Estimation with Incompletely Specified Loss Functions (the Case of Several Location Parameters)

LAWRENCE D. BROWN*

If $p \geq 3$ *separate* normal means are to be estimated, the usual, invariant estimator is "inadmissible" if the p separate loss functions can be pooled into a single loss function for the ensemble-of-means problem. Here, we formulate a version of the problem where the losses cannot be pooled in this way, and find a surprisingly weak necessary and sufficient condition under which the usual estimator is inadmissible in a sense appropriate to this formulation. We also examine the spherically symmetric case and find a different necessary and sufficient condition there.

1. INTRODUCTION

Suppose x_1 is an observation of a normal random variable with unknown mean μ_1 and known variance $\sigma_1^2 > 0$. The usual estimator of μ_1 is given by $\delta_0(x_1) = x_1$, i.e., by the observation itself. If the loss is measured by a "squared error form," $L(\mu_1, \delta) = c_1(\mu_1 - \delta)^2$, $c_1 > 0$, then this estimator is admissible. That is, there is no estimator δ such that

$$E_{\mu_1}\{(\mu_1 - \delta)^2\} \leq E_{\mu_1}\{(\mu_1 - \delta_0)^2\} = \sigma^2$$

with inequality for some value of μ_1.

The function $R_L(\mu, \delta) = E_\mu\{L(\mu, \delta)\}$ is called the risk function. If one looks at the statistical problem through the risk function, admissibility appears as a minimal requirement which an estimator must satisfy in order to be used—an experimenter should not use an inadmissible estimator.[1]

Suppose p such independent problems are treated at the same time and $p \geq 3$. The measure of loss which seems natural is the sum of the losses of the three independent problems, that is

$$L(\mu, \delta) = \sum c_i(\mu_i - \delta_i)^2 \ . \qquad (1.1)$$

(Here $\mu = (\mu_1, \cdots, \mu_p)'$, $\delta = (\delta_1, \cdots, \delta_p)'$, $c_i > 0$.) The intuitively natural estimator is again the observation itself, $\delta_0(x) = x$. (Here $x = (x_1, \cdots, x_p)'$.) This is the estimator which results from p independent applications of the one-dimensional estimator to the respective coordinate problems. The surprise is that this natural estimator is now inadmissible.

This surprise was first discoverd by Charles Stein [17] (in the case $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_p^2$). Estimators which provide a significant improvement over the usual estimator were later presented in James and Stein [13], Bhattacharya [4] Baranchik [2], Strawderman [19] and in Efron and Morris [10]. (See also Alam [1], Strawderman [20] and Bock [5], for related results.)

The inadmissibility theorem just described is surprising precisely because the various coordinate problems appear to be independent. It is not intuitively clear how the combining of several *independent* problems can make unacceptable a procedure which was acceptable when applied to each coordinate problem. In actual fact the p problems as just described *are not independent*. It is true that the random *variables* $X_1, \cdots, X_p$ are independent, but the p *problems* are linked by the fact that the variances $\sigma_1^2, \cdots, \sigma_p^2$ are measured on a common scale, as are the losses.

We illustrate this linkage by an example. Suppose the $p = 3$ separate problems involve measurements on

1. The percent of sulfur dioxide in a certain air sample,
2. The impedance in ohms of a certain electrical circuit, and
3. The yield in tons of wheat on a selected plot of land (having no connection with the air sample in (1)).

Suppose that squared error is the appropriate form of loss function for each of the three problems considered separately. (This supposition of squared error loss will be discussed in Section 6. It will be seen that it is a convenient assumption, but the basic results to follow do not depend on it in any important way.) The constants c_1, c_2 and c_3 which appear in these loss functions are not initially measured in the same units. In the example they are measured in units of $(\cdot)/\text{percent}^2$, $(\cdot)/\text{ohms}^2$, and $(\cdot)/\text{tons}^2$, respectively. *Pooling* of the loss into the single loss function $L(\mu, \delta) = \sum c_i(\delta_i - \mu_i)^2$ as we have just done *implies* a determination that errors of $c_1^{-\frac{1}{2}}$ percent, $c_2^{-\frac{1}{2}}$ ohms, and $c_3^{-\frac{1}{2}}$ tons, respectively, result in losses which are equivalent to the experimenter(s). The placing of losses on a single scale of measurement then enables the placing of the numbers σ_i^2 on that same scale.

* Lawrence D. Brown is professor, Statistics Center, Rutgers University, New Brunswick, N.J. 08903. This research was supported in part by NSF Research Grant MPS-72-05075-A02.

[1] This general statement is correct provided a minimal complete class exists, which is the case in all the problems discussed in this article. The statement also tacitly assumes that if an estimator is inadmissible the experimenter is easily able to describe and use one of the better estimators which exist.

© Journal of the American Statistical Association
June 1975, Volume 70, Number 350
Theory and Methods Section

To form the pooled loss function, the three separate problems must, therefore, be linked together through a realistic determination of the relative values of c_1, c_2 and c_3. Sometimes this can be done and sometimes it cannot.

Case 1. Particularly when the three experiments are performed by three different, unrelated experimenters it may be impossible to arrive at any determination whatsoever of appropriate relative values of the constants c_1, c_2 and c_3. In this case a conservative approach is to adopt a procedure which is satisfactory under each of the possible loss functions: $L_i(\mu, \delta) = c_i(\mu_i - \delta_i)^2$, $0 < c_i < \infty$, $i = 1, 2, 3$.

In this case we will say that the class, $\mathcal{L}$, of allowable loss functions includes the three functions L_i. And we will say that the procedure δ_0 is $\mathcal{L}$-admissible if there is no procedure δ such that $R_L(\mu, \delta) \leq R_L(\mu, \delta_0)$ for all $L \in \mathcal{L}$, with strict inequality for some μ and some $L \in \mathcal{L}$. In this language the "conservative approach" referred to previously amounts to requiring that the procedure to be used is $\mathcal{L}$-admissible.

Case 2. On the other hand, perhaps some determination of the relative values of c_1, c_2 and c_3 can be made. It is conceivable that a process of discussion among the experimenters and compromise for the general good might, for example, yield a mutually acceptable range of values for the ratios c_i/c_j, $i \neq j$. For example, it might be possible for all to agree that $c_i/c_j \leq K_{ij} < \infty$, $i, j = 1, 2, 3$. (Note that this indeed compromises among the conflicting goals of the three experimenters since the ith experimenter (who naturally feels that his experiment is much more important than the others!) desires that c_i/c_j should be large for $j \neq i$). In this situation the statistician need only adopt a procedure which is satisfactory over those loss functions agreed on by all concerned; those of the form

$$L(\mu, \delta) = \sum c_i(\mu_i - \delta_i)^2 \colon c_i/c_j \leq K_{ij} < \infty , \quad i, j = 1, \cdots, p . \tag{1.2}$$

In this case $\mathcal{L}$ consists of those loss functions described in (1.2). The formal definition of $\mathcal{L}$-admissibility is as before, and the statistician will want to use an $\mathcal{L}$-admissible procedure for this class of loss functions. After this formulation, we begin to approach the problems described in the preceding.

We begin by giving still another proof for Stein's classical inadmissibility result (see Theorem 1 in Section 3). In the course of this proof we develop a basic formula which we will later use for our results about $\mathcal{L}$-inadmissibility.

With this preparation we can prove the results concerning $\mathcal{L}$-admissibility in situations like those we have described. This is done in Theorem 2 in Section 4. It turns out that in Case 1 the usual estimator cannot be improved. This is not surprising. However, it did surprise us that in any less extreme case—as in Case 2—the usual estimator can be uniformly improved. This is true no matter how large are the chosen bounds, K_{ij}.

In Section 4 we only investigate the theoretical question of $\mathcal{L}$-admissibility of δ_0. In the case where δ_0 is not $\mathcal{L}$-admissible we do not give a realistic formula for an estimator which significantly improves on δ_0 over the class $\mathcal{L}$. Nor do we investigate the question of how much improvement is possible. These questions—while of interest—are beyond the scope of this article.

Section 5 treats a slightly different problem, which involves the same family of unknown distributions, but the class $\mathcal{L}$ is somewhat larger than (1.2). It contains all rotations of the functions described in (1.2). In this situation δ_0 may be $\mathcal{L}$-admissible, depending on the size of the constants, K_{ij}.

Section 6 describes several possible generalizations of the main result proved in Section 4. These generalizations should suffice to convince the reader that neither the assumptions made concerning normality nor those concerning a squared error form for the loss function are at all required for the validity of the $\mathcal{L}$-inadmissibility assertions of Section 4.

2. WEAK ADMISSIBILITY

Here we formulate in a very general setting the criterion described in Section 1, and then specialize this criterion to the situation described there.

As usual, suppose $\{P_\theta \colon \theta \in \Theta\}$ is a family of probability measures on some sample space X, $\mathcal{B}$. It is desired to make some decision concerning $\theta \in \Theta$; and the measurable space of possible decisions is A, $\mathcal{A}$. The set of measurable randomized decision procedures (transition functions from X to A) is denoted by $\mathfrak{D}$.

The usual formulation involves specification of a loss function $L\colon \Theta \times A \to [0, \infty]$ with $L(\theta, \cdot)$ measurable for $\theta \in \Theta$. In place of this we assume that $\mathcal{L}$ is a specified collection of such loss functions. For each $L \in \mathcal{L}$, let $R_L\colon \Theta \times \mathfrak{D} \to [0, \infty]$ denote the risk function corresponding to L, i.e.,

$$R_L(\theta, \delta) = \int \int L(\theta, a)\delta(da\,|\,x)P_\theta(dx) .$$

We say that a procedure δ is *weakly admissible relative to* $\mathcal{L}$ if $R_L(\theta, \delta') \leq R_L(\theta, \delta)$ for all $\theta \in \Theta$, $L \in \mathcal{L}$ implies $R_L(\theta, \delta') \equiv R_L(\theta, \delta)$, $\theta \in \Theta$, $L \in \mathcal{L}$. Correspondingly, we say that δ' is (weakly) better than δ relative to $\mathcal{L}$ if $R_L(\theta, \delta') \leq R_L(\theta, \delta)$, $\theta \in \Theta$, $L \in \mathcal{L}$ with strict inequality for some pair L, θ. For short we will write "$\mathcal{L}$-admissible," etc., in place of "weakly admissible relative to $\mathcal{L}$."

