The Annals of Statistics
1976, Vol. 4, No. 1, 33-50

TAIL MINIMAXITY IN LOCATION VECTOR PROBLEMS AND ITS APPLICATIONS

BY JAMES O. BERGER

Purdue University

Let $X = (X_1, \cdots, X_p)^t$, $p \geqq 3$, have density $f(x - \theta)$ with respect to Lebesgue measure. It is desired to estimate $\theta = (\theta_1, \cdots, \theta_p)^t$ under the loss $L(\delta - \theta)$. Assuming the problem has a minimax risk R_0, an estimator is defined to be tail minimax if its risk is no larger than R_0 outside some compact set. Under quite general conditions on f and L, sufficient conditions for an estimator to be tail minimax are given. A class of good tail minimax estimators is then developed and compared with the best invariant estimator.

1. Introduction. Let $X = (X_1, \cdots, X_p)^t$ be an observation from a p-dimensional random variable with density $f(x - \theta)$ with respect to Lebesgue measure. It is desired to estimate the location vector $\theta = (\theta_1, \cdots, \theta_p)^t$. The loss incurred in estimating θ by δ is $L(\delta - \theta)$, where $0 \leqq L(\cdot) < \infty$, $L(0) = 0$, and L is continuous.

It will be assumed that the best invariant estimator for θ (with respect to the translation group on R^p) exists and is unique. A simple reparameterization of the problem will then ensure that $\delta_0(X) = X$ is the best invariant estimator. It will also be assumed that δ_0 is a minimax estimator of θ. This additional assumption is fairly weak. Indeed if L is bounded or $L(x) \to \infty$ as $|x| \to \infty$ ($|x|$ is the usual Euclidean norm), then it can be concluded that δ_0 is minimax. (See Ferguson (1967) for similar results.)

For a measurable estimator $\delta(x) = (\delta_1(x), \cdots, \delta_p(x))^t$, define the risk function $R(\delta, \theta) = E_\theta L(\delta(X) - \theta)$, where E_θ stands for the expectation under θ. For convenience, define $\Delta_\delta(\theta) = R(\delta, \theta) - R(\delta_0, \theta)$. Noting that δ_0 has constant risk, it is clear that δ is a minimax estimator if and only if $\Delta_\delta(\theta) \leqq 0$ for all $\theta \in R^p$. The search for minimax estimators is thus equivalent to the search for estimators as good as or better than δ_0 (in terms of risks).

Stein (1955) showed that δ_0 is inadmissible for estimating a multivariate normal mean under squared error loss if $p \geqq 3$. Considerable effort has since been given to improving upon δ_0 in location vector problems. The theoretical questions have been answered quite thoroughly (in the case where there are no nuisance parameters) by Brown (1966). He has shown that in 3 or more dimensions δ_0 is inadmissible for an extremely wide variety of distributions and loss functions. Unfortunately, the theoretical results do not explicitly give estimators which are significantly better than δ_0.

Received August 1974; revised June 1975.

AMS 1970 *subject classifications.* Primary 62C99; Secondary 62F10, 62H99.

Key words and phrases. Minimax, tail minimax, location vector, best invariant estimator, risk function.

Finding minimax estimators of practical importance has been a topic receiving considerable treatment in the literature recently. Baranchik (1970), Strawderman (1971), Alam (1973) and others obtained classes of good minimax estimators of a normal mean with covariance matrix a multiple of the identity and squared error loss. Berger (1975) was able to extend the above results to the case Σ arbitrary (but known) and arbitrary quadratic loss. (Partial results in this direction had earlier been obtained by Bhattacharya (1966), Bock (1975), and Berger (1974c).) An extension in a different direction was made by Strawderman (1974) and Berger (1974d). They found classes of good minimax estimators for a wide variety of nonnormal distributions (but again only for quadratic loss).

The obvious weakness of the above practical results is that only quadratic loss is dealt with. Clearly, it would be desirable to determine how dependent these results are upon the assumption of quadratic loss, and to determine what should be done for other losses. Unfortunately, finding practical minimax estimators for nonquadratic loss appears to be an extremely difficult problem. Existing techniques do not offer hope of a solution.

Because of the above problem, this paper considers a subject we will call tail minimaxity. An estimator, δ, is said to be *tail minimax* if there exists $K > 0$ such that if $|\theta| > K$, then $\Delta_\delta(\theta) \leqq 0$. Classes of tail minimax estimators of θ will be developed for the general problem.

