Improving on Inadmissible Estimators in Continuous Exponential Families with Applications to Simultaneous Estimation of Gamma Scale Parameters

by

James Berger[1]
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #78-6

May 31, 1978

[1]Research supported by the John Simon Guggenheim Memorial Foundation, and by the National Science Foundation under Grant #MCS 76-06627A2.

Summary

A general technique is developed for improving upon inadmissible estimators of natural parameters (or integral powers thereof) from continuous exponential families. The technique is to reduce the problem to the study of a differential inequality. Typical differential inequalities are presented and solved.

Explicit results are given for simultaneous estimation of gamma scale parameters (and their inverses) for a variety of natural loss functions. Surprising behavior is observed for many of the estimators improving upon "standard" estimators.

For squared error loss (and any continuous exponential family) it is shown explicitely how to establish inadmissibility of an estimator and construct improved estimators.

1. Introduction

There has recently been considerable interest in improving upon standard estimators in multivariate estimation problems. This interest has been stimulated by the development of two powerful tools of analysis in Stein (1973) and Brown (1974). In Stein (1973) the use of integration by parts is shown (for the multivariate normal distribution) to lead to a relatively simple method of finding estimators improving upon the usual estimator (the sample mean) under squared error loss in three or more dimensions. This technique was shown to apply to general continuous exponential families by Hudson (1978). The result of the technique is essentially the representation of the risk of an estimator, δ, as the expected value of an expression involving δ and its derivatives. Indeed if $X = (X_1, X_2, \ldots X_p)$ has density $f(x|\theta)$ with respect to Lebesgue measure on R^p, and the loss in estimating $\psi(\theta)$ by $\delta(x)$ is $L(\delta, \psi(\theta))$, the representation for the risk, $R(\delta,\theta)$, of δ is of the form

$$R(\delta,\theta) = E_\theta\, L(\delta(X),\psi(\theta)) = \int L(\delta(x),\psi(\theta))f(x|\theta)dx = \int \mathcal{D}(\delta(x))f(x|\theta)dx,$$

where $\mathcal{D}(\delta(x))$ involves δ and its derivatives (but not θ). In comparing an estimator $\delta^*(x)$ with a "standard" estimator $\delta^o(x)$, if it happens that $\Delta(x) = \mathcal{D}(\delta^*(x)) - \mathcal{D}(\delta^o(x)) < 0$ for all x, then clearly

$$\Delta^*(\theta) = R(\delta^*,\theta) - R(\delta^o,\theta) = E_\theta[\Delta(X)] < 0$$

for all θ, so that δ^* is better than δ^o. The problem of improving upon an estimator δ^o can thus be dealt with by trying to find a solution, δ^*, to the differential inequality $\Delta(x) < 0$.

The importance of the study of such differential inequalities has been emphasized by Brown (1974), who developed a number of general techniques to use them in proving inadmissibility. (Stein (1965) earlier indicated their importance. See also Brown (1971 and 1975) and Berger (1976a, 1976b, and 1976c).)

This paper has two purposes. This first is to explain, for the general exponential family model, how the question of improving upon δ^o can be reduced to the study of a differential inequality $\Delta(x) < 0$, and to discuss and give solutions for the types of differential inequalities that are usually encountered. In so doing, a constructive general theory is developed for improving upon most inadmissible estimators of the natural parameters of an exponential family under squared error loss.

The second purpose is to apply the results to interesting practical problems, for the most part problems involving simultaneous estimation of gamma scale parameters (or their inverses). More precisely, assume that $X = (X_1,\ldots X_p)$ is observed, where the X_i are independent Gamma (α_i,θ_i) $(\alpha_i > 0,\ 0 < \theta_i < \infty)$ random variables, having density (on $(0,\infty)$)

$$f(x_i|\theta_i) = \theta_i^{\alpha_i} x_i^{(\alpha_i-1)} e^{-x_i\theta_i}/\Gamma(\alpha_i).$$

Although one example involving the estimation of $\theta = (\theta_1,\theta_2,\ldots\theta_p)$ will be given, we will for the most part discuss the estimation of the θ_i^{-1}. (The θ_i^{-1} are generally of greater interest, being for example multiples of the variances if the X_i are chi square.) Assume the loss in estimating $(\theta_1^{-1},\theta_2^{-1},\ldots\theta_p^{-1})$ by $\delta(x) = (\delta_1(x),\ldots\delta_p(x))$ is of the form

$$\text{(1.1)} \qquad L(\delta,\theta) = \sum_{i=1}^{p} \theta_i^m(1-\delta_i(x)\theta_i)^2.$$

The "standard" (best multiple of x_i) estimator of θ_i^{-1} for the loss $\theta_i^m(1-\delta_i\theta_i)^2$ is

$$\delta_i^o(x) = x_i/(\alpha_i+1).$$

It is thus natural in the simultaneous estimation problem to seek improvements upon $\delta^o(x) = (\delta_1^o(x),\ldots\delta_p^o(x))$.

Four choices of m in (1.1) will be considered: m=-2, m=-1, m=0, and m=1. The choice m=-2 corresponds to the usual sum of squares error loss (for estimating the θ_i^{-1}), while m=0 gives the standard invariant loss. m=-1 will be considered because it is the simplest of the four possibilities to deal with (and hence easiest to understand), and also may be of intrinsic interest as a compromise between the m=-2 and m=0 losses. The choice m=1 is examined because it leads to a different type of conclusion than the others. Indeed it will be seen that the cases $m < 0$, m=0, and $m > 0$ have considerably different solutions.

The results obtained are quite surprising. Three fairly prevalent beliefs about simultaneous estimation are:

(i) Improvement is usually obtained only in three or more dimensions (Clevenson and Zidek (1975), Berger (1976a), Berger (1976b), and Zaman (1977) being considered atypical situations);

(ii) Improvement is obtained by shrinking δ^o towards a point (or at least a subspace);

(iii) Somehow the improved estimators are really just taking advantage of some similarities between the coordinates (say in some empirical Bayes fashion).

The results obtained in this paper seem to violently contradict all three notions. First, improvement seems usually obtainable in just two or more dimensions. (Only the case m=0 requires at least three dimensions.) Second, the improved estimators need not shrink δ^o. Indeed to improve upon δ^o when $m < 0$ it seems necessary to expand δ^o towards infinity. Finally, an example will be given which seems to contradict (or at least makes meaninglessly vague) tenet (iii) above.

In Section 2 the basic tools for obtaining the differential inequality in continuous exponential families will be developed, and applied to the gamma problems. In Section 3 the most commonly encountered type of differential inequality will be presented, and a general solution obtained. Section 4 gives the specific solutions of the differential inequalities (and hence the estimators better than δ^o) for the gamma distribution and the four losses of interest. Section 5 gives a formulation and solution of the problem of improving upon an inadmissible estimator δ^o for arbitrary continuous exponential families and squared error loss. Section 6 consists of some concluding remarks.

2. Obtaining the Differential Inequality

Let X have density (with respect to Lebesgue measure on (a,b), a and b possibly infinite)

$$f(x|\theta) = \beta(\theta)t(x)\, e^{-\theta r(x)},$$

where $t(x) > 0$ and $r(x)$ is absolutely continuous (on (a,b) of course) with $r'(x) = \frac{d}{dx}\, r(x) > 0$.

Lemma 1. Assume that $h(x)$ is a real valued function satisfying

(i) $s(x) = t(x)h(x)/r'(x)$ is absolutely continuous on (a,b),

(ii) $\int_a^b |s'(x)|e^{-\theta r(x)}dx < \infty$, and

(iii) $\lim_{x\to a} [s(x)e^{-\theta r(x)}] = \lim_{x\to b} [s(x)e^{-\theta r(x)}] = 0$ for all θ in the natural parameter space.

Then $$E_\theta[\theta\, h(X)] = E_\theta\, [s'(X)/t(X)].$$

Proof. Note that

$$E_\theta[\theta h(X)] = \beta(\theta) \int_a^b [s(x)][\theta r'(x) e^{-\theta r(x)}]dx.$$

An integration by parts establishes the lemma.||

This is essentially the result in Hudson (1978), though stated here in a somewhat different way.

Lemma 2. Assume a finite indefinite integral, $g(x)$, of $[t(x) h(x)]$ exists, in the sense that

$$g(y_1) - g(y_2) = \int_{y_2}^{y_1} t(x)h(x)dx.$$

Assume that h satisfies

(i) $E_\theta[|h(X)|] < \infty$, and

(ii) $\lim_{x\to a} [g(x)e^{-\theta r(x)}] = \lim_{x\to b} [g(x)e^{-\theta r(x)}] = 0$, for all θ in the natural parameter space.

Then for $\theta \neq 0$, $$E_\theta[\theta^{-1}h(X)] = E_\theta[r'(X)g(X)/t(X)].$$

Proof. Again, just integrate by parts.||

For any loss which can be written in the form $L(\delta,\theta) = \Sigma\, \theta^{m_i} h_i(x)$ (the m_i being integers and the h_i functions of δ), the above lemmas can be used to represent the risk $R(\delta,\theta)$ as the expectation of a quantity involving only functions of x and not θ. If $m_i > 0$ for a particular term, Lemma 1 is applied repetitively (namely m_i times) to the term. If $m_i < 0$, Lemma 2 is applied $|m_i|$ times to the term. To obtain a differential inequality in the latter case, it is necessary to express all quantities in terms of the last indefinite integral, g, obtained. An example of this will be seen shortly.

Let us now return to the gamma problem discussed in Section 1, and derive the desired differential inequalities. The following corollaries to Lemmas 1 and 2 will be needed. In these corollaries, $X_1,\ldots X_p$ will be independent Gamma

(α_i, θ_i) random variables. For a function $h(x): R^p \to R^1$, let

$$h^{i(k)}(x) = \frac{\partial^k}{\partial x_i^k} h(x) \quad (h^{i(0)}(x) = h(x)),$$

providing the partial derivatives exist. Finally, define

$$s_{i,j}(x) = \sum_{k=0}^{j} \binom{j}{i} \alpha_i^{<j-k>} x_i^{(\alpha_i+k-j-1)} h^{i(k)}(x),$$

where $\alpha^{<\ell>} = (\alpha-1)(\alpha-2)\ldots(\alpha-\ell)$ $(\alpha^{<0>} = 1)$, and $\binom{j}{i} = j!/[i!(j-i)!]$.