Under suitable (and usually obtaining) conditions, for every inadmissible $\delta \in \mathfrak{D}$ there is an admissible $\delta' \in \mathfrak{D}$ which is better than δ. That is, the "weakly admissible" procedures form a "weakly complete" class. This result may be deduced from LeCam [15].

If it is decided *a priori* that the loss function is in the class $\mathcal{L}$, then as in the usual statistical formulation, weak admissibility relative to $\mathcal{L}$ is a minimal requirement for a usable procedure. The experimenter(s) should not use a weakly inadmissible procedure.

(A complementary notion would be "strong (or, uniform) admissibility." δ is "strongly admissible" relative to $\mathcal{L}$ if it is admissible for each $L \in \mathcal{L}$. This notion is less interesting since if $\mathcal{L}$ is very rich there may often be no nontrivial strongly admissible procedures. There are some possible types of admissibility which lie between weak and strong admissibility. We do not pursue any of these here.)

In the situations described in Section 1 and treated in Sections 3 and 4, the distributions P_θ are multivariate normal with expectation $\theta = \mu = (\mu_1, \cdots, \mu_p)(\Theta = R^p)$ and with independent coordinates having positive variances $\sigma_1^2, \cdots, \sigma_p^2$. In the first of the two situations discussed in Section 1, $\mathcal{L}$ is the class of loss functions

$$\mathcal{L} = \{L: L(\mu, \delta) = c_i(\delta_i - \mu_i)^2, 1 \leq i \leq p, c_i > 0\} \ . \tag{2.1}$$

We call this the situation of "total incompatibility" since it corresponds to the case where no meaningful compromise among the experimenters is possible. It should be easy to see that weak admissibility for $\mathcal{L}$ as above is equivalent to weak admissibility for the convex collection spanned by (2.1), namely,

$$\mathcal{L} = \{L: L(\mu, \delta) = \sum_{i=1}^{p} c_i(\delta_i - \mu_i)^2: \\ 0 \leq c_i < \infty, 1 \leq i \leq p\} \ . \tag{2.2}$$

In the case of "partial incompatibility," $\mathcal{L}$ is the class of loss functions satisfying (1.2).

3. STEIN'S INADMISSIBILITY RESULT

Before formulating and proving the results described in Section 1, let us examine certain properties of estimators for the classical normal distribution problem involving a single specified loss function

$$L(\mu, \delta) = \sum_{i=1}^{p} c_i(\delta_i - \mu_i)^2 \ .$$

In particular we study estimators similar to those proposed in [17] and [13].

We are interested here only in establishing admissibility or inadmissibility of the usual estimator in appropriate generalized senses. When the usual estimator is inadmissible, we are not interested here in describing acceptable (i.e., admissible) alternatives to it or even in finding estimators yielding a significantly large improvement. Some references which pursue such matters in the classical situation are mentioned in the introduction and results remain to be obtained in other situations.

It will suffice for our purposes to examine the behavior of estimators only when $\|x\|$ is large. And it is enough to know the behavior of their risk functions merely for $\|\mu\|$ large. First, we prove a result which shows why this is so in the classical situation, and then we proceed to the examination of Stein's [17] estimators. This result is based on the technique of randomizing the origin which is described in the proof following. (A like technique is used in classical proofs of the Hunt-Stein Theorem, but for a different purpose. (See [21].))

Proposition 1: Let $p \geq 2$. Let L be given, as in (1.1). Suppose there exists an estimator δ_1 whose risk

$$R(\mu, \delta_1) = E_\mu\{\sum_{i=1}^{p} c_i(\mu_i - \delta_{1i})^2\}$$

satisfies

$$\liminf_{\|\mu\| \to \infty} [(R(\mu, \delta_0) - R(\mu, \delta_1)) \cdot \|\mu\|^2)] > 0 \ . \tag{3.1}$$

and for some $B < \infty$

$$R(\mu, \delta_1) \leq B < \infty \ . \tag{3.2}$$

Then the usual estimator, $\delta_0(x) = x$, is inadmissible.

Proof: For convenience we let

$$R(\mu, \delta_0) = R_0 \text{ (a constant), and } \Delta(\mu) = R_0 - R(\mu, \delta_1)$$

We have

$$\liminf_{\mu \to \infty} \|\mu\|^2 \Delta(\mu) = a > 0 \ .$$

Hence, there is a constant, $r < \infty$, such that

$$\Delta(\mu) \geq a/2\|\mu\|^2 \quad \text{for} \quad \|\mu\| > r \ . \tag{3.3}$$

Let $q(t) = R_0 - B$ for $0 \leq t \leq r$ and $a/2t^2$ for $t > r$. Then $\Delta(\mu) \geq q(\|\mu\|)$.

Since the hypotheses of the proposition do not guarantee $\Delta(\mu) \geq 0$ everywhere, it need not be the case that δ_1 is itself a better estimator than δ_0. However, δ_1 may be used to produce a better estimator δ_2—by "randomizing the origin." Let $\delta_1{}^\theta(x) = \theta + \delta_1(x - \theta)$. $\delta_1{}^\theta$ is the estimator δ_1 with the origin placed at the point θ. Hence,

$$R(\mu, \delta_1{}^\theta) = R(\mu - \theta, \delta_1) \ ,$$

as can be checked directly from the definition of R.

Let θ be normally distributed with mean zero and variance-covariance matrix K^2I, $0 < K < \infty$. Let

$$\delta_2(x) = E(\delta_1{}^\theta(x))$$

where the expectation is taken (over θ) with respect to the above normal distribution. For convenience we suppress the dependence of δ_2 on the constant K. Apply Jensen's inequality:

$$\begin{aligned} R(\mu, \delta_2) &= E_\mu\{\sum_{i=1}^{p} c_i(\mu_i - \delta_{2i})^2\} \\ &= E_\mu\{\sum_{i=1}^{p} c_i(\mu_i - E(\delta_{1i}{}^\theta(x)))^2\} \\ &\leq E_\mu E\{\sum_{i=1}^{p} c_i(\mu_i - \delta_{1i}{}^\theta(x))^2\} \\ &= EE_\mu\{\sum c_i(\mu_i - \delta_{1i}{}^\theta(x))^2 | \theta\} \\ &= E(R(\mu - \theta, \delta_1)) \ . \end{aligned}$$

Thus, $R_0 - R(\mu, \delta_2) \geq E(\Delta(\mu - \theta)) \geq E(q(\|\mu - \theta\|))$.

Now,

$$(2\pi)^{p/2}K^pE(q(\|0-\theta\|))$$
$$=\int q(\|\theta\|)\exp\left(-\tfrac{1}{2}\|\theta\|^2/K^2\right)d\theta\to\infty \text{ as } K\to\infty$$

by the dominated convergence theorem. Hence, for K sufficiently large, say $K\geq K_0$, we have that $E(q(\|0-\theta\|))>0$.

Consider the distribution of the variable $T=\|\mu-\theta\|$, given the parameter μ. T^2/K^2 is a noncentral χ^2 variable with noncentrality parameter $\|\mu\|^2/2$. It is well known (and easily checked) that the distribution of T has a monotone likelihood ratio in the real parameter $\|\mu\|$. The function $q(t)$ crosses zero but once, and $q(t)>0$ for t sufficiently large. Hence, $E(q(\|\mu-\theta\|))=E_{\|\mu\|}(q(T))$ can cross zero at most once as a function of $\|\mu\|$; and this crossing, if it occurs, will be from negative values to positive ones. (See [14].) Since $E_{\|\mu\|=0}(q(T))>0$ for $K>K_0$, it must be that for any $K>K_0$ no crossing of zero occurs. We thus have for any $K>K_0$ $R_0-R(\mu,\delta_2)\geq E(q(\|\mu-\theta\|))>0$ for all μ, which is the desired result.

(We note here, and will later use without further proof, that this Proposition can be extended to cover the situations involving a class of loss functions and also the nonsquared error situations which arise later in this article. Conditions (3.1) and (3.2) are easily reinterpreted in such cases; they then remain as sufficient conditions for inadmissibility or $\mathcal{L}$-inadmissibility of the usual estimator. Note in particular that if a class of loss functions is involved, as in the later applications to $\mathcal{L}$-inadmissibility, then (3.1) and (3.2) must hold *uniformly* in all the normalized loss functions of this class in order for δ_2 to be better than δ_0 for all of these loss functions. More precisely, we will need that for $\|\mu\|>k_1$

$$(R(\mu,\delta_0)-R(\mu,\delta_1))\|\mu\|^2\geq k_2\sum_{i=1}^{p}c_i>0\ , \quad (3.1')$$

for all $c\in\mathcal{C}$ as defined in Section 4, and

$$R(\mu,\delta_1)\leq B\sum_{i=1}^{p}c_i<\infty \quad (3.2')$$

for all $c\in\mathcal{C}$.

To demonstrate inadmissibility of the usual estimator in the classical situation (fixed L), we thus need only to develop expressions for the asymptotic behavior (as $\|\mu\|\to\infty$) of the risk of alternative estimators to δ_0. We shall now do this. These expressions are the foundation for all of the subsequent results of this article.

