimator instead of the standard coordinatewise best equivariant estimator.

Example 2. This is an example in which a Pareto model seemed suitable. Table 4 below is copied from the Statistical Abstract of the United States (1983-84). For our purpose, we will only use the part of the table giving percent distributions of household incomes in four different regions of the country (northeast, north central, etc.) in 1982. The household income in each region is modeled as a Pareto variable and the shape and the scale parameters of each of these four Pareto distributions are estimated from the given data using the method of moments. Again, as in example 1, for risk calculations, the θ_i's and the α_i's are set equal to these method of moments estimates, which are listed below.

Table 3.

i	α_i	θ_i
1	2.8006	15372.54
2	2.7117	14857.52
3	2.7093	14007.58
4	2.7913	15847.87

Table 4.

Income, Expenditures, and Wealth

MONEY INCOME OF HOUSEHOLDS—PERCENT DISTRIBUTION BY MONEY INCOME LEVEL, BY SELECTED CHARACTERISTICS: 1982

[Households as of March 1983]

CHARACTERISTIC	Total households (1,000)	PERCENT DISTRIBUTION OF HOUSEHOLDS BY INCOME LEVEL (IN DOLLARS)								Median income (dollars)
		Under 5,000	5,000-9,999	10,000-14,999	15,000-19,999	20,000-24,999	25,000-34,999	35,000-49,999	50,000 and over	
Total [1]	**83,918**	**9.6**	**14.3**	**13.5**	**12.2**	**11.4**	**17.0**	**13.2**	**8.9**	**20,171**
White	73,182	8.2	13.5	13.2	12.3	11.7	17.6	13.9	9.6	21,117
Black	8,916	21.7	20.9	15.8	11.2	9.7	12.0	6.4	2.2	11,968
Spanish origin [2]	4,085	12.9	20.2	16.3	13.7	10.7	13.5	9.1	3.6	15,178
Northeast	17,926	8.7	15.3	12.6	11.7	11.3	17.4	14.1	9.2	20,707
North Central	21,331	9.5	13.6	12.8	12.0	12.0	18.2	13.9	8.1	20,820
South	28,120	11.7	14.5	14.4	12.5	11.4	15.7	11.6	8.2	18,591
West	16,541	7.3	13.9	13.7	12.3	11.0	16.9	14.1	10.7	21,192
Nonfarm	82,071	9.5	14.3	13.5	12.1	11.4	17.0	13.2	9.0	20,238
Farm	1,847	14.1	14.7	14.5	13.4	12.8	14.6	10.4	5.3	17,205
Marital status:										
Male householder	59,250	4.8	9.7	12.1	12.1	12.6	20.6	16.6	11.4	24,376
Married, wife present	47,720	2.9	8.0	11.4	11.8	12.8	21.7	18.5	12.9	26,164
Married, wife absent	1,356	15.2	16.7	17.6	10.2	10.5	15.2	8.3	6.4	15,207
Widowed	1,704	16.6	28.1	16.3	8.9	10.4	8.6	6.8	4.3	11,319
Divorced	3,208	10.5	14.6	12.3	13.2	13.8	18.5	9.7	7.3	19,766
Single (never married)	5,263	12.1	14.5	16.2	15.6	11.8	16.3	9.2	4.4	17,150
Female householder	24,668	21.1	25.2	16.7	12.4	8.6	8.3	5.0	2.7	10,970
Married, husband present	2,187	4.1	10.1	12.6	13.4	13.4	15.9	17.0	13.4	23,396
Married, husband absent	2,407	33.3	27.3	15.0	10.9	5.6	5.2	2.0	.8	8,252
Widowed	9,245	25.2	33.8	15.2	9.5	6.1	5.4	3.2	1.6	8,076
Divorced	5,645	17.4	20.1	19.8	15.2	10.6	10.7	4.3	2.0	13,033
Single (never married)	5,183	19.5	21.0	18.4	14.5	10.1	9.1	5.4	2.0	12,448
Age of householder:										
15-24 years	5,695	14.8	19.4	20.1	14.9	13.3	11.6	4.3	1.5	13,816
25-34 years	19,104	7.2	10.0	13.7	15.0	14.5	21.3	13.1	5.0	21,281
35-44 years	16,020	6.0	8.0	9.9	10.7	12.0	21.8	18.9	12.5	26,370
45-54 years	12,354	6.5	8.0	9.3	9.8	10.4	19.0	19.9	17.1	27,985
55-64 years	13,074	8.6	12.0	13.0	11.8	10.6	16.9	14.9	12.3	22,075
65 years and over	17,671	16.7	29.2	17.6	11.4	8.4	8.1	4.8	3.7	11,041
Size of household:										
One person	19,250	22.7	27.3	16.7	12.1	8.4	8.0	3.0	1.7	9,984
Two persons	26,439	6.6	13.3	15.7	13.9	12.6	17.2	12.5	8.2	20,201
Three persons	14,793	6.5	9.3	11.4	12.0	12.1	20.3	17.6	10.8	24,445
Four persons	13,303	4.0	6.9	9.0	10.9	12.5	22.8	19.8	14.0	27,617
Five persons	6,105	4.4	8.2	10.0	9.1	12.6	21.9	19.1	14.6	27,176
Six persons	2,460	5.1	9.6	10.1	10.4	10.4	18.6	20.0	15.8	27,191
Seven persons or more	1,568	4.7	12.4	12.8	11.3	10.7	17.0	17.4	13.6	23,717
Education attainment of householder:										
Elementary school	13,396	21.4	28.2	18.0	11.0	7.9	8.2	3.9	1.5	10,097
Less than 8 years	7,226	24.1	29.1	18.2	10.6	6.9	6.8	3.0	1.1	9,327
8 years	6,170	18.2	27.1	17.8	11.5	9.0	9.6	4.9	1.8	11,192
High school	40,217	10.0	15.2	15.1	13.5	12.4	17.4	11.8	4.8	18,527
1-3 years	11,131	16.0	22.2	16.3	12.3	10.3	12.7	7.4	2.8	13,483
4 years	29,086	7.7	12.6	14.6	14.0	13.1	19.1	13.4	5.5	20,391
College	30,305	3.9	6.9	9.4	10.8	11.7	20.3	19.2	17.6	28,216
1-3 years	13,371	5.4	9.8	12.8	12.5	13.3	21.2	16.3	8.7	23,524
4 years	9,140	2.9	5.5	7.6	10.6	11.2	20.7	21.4	20.0	30,613
Tenure:										
Owner occupied	54,494	6.4	11.1	11.5	11.2	11.7	19.2	16.8	12.2	24,148
Renter occupied	27,834	15.2	20.1	17.1	13.9	11.2	13.0	6.5	2.8	14,199
Occupier paid no cash rent	1,589	23.8	22.5	16.9	15.0	7.1	7.7	5.2	1.9	10,873