Corollary 1. Assume m is a positive integer and $h(x): R^p \to R^1$ satisfies for all $0 < x_k < \infty (k \neq i)$ and $j = 0, 1, \ldots m-1$,

(i) $s_{i,j}(x)$ is absolutely continuous as a function of x_i on $(0,\infty)$,

(ii) $E_\theta[|s_{i,j}^{i(1)}(X)|X_i^{(1-\alpha_i)}] < \infty$, and

(iii) $\lim_{x_i \to 0} [s_{i,j}(x)e^{-x_i\theta_i}] = \lim_{x_i \to \infty} [s_{i,j}(x)e^{-x_i\theta_i}] = 0$ for all $0 < \theta_i < \infty$.

Then

$$E_\theta[\theta_i^m h(X)] = E_\theta[s_{i,m}(X)X_i^{(1-\alpha_i)}] = \sum_{k=0}^{m} \binom{m}{k} \alpha_i^{<m-k>} E_\theta[h^{i(k)}(X)X_i^{(k-m)}].$$

Proof. The proof will be by induction on m. By condition (ii) it will always be possible at all stages to reverse orders of integration so that the inner integral of $E_\theta[\theta_i^m h(X)]$ is over x_i. Lemma 1 will then be applied to this inner integral with $t(x_i) = x_i^{(\alpha_i-1)}$ and $r(x_i) = x_i$.

For m=1, Corollary 1 is just a restatement of Lemma 1, noting that

$$s_{i,0}^{i(1)}(x) = \frac{\partial}{\partial x_i} s_{i,0}(x) = \frac{\partial}{\partial x_i}[h(x)x_i^{(\alpha_i-1)}] = h^{i(1)}(x)x_i^{(\alpha_i-1)} + (\alpha_i-1)h(x)x_i^{(\alpha_i-2)}$$
$$= s_{i,1}(x).$$

If the corollary is true for (m-1), one has that

$$E_\theta[\theta_i^m h(X)] = E_\theta[\theta_i s_{i,(m-1)}(X)X_i^{(1-\alpha_i)}].$$

It can be checked that $\frac{\partial}{\partial x_i} s_{i,(m-1)}(x) = s_{i,m}(x)$. Hence applying Lemma 1 to the function $s_{i,(m-1)}(x)x_i^{(1-\alpha_i)}$ and using the induction hypothesis gives the desired result. ||

Corollary 2. Assume a finite indefinite integral, $g_i(x)$, of $[x_i^{(\alpha_i-1)}h(x)]$ over x_i exists, in the sense that for all $0 < x_k < \infty$ $(k \neq i)$

$$g_i(x_1,\ldots x_{i-1},y_1,x_{i+1},\ldots x_p) - g_i(x_1,\ldots x_{i-1},y_2,x_{i+1},\ldots x_p) = \int_{y_2}^{y_1} x_i^{(\alpha_i-1)} h(x)dx_i .$$

Assume also that

(i) $E_\theta[|h(X)|] < \infty$, $E_\theta[|g_i(X)|X_i^{(1-\alpha_i)}] < \infty$, and

(ii) $\lim_{x_i \to 0} [g_i(x)e^{-x_i\theta_i}] = \lim_{x_i \to \infty} [g_i(x)e^{-x_i\theta_i}] = 0$ for all $0 < \theta_i < \infty$.

Then

$$E_\theta[\theta_i^{-1}h(X)] = E_\theta[X_i^{(1-\alpha_i)} g_i(X)].$$

Proof. Rearranging orders of integration so that the inside integral is over x_i, and applying Lemma 2 (with $t(x) = x_i^{(\alpha_i-1)}$ and $r(x_i) = 1$) gives the desired result. ||

We now proceed with the derivation of the differential inequalities for the gamma problems.

Case 1. $L(\delta,\theta) = \sum_{i=1}^{p} \theta_i^{-1}(1-\delta_i\theta_i)^2$ (i.e. m=-1).

This case gives the simplest differential inequality and so will be considered first. Note that

$$E_\theta[\theta_i^{-1}(1-\delta_i(X)\theta_i)^2] = \theta_i^{-1} - 2E_\theta[\delta_i(X)] + E_\theta[\theta_i\delta_i^2(X)].$$

Applying Corollary 1 with m=1 and $h(x) = \delta_i^2(x)$ gives

$$E_\theta[\theta_i\delta_i^2(X)] = \sum_{k=0}^{1} \alpha_i^{<1-k>} E_\theta[X_i^{(k-1)}h^{i(k)}(X)] = E_\theta[(\alpha_i-1)X_i^{-1}\delta_i^2(X) + 2\delta_i(X)\delta_i^{i(1)}(X)].$$

(The conditions needed for the corollary to apply will be satisfied by the estimators we end up using. The verification will be easy and so no further mention of the conditions will be made unless they impose some restrictions on the solutions.) Hence

$$(2.1)\quad R(\delta,\theta) = E_\theta L(\delta(X),\theta) = \sum_{i=1}^{p} \theta_i^{-1} + E_\theta\{\sum_{i=1}^{p}[-2\delta_i(X) + (\alpha_i-1)X_i^{-1}\delta_i^2(X) + 2\delta_i(X)\delta_i^{i(1)}(X)]\}.$$

For a competitor $\delta^*(x)$ to $\delta^o(x)$ (recall $\delta_i^o(x) = x_i/(\alpha_i+1)$), define

$$\Delta^*(\theta) = R(\delta^*,\theta) - R(\delta^o,\theta).$$

A simple calculation gives that $R(\delta^o,\theta) = \sum_{i=1}^{p} [(\alpha_i+1)\theta_i]^{-1}$. Writing

$$\delta_i^*(x) = \frac{x_i}{(\alpha_i+1)}\ (1+\phi_i(x))$$

and using (2.1), a little algebra then gives that

$$\Delta^*(\theta) = E_\theta\{\sum_{i=1}^{p} \frac{2X_i^2\phi_i^{i(1)}(X)}{(\alpha_i+1)^2}(1+\phi_i(X)) + \sum_{i=1}^{p}\frac{X_i\phi_i^2(X)}{(\alpha_i+1)}\}.$$

Hence if we can find a solution $\phi = (\phi_1,\ldots\phi_p)$ to the differential inequality

$$(2.2)\quad \Delta_{-1}(x) = \sum_{i=1}^{p}\frac{2x_i^2\phi_i^{i(1)}(x)}{(\alpha_i+1)^2} + \sum_{i=1}^{p}\frac{x_i\phi_i^2(x)}{(\alpha_i+1)} + \sum_{i=1}^{p}\frac{2x_i^2\phi_i^{i(1)}(x)\phi_i(x)}{(\alpha_i+1)^2} \le 0$$

for all $0 < x_i < \infty$ $(i=1,\ldots p)$, with strict inequality for some set of x with positive measure, then $\Delta^*(\theta) < 0$ for all θ and δ^* is better than δ^o.

<u>Case 2.</u> $L(\delta,\theta) = \sum_{i=1}^{p}(1-\delta_i\theta_i)^2$ (i.e. m=0).

This is the natural scale invariant loss for the problem. Using the usual log transform, this problem could be transformed to a location parameter problem. It is easier, however, to deal directly with the untransformed problem.

Noting that

$$E_\theta[(1-\delta_i(X)\theta_i)^2] = 1-2E_\theta[\theta_i\delta_i(X)] + E_\theta(\theta_i^2\delta_i^2(X)],$$

Corollary 1 can be applied to give

$$E_\theta[\theta_i\delta_i(X)] = \sum_{k=0}^{1}\alpha_i^{<1-k>}E_\theta[\delta_i^{i(k)}(X)X_i^{(k-1)}] = E_\theta[(\alpha_i-1)X_i^{-1}\delta_i(X) + \delta_i^{i(1)}(X)],$$

and

$$E_\theta[\theta_i^2\delta_i^2(X)] = \sum_{k=0}^{2}\binom{2}{k}\alpha_i^{<2-k>}\, E_\theta[\{\delta_i^2(X)\}^{i(k)}X_i^{(k-2)}]$$

$$= E_\theta[(\alpha_i-1)(\alpha_i-2)X_i^{-2}\delta_i^2(X)+4(\alpha_i-1)X_i^{-1}\delta_i(X)\delta_i^{i(1)}(X)+2\delta_i(X)\delta_i^{i(2)}(X)+2\{\delta_i^{i(1)}\}^2].$$

Defining δ^* and Δ^* as before, a calculation then gives that

$$\Delta^*(\theta) = E_\theta\{\Delta_0(X)\},$$

where

$$(2.3)\quad \Delta_0(x) = 2\sum_{i=1}^{p}\frac{x_i\phi_i^{i(1)}(x)}{(\alpha_i+1)} + \sum_{i=1}^{p}\frac{\alpha_i\phi_i^2(x)}{(\alpha_i+1)} + 4\sum_{i=1}^{p}\frac{x_i\phi_i(x)\phi_i^{i(1)}(x)}{(\alpha_i+1)} + 2\sum_{i=1}^{p}\frac{x_i^2}{(\alpha_i+1)^2}[\phi_i^{i(2)}(x)+\phi_i(x)\phi_i^{i(2)}(x)+\{\phi_i^{i(1)}(x)\}^2].$$

A solution ϕ to $\Delta_0(x) \le 0$ is thus sought.

<u>Case 3.</u> $L(\delta,\theta) = \sum_{i=1}^{p}\theta_i(1-\delta_i\theta_i)^2$ (i.e. m=1).

A fairly lengthy calculation using Corollary 1 gives that $\Delta^*(\theta) = E_\theta\{\Delta_1(X)\}$, where

$$(2.4)\quad \Delta_1(x) = 2\sum_{i=1}^{p}\frac{\alpha_i\phi_i^{i(1)}(x)}{(\alpha_i+1)} + \sum_{i=1}^{p}\frac{\alpha_i(\alpha_i-1)\phi_i^2(x)}{(\alpha_i+1)x_i} + 4\sum_{i=1}^{p}\frac{x_i\phi_i^{i(2)}(x)}{(\alpha_i+1)} + 6\sum_{i=1}^{p}\frac{\alpha_i\phi_i(x)\phi_i^{i(1)}(x)}{(\alpha_i+1)} + 6\sum_{i=1}^{p}\frac{x_i[\{\phi_i^{i(1)}(x)\}^2+\phi_i(x)\phi_i^{i(2)}(x)]}{(\alpha_i+1)} + 2\sum_{i=1}^{p}\frac{x_i^2[\phi_i^{i(3)}(x)+3\phi_i^{i(1)}(x)\phi_i^{i(2)}(x)+\phi_i(x)\phi_i^{i(3)}(x)]}{(\alpha_i+1)^2}.$$

<u>Case 4.</u> $L(\delta,\theta) = \sum_{i=1}^{p}\theta_i^{-2}(1-\delta_i\theta_i)^2$ (i.e. m=-2).