Let $\delta(x)=x-h(x)$ where $h=(h_1,\cdots,h_p)$. Let $\Delta_i(\mu)=\sigma_i^2-E(\delta_i(x)-\mu_i)^2$, so that

$$R_0-R(\mu,\delta)=\sum_{i=1}^{p}c_i\Delta_i(\mu)\ .$$

Let $z_i=x_i-\mu_i$ and $z=(z_1,\cdots,z_p)$. z_i is normal with mean zero and variance σ_i^2. Then

$$\begin{aligned}\Delta_i(\mu)&=\sigma_i^2-E_\mu((\delta_i(x)-\mu_i)^2)\\&=\sigma_i^2-E(z_i-h_i(\mu+z))^2\\&=+2E(z_ih_i(\mu+z))-E(h_i^2(\mu+z))\ .\end{aligned}$$

The Fundamental Expressions: Suppose h_i is continuously differentiable. Then

$$h_i(\mu+z)=h_i(\mu)+\sum_{j=1}^{p}z_jh_{ij}(\mu)+e_i(\mu,z)$$

where $h_{ij}=(\partial/\partial j)h_i$ and e_i is the appropriate error term. Since $E(z_i)=0$, and $E(z_iz_j)=\sigma_i^2$ for $i=j$, and $=0$ for $i\neq j$ we have

$$\Delta_i(\mu)=+2h_{ii}(\mu)\sigma_i^2-E(h_i^2(\mu+z))+e_i'(\mu)\ ,$$

where $e_i'(\mu)$ represents the appropriate error term. If we further approximate $E(h_i^2(\mu+z))$ by $h_i^2(\mu)$ and rewrite the error term as $e_i''(\mu)$, we have

$$\Delta_i(\mu)=2\sigma_i^2h_{ii}(\mu)-h_i^2(\mu)+e_i''(\mu)\ . \quad (3.4)$$

For L as given by (1.1) this yields the *fundamental expression*:

$$\Delta(\mu)=D(\mu)+e''(\mu) \quad (3.5)$$

where

$$D_h(\mu)=D(\mu)=2\sum c_i\sigma_i^2h_{ii}(\mu)-\sum c_ih_i^2(\mu)\ . \quad (3.6)$$

According to Proposition 1, to show that δ_0 is inadmissible it suffices to find a smooth vector valued function h for which

$$\liminf_{\|\mu\|\to\infty}\|\mu\|^2D(\mu)>0 \quad (3.7)$$

and for which the error term e'' is small enough to be insignificant.

The Error Term: In the preceding we have refrained from a detailed computation of the error terms e, e', and e''. The necessary computations are similar to those which appear in [6, Sec. 3]. To view these terms qualitatively, note first that the standard estimates for e_i involve the second derivatives of h_i near μ (assuming h is sufficiently smooth). From these it is to be expected that

$$e_i'(\mu)=O\Big(\sum_{i,j,k}|(\partial/\partial k)h_{ij}(\mu)|\Big)$$

The solutions of (3.7) which we shall encounter are mainly of the form $h_i=\gamma_i/\|\mu\|$ where γ_i is a smooth bounded function. For convenience write $h_i\leftrightarrow(1/\|\mu\|)$ for such a function. In addition the γ_i will satisfy that their nth derivatives are $O(\|\mu\|^{-n})$, $n=1,2$. Then $h_{ii}\leftrightarrow(1/\|\mu\|^2)$ and $(\partial/\partial k)h_{ij}\leftrightarrow(1/\|\mu\|^3)$. Hence, $\|\mu\|^2\cdot e_i'(\mu)\to0$ as $\|\mu\|\to\infty$ so that $e_i'(\mu)$ is of a smaller order than the terms in (3.7).

Also, $e_i''(\mu)=e_i'(\mu)+E(h_i^2(\mu+z))-h_i^2(\mu)$. For h as before, $[E(h_i^2(\mu+z))-h_i^2(\mu)]\leftrightarrow(1/\|\mu\|^4)$ so that e_i'' is also insignificant relative to the terms in (3.7). For h as above (3.7) itself does indeed imply the key condition (3.1) of Proposition 1.

These considerations lead to the following inadmissibility theorem, which generalizes the result in [17] and is included in Theorem 3.1.1 of [6]. (See also [4].)

Theorem 1: Suppose the loss function is given by (1.1) and $p \geq 3$. Then δ_0 is inadmissible.

Proof: Let $\delta(x) = x - h(x)$ as before where the coordinates h_i of h are smooth bounded functions satisfying

$$h_i(x) = \epsilon x_i / c_i \sigma_i^2 \|x\|^2 \quad \text{for} \quad \|x\| > 1 \ .$$

(In the case where $c_i \sigma_i^2 = k$, $1 \leq i \leq p$, then, for $\|x\| > 1$, $\delta(x) = (1 - \epsilon/k\|x\|^2)x$.) Computing $h_{ii}(\mu)$ yields

$$D(\mu) = \frac{\epsilon}{\|\mu\|^2}\left(\sum_{i=1}^{p} 2\left(1 - \frac{2\mu_i^2}{\|\mu\|^2}\right) - \sum_{i=1}^{p} \frac{\epsilon \mu_i^2}{c_i \sigma_i^4 \|\mu\|^2}\right), \quad \|\mu\| > 1 \ .$$

Now,

$$\sum_{i=1}^{p} \mu_i^2 / \|\mu\|^2 = 1 \ .$$

Thus if we choose $\epsilon = \eta \min \{c_i \sigma_i^4 \colon 1 \leq i \leq p\}$ with $0 < \eta < 2(p - 2)$, the preceding becomes

$$D(\mu) \geq (\epsilon/\|\mu\|^2)(2(p - 2) - \eta) > 0 \ , \quad \|\mu\| > 1 \ .$$

Condition (3.7) is therefore satisfied. Since h is bounded, Condition (3.2) of Proposition 1 is also satisfied.

The error term e'' is of the type described so that (3.7) implies Condition (3.1) of Proposition 1. A randomization of the origin as described in that Proposition therefore yields an estimator better than δ_0.

Remark: Since the constant K which is used in the randomization described in Proposition 1 is not given by an explicit formula, the preceding process does not explicitly describe an estimator δ_2 which is better than δ_0. However, it is evident that the better estimator, δ_2, constructed by the recipe in Proposition 1 from the estimator $\delta(x) = x - h(x)$ of the preceding proof will satisfy

$$\begin{aligned} \delta_2(x) = x - h'(x), \quad \text{where} \quad h_i'(x) &= h_i(x) + o(1/\|x\|) \\ &= \epsilon x_i / c_i \sigma_i^2 \|x\|^2 + o(1/\|x\|), \quad \text{as} \quad \|x\| \to \infty \ . \end{aligned}$$

Hence we know the asymptotic form of δ_2 as $\|x\| \to \infty$. When $c_i \sigma_i^2 \equiv k$ this form agrees with those given in [17] and [13].

4. THE MAIN RESULT

We now turn to the general problem described in Section 1. To be precise, let $\mathcal{C}$ be a subset of the nonnegative quadrant of R^p and let

$$\mathcal{L}(\mathcal{C}) = \{L \colon L(\mu, \delta) = \sum_{i=1}^{p} c_i(\delta_i - \mu_i)^2 \ ; \quad c = (c_1, \cdots, c_p) \in \mathcal{C}\} \ .$$

Let $\mathcal{C}^*$ denote the closed convex cone generated by $\mathcal{C}$, except for the origin. *We emphasize that* $0 \notin \mathcal{C}^*$.

As before, the observations are independent normal with unknown means and with known variances σ_i^2. (See Section 6 for generalizations.)

The following simple proposition will be very useful in our discussion. Its conclusion is true in situations much more general than the one we treat here.

Proposition 2: δ_0 is weakly admissible relative to $\mathcal{L}(\mathcal{C})$ if and only if it is weakly admissible relative to $\mathcal{L}(\mathcal{C}^*)$.

Proof: Suppose $R_L(\mu, \delta) \leq R_L(\mu, \delta_0)$ for every $L \in \mathcal{L}(\mathcal{C})$, $\mu \in R^p$. Let $\mathcal{C}'$ denote the convex cone generated by $\mathcal{C}$. Let $L' \in \mathcal{L}(\mathcal{C}')$. Then $L' = \sum_{i=1}^{p} a_i L_i$ for some $a_i \geq 0$, $L_i \in \mathcal{L}(\mathcal{C})$. Hence, $R_{L'}(\mu, \delta) \leq R_{L'}(\mu, \delta_0) < \infty$.

Suppose L'' corresponds to a point in the closure of $\mathcal{C}'$. We write $L'' = \lim L_i'$, where $L_i' \in \mathcal{L}(\mathcal{C}')$. It is then straightforward to conclude that

$$R_{L''}(\mu, \delta) = \lim_{i \to \infty} R_{L_i'}(\mu, \delta) \leq R_L(\mu, \delta_0) \ .$$

Inadmissibility of δ_0 relative to $\mathcal{L}(\mathcal{C})$ thus implies inadmissibility relative to $\mathcal{L}(\mathcal{C}^*)$.

Conversely, suppose δ_0 is inadmissible relative to $\mathcal{L}(\mathcal{C}^*)$. Then there is a procedure δ such that $R_L(\mu, \delta) \leq R_L(\mu, \delta_0)$ for all $L \in \mathcal{L}(\mathcal{C}^*)$ and strict inequality holds for some μ and some $L \in \mathcal{L}(\mathcal{C}^*)$, call them μ_0, L_0. Clearly $R_L(\mu, \delta) \leq R_L(\mu, \delta_0) \leq R_L(\mu_0, \delta_0) < \infty$ for all $L \in \mathcal{L}(\mathcal{C})$. Furthermore, $L_0 = \lim L_i$ where $L_i \in \mathcal{L}(\mathcal{C}') \subset \mathcal{L}(\mathcal{C}^*)$. We can conclude that

$$\lim_{i \to \infty} R_{L_i}(\mu_0, \delta) = R_{L_0}(\mu_0, \delta) < R_{L_0}(\mu_0, \delta_0) = \lim_{i \to \infty} R_{L_i}(\mu_0, \delta_0) \ .$$

Thus δ_0 is also inadmissible relative to $\mathcal{C}$. The proof is complete.