Tail minimaxity is a useful concept for several reasons. First, since a minimax estimator must clearly be tail minimax, it provides a necessary condition for minimaxity that turns out to be quite easy to check. Related to this is its usefulness in suggesting what estimators may actually be minimax. For example, it was this theory which suggested the class of minimax estimators found in Berger (1974c). Finally, tail minimaxity will be discussed as a criterion on its own merits. Considerable evidence will be presented which indicates that reasonable tail minimax estimators are usually nearly minimax, and even if not seem considerably more desirable than δ_0 in terms of risks. The intuitve justification for tail minimax estimators is as follows. It is well known that "Stein type" estimators, which improve upon δ_0, pull the usual estimate $\delta_0(X) = X$ in towards zero. (This behavior is clear for the typical Stein estimator $\delta(X) = (1 - |X|^{-2})^+X$, where $^+$ stands for the positive part.) Estimators which behave in this fashion will clearly have risk smaller than δ_0 in a neighborhood of $\theta = 0$. If they are also tail minimax, their risks will be as small as δ_0 for large values of $|\theta|$. It can be hoped that the risk behavior, for θ in the midrange, is reasonable. Section 3, in dealing with applications, presents numerical evidence supporting this belief.

As a final comment, note that the multiobservational situation can be subsumed into the general framework of this paper. If a sample $X^1, X^2, \cdots, X^n$ is taken, and an invariant estimator $\delta_1(X^1, \cdots, X^n)$ is to be used, merely set $X = \delta_1$ and let f be the density of X. Hopefully, estimators improving upon δ_1 can then be found.

2. Theoretical results. This section develops the major results on tail minimaxity. The analysis is accomplished by using the methods of Brown (1974) to approximate $\Delta_\delta(\theta)$ for large values of $|\theta|$. Conditions for tail minimaxity can then be easily derived. A relatively simple class of tail minimax estimators is also developed.

Before proceeding, some needed notation will be given. For an arbitrary estimator δ, let

$$\gamma(x) = (\gamma_1(x), \cdots, \gamma_p(x))^t = \delta(x) - x .$$

If $h: R^p \to R^1$ is a function with the appropriate number of derivatives, let

$$h^{(i)}(x) = \frac{\partial}{\partial x_i} h(x) , \qquad h^{(i,j)}(x) = \frac{\partial^2}{\partial x_i \, \partial x_j} h(x) , \quad \text{etc.} ,$$

and let $\nabla h(x) = (h^{(1)}(x), \cdots, h^{(p)}(x))$ denote the gradient of h. Let $J_\gamma(x)$ be the Jacobian matrix of γ. Thus the (i, j) entry of $J_\gamma(x)$ is $\gamma_i^{(j)}(x)$. The usual "o" (little oh) and "O" (big oh) notation will be used. For a $p \times p$ matrix A, let $\operatorname{tr} A$ denote the trace of A, and $\operatorname{ch}_{\max}[A]$ denote the maximum characteristic root of A. The letter K will be used as a generic constant throughout the paper.

A fairly large number of assumptions are needed. They are not the most general possible, in that many of the more technical assumptions could undoubtedly be weakened. They are fairly easy to verify in their present form, however, and do cover many of the situations that occur in practice. The first 7 assumptions deal with f and L. The last 2 assumptions concern the estimator $\delta(x) = x + \gamma(x)$.

ASSUMPTIONS.

1. L has continuous third order partial derivatives, except possibly at zero.
2. $E_0 L^{(i)}(X) = 0$ for $1 \leqq i \leqq p$.
3. The $p \times p$ matrix $\mathscr{L}$, with elements $l_{ij} = E_0 L^{(i,j)}(X)$, is positive definite.
4. The $p \times p$ matrix M, with elements $m_{ij} = E_0[L^{(i)}(X)X_j]$, has positive eigenvalues.
5. If $1 \leqq i \leqq p$, $1 \leqq j \leqq p$, $1 \leqq k \leqq p$, then
   (a) $E_0|X|^3 < \infty$,
   (b) $E_0[|X|^n|L^{(i)}(X)|] < \infty$ for $0 \leqq n \leqq 3$,
   (c) $E_0[|X|^n|L^{(i,j)}(X)|] < \infty$ for $0 \leqq n \leqq 4$,
   (d) $E_0[|X|^n|L^{(i,j,k)}(X)|] < \infty$ for $0 \leqq n \leqq 3$.
6. (a) If $|y| \leqq K_1$, then there exist K_2 and K_3 such that $L(x + y) \leqq K_2 + K_3 L(x)$ for all $x \in R^p$.
   (b) There exist constants K_4 and K_5 such that if $|y| < |x|/2$, then $|L^{(i,j,k)}(x + y)| \leqq K_4 + K_5|L^{(i,j,k)}(x)|$.
7. Let $Q(\varepsilon) = \{x \in R^p : |x| < \varepsilon\}$. Assume that as $\varepsilon \to 0$
   (a) $\int_{Q(\varepsilon)} f(x)\,dx = o(\varepsilon^2)$,
   (b) $\int_{Q(\varepsilon)} |L^{(i)}(x)| f(x)\,dx = o(\varepsilon)$,
   (c) $\int_{Q(\varepsilon)} |L^{(i,j)}(x)| f(x)\,dx = o(1)$.