This "standard" squared error loss turns out to be the trickiest to work with because Corollary 2 must be used. Since

$$(2.5)\quad E_\theta[\theta_i^{-2}(1-\delta_i(X)\theta_i)^2] = \theta_i^{-2}-2E_\theta[\theta_i^{-1}\delta_i(X)]+E_\theta[\delta_i^2(X)],$$

Corollary 2 can be applied to give

$$E_\theta[\theta_i^{-1}\delta_i(X)] = E_\theta[X_i^{(1-\alpha_i)} g_i(X)],$$

where $g_i(x)$ is an indefinite integral (over x_i) of $[x_i^{(\alpha_i-1)}\delta_i(x)]$ as defined in Corollary 2. Note that

$$\delta_i(x) = g_i^{i(1)}(x)x_i^{(1-\alpha_i)},$$

so that (2.5) can be rewritten

$$(2.6) \quad E_\theta[\theta_i^{-2}(1-\delta_i(X)\theta_i)^2] = \theta_i^{-2}+E_\theta\{-2g_i(X)X_i^{(1-\alpha_i)}+[g_i^{i(1)}(X)X_i^{(1-\alpha_i)}]^2\}.$$

(Note that it is important to write all quantities in terms of g_i so as to obtain a differential inequality.) Defining $\delta_i^*(x) = (\alpha_i+1)^{-1}x_i(1+\phi_i(x))$ as before, g_i^* as the corresponding indefinite integral, and

$$h_i(x) = (\alpha_i+1)g_i^*(x)-x_i^{(\alpha_i+1)}/(\alpha_i+1),$$

calculation using (2.6) shows that

$$E_\theta[\theta_i^{-2}(1-\delta_i^*(X)\theta_i)^2] = \theta_i^{-2}-E_\theta\{\frac{x_i^2}{(\alpha_i+1)^2}\}+ E_\theta\{\frac{-2h_i(X)}{(\alpha_i+1)X_i^{(\alpha_i-1)}} + \frac{2h_i^{i(1)}(X)}{(\alpha_i+1)^2X_i^{(\alpha_i-2)}}$$

$$+ \frac{[h_i^{i(1)}(X)]^2}{(\alpha_i+1)^2X_i^{2(\alpha_i-1)}}\}.$$

Defining $\Delta^*(\theta)$ as before, it follows that

$$\Delta^*(\theta) = E_\theta\{\Delta_{-2}(X)\},$$

where

$$(2.7) \quad \Delta_{-2}(x) = -2\sum_{i=1}^p \frac{h_i(x)}{(\alpha_i+1)x_i^{(\alpha_i-1)}} + 2\sum_{i=1}^p \frac{h_i^{i(1)}(x)}{(\alpha_i+1)^2x_i^{(\alpha_i-2)}} + \sum_{i=1}^p \frac{[h_i^{i(1)}(x)]^2}{(\alpha_i+1)^2x_i^{2(\alpha_i-1)}}.$$

Again, a solution to $\Delta_{-2}(x) \le 0$ is sought. Note for future reference that

$$(2.8) \quad h_i^{i(1)}(x) = (\alpha_i+1)g_i^{*i(1)}(x)-x_i^{\alpha_i} = (\alpha_i+1)\delta_i^*(x)x_i^{(\alpha_i-1)}-x_i^{\alpha_i} = x_i^{\alpha_i}\phi_i(x).$$

3. Heuristic Solution of the Differential Inequality

An examination of the differential inequalities $\Delta_m(x) \le 0$ (see (2.2), (2.3), (2.4), and (2.7)) shows that a wide range of expressions can be encountered. Finding a general solution seems very difficult. On the other hand, it often happens that only a few terms of the differential expression are important, in the sense that they determine the basic nature of the solution. Indeed, in the expressions (2.2), (2.3), and (2.4) it will be seen that the first two terms are dominant. ((2.7) is a special case that will be discussed later.) These first two terms are of the form

$$(3.1) \quad \sum_{i=1}^{p} \alpha_i x_i^{(1-m)} \phi_i^{i(1)}(x) + \sum_{i=1}^{p} b_i x_i^{-m} \phi_i^2(x).$$

Rough heuristic arguments can be given which show, for any m, that the differential expression $\Delta_m(x)$ resulting from the application of Corollaries 1 and 2 should behave as in (3.1). The basic idea is that multiplying a term of the loss by θ_i or θ_i^{-1} is roughly equivalent to multiplying the corresponding term of the differential expression by c_i/x_i or x_i/c_i, respectively. No attempt will be made to make such heuristics precise, since (3.1) will be used simply as a guide in choosing solutions to the actual differential inequalities encountered.

The expression (3.1) is quite specific to the distribution and losses involved. It is, however, a typical special case of the more general form

$$(3.2) \quad \Delta(x) = \psi(x) \sum_{i=1}^{p} v_i(x_i) \phi_i^{i(1)}(x) + \sum_{i=1}^{p} w_i(x) \phi_i^2(x).$$

Many simultaneous estimation problems encountered will have an expression of this form dominant in the resultant differential inequality. The functions ψ, v_i, and w_i can be quite arbitrary, except that usually $w_i(x) \ge 0$ and $\psi(X) > 0$ with probability one. This will be assumed in the following analysis.

To find solutions ϕ to the inequality $\Delta(x) < 0$ ($\Delta(x)$ as in (3.2)), let $g_i(x_i)$ be an indefinite integral of $[1/v_i(x_i)]$ (so that $g_i'(x_i) = 1/v_i(x_i)$), and find numbers $b \geq 0$, $d_j \geq 0$ and β_j such that

$$\frac{w_i(x)g_i^2(x_i)}{\psi(x)(b+\sum_{j=1}^{p} d_j|g_j(x_j)|^{\beta_j})} \leq K < \infty, \qquad i=1,\ldots p.$$

(It is assumed that ψ, v_i, and w_i are such that this can be done.)

Theorem 1. If $p > \max\{\beta_i\}$ and $0 < c < (p-\max\{\beta_i\})/(pK)$, then

$$\phi_i(x) = \frac{-cg_i(x_i)}{b+\sum_{j=1}^{p} d_j|g_j(x_j)|^{\beta_j}} \qquad (i=1,\ldots p) \tag{3.3}$$

is a solution to $\Delta(x) < 0$.

Proof. Let D denote the denominator in (3.3). Clearly

$$\phi_i^{i(1)}(x) = \frac{-cg_i'(x_i)}{D} + \frac{cg_i(x_i)d_i|g_i(x_i)|^{(\beta_i-1)}\{\mathrm{sgn}[g_i(x_i)]\}g_i'(x_i)}{D^2}$$

$$= \frac{-c}{v_i(x_i)D} + \frac{c\beta_i d_i|g_i(x_i)|^{\beta_i}}{v_i(x_i)D^2} .$$

Since, by assumption,

$$\sum_{i=1}^{p} \frac{w_i(x)\phi_i^2(x)}{\psi(x)} = (\frac{c^2}{D})\sum_{i=1}^{p} \frac{w_i(x)g_i^2(x_i)}{\psi(x)D} \leq \frac{c^2pK}{D},$$

it is clear that

$$\Delta(x) = \psi(x)[\sum_{i=1}^{p} v_i(x)\phi_i^{i(1)}(x) + \sum_{i=1}^{p} w_i(x)\phi_i^2(x)/\psi(x)]$$

$$= \psi(x)[\frac{-cp}{D} + (\frac{c}{D})\frac{\sum_{i=1}^{p}\beta_i d_i|g_i(x_i)|^{\beta_i}}{D} + \frac{c^2pK}{D}]$$

$$= \psi(x)\left[\frac{-cp}{D} + \frac{c\max\{\beta_i\}}{D} + \frac{c^2pK}{D}\right]$$

$$= \frac{-c\psi(x)}{D}[p-\max\{\beta_i\}-cpK] < 0. \quad ||$$

It is generally desirable to choose the β_i in the above theorem as small as possible to maximize the improvement in risk and minimize the number of dimensions needed for improvement. The d_i are arbitrary positive constants often chosen for convenience, as will be seen in the following examples. The constants c and b will typically have to be adjusted when dealing with the actual differential inequality (of which $\Delta(x)$ is, hopefully, an approximation) to take care of nondominant terms. Examples of this will be seen in Section 4.

Example 1. Let $\psi(x)=2$, $v_i(x_i)=a_ix_i^{(1-m)}$ and $w_i(x)=b_ix_i^{-m}$ ($a_i > 0$ and $b_i > 0$), and assume a solution to $\Delta(x) < 0$ is sought for $0 < x_i < \infty$, $i=1,\ldots p$. (This, recall, is (heuristically) the situation for the gamma problem.) Clearly

$$g_i(x_i) = \int \frac{1}{a_ix_i^{(1-m)}}\,dx_i = \begin{cases} \dfrac{x_i^m}{ma_i} & \text{if } m \neq 0 \\[2ex] \dfrac{\log(x_i)}{a_i} & \text{if } m = 0 \end{cases}.$$

Also

$$\frac{w_i(x)g_i^2(x_i)}{\psi(x)(b+\sum_{j=1}^{p} d_j|g_j(x_j)|^{\beta_j})} = \begin{cases} \dfrac{b_ix_i^{-m}x_i^{2m}}{2(ma_i)^2[b+\Sigma d_j\{x_j^m/(ma_j)\}^{\beta_j}]} & \text{if } m \neq 0 \\[2ex] \dfrac{b_i(\log x_i)^2}{2a_i^2[b+\Sigma d_ja_j^{-1}|\log x_j|^{\beta_j}]} & \text{if } m = 0 \end{cases}.$$

This quantity is bounded if $\beta_i \geq 1$ $(m \neq 0)$ or $\beta_i \geq 2$ $(m=0)$. Since small β_i are desirable, this suggests choosing $\beta_i=1$ if $m \neq 0$ and $\beta_i=2$ if $m=0$. Theorem 1 then gives solutions to $\Delta(x) < 0$ if $p \geq 2$ $(m \neq 0)$ or $p \geq 3$ $(m=0)$.