[1] Includes other races not shown separately. [2] Persons of Spanish origin may be of any race.

Source: U.S. Bureau of the Census, *Current Population Reports*, series P-60, No. 140.

Using (2.10) and the fact that $R(\underset{\sim}{\theta},\delta_0) = \sum_{j=1}^{4} \frac{\theta_j^2}{(\alpha_j-1)^2}$, one has,

$R(\underset{\sim}{\theta},\delta_0) = 293658672.5$, and $-\Delta(\underset{\sim}{\theta}) = 69844051.39$, thus giving 23.78% risk-improvement in as few as 4 dimensions.

5. Final remarks

The main result of this paper implies that the estimates for the gamma scale-parameters proposed in DasGupta (1984) are robust and work equally fine for arbitrary scale-parameters. Hopefully, however, the point has been made that the important problem is to build up estimators allowing incorporation of prior knowledge and providing significant improvement on desirable parts of the parameter space. In our opinion, minimaxity and (or) uniform domination may have to be sacrificed to meet this important need. It was unfortunate that we were unable to find any reasonable Bayesian or empirical Bayesian interpretations of the estimates in Theorem 2.1 outside of the gamma family. Again, broader classes of improved estimates may be needed for this. See Berger and DasGupta (1985) for results in the gamma case. Finally, although the results of this paper are stated for one observation from each coordinate distribution, the multiple observation case, theoretically, can be treated equally well by restricting to the distributions of the coordinatewise Pitman estimates for which θ_i's are still scale-parameters.

Acknowledgement

I am grateful to James Berger, the editors, and the referee for their highly perceptive and constructive suggestions. Sincere thanks are also due to Herman Rubin for graciously helping at various stages

References

Berger, James (1976). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Mult. Anal. 6 256-264.

Berger, James (1980a). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Ann. Statist. 8 716-761.

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

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

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

Berger, James and Berliner, L. M. (1983). Bayesian input in Stein estimation and a new minimax empirical Bayes estimator. (To appear)

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

Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. Ann. Math. Stat. 37 1087-1136.

DasGupta, Anirban (1984). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses. (To appear in Ann. Statist.).

Hwang, J. T. (1983). Universal domination and stochastic domination: estimation simultaneously under a broad class of loss functions. Ann. Statist. 13 295-314.

Proschan, F. (1963). Theoretical explanation of observed decrease failure rate. Technometrics. 5, 375-384.

Shinozaki, Nobuo (1984). Simultaneous estimation of location parameters under quadratic loss. Ann. Statist. 12 322-335.

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.