8. $|\gamma(x)| < \beta < \infty$ for all $x \in R^p$.
9. There exist $\alpha > 0$ and $T > 0$ such that if $|x| > T$, then
   (a) $|\gamma(x)| \leqq \alpha/|x|$,
   (b) $\gamma_i(x)$ has all second order partial derivatives, $1 \leqq i \leqq p$,
   (c) $|\gamma_i^{(j)}(x)| = o(|x|^{-1})$, $1 \leqq i \leqq p$, $1 \leqq j \leqq p$,
   (d) $|\gamma_i^{(j,k)}(x)| = o(|x|^{-2})$, $1 \leqq i \leqq p$, $1 \leqq j \leqq p$, $1 \leqq k \leqq p$.

Discussion of Assumptions.

1. The possible nondifferentiability of L at zero causes complications in the analysis but is allowed so that important losses such as $L(x) = |x|$ can be considered.

2 and 3. These assumptions are very closely related to the assumption that the best invariant estimator is $\delta_0(X) = X$. To see this, note that by the definition of the best invariant estimator, the function $g(c) = \int L(x + c)f(x)\,dx$ is minimized at $c = (c_1, \cdots, c_p) = 0$. Under appropriate conditions, this implies that $0 = g^{(i)}(0) = E_0 L^{(i)}(X)$, and that the matrix $\mathscr{L}$, with elements $l_{ij} = E_0 L^{(i,j)}(X) = g^{(i,j)}(0)$, is positive definite (in order for the critical point to be a unique local minimum).

4. This assumption guarantees that it is desired to estimate the full location vector θ, rather than certain coordinates of θ. (See Berger (1974a) and Berger (1974b) for a discussion of the latter type of problem.) The assumption, itself, is really quite weak. It holds in all practical examples of which the author is aware (providing the entire location vector is of interest). Three common situations in which it is satisfied are:

(i) If f is a p-variate normal density with known covariance matrix Σ, it is shown in Section 3 that $M = \mathscr{L}\Sigma$, which clearly has positive eigenvalues.

(ii) If L is the quadratic loss $L(x) = x^t Q x$, Q positive definite, an easy calculation shows that $M = 2Q\Sigma$, where Σ is the covariance matrix of f.

(iii) If $L(x) = \sum_{i=1}^p h_i(x_i)$, where $h_i(x_i)$ is strictly increasing in $|x_i|$, and $X_1, \cdots, X_p$ are all independent, then $m_{ii} = \int x_i L^{(i)}(x) f(x)\,dx = \int x_i h_i'(x_i) f(x)\,dx > 0$, $1 \leqq i \leqq p$; $m_{ij} = E_0[X_j h_i'(X_i)] = E_0[X_j]E_0[h_i'(X_i)] = 0$ if $i \neq j$ (by Assumption 2).

5. As an example, note that if L is quadratic loss, then 4 absolute moments of the density are required. Because of the possible nondifferentiability of L at zero, care must be taken in checking that the integrals in (b), (c), and (d) are finite around zero. For $p \geqq 3$, the conditions are usually satisfied. For example, if f is bounded, $L(x) = |x|^a$ $(a > 0)$, and $p \geqq 3$, it can be easily checked that the integrals over compact neighborhoods of zero are finite.

6. This assumption can be easily verified for most loss functions. If $L(x) = |x|^a$ $(a > 0)$, for example, verification is straightforward.

7. This assumption will usually be satisfied if $p \geqq 3$. Part (a), for example, only requires f to be bounded in a neighborhood of zero (in addition to $p \geqq 3$). It can be checked that part (b) is satisfied for $L(x) = |x|^a$ $(a > 0)$, and $p \geqq 3$. Part (c) actually follows from Assumption 4, but is included here for simplicity.