A better bound for c than that in Theorem 1 can be obtained if $d_i=b_i/(ma_i)$ $(m \neq 0)$ or $d_i=b_i/a_i$ $(m=0)$ are chosen in (3.3). Direct calculation as in Theorem 1 then shows that

$$\Delta(x) \leq \begin{cases} \dfrac{-c(p-1-c/2)}{[b+\sum_{i=1}^{p} b_i x_i^m/(ma_i)^2]} & \text{if } m \neq 0 \\[2ex] \dfrac{-c(p-2-c/2)}{[b+\sum_{i=1}^{p} b_i(\log x_i)^2/a_i^2]} & \text{if } m = 0 \end{cases}$$

The conclusion is that $\Delta(x) < 0$ for

$$\phi_i(x) = \begin{cases} \dfrac{-cx_i^m}{ma_i[b+\sum_{j=1}^{p} b_j x_j^m/(ma_j^{\,2})]} & \text{if } m \neq 0,\ b \geq 0,\ p \geq 2,\ 0 < c < 2(p-1) \\[2ex] \dfrac{-c(\log x_i)}{a_i[b+\sum_{j=1}^{p} b_j(\log x_j)^2/a_j^2]} & \text{if } m = 0,\ b \geq 0,\ p \geq 3,\ 0 < c < 2(p-2) \end{cases} \tag{3.4}$$

In the gamma problem it will be estimators $\delta_i^*(x) = (\alpha_i+1)^{-1}x_i(1+\phi_i(x))$, with the ϕ_i as in (3.4), that will be shown to be better than δ^o.

Several interesting observations about these estimators δ^* can be made. First, p=2 seems to be the standard minimal dimension in which improvement upon δ^o is possible. Three dimensions appear needed only for the loss corresponding to m=0. Also of interest is the form of the estimators, themselves. The "correction" to $\delta_i^o(x)$ is

$$\delta_i^*(x) - \delta_i^o(x) = (\alpha_i+1)^{-1}x_i\phi_i(x).$$

If $m > 0$ this correction is always negative (i.e. δ_i^o is pulled towards zero); if $m < 0$ the correction is always positive (i.e. δ_i^o is pulled towards infinity!); while if $m=0$ the estimator corrects δ_i^o negatively if $x_i > 1$ and positively if $x_i < 1$ (i.e. pulls δ_i^o towards $(\alpha_i+1)^{-1}$). This indicates that shrinking towards a point seems to be the exception rather than the rule, and that usually improved simultaneous estimators will pull to, or away from, a boundary of the parameter space. It also seems reasonable to believe that the need for $p=3$ when $m=0$ is due to the fact that the estimator can't decide which way to correct δ^o.

For $m > 0$, the $\phi_i(x)$ in (3.4) are very similar to correction factors obtained by Clevenson and Zidek (1975) for estimating Poisson means λ_i under the loss $\sum_{i=1}^{p} \lambda_i^{-1}(\delta_i - \lambda_i)^2$. For $m=0$, the ϕ_i are similar to the correction factors in Peng (1978) for estimating Poisson means under the loss $\sum_{i=1}^{p} (\delta_i - \lambda_i)^2$. The ϕ_i for $m=0$ are also related to "Stein type" estimators, as can be seen by letting $y_i = (\log x_i)$. (This is to be expected from Brown (1966), since the transformed problem induced by the log transform is a location vector problem when $m=0$.) For $m < 0$, however, the estimators determined by (3.4) seem completely novel. When $m=-2$, for example, (corresponding to standard squared error loss) the suggested estimator for θ_i is

$$\delta_i^*(x) = \frac{x_i}{(\alpha_i+1)} + \frac{c/x_i}{2(\alpha_i+1)a_i[b + \sum_{j=1}^{p} b_j (2a_j)^{-2} x_j^{-2}]} .$$

<u>Example 2.</u> Assume $X_1, \ldots X_p$ are independently $N(\theta_i, 1)$ (i.e. normal with mean θ_i and variance 1), and that it is desired to estimate $\theta = (\theta_1, \ldots \theta_p)$ under loss $L(\delta, \theta) = \sum_{i=1}^{p} (\delta_i - \theta_i)^2$. It can easily be calculated using Lemma 1 that if $\delta_i(x) = x_i + \phi_i(x)$, then

$$R(\delta, \theta) = p + E_\theta\{2 \sum_{i=1}^{p} \phi_i^{i(1)}(x) + \sum_{i=1}^{p} \phi_i^2(x)\}.$$

(This result was obtained by Stein (1973).) Noting that $\delta^{o}(x) = x$ has risk $R(\delta^{o},\theta) = p$, it is clear that an improved estimator will have been found if a solution to

$$(3.5) \qquad \Delta(x) = 2\sum_{i=1}^{p}\phi_i^{i(1)}(x) + \sum_{i=1}^{p}\phi_i^2(x) < 0$$

is obtained. This is the form (3.3) with $\psi(x) = 2$, $v_i(x_i) = 1$, and $w_i(x) = 1$. Calculating

$$g_i(x_i) = \int \frac{1}{v_i(x_i)}\, dx_i = x_i,$$

and using $\beta_i = 2$ as suggested by Theorem 1, the indicated solutions to (3.5) are

$$\phi_i(x) = \frac{-cx_i}{b+\Sigma x_j^2} .$$

A direct calculation gives that these are indeed solutions for $b \geq 0$ and $0 < c < 2(p-2)$. Note that $p > 2 = \beta_i$ is required as indicated by Theorem 1. The above choice of the ϕ_i gives rise to a standard "Stein estimator" for a multivariate normal mean.

Example 3. In the previous two examples, the differential inequality involved similar terms. To emphasize that this need not be so, we consider an interesting combination of the two previous examples. Indeed consider the differential inequality

$$(3.6) \qquad \Delta(x) = \sum_{i=1}^{2}[2\phi_i^{i(1)}(x) + \phi_i^2(x)] + \frac{2x_3^{\ 2}}{(\alpha+1)^2}\ \phi_3^{3(1)}(x) + \frac{x_3\phi_3^2(x)}{(\alpha+1)} .$$

(These are the dominant terms arising from the simultaneous estimation of two normal means under squared error loss and one normalized gamma mean under loss $\theta^{-1}(1-\delta\theta)^2$.) Even though the third coordinate terms are quite different, the solution given in Theorem 1 to $\Delta(x) < 0$ is still valid. Indeed a calculation shows that for $b \geq 0$ and $0 < c < 2$ solutions to $\Delta(x) < 0$ are given by

(3.7)
$$\phi_i(x) = \begin{cases} \dfrac{-cx_i}{b+x_1^2+x_2^2+(\alpha+1)^3/x_3} & \text{if } i = 1 \text{ or } 2 \\[2ex] \dfrac{c(\alpha+1)^2/x_3}{b+x_1^2+x_2^2+(\alpha+1)^3/x_3} & \text{if } i = 3 \end{cases}.$$

The implication of this is that in seeking improved simultaneous estimators, it doesn't matter what problems are combined. More will be said of this later.

4. Results for the Gamma Family

Using the results of Sections 2 and 3, we now proceed with a rigorous analysis of simultaneous estimation for the gamma problem and the four losses of interest.

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

By the results of Sectin 2, an estimator better than δ^o will have been determined if (2.2) can be solved. The heuristic solution to (2.2) is given in (3.4), namely

(4.1)
$$\phi_i(x) = \frac{c(\alpha_i+1)^2/x_i}{b+\sum_{j=1}^{p}(\alpha_j+1)^3/x_j}$$

$(m=-1,\ a_i=1/(\alpha_i+1)^2,\ b_j=(\alpha_j+1))$. Noting that

$$\phi_i^{i(1)}(x) = \frac{-c(\alpha_i+1)^2 x_i^{\ 2}}{b+\Sigma(\alpha_j+1)^3/x_j} + \frac{c(\alpha_i+1)^5/x_i^3}{(b+\Sigma(\alpha_j+1)^3/x_j)^2} < 0,$$

it is clear that

$$\Delta_{-1}(x) = 2\sum_{i=1}^{p} \frac{x_i^2}{(\alpha_i+1)^2} \phi_i^{i(1)}(x) + \sum_{i=1}^{p} \frac{\hat{x}_i}{(\alpha_i+1)} \phi_i^2(x) + 2\Sigma \frac{x_i^2}{(\alpha_i+1)^2} \phi_i^{i(1)}(x)\phi_i(x)$$

$$< 2\sum_{i=1}^{p} \frac{x_i^2}{(\alpha_i+1)^2} \phi_i^{i(1)}(x) + \sum_{i=1}^{p} \frac{x_i}{(\alpha_i+1)} \phi_i^2(x).$$

But this is the basic expression which was analyzed in Example 1 of Section 3. The conclusion is that $\Delta_{-1}(x) < 0$ for the solution (4.1) with $b \geq 0$ and $0 < c < 2(p-1)$. Thus a better estimator than δ^o for $p \geq 2$ is

$$\delta_i^*(x) = \frac{x_i}{(\alpha_i+1)}(1+\phi_i(x)) = \frac{x_i}{(\alpha_i+1)} + \frac{c(\alpha_i+1)}{b+\sum_{j=1}^{p}(\alpha_j+1)^3/x_j} .$$

This estimator can easily be shown to satisfy the conditions of Corollary 1 (which was used in the development of $\underline{\Delta}_1(x)$). Natural choices of b and c are b=0 and c=(p-1).

<u>Case 2</u>. $L(\delta,\theta) = \sum_{i=1}^{p} (1-\delta_i\theta_i)^2$.

It is now necessary to find a solution to $\Delta_0(x) < 0$, where $\Delta_0(x)$ is given by (2.3). The solutions suggested by (3.4) for the first two terms of (2.3) are

$$\phi_i(x) = \frac{-c(\alpha_i+1)(\log x_i)}{b+\sum_{j=1}^{p}\alpha_j(\alpha_j+1)(\log x_j)^2}$$

($a_i=1/(\alpha_i+1)$, m=0, and $b_j = \alpha_j/(\alpha_j+1)$). Using this choice in (2.3) and defining for convenience, $y_i=(\log x_i)$ and $D = b+\sum_{j=1}^{p}\alpha_j(\alpha_j+1)(\log x_j)^2$, a calculation gives

$$(4.2)\quad \Delta_U(x) = \frac{-2c}{D}\{p - \frac{(2+c/2)\Sigma\alpha_i(\alpha_i+1)y_i^2}{D} - \Sigma\frac{1}{(\alpha_i+1)}$$

$$+ \frac{4c}{D}\{\frac{3\Sigma\alpha_i y_i}{D} - \frac{\Sigma\alpha_i y_i^2}{D} - \frac{4\Sigma\alpha_i^2(\alpha_i+1)y_i^3}{D^2}\}$$

$$+ \frac{2c^2}{D}\{\frac{\Sigma(2\alpha_i+1)y_i}{D} - \frac{10\Sigma\alpha_i(\alpha_i+1)y_i^2}{D^2} - \frac{2\Sigma(2\alpha_i+1)\alpha_i(\alpha_i+1)y_i^3}{D^2}$$

$$+ \frac{12\Sigma\alpha_i^2(\alpha_i+1)^2 y_i^4}{D^3} + \frac{p}{D}\}.$$

In simplifying this expression, the following two lemmas are useful. Their proofs are standard calculations using Lagrange multipliers and will be omitted.