We then have the following main result.

Theorem 2: Let $\mathcal{L} = \mathcal{L}(\mathcal{C})$ as above. Then δ_0 is admissible relative to $\mathcal{L}$ iff for every subset of coordinate indices $1 \leq i_1 < \cdots < i_\pi \leq p$ either there are two indices—i_k, i_ℓ, say—and a point $c \in \mathcal{C}^*$ with $c_{i_j} = 0$ for $j \neq k$, $\ell (1 \leq j \leq \pi)$ and $\max (c_{i_k}, c_{i_\ell}) > 0$, or $c_{i_j} \equiv 0$, $j = 1, \cdots, \pi$, for every $c \in \mathcal{C}^*$.

Interpretory Remarks: James and Stein [13] showed that in dimension $p = 2$, δ_0 is admissible relative to any given $L \in \mathcal{L}$, and therefore admissible relative to $\mathcal{L}$ itself. The theorem says this is essentially the only situation in which δ_0 is admissible relative to $\mathcal{L}$—for δ_0 to be admissible relative to $\mathcal{L}$ it must be that in every nontrivial π dimensional subproblem, $\mathcal{L}(\mathcal{C}^*)$ contains a loss function with at most two nonzero coefficients within that subproblem.

In the interpretation of Section 1, Theorem 2 shows that for $p \geq 3$ δ_0 is admissible in the case of total incompatibility (2.1) of the loss functions,[2] but not much more generally. In particular, in the case of partial incompatibility as specified in (1.2), δ_0 is not admissible relative to $\mathcal{L}$.

Proof: In preparation for the proof of sufficiency of the condition of the theorem, we review some known admissibility results. Suppose for some procedure δ and for some $1 \leq k < \ell \leq p$ and some $c_k > 0$, $c_\ell \geq 0$,

$$E_\mu(c_k(\mu_k - \delta_k)^2 + c_\ell(\mu_\ell - \delta_\ell)^2) \leq c_k \sigma_k^2 + c_\ell \sigma_\ell^2 \quad (4.1)$$

[2] A referee has pointed out that this result is in [12, p. 704].

for all $\mu \in R^p$. Then, it is known that $\delta_k(x_1, \cdots, x_p) = x_k$ a.e. (w.r.t. Lebesgue measure on R^p). The reasoning for this result is as follows.

For notational convenience suppose $k = 1$, $\ell = 2$. Fix $\mu_3, \cdots, \mu_p$. Let $\delta^*(x_1, x_2) = (\delta_1^*(x_1, x_2), \delta_2^*(x_1, x_2))$ be defined by $\delta_i^*(x_1, x_2) = E_\mu(\delta_i(x) | x_1, x_2)$, $i = 1, 2$. Since the coordinates $\{X_i\}$ are independent, δ^* depends only on $x_1, x_2, \mu_3, \cdots, \mu_p$, as indicated by the notation. Then

$$E_{\mu_1,\mu_2}(c_1(\mu_1 - \delta_1^*)^2 + c_2(\mu_2 - \delta_2^*)^2) \leq c_1\sigma_1^2 + c_2\sigma_2^2 \quad \text{for all } \mu_1, \mu_2 \in R^2 . \quad (4.2)$$

In short, δ^* is at least as good as δ_0 for the two-dimensional problem defined by X_1, X_2. By [13, Theorem 2], δ_0 is admissible in this two-dimensional problem; thus the two sides of (4.2) must actually be equal. (See [13, p. 375] for a summary of this argument.) By Jensen's inequality the estimator $\frac{1}{2}(\delta^* + \delta_0)$ will be strictly better than δ^* and δ_0 unless $\delta_1^* = (\delta_0)_1 = x_1$ a.e. (See [11, (5.16)–(5.18)] for a similar use of Jensen's inequality.) This would contradict the admissibility result previously quoted. Hence, $\delta_1^* = x_1$ a.e. This result must be true no matter what fixed value of $\mu_3, \cdots, \mu_p$ was used in the definition of δ^*. Hence,

$$E_{\mu_3,\ldots,\mu_p}(\delta_1(a, b, x_3, \cdots, x_p) - a) = 0$$

for all $\mu_3, \cdots, \mu_p$ and almost all $a, b \in R^2$. By the completeness of the family of normal distributions this implies $\delta_1(x) = x_1$ a.e. on R^p, as claimed in the first paragraph of the proof.

We can now prove sufficiency of the condition of the theorem. Suppose δ is as good as δ_0 with respect to $\mathfrak{L}$. By the condition there is a $c \in \mathcal{C}^*$ and two indices i_1 and i_2 such that $c_j = 0$ $j \neq i_1, i_2$ and $c_{i_1} > 0$. (The possibility $c_i \equiv 0$ $i = 1, \cdots, p$ for all $c \in \mathcal{C}^*$ is ruled out by the definition of $\mathcal{C}^*$.) For notational convenience suppose $i_1 = 1$. Then, by the preceding remarks, $\delta_1(x) = x_1$ a.e.

Now consider the collection of remaining coordinates $2, 3, \cdots, p$. If $c_i \equiv 0$ for $i = 2, 3, \cdots, p$ for all $c \in \mathcal{C}^*$ then for all $L \in \mathfrak{L}$, $R_L(\mu, \delta) = E_\mu(c_1(\delta_1(x) - \mu)^2) = c_1\sigma_1^2 = R_L(\mu, \delta_0)$. Hence, δ_0 is admissible with respect to $\mathfrak{L}$, and the proof is completed here. If $c_i \not\equiv 0$, $i = 2, \cdots, p$ for all $c \in \mathcal{C}^*$ then by the condition of the theorem there exist two co-ordinates $2 \leq i_k, i_\ell \leq p$ and $c \in \mathcal{C}^*$ such that $c_{i_k} > 0$ and $c_j = 0$ for $j \neq i_k, i_\ell$. For notational convenience assume here that $i_k = 2$. Then, as before, $\delta_2(x) = x_2$. In this case, proceed by induction as follows.

Consider the collection of remaining coordinates $3, \cdots, p$. Proceed as in the preceding paragraph. Continue the process by induction. Either the process will terminate because all the p coordinates have been exhausted, or because at some stage the set of remaining coordinates. $i_1, \cdots, i_\pi$, satisfy $c_{i_j} \equiv 0$, $j = 1, \cdots, \pi$ for all $c \in \mathcal{C}^*$. In the former case we will have $\delta_i(x) = x_i$ a.e., so that $R_L(\mu, \delta) = R_L(\mu, \delta_0)$ for all $L \in \mathfrak{L}$. In the latter case $R_L(\mu, \delta) = R_L(\mu, \delta_0)$ by the reasoning of the preceding paragraph. This proves the sufficiency of the condition of the theorem.

To show necessity of the condition in the theorem we will, under the failure of this condition, formulate an estimator which dominates the usual estimator for all $L \in \mathfrak{L}(\mathcal{C}^*)$.

First, some preliminary remarks concerning the condition in the theorem. When the condition fails, there are $\pi \geq 3$ indices $i_1 < \cdots < i_\pi$ such that for every $c \in \mathcal{C}^*$ three of the coordinates, say $c_{i_{k_1}}$, $c_{i_{k_2}}$, and $c_{i_{k_3}}$, are nonzero. For notational simplicity, and without loss of generality, we may restrict attention to these coordinates and hence assume that $p = \pi$. For any $c \in \mathcal{C}^*$, let $c^{(k)}$ denote the kth largest among the numbers $(c_1, c_2, \cdots, c_p)$. By assumption if $c \in \mathcal{C}^*$ then $c^{(1)} \geq c^{(2)} \geq c^{(3)} > 0$. Let

$$B' = \sup \{c^{(1)}/c^{(3)}: c \in \mathcal{C}^*\} .$$

Since $\mathcal{C}^*$ is a closed cone and $c^{(3)} > 0$ for $c \in \mathcal{C}^*$, it follows that $B' < \infty$.

The remainder of the proof consists in first finding functions $\{h_i\}$ which satisfy (3.7) uniformly in $\|\mu\|$ and $c \in \mathcal{C}^*$. (The exact definition of what is meant here by "uniformly" is described by (3.1′) which follows the proof of Proposition 1. It is also necessary that (3.2′) be satisfied.) It is then necessary in the proof to check that the error terms indicated in (3.5) are uniformly negligible in the appropriate sense for these functions $\{h_i\}$. The functions $\{h_i\}$ which we will find depend on $\mathcal{C}$ only through the constant B'; and the condition $B' < \infty$ is what allows for the uniformity in (3.7) and in (3.5) referred to before.

The remainder of the proof is deferred to the appendix, where we carry out the steps referred to in the preceding. The functions $\{h_i\}$ which are defined there were found by a process of trial and error. We have not been able to develop any intuitive explanation of their nature or properties.

5. SPHERICALLY SYMMETRIC ESTIMATORS

We began this investigation in the hope that the usual estimator would be found to be $\mathfrak{L}$-admissible in many nonextreme situations. Theorem 2 shows that this is not the case—at least not for the formulation examined in Section 4. We are then led to ask whether some other description of the class $\mathfrak{L}$ leads to $\mathfrak{L}$-admissibility of δ_0 in nonextreme situations. This is indeed so; however, the new class $\mathfrak{L}$ involved here does not have the same justification as a compromise among competing experimenters as does the class of loss functions discussed in Section 1 and examined in Section 4. The result is described in Corollary 1 to Theorem 3.