Lemma 1. For $a_i > 0$, $c_i > 0$, and $b > 0$, the expression

$$\frac{\sum_{i=1}^{p} a_i y_i}{b + \sum_{i=1}^{p} c_i y_i^2}$$

is maximized at

$$y_i = \frac{a_i}{c_i}\left(\frac{b}{\sum_{j=1}^{p} a_j^2/c_j}\right)^{1/2}, \qquad i=1,\ldots p,$$

achieving a maximum value of

$$\frac{(\sum_{j=1}^{p} a_j^2/c_j)^{1/2}}{2b^{1/2}}.$$

Lemma 2. For $b > 0$ and $z_i \geq 0$ $(i=1,\ldots p)$, the expression

$$\sum_{j=1}^{p} z_j^2/(b + \sum_{j=1}^{p} z_j)^3$$

is maximized when $z_j = 2b/p$ and has a maximum value of $4/(27bp)$.

Denote, for convenience, the three terms on the right hand side of (4.2) as T_1, T_2, and T_3 respectively. Since

$$\frac{\Sigma\alpha_i(\alpha_i+1)y_i^2}{D} = \frac{\Sigma\alpha_i(\alpha_i+1)y_i^2}{b+\Sigma\alpha_i(\alpha_i+1)y_i^2} \leq 1,$$

it is clear that

$$(4.3) \qquad T_1 \leq \frac{-2c}{D}\{p - (2+c/2) - \Sigma\frac{1}{(\alpha_i+1)}\}.$$

To bound T_2, note that for $(i=1,\dots p)$

$$\alpha_i|y_i| \geq \frac{\alpha_i^2(\alpha_i+1)|y_i|^3}{b + \sum_{j=1}^{p}\alpha_j(\alpha_j+1)y_j^2} .$$

Hence

$$\frac{3\Sigma\alpha_i y_i}{D} - \frac{4\Sigma\alpha_i^2(\alpha_i+1)y_i^3}{D^2} \leq \frac{3\Sigma\alpha_i|y_i|}{D},$$

and so

$$T_2 \leq \frac{12c}{D}\{\frac{\Sigma\alpha_i|y_i|}{D}\}.$$

Lemma 1 can be applied to this last expression (with $a_i=\alpha_i$ and $c_i=\alpha_i(\alpha_i+1)$) to give

$$(4.4) \qquad T_2 \leq \frac{12c}{D}\{\frac{[\Sigma\alpha_i/(\alpha_i+1)]^{1/2}}{2b^{1/2}}\} = \frac{6c[\Sigma\alpha_i/(\alpha_i+1)]^{1/2}}{Db^{1/2}} .$$

To deal with T_3, note first that for $i=1,\dots p$,

$$\alpha_i(\alpha_i+1)y_i^2 > \frac{\alpha_i^2(\alpha_i+1)^2y_i^4}{D} .$$

Using this with Lemma 2 (setting $z_i = \alpha_i(\alpha_i+1)y_i^2$) gives

$$(4.5) \qquad \frac{-10\Sigma\alpha_i(\alpha_i+1)y_i^2}{D^2} + \frac{12\Sigma\alpha_i^2(\alpha_i+1)^2y_i^4}{D^3} \leq \frac{2\Sigma\alpha_i^2(\alpha_i+1)^2y_i^4}{D^3} \leq \frac{8}{27bp} .$$

Observe next that for $i=1,\ldots p$,

$$(2\alpha_i+1)|y_i| > \frac{(2\alpha_i+1)\alpha_i(\alpha_i+1)|y_i|^3}{D},$$

which, with Lemma 1, shows that

$$(4.6)\quad \frac{\Sigma(2\alpha_i+1)y_i}{D} - \frac{2\Sigma(2\alpha_i+1)\alpha_i(\alpha_i+1)y_i^3}{D^2} \le \frac{\Sigma(2\alpha_i+1)|y_i|}{D}$$

$$\le \frac{[\Sigma(2\alpha_i+1)^2/\{\alpha_i(\alpha_i+1)\}]^{1/2}}{2b^{1/2}}$$

$$= \frac{[4p+\Sigma 1/\{\alpha_i(\alpha_i+1)\}]^{1/2}}{2b^{1/2}}.$$

Using (4.5) and (4.6), together with the observation that $p/D \le p/b$, it follows that

$$(4.7)\quad T_3 \le \frac{2c^2}{D}\ \{\frac{p}{b} + \frac{8}{27bp} + \frac{[4p+\Sigma 1/\{\alpha_i(\alpha_i+1)\}]^{1/2}}{2b^{1/2}}\}.$$

Combining (4.3), (4.4), and (4.7), the following bound is obtained.

$$(4.8)\quad \Delta_0(x) \le \frac{-2c}{D}\ [\ \{p-2-\Sigma\frac{1}{(\alpha_i+1)} - \frac{3[\Sigma\alpha_i/(\alpha_i+1)]^{1/2}}{b^{1/2}}\}$$

$$- \frac{c}{2}\ \{1+\frac{2p}{b} + \frac{16}{27bp} + \frac{[4p+\Sigma\{\alpha_i(\alpha_i+1)\}^{-1}]^{1/2}}{b^{1/2}}\}].$$

Equality can hold only if all $x_i=1$ (i.e. $y_i=0$). Hence we have a solution (with probability one) to $\Delta_0(x) < 0$, providing

$$(4.9)\quad 0 < c \le \frac{2\{p-2-\Sigma(\alpha_i+1)^{-1}-3b^{-1/2}[\Sigma\alpha_i/(\alpha_i+1)]^{1/2}\}}{1+(2p/b)+(16/[27bp])+b^{-1/2}[4p+\Sigma\{\alpha_i(\alpha_i+1)\}^{-1}]^{1/2}}.$$

By choosing b large enough, it is clear that a solution always exists providing

$$p > 2+\Sigma(\alpha_i+1)^{-1},$$

which if $\alpha_i \ge 3$ will be satisfied for $p \ge 3$.

For c satisfying (4.9) (note $b > 0$ is necessary), the estimator improving upon δ^o is

$$\delta_i^*(x) = \frac{x_i}{(\alpha_i+1)} \left(1 - \frac{c(\alpha_i+1)\log x_i}{b+\Sigma\alpha_j(\alpha_j+1)(\log x_j)^2}\right).$$

This estimator essentially shrinks towards $(\alpha_i+1)^{-1}$ (attained when $x_i=1$). To achieve shrinkage towards a point $\gamma_i/(\alpha_i+1)$, simply replace $(\log x_i)$ above by $[(\log x_i)-(\log \gamma_i)]$. It is easy to check that none of the calculations leading up to (4.8) are affected by this change. Hence such an estimator still improves upon δ^o, and allows shrinkage towards apriori suspected values of θ_i^{-1}.

A final observation is that $\delta_i^*(x)$ should clearly never be allowed to be negative. One method of preventing this is to truncate the estimator at zero. Alternatively, b could be chosen to ensure that $\delta_i^*(x)$ is never negative. Indeed, using Lemma 1 it can be checked that

$$\frac{c(\alpha_i+1)|\log x_i|}{b+\Sigma\alpha_j(\alpha_j+1)(\log x_j)^2} \le \frac{c(1+\alpha_i^{-1})^{1/2}}{2b^{1/2}},$$

so that choosing $b \ge (c^2/4)\,(1+1/\min\{\alpha_i\})$ will suffice.

Case 3. $L(\delta,\theta) = \sum_{i=1}^{p} \theta_i(1-\delta_i\theta_i)^2.$

It is desired to find a solution to $\Delta_1(x) < 0$, where $\Delta_1(x)$ is given by (2.4). The solutions suggested by (3.4) for the first two terms of (2.4) are (choosing $b=0$, $a_i=\alpha_i/(\alpha_i+1)$, and $m=1$)

$$\phi_i(x) = \frac{-c(\alpha_i+1)x_i/\alpha_i}{\sum_{j=1}^{p} d_j x_j} \quad . \tag{4.10}$$

(It will be convenient to choose the d_j in a manner somewhat different than suggested by (3.4).) In order that the resultant estimator satisfies the conditions of Corollary 1 of Section 2 (so that the derivation of (2.4) is valid), it is necessary to assume that $\alpha_i > 1$, $i=1,\ldots p$.

Defining $D = \sum_{j=1}^{p} d_j x_j$, a lengthy calculation using (4.10) in (2.4) gives

$$(4.11) \quad \Delta_1(x) = \frac{-2(p-1)c}{D} + \frac{c^2\Sigma(\alpha_i+1)(\alpha_i+2)(\alpha_i+3)\alpha_i^{-2}x_i}{D^2} + \frac{8c\Sigma d_i x_i/\alpha_i}{D^2} + T_1+T_2+T_3+T_4,$$

where

$$T_1 = \frac{-8c\Sigma d_i^2 x_i^2/\alpha_i}{D^3} - \frac{24c^2\Sigma d_i^3 x_i^4/\alpha_i^2}{D^5},$$

$$T_2 = -12c\left\{\frac{\Sigma d_i^2 x_i^2/[\alpha_i(\alpha_i+1)]}{D^3} - \frac{\Sigma d_i^3 x_i^3/[\alpha_i(\alpha_i+1)]}{D^4}\right\},$$

$$T_3 = -6c^2\left\{\frac{\Sigma d_i(\alpha_i+1)x_i^2\alpha_i}{D^3} + \frac{4\Sigma d_i(\alpha_i+1)x_i^2/\alpha_i^2}{D^3} - \frac{3\Sigma d_i^2(\alpha_i+1)x_i^3/\alpha_i^2}{D^4}\right\},$$

$$T_4 = 12c^2\left\{\frac{3\Sigma d_i^2 x_i^3/\alpha_i^2}{D^4} - \frac{\Sigma d_i x_i^2/\alpha_i^2}{D^3}\right\}.$$

Noting that for $K_i > 0$, $d_i^2 x_i^2 K_i \geq d_i^3 x_i^3 K_i/D$, it is clear that

$$T_2 \leq 0,$$

$$(4.12) \quad T_3 \leq -6c^2\left\{\frac{\Sigma d_i(\alpha_i+1)x_i^2/\alpha_i}{D^3} + \frac{\Sigma d_i(\alpha_i+1)x_i^2/\alpha_i^2}{D^3}\right\} = \frac{-24c^2\Sigma d_i(\alpha_i+1)^2 x_i^2/\alpha_i^2}{D^3},$$

and

$$T_4 \leq \frac{24c^2\Sigma d_i^2 x_i^3/\alpha_i^2}{D^4} - \frac{24c^2\Sigma d_i x_i^2/\alpha_i^2}{D^3}.$$

Hence

$$(4.13) \quad T_3 + T_4 \leq -6c^2\left\{\frac{\Sigma d_i(\alpha_i+1)^2 x_i^2/\alpha_i^2}{D^3} - \frac{4\Sigma d_i x_i^2/\alpha_i^2}{D^3}\right\} = \frac{-6c^2\Sigma d_i[(\alpha_i+1)^2-4]x_i^2/\alpha_i^2}{D^3}.$$

At this point, two lemmas are needed. Their proofs are again simple Lagrange multiplier arguments.

Lemma 3. For $a_i > 0$, $b_i > 0$, and $x_i > 0$ $(i=1,\ldots p)$, the expression

$$(\sum_{i=1}^{p} a_i x_i^2)/(\sum_{i=1}^{p} b_i x_i)^2$$

is minimized when x_i is a multiple of b_i/a_i, attaining a minimum value of

$$1/(\sum_{i=1}^{p} b_i^2/a_i).$$

Lemma 4. For $a_i > 0$, $b_i > 0$, and $x_i > 0$ $(i=1,\ldots p)$, the expression

$$(\sum_{i=1}^{p} a_i x_i^4)/(\sum_{i=1}^{p} b_i x_i)^4$$

is minimized when x_i is a multiple of $(b_i/a_i)^{1/3}$, attaining a minimum value of

$$1/(\sum_{i=1}^{p} b_i^{4/3}/a_i^{1/3})^3.$$

Applying these lemmas to T_1 and the last expression in (4.13), it can be concluded that

$$(4.14)\quad T_1+T_2+T_3+T_4 \le \frac{-8c}{D(\Sigma\alpha_i)} - \frac{24c^2}{D(\Sigma[d_i\alpha_i^2]^{1/3})^3} - \frac{6c^2}{D(\Sigma d_i\alpha_i^2/[(\alpha_i+1)^2-4])} .$$

Returning to the second term of (4.11), it seems natural to define

$$d_i = (\alpha_i+1)(\alpha_i+2)(\alpha_i+3)/\alpha_i^2.$$

Noting, finally that

$$\frac{\Sigma d_i x_i/\alpha_i}{D} \lessgtr \frac{1}{\min\{\alpha_i\}} ,$$

it can be concluded that

$$\Delta_1(x) \le \frac{-2c}{D} \{[p-1 - \frac{4}{\min\{\alpha_i\}} + \frac{4}{(\Sigma\alpha_i)}] - \frac{c}{2} [1-6(\Sigma \frac{(\alpha_i+1)(\alpha_i+2)}{(\alpha_i-1)})^{-1}$$

$$- 24(\Sigma[(\alpha_i+1)(\alpha_i+2)(\alpha_i+3)]^{1/3})^{-3}]\}.$$

The ϕ_i in (4.10) are thus solutions to $\Delta_1(x) < 0$ for

$$(4.15) \quad 0 < c < \frac{2[p-1-4/\min\{\alpha_i\}+4/(\Sigma\alpha_i)]}{1-6(\Sigma \frac{(\alpha_i+1)(\alpha_i+2)}{(\alpha_i-1)})^{-1} - 24(\Sigma[(\alpha_i+1)(\alpha_i+2)(\alpha_i+3)]^{1/3})^{-3}}.$$

(It can be checked for $\alpha_i > 1$ that the denominator above is always positive.) To have a solution we thus need

$$p > 1 + 4[\frac{1}{\min\{\alpha_i\}} - \frac{1}{(\Sigma\alpha_i)}].$$

If $\alpha_i > 4$ $(i=1,\ldots p)$, it follows that solutions will exist when $p \geq 2$. If the α_i are all equal, the numerator in (4.15) can be rewritten $(p-1)(1-4/[p\alpha])$, so solutions will exist for $p > 1$ and $p > \frac{4}{\alpha}$. Thus $\alpha=3$ will suffice to ensure the existence of solutions for $p \geq 2$.

The estimator improving upon δ^o is

$$\delta_i^*(x) = \frac{x_i}{(\alpha_i+1)} \left(1 - \frac{c(\alpha_i+1)x_i/\alpha_i}{\sum_{j=1}^{p} (\alpha_i+1)(\alpha_i+2)(\alpha_i+3)x_i/\alpha_i^2}\right).$$

As before, this estimator should be truncated at zero (if necessary) to prevent possible negative values.

<u>Case 4</u>. $L(\delta,\theta) = \sum_{i=1}^{p} \theta_i^{-2}(1-\delta_i\theta_i)^2.$

This case requires a different technique of analysis, due to the fact that Corollary 2 of Section 2 was used to derive $\Delta_{-2}(x)$ in (2.7). The difficulty is that we can't work directly with the ϕ_i, but instead must deal with the functions h_i.

A natural approach is to choose ϕ_i as suggested by (3.4), determine the h_i from the ϕ_i, and see if $\Delta_{-2}(x) < 0$ for this choice. Unfortunately, the analysis becomes too difficult and an alternative approach is needed. One that suggests itself is to try and choose the h_i so that $\Delta_{-2}(x)$ is relatively easy to deal with. Such a simplifying choice is

$$h_i(x) = (\alpha_i+1)^2 x_i^{(\alpha_i-1)} h(x),$$

where h is a function to be determined later. For this choice, (2.7) reduces to

$$(4.16)\quad \Delta_{-2}(x) = -4ph(x) + 2\Sigma x_i h^{i(1)}(x) + \Sigma(\alpha_i+1)^2[h^{i(1)}(x)]^2 + \Sigma(\alpha_i^2-1)^2 x_i^{-2} h^2(x) + 2h(x)\Sigma(\alpha_i+1)^2(\alpha_i-1)x_i^{-1}h^{i(1)}(x).$$

Using (2.8), a calculation gives that

$$(4.17)\quad \phi_i(x) = x_i^{-\alpha_i} h_i^{i(1)}(x) = (\alpha_i+1)^2(\alpha_i-1)x_i^{-2}h(x) + (\alpha_i+1)^2 x_i^{-1} h^{i(1)}(x).$$

The heuristics of Section 3 suggest that $\phi_i(x)$ should be as in (3.4), i.e.

$$\phi_i(x) \cong \frac{cx_i^{-2}}{2a_i[b+\Sigma d_j x_j^{-2}]} .$$

The first term on the right hand side of (4.17) is like this if

$$h(x) = \frac{c}{b+\sum_{j=1}^{p} d_j x_j^{-2}}$$

is chosen. For this choice of h, (4.16) becomes (letting D denote the denominator of h)

$$\Delta_{-2}(x) = \frac{-4pc}{D} + \frac{4c\Sigma d_i x_i^{-2}}{D^2} + \frac{4c^2\Sigma(\alpha_i+1)^2 d_i^2 x_i^{-6}}{D^4} + \frac{c^2\Sigma(\alpha_i^2-1)^2 x_i^{-2}}{D^2} + \frac{4c^2\Sigma(\alpha_i+1)^2(\alpha_i-1)d_i x_i^{-4}}{D^3}$$

$$< \frac{-4pc}{D} + \frac{4c}{D} + \frac{4c^2\Sigma(\alpha_i+1)^2 x_i^{-2}}{D^2} + \frac{c^2\Sigma(\alpha_i^2-1)^2 x_i^{-2}}{D^2} + \frac{4c^2\Sigma(\alpha_i+1)^2(\alpha_i-1)x_i^{-2}}{D^2}$$

$$= \frac{-4(p-1)c}{D} + \frac{c^2\Sigma(\alpha_i+1)^4 x_i^{-2}}{D^2} .$$

If $d_i = (\alpha_i+1)^4$ is now chosen, it follows that

$$\Delta_{-2}(x) < \frac{-c}{D}\{4(p-1)-c\} .$$

The given h is thus a solution to $\Delta_{-2}(x) < 0$ for

$$0 < c \leq 4(p-1).$$

Clearly $p \geq 2$ suffices to give a solution.

Using (4.17), the estimator better than δ^o can be calculated to be

$$\delta_i^*(x) = \frac{x_i}{(\alpha_i+1)}(1+\phi_i(x)) = \frac{x_i}{(\alpha_i+1)} + \frac{c(\alpha_i^2-1)x_i^{-1}}{[b+\Sigma(\alpha_j+1)^4 x_j^{-2}]} + \frac{2c(\alpha_i+1)^5 x_i^{-3}}{[b+\Sigma(\alpha_j+1)^4 x_j^{-2}]^2} .$$

For this estimator and the corresponding h_i, it can be checked that the conditions of Corollary 2 of Section 2 (needed to derive $\Delta_{-2}(x)$) are satisfied. Note that for large α_i or large p, the third term of δ_i^* is likely to be considerably smaller than the second term (which was the correction factor derived from the heuristics).

5. Simultaneous Estimation Under Squared Error Loss

Usually the heart of a theory lies in methods and examples, which are what the last three sections concentrated on. It is possible, however, to construct a fairly elegant general theory, using the results in Sections 2 and 3. This will be demonstrated here for the simplest situation - squared error loss.