Let the observations be multivariate normal, as previously, and also let the loss function be $\sum c_i(\delta_i - \mu_i)^2$, $c \in \mathcal{C}$, as before. Note that if δ is spherically symmetric, i.e.,

$$\delta(x) = (1 - g(\|x\|))x ,$$

and $h(x) = g(\|x\|)x$ has the particularly simple form

$h(x) = \epsilon x/\|x\|^2$, then

$$\|\mu\|^2 D(\mu) = 2\epsilon \sum_{i=1}^{p} (c_i\sigma_i^2(1 - 2\mu_i^2/\|\mu\|^2) - \epsilon c_i\mu_i^2/2\|\mu\|^2) \ . \quad (5.1)$$

We have the following result.

Theorem 3: There exists a spherically symmetric estimator $\delta \neq \delta_0$ such that $R_L(\mu, \delta) \leq R_L(\mu, \delta_0)$ for all μ and all $L \in \mathcal{L}(\mathcal{C})$ if for all $c \in \mathcal{C}^*$, $1 \leq k \leq p$, the inequality

$$\sum_{j=1}^{p} c_j\sigma_j^2 > 2c_k\sigma_k^2 \quad (5.2)$$

is satisfied. Conversely, if $\sigma_1^2 = \cdots = \sigma_p^2$, then (5.2) is also a necessary condition for the existence of the $\delta \neq \delta_0$ described above.

[Condition (5.2) is most easily interpreted in the case where $\sigma_1^2 = \cdots = \sigma_p^2$. It then becomes that the largest among the coefficients $c_1, \cdots, c_p$ must be strictly smaller than the sum of the remaining coefficients, for all $c \in \mathcal{C}^*$.] M.E. Bock [5] has independently obtained this same result by an entirely different method of proof.

Proof: Let

$$\alpha = \inf_{c\in\mathcal{C}} [\sum c_i\sigma_i^2 - 2 \max \{c_i\sigma_i^2 : 1 \leq i \leq p\}]/\sum c_i \ .$$

Using (5.2) and the fact that $\mathcal{C}^*$ is a closed cone yields that $\alpha > 0$. Substituting the condition (5.2) into (5.1) yields

$$\|\mu\|^2 D(\mu) \geq 2\epsilon(\sum_{i=1}^{p} c_i\sigma_i^2 - 2 \max \{c_i\sigma_i^2 : 1 \leq i \leq p\} - (\epsilon/2) \sum_{i=1}^{p} c_i\mu_i^2/\|\mu\|^2)$$
$$\geq 2\epsilon(\alpha - \epsilon/2) \sum_{i=1}^{p} c_i \ .$$

Hence, choosing $\epsilon < 2\alpha$, randomizing the origin, and verifying the insignificance of the error term yields the existence of the desired estimator δ.

The converse result is proved by an application of the multivariate Cramér-Rao inquality. The following argument is patterned after that in [17, p. 202–3].

Applying the multivariate Cramér-Rao inequality yields the following necessary condition for an estimator to be better than the usual estimator:

$$2\sum c_i\sigma_i^2 b_{ii}(\mu) - \sum_{i,j} c_i\sigma_i^2 b_{ij}^2(\mu) - \sum c_i b_i^2(\mu) \geq 0 \ , \quad (5.3)$$

where b is the negative of the bias:

$$b(\mu) = \mu - E_\mu(\delta(x)) \ , \quad (5.4)$$

and

$$b_{ij}(\mu) = (\partial/\partial\mu_j)b_i(\mu) \ .$$

Dropping the b_{ij}^2 terms reduces (5.3) to the necessary condition:

$$2 \sum c_i\sigma_i^2 b_{ii}(\mu) - \sum c_i b_i^2(\mu) \geq 0 \ . \quad (5.5)$$

Hence, δ_0 will be shown to be admissible if it can be shown that there is no nontrivial function $b(\mu)$ which satisfies (5.5).

[Note the similarity of (5.5) to the basic equation $\Delta(\mu) \geq 0$, as in (3.5). The two equations are formally identical, except that $b = E_\mu(h)$. This similarity is much more than a coincidence. See the remarks in Section 7.]

In this theorem, b is spherically symmetric. Let $t = \|\mu\|^2$ and $b(\mu) = (\psi(\|\mu\|^2)/\|\mu\|^2)\mu = (\psi(t)/t)\mu$. For notational simplicity, suppose $c \in \mathcal{C}^*$ satisfies

$$(\sum c_i\sigma_i^2 - c_1\sigma_1^2)/\sum c_i = \alpha \leq 0 \ .$$

Suppose $\psi(t_0) \geq 0$ for some $t_0 > 0$. Let

$$\mu = (\|\mu\|, 0, \cdots, 0) = (\sqrt{t}, 0, \cdots, 0) \ .$$

In terms of t and ψ, at this $c \in \mathcal{C}^*$, and at such values of μ, (5.5) becomes

$$2\alpha(\sum c_i)\psi(t)/t + 4c_1\sigma_1^2\psi'(t) - c_1\psi^2(t)/t \geq 0 \ . \quad (5.6)$$

Then (5.6) implies

$$4\psi'(t) - \psi^2(t)/\sigma^2 t \geq 0 \quad \text{for} \quad t \geq t_0 \ . \quad (5.7)$$

The argument in [17, p. 203] shows that (5.6) implies $\psi(t) \leq 0$ for $t \geq t_0$. Hence $\psi(t) = 0$ for $t \geq t_0$.

Suppose, on the other hand, $\psi(t_0) \leq 0$ for some $t_0 > 0$. Evaluate (5.5) at $\mu = (\|\mu\|/\sqrt{p}, \|\mu\|/\sqrt{p}, \cdots, \|\mu\|/\sqrt{p})$. It becomes

$$(2 - 4/p)(\sum c_i\sigma_i^2)\psi(t)/t + (4/p)(\sum c_i\sigma_i^2)\psi'(t) - (\sum c_i/p)\psi^2(t)/t \geq 0 \ . \quad (5.8)$$

In (5.8), $\psi(t) < 0$ for $t = t_0$, and $(2 - 4/p) \geq 0$ imply that $\psi'(t) \geq 0$ for $t < t_0$ and, hence, $\psi(t) < 0$ for $t \leq t_0$. Thus

$$\psi'(t) - (\psi^2(t)/t)(\sum c_i/4 \sum c_i\sigma_i^2) \geq 0 \quad \text{for} \quad t \leq t_0 \quad (5.9)$$

Again, the argument in [17, p. 203] shows that this implies $\psi(t) \geq 0$ for $t \leq t_0$. Hence, $\psi(t) = 0$ also for $t \leq t_0$.

In summary, we have that (5.5) and b differentiable and spherically symmetric imply that $b \equiv 0$. This implies that $\delta(x) = \delta_0(x)$ a.e., which is the desired result.

The setting of the corollary is as follows: The observations are normal as before but with $\sigma_1^2 = \cdots = \sigma_p^2$, and now the class $\mathcal{L}$ consists of all quadratic forms in $(\mu - \delta)$ with specified eigenvalues. Thus

$$\mathcal{L} = \mathcal{L}^*(\mathcal{C}) = \{(\mu - \delta)'M(\mu - \delta) : M = Q'CQ \text{ where } Q \text{ is orthogonal and } C = \begin{bmatrix} c_1 & & 0 \\ & \ddots & \\ 0 & & c_p \end{bmatrix} \text{ with } (c_1, \cdots, c_p) \in \mathcal{C}\} \ .$$

For convenience in stating the following corollary, let $\sigma^2 = \sigma_1^2 = \cdots = \sigma_p^2$ and for $c \in \mathcal{C}^*$ let $c^{(1)}, \cdots, c^{(p)}$ denote the coordinates of c arranged in descending order.

Corollary 1: δ_0 is $\mathcal{L}^* = \mathcal{L}^*(\mathcal{C})$ admissible iff for some $c \in \mathcal{C}^*$

$$c^{(1)} \geq \sum_{i=2}^{p} c^{(i)} \ . \quad (5.10)$$

[This can be expressed by saying that δ_0 is $\mathcal{L}$-admissible iff some loss function in $\mathcal{L}^*$ weights its principle coordinate direction at least as heavily as all the other directions combined.]

Proof: If δ_0 is $\mathcal{L}^*$ inadmissible there must (by invariance, and compactness of the rotation group) be a spherically symmetric estimator δ which is better for all $L \in \mathcal{L}$. (The invariance argument includes the fact that $\mathcal{L}^*$ is invariant under the rotation group.) Theorem 3 tells whether such an estimator δ exists, and (5.10) is merely a re-expression of (5.2) appropriate to the current situation. The converse also follows immediately from Theorem 3.

An alternate interpretation: There is an alternate interpretation of Theorem 3 and its corollary which may also be of interest, and which is inspired by some of the considerations in [9].

As in Corollary 1, suppose $\sigma^2 = \sigma_1^2 = \cdots = \sigma_p^2$. Suppose the estimate δ of the mean vector μ is to be used as follows:

Some linear functional $B\mu$ is to be estimated where B is a $(k \times p)$ matrix and, of course, μ is $(p \times 1)$. This functional is to be estimated by $B\delta$ (δ is $(p \times 1)$); and the loss is measured by squared error in k-dimensions. That is,

$$L^{(B)}(\mu, \delta) = \|B\mu - B\delta\|^2 .$$

Suppose the matrix B is not given precisely at the time the estimate δ is to be made; or, what amounts mathematically to the same thing, that the same estimate, δ, is to be used in the preceding manner for estimating $B\mu$ for a variety of different matrices B. We will suppose B is given *a priori* to be a specified class, denoted by $\mathcal{B}$.

The appropriate generalized admissibility concept is thus that δ_0 is $\mathcal{B}$-admissible if the inequality

$$E_\mu(L^{(B)}(\mu, \delta)) = R^{(B)}(\mu, \delta) \leq R^{(B)}(\mu, \delta_0)$$

for all μ and all $B \in \mathcal{B}$ implies that

$$R^{(B)}(\mu, \delta) = R^{(B)}(\mu, \delta_0)$$

for all μ and all $B \in \mathcal{B}$.