We will seek an improvement upon an estimator δ^o (which could be any estimator) in the general continuous exponential family setup. Thus assume $X = (X_1,\ldots Xp)$ is observed, where the X_i are independent with densities

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

with respect to Lebesgue measure on R^1. Assume it is desired to estimate

$\theta = (\theta_1, \ldots \theta_p)$ under loss

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

Using Lemma 1 of Section 2, it can be shown that (under the appropriate conditions)

$$R(\delta,\theta) = E_\theta[L(\delta(X),\theta)] = \sum_{i=1}^{p} \theta_i^2 - 2\sum_{i=1}^{p} E_\theta[\theta_i \delta_i(X)] + \sum_{i=1}^{p} E_\theta[\delta_i^2(X)]$$

$$= \sum_{i=1}^{p} \theta_i^2 - 2\sum_{i=1}^{p} E_\theta\left[\frac{t_i'(X_i)\delta_i(X)}{t_i(X_i)r_i'(X_i)} + \frac{\delta_i^{i(1)}(X)}{r_i'(X_i)} - \frac{\delta_i(X)r_i''(X_i)}{\{r_i'(X_i)\}^2}\right] + \sum_{i=1}^{p} E_\theta[\delta_i^2(X)].$$

Writing a competitor to δ^o (componentwise) as

(5.1) $$\delta_i^*(x) = \delta_i^o(x) - q_i(x)\phi_i(x),$$

a little algebra then gives that

(5.2) $$\Delta(\theta) = R(\delta^*,\theta) - R(\delta^o,\theta)$$

$$= \sum_{i=1}^{p} E_\theta\left[\frac{2\phi_i(X)q_i(X)}{r_i'(X_i)}\left\{\frac{q_i^{i(1)}(X)}{q_i(X)} - \delta_i^o(X)r_i'(X_i) + \frac{t_i'(X_i)}{t_i(X_i)} - \frac{r_i''(X_i)}{r_i'(X_i)}\right\}\right]$$

$$+ E_\theta\left[\sum_{i=1}^{p} \frac{2q_i(X)\phi_i^{i(1)}(X)}{r_i'(X_i)} + \sum_{i=1}^{p} q_i^2(X)\phi_i^2(X)\right].$$

To simplify this, define $s_i(x)$ as an indefinite integral (with respect to x_i) of $[\delta_i^o(x)r_i'(x_i)]$, and let

(5.3) $$q_i(x) = \frac{r_i'(x_i)e^{s_i(x)}}{t_i(x_i)}.$$

It is easy to check that the first term on the right hand side of (5.2) is then zero, so that

$$\Delta(\theta) = E_\theta[\Delta(X)],$$

where

$$\Delta(x) = \sum_{i=1}^{p} \frac{2q_i(x)\phi_i^{i(1)}(x)}{r_i'(x_i)} + \sum_{i=1}^{p} q_i^2(x)\phi_i^2(x).$$

To find an estimator δ^* better than δ^o, we seek a solution to $\Delta(x) < 0$. Providing

(5.4) $$q_i(x) = \tau(x)h_i(x_i)$$

(without loss of generality assuming that $\tau(x) > 0$) this is exactly the problem solved in Section 3 (with $\psi(x) = 2\tau(x)$ and $v_i(x_i) = h_i(x_i)/r_i'(x_i)$). The answer found there was to calculate the indefinite integrals

(5.5) $$g_i(x_i) = \int \frac{1}{v_i(x_i)}\,dx_i = \int \frac{r_i'(x_i)}{h_i(x_i)}\,dx_i,$$

find constants $b \geq 0$ and β_j such that (for i=1,...p)

(5.6) $$\frac{q_i^2(x)g_i^2(x_i)}{[2\tau(x)][b+\sum_{j=1}^{p}|g_j(x_j)|^{\beta_j}]} = \frac{\tau(x)h_i^2(x_i)g_i^2(x_i)}{2[b+\sum_{j=1}^{p}|g_j(x_j)|^{\beta_j}]} \leq K < \infty,$$

and use (for $p > \max\{\beta_i\}$ and $c > 0$ small enough) functions of the form

(5.7) $$\phi_i(x) = \frac{-cg_i(x_i)}{b+\sum_{j=1}^{p} d_j|g_j(x_j)|^{\beta_j}},$$

where the d_j are convenient constants. (Implicit in all these calculations are certain conditions on the functions involved; namely the existence of all necessary functions, derivatives, and integrals, and the conditions of Lemma 1 (Section 2) as applied to $h(x) = \delta_i^o(x)$ and $h(x) = \delta_i^*(x) = \delta_i^o(x) - q_i(x)\phi_i(x)$, $i=1,\ldots p$. No attempt is made to write down general sets of conditions, since verification in specific instances is usually very easy.) Some examples of this technique should prove enlightening.

Example 1. Assume the X_i are Gamma(α_i, θ_i) with $\alpha_i > 2$ $(i=1,\ldots p)$. (Note we are now trying to estimate the θ_i themselves, not the θ_i^{-1}.) Here $r_i'(x_i) = 1$ and $t_i(x_i) = x_i^{(\alpha_i-1)}$. Consider the "standard" (i.e. best multiple of x_i^{-1}) estimator defined by $\delta_i^o(x) = (\alpha_i-2)x_i^{-1}$. A calculation using (5.3) shows that $s_i(x) = (\alpha_i-2)(\log x_i)$ and $q_i(x) = x_i^{-1}$. As in (5.4) and (5.5) define $\tau(x) = 1$, $h_i(x_i) = x_i^{-1}$, and

$$g_i(x_i) = \int \frac{r_i'(x_i)}{h_i(x_i)} dx_i = \int x_i dx_i = x_i^2/2.$$

It is easy to check that (5.6) is satisfied for $\beta_i=1$ $(i=1,\ldots p)$. Hence solutions to $\Delta(x) < 0$ are given in (5.7) (for small enough c and $p > 1$). Indeed, choosing

$$\phi_i(x) = \frac{-cx_i^2}{b+\sum_{j=1}^{p} x_j^2},$$

a simple calculation verifies that $\Delta(x) < 0$ providing $b \geq 0$, $p \geq 2$, and $0 < c < 4(p-1)$. The estimators improving upon δ^o are thus given (componentwise) by

$$\delta_i^*(x) = \delta_i^o(x) - q_i(x)\phi_i(x) = \frac{(\alpha_i-2)}{x_i} + \frac{cx_i}{(b+\sum_{j=1}^{p} x_j^2)}.$$

Example 2. Assume the X_i are $\mathcal{N}(\theta_i,1)$, so that $r_i'(x_i) = -1$ and $t_i(x_i)=\exp\{-x_i^2/2\}$. Assume

$$\delta^o(x) = x - x\varphi(|x|^2),$$

where φ is a real valued function. Then

$$s_i(x) = \int -\delta_i^o(x)dx_i = \frac{-x_i^2}{2} + \frac{\lambda(|x|^2)}{2},$$

where $\lambda(z)$ is an indefinite integral of $\varphi(z)$. (Assume that λ exists for all $0 \le z < \infty$.) Hence by (5.3)

$$q_i(x) = -\exp\{\lambda(|x|^2)/2\}.$$

This is of the form (5.4) with $\tau(x) = \exp\{\lambda(|x|^2)/2\}$ and $h_i(x_i) = -1$. Hence define

$$g_i(x_i) = \int \frac{r_i'(x_i)}{h_i(x_i)} dx_i = \int (1)dx_i = x_i.$$

It is now necessary to find $b \ge 0$ and β_i such that (for i=1,...p)

$$\text{(5.8)} \qquad \frac{\tau(x)h_i^2(x_i)g_i^2(x_i)}{2[b+\sum_{j=1}^{p}|g_j(x_j)|^{\beta_j}]} = \frac{\exp\{\lambda(|x|^2)/2\}x_i^2}{2[b+\sum_{j=1}^{p}|x_j|^{\beta_j}]} \le K < \infty.$$

Assuming this can be done, estimators better than δ^o are (componentwise)

$$\text{(5.9)} \qquad \delta_i^*(x) = \delta_i^o(x) - q_i(x)\phi_i(x) = x_i - x_i\varphi(|x|^2) + \exp\{\lambda(|x|^2)/2\}\phi_i(x),$$

where the ϕ_i are given by (5.7) (for appropriate c) and $p > \max\{\beta_i\}$.

From Lemma 1 it is easy to check that the conditions needed for this analysis are

(5.10) (i) $\delta^o(x)$ and $q_i(x)\phi_i(x)$ are absolutely continuous in all coordinates; and

(ii) $E_0[|X_i\delta_i(X)| + |\delta_i^{i(1)}(X)|] < \infty$, and

$E_0[|X_i||q_i(X)\phi_i(X)| + |q_i'(X)\phi_i(X) + \phi_i^{i(1)}(X)q_i(X)|] < \infty$, for i=1,...p.

It remains only to check (5.8). A more convenient bound to work with can be obtained by noting that for reasons of symmetry it is desirable to choose all $\beta_j = \beta$. Note also that

(5.11) $$\min\{1, p^{(1-\beta/2)}\} \le \sum_{i=1}^{p} \left(\frac{|x_i|}{|x|}\right)^{\beta} \le p.$$

Recall that for (5.7) to be a solution it is necessary to have $p > \max\{\beta_i\} = \beta$, so from (5.11) it follows that for any feasible $0 \le \beta < p$,

$$K_1(p)|x|^{\beta} \le \sum_{i=1}^{p} |x_i|^{\beta} \le p|x|^{\beta},$$

where $K_1(p) > 0$. Noting also that summing over i in (5.8) does not qualitatively affect the bound (if each term is bounded the sum is, and if the sum is bounded each term is) it follows that a bound equivalent to (5.8) is

$$\frac{\exp\{\lambda(|x|^2)/2\}|x|^2}{bK_2(p) + |x|^{\beta}} \le K_3(p) < \infty,$$

where $K_2(p) > 0$ and $K_3(p) > 0$. Since $\lambda(z)$ is finite and continuous for $0 \le z < \infty$ (the finiteness by earlier assumption), it is clear for a fixed p that the above expression can be unbounded only as $|x| \to \infty$ (if $b > 0$). The verification of inadmissibility thus reduces to showing that

(5.12) $$\lim_{z\to\infty} z^{(1-\beta/2)}\exp\{\lambda(z)/2\} < \infty .$$

Note that to do this it is only necessary to know $\lambda(z)$ as $z\to\infty$.

Example 2 (a). $\delta^{o}(x)=(1-d)x$ $(d < 1)$. Here $\varphi(z)=d$ so $\lambda(z)= \int\varphi(z)dz=zd$ Clearly for $0 < d < 1$, (5.12) cannot be satisfied for any β. (This indicates the admissibility of δ^{o} for this choice of d, a well known result.) If $d < 0$, (5.12) is satisfied for $\beta=0$. Hence choosing $\phi_i(x)=-cx_i$, it follows that δ^* in (5.9) improves upon δ^{o} for appropriate c. (The conditions in (5.10) are easy to verify.) This, of course, is well known also. If $c=0$, (5.12) will be satisfied for $\beta=2$.