In the following corollary we are interested in such a situation when $\mathcal{B}$ has the form:

$$\mathcal{B} = \bigcup_{k=1}^{\infty} \{B\colon B \text{ is } (k \times p) \text{ with the eigenvalues of } B'B \text{ being } c_1, \cdots, c_k \in \mathcal{C}_k\} .$$

Here $\mathcal{C}_k$ is a specified subset of R^k, and $\mathcal{C}_k^*$ will be defined as before. Note that some (but not all) of the sets $\mathcal{C}_k$ in the preceding formulation may be empty. (Note also that if $k > p$ then for every $c \in \mathcal{C}_k$ some of the coordinates $c_1, \cdots, c_k$ must be zero. This is not true of $\mathcal{C}_k^*$. These peculiarities do not interfere with the result which follows.)

[In [9] this situation was considered with $k = 1$ only (i.e., $\mathcal{C}_k = \phi$ for $k \geq 2$), but average admissibility of the estimators with respect to Haar measure on $\{Q\}$ was considered, rather than $\mathcal{B}$ admissibility.]

Corollary 2: The usual estimator, $\delta_0(x) = x$, is $\mathcal{B}$-admissible in the preceding interpretation iff $\mathcal{C}_1 \neq \phi$ or $\mathcal{C}_2 \neq \phi$, or for some $k \geq 3$ and some $c = (c_1, \cdots, c_k) \in \mathcal{C}_k^*$

$$c^{(1)} \geq \sum_{i=2}^{k} c^{(i)} . \tag{5.11}$$

Proof: Note that

$$R^{(B)}(\mu, \delta) = E_\mu[(\delta - \mu)'B'B(\delta - \mu)] .$$

Hence, δ_0 is $\mathcal{B}$-admissible iff it is $\mathcal{L}$-admissible for $\mathcal{L} = \{(\delta - \mu)'B'B(\delta - \mu)\colon B \in \mathcal{B}\}$. Note also that if $B \in \mathcal{B}$, then $BQ \in \mathcal{B}$ for Q $p \times p$ orthonormal. Hence, $\mathcal{L}$ is of the form $\mathcal{L}^*(\mathcal{C})$ treated in Corollary 1, where $\mathcal{C}$ depends on the sets $\mathcal{C}_1, \mathcal{C}_2, \cdots$. It is then a matter of algebra to check that Condition (5.10) is satisfied iff $\mathcal{C}_1 \neq \phi$ or $\mathcal{C}_2 \neq \phi$, or Condition (5.11) is satisfied.

6. GENERALIZATIONS

Theorems 2 and 3 can be generalized in many directions. The possibility for such generalizations can be easily seen if one notes that the fundamental expressions (3.5)–(3.7) are really the beginning of a Taylor series expansion for $R_0 - R_L(\mu, \delta)$; and that given these expressions the proof of these theorems involves only the first few moments of the distributions involved. The following comments deal specifically with Theorem 2, and amplify the preceding observations. Similar considerations would apply to the sufficiency part of Theorem 3 and its corollary. (The reader may wish to consult [8] for some other facets of the method indicated by these remarks.)

Theorem 2 dealt with the situation where the observations $x_1, \cdots, x_p$ are independent normally distributed variables with location parameters (means) $\mu_1, \cdots, \mu_p$; and where the loss function is of the form $\sum c_i L_i(\mu_i - \delta_i)$ with each $L_i(t) = t^2$. These assumptions of normality and of quadratic loss are in no way essential for the validity of Theorem 2. The following remarks should make this clear without the necessity for further formal arguments.

Remark 1: The assumption of normality is not necessary in Theorem 2. The proof of this theorem involves only certain moments of the variables X_i. It will suffice if the observations $x_1, \cdots, x_p$ are independent observations, and have probability distributions $F_i(x_i - \mu_i)$, respectively, with $E_0 X_i = 0$ and where the second moment of F_i is known (i.e., $\operatorname{Var}(X_i) = \sigma_i^2$) and the fifth moment of F_i is bounded (so that the error terms, $e''(\mu)$, are sufficiently small).

Remark 2: The assumption in Remark 1 of a single set of observations $x_1, \cdots, x_p$—rather than repeated independent observations $\{x_{ij}\}$, $j = 1, \cdots, J(i)$, $i = 1, \cdots, p$ with each x_{ij} coming from the distribution F_i—is unnecessary. The situation of repeated observations can be handled in the usual manner by conditioning on the value of a maximal invariant, as was done, e.g., in Brown ([6]; see especially Ch. 3).

Remark 3: The assumption in Theorem 2 that the loss functions $L_i = L_i(\mu_i - \delta_i)$ be quadratic is also not necessary. An extension of Theorem 2 is valid also for nonquadratic loss functions.

Relaxation of this assumption does require slightly more care than the relaxation of the normality assumption. In essence the relaxation depends on two observations.

(i) Suppose X_i has the distribution $F_i(x_i - \mu_i)$ as in Remark 1 and $\delta_{oi}(x_i) = x_i$ is the best invariant estimator of μ_i for the given location invariant loss function $L_i = L_i(\mu_i - \delta_i)$. (This assumption involves no real loss of generality.) If L_i possesses a continuous first derivative, then

$$E_0(L_i'(-x_i)) = 0 \ , \tag{6.1}$$

provided that

$$\int \sup_{|\epsilon|<\epsilon_1} L_i'(-t - \epsilon)F_i(dt) < \infty \ , \quad \text{for some } \epsilon_1 > 0 \ .$$

If the second and third derivatives of L_i also exist, then letting $\delta_i(x) = x_i - h_i(x)$ and expanding L_i and h_i in a Taylor series yields

$$\begin{aligned}\Delta_i(\mu) &= E(L_i(\mu_i - x_i)) - E(L_i(\mu_i - \delta_i(x))) \\ &= 2h_{ii}(\mu)E_0(-x_iL_i'(-x_i)) - h_i^2(\mu)E_0(L_i''(-x_i)) + e_i(\mu) \ .\end{aligned}$$

If $L_i'(t) < 0$ for $t < a$ and $L_i'(t) > 0$ for $t > a$, it follows from (6.1) that $E(-x_iL_i'(-x_i)) > 0$. Under reasonable assumptions concerning the existence of moments and the growth of L and L', an h of the form used in Theorem 2 can then be chosen to yield $e_i(\mu) = o(1/\|\mu\|^2)$ and

$$\liminf_{\|\mu\|\to\infty} \|\mu\|^2 \sum c_i\Delta_i(\mu) > 0$$

uniformly over $\mathcal{C}^*$ in the sense of (3.1′) of Section 2 where $\mathcal{C}^*$ is as in the hypothesis of Theorem 2.

(ii) If L_i is convex an analog of Theorem 2 follows immediately from the preceding remarks by using an appropriate modification of Proposition 1. If L_i is not convex the use of Jensen's inequality in the proof of Proposition 1 is of course not valid, but a useful version of Proposition 1 can still be proved in which the new estimator constructed is a randomized estimator.

Remark 4: In the preceding considerations we have assumed that the distributions of $X_1, \cdots, X_p$ are known, except for the unknown location parameter. It is fairly clear that a more general result is possible, although a precise formulation is beyond the intended scope of this article. For example, suppose $X_1, \cdots, X_p$ are $N(\mu, \sigma_i^2)$ as in Theorem 2, except that the variances, σ_i^2, are unknown; but that, as usual, one has available estimates s_i^2 of σ_i^2 where s_i^2 is independently distributed as $(\sigma_i^2\chi_{m_i}^2)/m_i$ $(i = 1, \cdots, p)$. One may then substitute these estimates for the true values in the definition of δ in Theorem 2. For $m_1, \cdots, m_p$ sufficiently large, this estimator will be better than δ_0 for all $L \in \mathfrak{L}(\mathcal{C}^*)$ when the hypothesis of Theorem 2 is satisfied. (This result can be rigorously established since the expressions for $R_L(\mu, \delta)$ in Theorem 2 can be shown to be the limit as $\inf\{m_i: 1 \le i \le p\} \to \infty$ of the risk function in the current situation. However, it is not clear how large the m_i need to be for this result to be valid.) Computations (for the case of a single specified loss function) which are mathematically similar appear in [22].

7. REMARKS

By differentiating inside an integral sign as in [7, Formula (1.2.2)], Stein [18] has recently shown in the multivariate normal situation of Section 2 that

$$\Delta(\mu) = E_\mu(D_h(\mu)) \tag{7.1}$$

where D is defined in (3.6). (See also [16].) Hence, to prove inadmissibility in Theorem 1, it suffices to find h so that $D_h(\mu) > 0$ for all μ. A similar comment holds for Theorem 2.

Furthermore, since D_h is concave in h, a version of Proposition 1 is valid for the operator D:

Let $p \ge 2$. If for some bounded h

$$\liminf_{\|\mu\|\to\infty} \|\mu\|^2 D_h(\mu) > 0 \tag{7.2}$$

then there is bounded h^* such that $D_{h^*}(\mu) > 0$ for all μ. (In fact h^* can be taken to be $E^\theta(h(\mu - \theta))$ where θ is as in the proof of Proposition 1.)

It should be clear that we could have rewritten Sections 3–5 so that the proofs would depend on (7.2) rather than on (3.5). This would have eliminated the necessity of discussing the error terms in Section 3 and in the proof of Theorem 2. On the other hand, the argument based on (7.2) seems to be very dependent on the exponential character of the normal family of distributions. The generalizations indicated in Section 6 do not seem to be derivable by means of this argument.