This corresponds to the basic situation in which Stein estimators improve upon $\delta^o(x)=x$. If $c > 1$ the conditions in (5.10) will be violated (though δ^o is clearly inadmissible), so the method will not work.

Example 2(b). $\delta^o(x)=\{1-a/(d+|x|^2)\}x$. Here $\varphi(z)=a/(d+z)$, so $\lambda(z)=\int\varphi(z)dz=a\log(d+z)$. Clearly

$$\lim_{z\to\infty} z^{(1-\beta/2)}\exp\{\lambda(z)/2\} = \lim_{z\to\infty} z^{(1-\beta/2)}(d+z)^{a/2} = \lim_{z\to\infty} z^{(1+a/2-\beta/2)}.$$

This is finite (and hence (5.12) is satisfied) for $\beta = (a+2)$. Recalling that this leads to an improved estimator only for $p > \beta = (a+2)$, this means that δ^o is inadmissible if $a < (p-2)$. (Again the conditions in (5.10) are easy to check.)

If one were solely interested in proving inadmissibility in Example 2, the results of Brown (1971) would apply (in for the most part greater generality). The nice features of the approach in Example 2 are that it is easy, and that it is constructive, with an explicit improved estimator being determined. Note that this method is not just for "simultaneous estimation", in that it applies to one dimensional problems also.

Unfortunately there are certain inadequacies of the method. The calculation of the q_i and the g_i, and the verification of (5.6) can be analytically difficult in some cases. Also, the improved estimators obtained can be unwieldy and are in no sense necessarily optimal improvements.

6. Conclusions and Generalizations

The results of the preceeding sections provide support for the statements made in the introduction. Improvement upon δ^o was most commonly obtainable in two dimensions, with only Case 2 of Section 4 and the standard Normal situation requiring at least three dimensions. Also, in Cases 1 and 4 of Section 4 and Example 2 of Section 5 the improved estimators, δ^*, corrected δ^o by shifting towards infinity, a rather surprising phenomenon. Finally, consider the following example, which undermines any intuitive Bayesian or empirical Bayesian explanations of the "Stein effect".

Example 3 (continued). Example 3 of Section 3 involved the analysis of the differential expression (3.6), which arises as follows. Assume X_1 and X_2 are independent $\mathcal{N}(\theta_i,1)$ random variables, and it is desired to estimate (θ_1,θ_2) under the loss $[(\delta_1-\theta_1)^2 + (\delta_2-\theta_2)^2]$. The standard estimator $\delta_i^o(x) = x_i$ (i=1,2) is known (James and Stein (1960)) to be admissible. If a competitor $\delta_i^*(x)=x_i+\phi_i(x)$ (i=1,2) to (δ_1^o,δ_2^o) is considered, Example 2 of Section 3 shows that

$$\sum_{i=1}^{2} [2\phi_i^{i(1)}(x) + \phi_i^2(x)]$$

is the integrand of the difference in risk between δ_i^* (i=1,2) and δ_i^o(i=1,2).

If (independently) X_3 is Gamma (α,θ_3), and it is desired to estimate $1/\theta_3$ under loss $\theta_3^{-1}(1-\delta_3\theta_3)^2$, it is known that $\delta_3^o(x_3)=x_3/(\alpha+1)$ is admissible. (Hodges and Lehmann (1951).) For a competitor $\delta_3^*(x)=(\alpha+1)^{-1}x_3(1+\phi_3(x))$ to $\delta_3^o(x)$, the analysis in Case 1 of Section 4 shows that the integrand of the difference in risk between $\delta_3^*(x)$ and $\delta_3^o(x)$ is bounded above by

$$2(\alpha+1)^{-2}x_3^2\phi_3^{3(1)}(x) + (\alpha+1)^{-1}x_3\phi_3^2(x),$$

providing $\phi_3^{3(1)}(x) \leq 0$.

Letting $\delta^* = (\delta_1^*,\delta_2^*,\delta_3^*)$, $\delta^o = (\delta_1^o,\delta_2^o,\delta_3^o)$, $X = (X_1,X_2,X_3)$, $\theta = (\theta_1,\theta_2,\theta_3)$, and assuming that $\phi_3^{3(1)}(x) \leq 0$, it follows that in estimating $(\theta_1,\theta_2,\theta_3^{-1})$ under the overall loss $[(\delta_1-\theta_1)^2+(\delta_2-\theta_2)^2+\theta_3^{-1}(1-\delta_3\theta_3)^2]$,

$$\Delta(\theta) = R(\delta^*,\theta) - R(\delta^o,\theta) \leq E_\theta[\Delta(X)],$$

where $\Delta(x)$ is given by (3.6). The solutions to $\Delta(x) \leq 0$ which are given in (3.7) can be shown to satisfy all necessary conditions of the derivation provided $b > 0$. Hence an estimator better than δ^o is (for $b > 0$, $0 < c < 2$)

$$\delta_i^*(x) = \begin{cases} x_i - \dfrac{cx_i}{b+x_1^2+x_2^2+(\alpha+1)^3/x_2} & \text{if } i=1 \text{ or } 2 \\[2ex] \dfrac{x_3}{(\alpha+1)} + \dfrac{c(\alpha+1)}{b+x_1^2+x_2^2+(\alpha+1)^3/x_2} & \text{if } i=3. \end{cases}$$

The implications of this are interesting. First, two completely unrelated problems, one involving a two dimensional normal mean and one a gamma scale parameter, can be combined to obtain an estimator improving upon admissible estimators in each separate problem. Secondly, the improved estimator treats the coordinates quite differently. In the first two coordinates $\delta_i^o(x) = x_i$ is shrunk towards zero, while for the last coordinate $\delta_3^o(x)$ is shifted towards infinity.

This example will be very difficult to explain using any Bayesian, empirical Bayesian, or other intuitive arguments. The point is not that such arqumeñts are not useful or enlightening in other situations, but that the basic Stein effect obtainable in simultaneous estimation appears to be a more basic and general phenomenon.

A number of questions remain concerning the actual application of the estimators found in Section 4 to practical problems involving simultaneous estimation of gamma parameters. One such question is whether or not an improved estimator can be found which will work well for several or all of the losses considered. Unfortunately, the drastic differences between the estimators that were obtained for the different losses indicate that hoping for a positive answer to the question may be somewhat optimistic.

A related and useful generalization is to weighted losses of the form

$$(6.1) \qquad L(\delta, \theta) = \sum_{i=1}^{p} q_i \theta_i^m (1-\delta_i \theta_i)^2,$$

where the q_i are positive constants designed to reflect the relative importance of the various coordinates. There are two possible methods of dealing with such a loss.

The first is to include the q_i in the differential expressions $\Delta(x)$, and then solve $\Delta(x) < 0$ along the same lines as before. The second method of dealing with the q_i is simpler, and so is used here.

This second method is discussed in Berger (1977b), and involves a decomposition to similar subproblems. Begin by ordering the q_i, say so that $q_1 \geq q_2 \dots \geq q_p$. Consider the subproblem of estimating $(\theta_1^{-1}, \dots \theta_j^{-1})$ under loss $\sum_{i=1}^{j} \theta_i^m (1-\delta_i\theta_i)^2$, and let $\delta^{(j)}(x_1, \dots x_j)$ be an improved estimator found in Section 4 (with $p=j$ and $x = (x_1, \dots x_j)$). If $j=1$ or the α_i are such that no improved estimator was obtained, let $\delta^{(j)} = \delta^{o(j)} = ((\alpha_1+1)^{-1}x_1, \dots (\alpha_j+1)^{-1}x_j)$. Finally, define for $i=1, \dots p$, and $j=1, \dots p$,

$$\beta_i^j = \begin{cases} 0 & \text{if } j < i \\ (q_j - q_{j+1})/q_i & \text{if } j \geq i, \end{cases}$$

where $q_{(p+1)}$ is defined to be zero. Berger (1977b) then proves that an estimator better than δ^o in the original problem with loss (6.1) is given coordinatewise by

$$\delta_i^*(x) = \sum_{j=i}^{p} \beta_i^j \delta_i^{(j)}(x)$$

(providing that in at least one of the subproblems, $\delta^{(j)}$ is not identically equal to $\delta^{o(j)}$).

Another problem of interest is that of the incorporation of prior information. In simultaneous estimation of normal means, it has been observed (see Berger (1977a) and Berger (1977b)) that to obtain significant practical improvement upon δ^o it is usually necessary to incorporate prior information. In the simplest case, this can be accomplished by shrinking δ^o towards an apriori "most likely" parameter value. It is, unfortunately, not clear how such prior information can be incorporated into the gamma estimators δ^*, except for Case 2 in which shrinkage towards an arbitrary point is possible. Perhaps a broader class of solutions to $\Delta(x) < 0$ is needed, allowing a solution corresponding to possible prior information to be chosen.

References

[1] Berger, J. (1976a). Inadmissibility results for generalized Bayes estimators of coordinates of a location vector. Ann. Statist. 4, 302-333.

[2] Berger, J. (1976b). Inadmissibility results for the best invariant estimator of two coordinates of a location vector. Ann. Statist. 4, 1065-1076.

[3] Berger, J. (1976c). Tail minimaxity in location vector problems and its applications. Ann. Statist. 4, 33-50.

[4] Berger, J. (1977a). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Technical Report #480, Dept. of Statistics, Purdue University.

[5] Berger, J. (1977b). Multivariate estimation with nonsymmetric loss functions. Technical Report #517, Dept. of Statistics, Purdue University.

[6] Berger, J. (1978). Minimax estimation of a multivariate normal mean under polynomial loss. J. Multivariate Anal., June 1978.

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

[8] Brown, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. Ann. Math. Statist. 42, 855-903.

[9] Brown, L. D. (1974). An heuristic method for determining admissibility of estimators-with applications. Technical Report, Rutgers University.

[10) Brown, L. D. (1975). Estimation with incompletely specified loss functions. J. Amer. Statist. Assoc. 70, 417-427.

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

[12] Hodges, J. L., Jr., and Lehmann, E. L. (1951). Some applications of the Cramer-Rao inequality. Proc. Second Berkeley Symp. Math. Statist. Prob. 13-22. Univ. of California Press.

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

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

[15] Peng, J. C. (1978). Simultaneous estimation of the parameters of independent Poisson distributions. Ann. Statist. To appear.

[16] Stein, C. (1965). Approximation of improper prior measures by probability measures. Bayes, Bernouilli, Laplace. Springer-Verlag, Berlin.

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

[18] Zaman, A. (1977). The single period control problem in econometrics. Ph.D. Thesis, Stanford University.