APPENDIX: PROOF OF THEOREM 2, COMPLETED:

Let $B = 2B' \max\{\sigma_i^2/\sigma_j^2: i, j = 1, \cdots, p\}$, $B \ge 2$. For given $0 < b < \infty$, define

$$\eta = \eta(x) = (\sum_{i=1}^{p} |x_i|^{1/b})^b \ ,$$

$$\xi_j = \xi_j(x) = |x_j|^{1/b}/\eta^{1/b}(x) = |x_j|^{1/b}/\sum |x_i|^{1/b} \ ,$$

$$\ell(\xi) = \begin{cases}(1/3 - \xi) + 1/12B & 0 \le \xi \le 1/3 \\ (1/3 - \xi)/B + 1/12B & 1/3 \le \xi \le 1\end{cases} ,$$

$$w(\xi) = \int_0^\xi \ell(t)t^{(b-1)}(1 - t)^{b-1}dt \ .$$

Note that $\ell(t) \le 0$ for $t \ge 5/12$, $\ell(t) > 0$ for $t < 1/3$, and $\ell(\cdot)$ is convex. Note also that $\ell(0) + \ell(1) > 0$. If follows that it is possible to choose b so that $w(0) = 0 = w(1)$. (Note $b = b(B) \to \infty$ as $B \to \infty$.)

Let

$$h_i(x) = \epsilon w(\xi_i) \operatorname{sgn} x_i/\eta(x)(1 - \xi_i)^b \ .$$

Since $w(\xi) = 0(\xi^b(1 - \xi)^b)$, h_i is bounded and continuous outside a neighborhood of $x = 0$. The constant, ϵ, will be determined later. Then let

$$\delta(x) = \begin{cases}x & \text{for } \|x\| \le 1 \\ x - h(x) & \text{for } \|x\| > 1 \text{ where } h = (h_1, \cdots, h_p)\end{cases} .$$

Now,

$$(\partial/\partial x_i)h_i(x) = h_{ii}(x) = (\epsilon/b\eta^2(x))\ell(\xi_i) \ .$$

The basic formula (3.6) in this case is thus

$$D(\mu) = \epsilon\eta^{-2}(\mu)b^{-1}(2\sum_{i=1}^{p} c_i\sigma_i^2\ell(\xi_i(\mu) - \epsilon b\sum_{i=1}^{p} c_i w^2(\xi_i)/(1-\xi_i)^{2b}) \;.$$

The crucial property of the functions $\ell(\cdot)$ is that since $\sum \xi_i = 1$, $p \geq 3$, and $c^{(1)} \max\{\sigma_i^2\} < Bc^{(3)} \min\{\sigma_i^2\}$,

$$\sum c_i\sigma_i^2\ell(\xi_i) \geq \sum c_i \min\{\sigma_i^2\}/12B > 0 \quad \text{for all } c \in \mathcal{C}^* \;. \tag{A.1}$$

To verify (A.1), consider the problem of minimizing $\sum_{i=1}^{p}\alpha_i\ell(\xi_i)$ subject to the restrictions $\alpha_i \geq 0$, $\sum\alpha_i = 1$ and $\alpha^{(1)} < B\alpha^{(3)}$ where $\alpha^{(1)} \geq \alpha^{(2)} \geq \cdots \geq \alpha^{(p)}$. Fix ξ. For notational convenience suppose $\xi_1 \leq \xi_2 \leq \cdots \leq \xi_p$. Since ℓ is decreasing, the minimum of $\sum\alpha_i\ell(\xi_i)$ subject to these restrictions occurs when α_1, α_2, α_3 are as large as possible in that order—to be specific, when $\alpha_1 = B(2+B)$, $\alpha_2 = \alpha_3 = 1/(2+B)$, $\alpha_4 = \cdots = \alpha_p = 0$. Now, fix $\alpha_1, \cdots, \alpha_p$ as specified here, and let ξ vary subject to the restriction $\xi_1 \geq \cdots \geq \xi_p$. For a given value of $\xi_2 + \xi_3 = v$ the minimum of $(1/(1+2B))(\ell(\xi_2) + \ell(\xi_3))$ occurs when $\xi_2 = \xi_3 = v/2 \leq \xi_1$ since ℓ is convex. Consider such ξ's. Since $\sum\xi_i = 1$ we have that $\xi_2 = \xi_3 \leq 1/3$. Substituting in the definition of ℓ yields

$$\begin{aligned}\sum\alpha_i\ell(\xi_i) &\geq \frac{1/3-\xi_1}{B}\cdot\frac{B}{2+B} + \frac{2(1/3-v/2)}{2+B} + \frac{\xi_1+v}{12B} \\ &= \frac{1}{2+B} - (\xi+v)\left(\frac{1}{2+B} - \frac{1}{12B}\right) \\ &\geq 1/12B \text{ since } 0 \leq \xi + v \leq 1 \;.\end{aligned}$$

Letting $\alpha_i = c_i\sigma_i^2/\sum c_i\sigma_i^2$ yields (A.1).

Recall that $w^2(\xi)/(1-\xi)^{2b} = O(1)$. Hence ϵ may certainly be chosen sufficiently small so that

$$D(\mu) \geq K\eta^{-2}(\mu)\sum_{i=1}^{p} c_i > 0$$

for all $c \in \mathcal{C}^*$ for $\|\mu\|$ sufficiently large. Here, K is an appropriate positive constant which depends on p, B, $\min\{\sigma_i^2\}$, and b, but not on $c \in \mathcal{C}^*$. Since $\eta^2(\mu) = O(\|\mu\|^2)$, it follows that

$$\liminf_{\|\mu\|\to\infty} \|\mu\|^2 D(\mu) \geq K'\sum c_i > 0 \;,$$

as required in (3.7) and (3.1). Since $\|x - \delta(x)\|$ is bounded, (3.2′) is satisfied.

Additional computations show that the error term $e''(\mu)$ satisfies $e''(\mu) = c^{(1)}o(\|\mu\|^{-2})$ as $\|\mu\| \to \infty$. It follows that randomization of the origin yields an estimator which dominates the usual estimator for all $L \in \mathcal{L}(\mathcal{C}^*)$. We now summarize these computations.

Error terms: We now turn to the error term $e_1'(\mu)$, as defined before (3.4), in the situation of Theorem 2. By definition

$$e_1'(\mu) = 2E[(h_1(x) - h_1(\mu) - h_{11}(\mu)(x_1-\mu_1))(x_1-\mu_1)] \;.$$

To get a satisfactory estimate of this quantity using an elementary Taylor series approach seems to require special care because $h_{1j}(t)$ is not continuous when some $t_j = 0$ for $j \neq 1$. A minor modification of h might avoid this strictly technical difficulty and still allow for the essential part of the proof of Theorem 2. However, we will take a different approach, using the given h, and capitalizing on its symmetry properties and the independence of the coordinates of X. Fix μ. Let $Y = X - \mu$ and consider the function

$$g(w) = 2E\{h_1(\mu + wy)y_1\} \;.$$

This function is differentiable on $0 \leq w \leq 1$ and can be differentiated inside the expectation and $e_1'(\mu) = g(1) - \sigma^2 h_{11}(\mu)$. By Taylor's theorem, $g(1) = g(0) + g'(\hat{w})$ where $0 < \hat{w} < 1$. $g(0) = 0$. Thus, after some calculation,

$$e_1'(\mu) = 2E[(h_{11}(\mu+\hat{w}y) - h_{1i}(\mu))y_1^2\sum_{j=2}^{p} + h_{1j}(\mu+\hat{w}y)y_1y_j] \;; \tag{A.2}$$

h_{11} has been previosly computed. We have

$$\begin{aligned}\eta^2(\mu)|h_{11}(\mu+\hat{w}y) - h_{11}(\mu)|y_1^2 &\\ \leq (\epsilon y_1^2/b)[|\ell(\xi_1(\mu)) - \ell(\xi_1(\mu+\hat{w}y))| &\\ + \ell(\xi_1(\mu+\hat{w}y))|1 - \eta^2(\mu)/\eta^2(\mu+\hat{w}y)|] \;. &\end{aligned} \tag{A.3}$$

Now,

$$\begin{aligned}&\int y_1^2|1-\eta^2(\mu)/\eta^2(\mu+\hat{w}y)|\exp(-\sum\sigma_i^2y_i^2/2) \\ &= \left\{\int_{\|\mu+\hat{w}y\|>1} + \int_{\|\mu+\hat{w}y\|\leq 1}\right\} y_1^2|1-\eta^2(\mu)/\eta^2(\mu+\hat{w}y)| \\ &\qquad\cdot\exp(-\sum\sigma_i^2y_i^2/2)dy\end{aligned}$$

$\to 0$ as $\|\mu\| \to \infty$ uniformly on $0 \leq \hat{w} \leq 1$, since in the first integral after the equality above $y_1^2|1-\eta^2(\mu)/\eta^2(\mu+\hat{w}y)|$ is bounded and for each y tends to zero uniformly on $0 \leq \hat{w} \leq 1$ as $\|\mu\| \to \infty$; and since the second integral above is less than

$$\eta^2(\mu)(\sup_{\|\mu+\hat{w}y\|<1} y_1^2\exp(-\sum\sigma_i^2y_i^2/2)) \cdot\int_{\|\mu+\hat{w}y\|<1}(1+\eta^{-2}(\mu+\hat{w}y))dy \to 0$$

uniformly on $0 \leq \hat{w} \leq 1$ as $\|\mu\| \to \infty$.

Let $I_1(\hat{w},\mu) = E$ (r.h.s. (A.3)). Apply the bounded convergence theorem on the first term of I_1 and the preceding arguments on the second. This yields $I_1(\hat{w},\mu) \to 0$ uniformly on $0 \leq \hat{w} \leq 1$ as $\|\mu\| \to \infty$. This part of the error term $e_i(\mu)$ is therefore insignificant, as desired.

If $y = (y_1, \cdots, y_p)$ let $y^* = (-y_1, y_2, \cdots, y_p)$. Assume without loss of generality that $\mu_1 \geq 0$. Then, by the symmetry of the normal distribution

$$\begin{aligned}&|2E(\eta^2(\mu)h_{12}(\mu+\hat{w}y)y_1y_2)| \\ &= 2|E\{\eta^2(\mu)y_1y_2(h_{12}(\mu+\hat{w}y) - h_{12}(\mu+\hat{w}y^*))|y_1>0\}| \;.\end{aligned} \tag{A.4}$$

Since

$$h_{12}(z) = \frac{\epsilon w(\xi_1(z))\xi_2^{1-b}(z)}{(1-\xi_1(z))^b\eta^2(z)}\operatorname{sgn} z_2 \;,$$

we have

$$\begin{aligned}&\eta^2(\mu)|h_{12}(\mu+\hat{w}y) - h_{12}(\mu+\hat{w}y^*)| \\ &= \Bigg|\frac{\epsilon w(\xi_1(\mu+\hat{w}y))}{(1-\xi_1(\mu+\hat{w}y))^b\eta^2(\mu+\hat{w}y)}(\xi_2^{1-b}(\mu+\hat{w}y) \\ &\quad - \xi_2^{1-b}(\mu+\hat{w}y^*)) + \epsilon\xi_2^{1-b}(\mu+\hat{w}y^*)\left(\frac{w(\xi_1(\mu+\hat{w}y))}{(1-\xi_1(\mu+\hat{w}y))^b}\right. \\ &\quad \left. - \frac{w(\xi_1(\mu+\hat{w}y^*))}{(1-\xi_1(\mu+\hat{w}y^*))^b}\right)\frac{\eta^2(\mu)}{\eta^2(\mu+\hat{w}y)} + \left(\frac{\eta^2(\mu)}{\eta^2(\mu+\hat{w}y)}\right. \\ &\quad \left. - \frac{\eta^2(\mu)}{\eta^2(\mu+\hat{w}y^*)}\right)\epsilon\xi_2^{1-b}(\mu+\hat{w}y^*)\frac{w(\xi_1(\mu+\hat{w}y^*))}{(1-\xi_1(\mu+\hat{w}y^*))^b}\Bigg| \;.\end{aligned} \tag{A.5}$$

Observe that

$$(\partial/\partial z_1)\xi_2^{1-b}(z) = (1/b - 1)\xi_1^{1-b}(z)\xi_2^{1-b}(z)\operatorname{sgn} z_1/\eta(z)$$

and that $\xi_2(z)$ is an even function of z_1. Hence, for $\mu_1 \geq 0$, $y_1 > 0$, and $(\mu+\hat{w}y)_2 \neq 0$, we have

$$\begin{aligned}&|\xi_2^{1-b}(\mu+\hat{w}y) - \xi_2^{1-b}(\mu+\hat{w}y^*)| \\ &\leq (1-1/b)\xi_1^{1-b}(\mu+\hat{w}y^*)\xi_2^{1-b}(\mu+\hat{w}y)(2\hat{w}y_1)/\eta(\mu+\hat{w}y^*) \;.\end{aligned}$$

Note also that on this same region $\eta^2(\mu) < \eta^2(\mu+\hat{w}y)$, and recall that $w(\xi_1(\mu+\hat{w}y))/(1-\xi_1(\mu+\hat{w}y))^b \leq K\xi_1^b(\mu+\hat{w}y) \leq K\xi_1^b(\mu+\hat{w}y^*)$ on this region for some appropriate constant, $K < \infty$. Hence,

$$\begin{aligned}&E\left(\left|y_1y_2\frac{\epsilon w(\xi_1(\mu+\hat{w}y))\eta^2(\mu)}{(1-\xi_1(\mu+\hat{w}y))^b\eta^2(\mu+\hat{w}y)}(\xi_2^{1-b}(\mu+\hat{w}y)\right.\right. \\ &\quad \left.\left. - \xi_2^{1-b}(\mu+\hat{w}y^*))\right| \;|y_1>0\right) \\ &\leq E\left(\left|y_1y_2\epsilon\left(1-\frac{1}{b}\right)K\xi_2^{1-b}(\mu+\hat{w}y)2\hat{w}y_1/\eta(\mu+\hat{w}y^*)\right| \;|y_1>0\right) \\ &= 2K\epsilon E\left(\left|y_1^2y_2\frac{|(\mu+\hat{w}y^*)_2|^{1/b-1}\hat{w}}{\eta(\mu+\hat{w}y^*)^{1/b}}\right| \;|y_1>0\right)\end{aligned}$$

$\to 0$ as $\|\mu\| \to \infty$, since $E(|(\mu + \hat{w}y^*)_2|^{(1/b)-1}\hat{w}) < K' < \infty$ for all values of μ, $\hat{w} > 0$, for some appropriate constant, $K' < \infty$.

Expectations appropriate to the remaining two terms on the right of (A.5) can be similarly shown to tend to zero as $\|\mu\| \to \infty$. Substituting in (A.4) thus yields that this part of the error term $e_1'(\mu)$ is also insignificant. The remaining terms in (A.2) (for $j = 3, \cdots, p$) are exactly like the preceding term. Hence, $\eta^2(\mu)e_1'(\mu) \to 0$ as $\|\mu\| \to \infty$.

The remaining part of the error involves terms of the form $E_\mu(h_1^2(x) - h_1^2(\mu))$. Now,

$$|h_1^2(x) - h_1^2(\mu)| \leq k \left|\frac{1}{\eta^2(x)} - \frac{1}{\eta^2(\mu)}\right| + \frac{1}{\eta^2(\mu)}\left(\frac{w^2(\xi_1(x))}{(1 - \xi_1(x))^{2b}} - \frac{w^2(\xi_1(\mu))}{(1 - \xi_1(\mu))^{2b}}\right).$$

From this expression it is easily verified as before that $E_\mu(h_1^2(x) - h_1^2(\mu)) = o(\|\mu\|^{-2})$ as $\|\mu\| \to \infty$. This result combined with the estimate for $e_i'(\mu)$ yield that $e''(\mu) = c^{(1)}o(\|\mu\|^{-2})$, as claimed in the proof of Theorem 2.

[Received May 1973. Revised November 1974.]

REFERENCES

[1] Alam, K., "A Family of Admissible Minimax Estimators of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, 1 (May 1973), 517–25.

[2] Baranchik, A.J., "A Family of Minimax Estimators of the Mean of a Multivariate Normal Distribution," *Annals of Mathematical Statistics*, 41 (April 1970), 642–5.

[3] Berk, R. and Bickel, P., "On Invariance and Almost Invariance," *Annals of Mathematical Statistics*, 39 (October 1968), 1573–6.

[4] Bhattacharya, P.K., "Estimating the Mean of a Multivariate Normal Population with General Quadratic Loss Function," *Annals of Mathematical Statistics*, 37 (December 1966), 1819–25.

[5] Bock, M.E., "Minimax Estimators of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, 3 (January 1975), 209–18.

[6] Brown, L., "On the Admissibility of Invariant Estimators of One or More Location Parameters," *Annals of Mathematical Statistics*, 37 (October 1966), 1087–136.

[7] ———, "Admissible Estimators, Recurrent Diffusions and Insoluble Boundary Value Problems," *Annals of Mathematical Statistics*, 42 (June 1971), 855–903.

[8] ———, "Taylor Series and Admissible Estimators, a Heuristic Approach—with Applications," special invited paper, delivered to Institute of Mathematical Statistics, May 1973. Submitted to *Annals of Statistics*.

[9] Efron, B. and Morris, C., "Limiting the Risk of Bayes and Empirical Bayes Estimators—Part II: The Empirical Bayes Case," *Journal of the American Statistical Association*, 67 (March 1972), 130–9.

[10] ——— and Morris, C., "Stein's Estimation Rule and Its Competitors—An Empirical Bayes Approach," *Journal of the American Statistical Association*, 68 (March 1973), 117–30.

[11] Farrell, R., "Estimators of a Location Parameter in the Absolutely Continuous Case," *Annals of Mathematical Statistics*, 35 (September 1964), 949–99.

[12] Fintushal, S.M., "Admissibility of Pitman's Estimator for a Multidimensional Location Parameter," *Theory of Probability and Its Applications*, 16, No. 4 (1971), 702–7.

[13] James, W. and Stein, C., "Estimation with Quadratic Loss," *Proceedings of 4th Berkeley Symposium on Probability and Statistics*, I (1960), 367–79.

[14] Karlin, S., "Polya Type Distributions II," *Annals of Mathematical Statistics*, 28 (May 1957), 281–308.

[15] LeCam, L., "An Extension of Wald's Theory of Statistical Decision Functions," *Annals of Mathematical Statistics*, 26 (March 1955), 69–81.

[16] Morris, C. and Efron, B., "Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution," submitted to *Annals of Statistics* (1974).

[17] Stein, C., "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," *Proceedings 3rd Berkeley Symposium on Probability and Statistics*, I (1955), 197–206.

[18] Stein, C., "Estimation of the Mean of a Multivariate Normal Distribution," Technical Report No. 48, Department of Statistics, Stanford University, 1973.

[19] Strawderman, W., "Proper Bayes Estimators of the Multivariate Normal Mean," *Annals of Mathematical Statistics*, 42 (February 1971), 385–8.

[20] ———, "Proper Bayes Minimax Estimators of the Multivariate Normal Mean Vector for the Case of Common Unknown Variances," *Annals of Statistics* (November 1973), 1189–1194.

[21] Wesler, O., "Invariance Theory and a Modified Minimax Principle," *Annals of Mathematical Statistics*, 30 (March 1959), 1–20. (See correction in Berk and Bickel [3, p. 1576].)

[22] Wind, S.L., "Stein-James Estimators of a Multivariate Location Parameter," *Annals of Mathematical Statistics*, 43 (February 1972), 340–7.