8. This assumption is somewhat restrictive in that it does not allow consideration of basic Stein type estimators such as $(1 - |x|^{-2})x$. (Clearly $\gamma(x) = -x/|x|^2$ is unbounded.) It is well known, however, that such estimators can be significantly improved upon by "positive part versions" such as $\delta(x) = (1 - |x|^{-2})^+x$. It is easy to check that this estimator has bounded γ.

9. All minimax estimators of practical interest, of which the author is aware, satisfy this condition.

THEOREM 1. *Assume Assumptions* 1 *through* 9 *hold. It follows that there exists* $K > 1$ *such that if* $|\theta| > K$, *then*

$$\Delta_\delta(\theta) = \operatorname{tr}(J_\gamma(\theta)M^t) + \tfrac{1}{2}\gamma^t(\theta)\mathscr{L}\gamma(\theta) + o(|\theta|^{-2}).$$

PROOF. Much of the proof is patterned after Brown (1974). A fair number of details are included since the possible nondifferentiability of L at zero complicates matters, and since similar results sketched in Brown (1974) have not been worked out in detail.

Define $V = \{x: |x - \theta| > 3\alpha/|\theta|\}$, where α is from Assumption 9. Clearly

$$\begin{aligned}(2.1)\qquad \Delta_\delta(\theta) &= \int [L(\delta(x) - \theta) - L(x - \theta)]f(x - \theta)\,dx \\ &= \int_V [L(\gamma(x) + x - \theta) - L(x - \theta)]f(x - \theta)\,dx \\ &\quad + \int_{V^c} [L(\gamma(x) + x - \theta) - L(x - \theta)]f(x - \theta)\,dx .\end{aligned}$$

Consider first the integral over V^c above. Using Assumption 8, it is clear that if $x \in V^c$ and $|\theta| > 1$, then $|\gamma(x) + x - \theta|$ and $|x - \theta|$ are bounded. The continuity of L and Assumption 7(a) thus imply that

$$\begin{aligned}(2.2)\qquad &\int_{V^c} [L(\gamma(x) + x - \theta) - L(x - \theta)]f(x - \theta)\,dx \\ &\qquad \leqq K\int_{V^c} f(x - \theta)\,dx = o(|\theta|^{-2}) .\end{aligned}$$

Consider next the integral over V in (2.1). Note first that if $|x - \theta| \geqq \beta$, then Assumption 8 implies that $|\gamma(x)| < \beta \leqq |x - \theta|$. If $|x - \theta| < \beta$ and $|\theta| > 2\beta + T$, then by Assumption 9(a), $|\gamma(x)| \leqq \alpha/|x| \leqq \alpha/(|\theta| - \beta) \leqq 2\alpha/|\theta|$. Combining these two observations, it is clear that

$$(2.3)\qquad |\gamma(x)| < |x - \theta| , \qquad \text{if} \quad x \in V \quad \text{and} \quad |\theta| > 2\beta + T .$$

Assuming $|\theta| > 2\beta + T$, it follows from (2.3) that if $x \in V$, then the line between $(x - \theta)$ and $(\gamma(x) + x - \theta)$ does not contain zero. Assumption 1 thus implies that $L(\gamma(x) + x - \theta)$ can be expanded in a Taylor series about $(x - \theta)$, up to fourth order terms. Using this expansion and rearranging terms gives

$$\begin{aligned}(2.4)\qquad &\int_V [L(\gamma(x) + x - \theta) - L(x - \theta)]f(x - \theta)\,dx \\ &\qquad = I_1 + I_2 + I_3 + I_4 + I_5 ,\end{aligned}$$

where

$$I_1 = \sum_{i=1}^p \int_{R^p} L^{(i)}(x - \theta)\gamma_i(x)f(x - \theta)\,dx ,$$

$$I_2 = \tfrac{1}{2}\sum_{i=1}^p \sum_{j=1}^p \int_{R^p} L^{(i,j)}(x - \theta)\gamma_i(x)\gamma_j(x)f(x - \theta)\,dx ,$$

$$I_3 = \tfrac{1}{6} \textstyle\sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p \int_V L^{(i,j,k)}(t\gamma(x) + x - \theta)\gamma_i(x)\gamma_j(x)\gamma_k(x) f(x-\theta)\,dx\,, \quad \text{where} \quad 0 \leqq t(x,\theta) \leqq 1\,,$$

$$I_4 = -\textstyle\sum_{i=1}^p \int_{V^c} L^{(i)}(x-\theta)\gamma_i(x) f(x-\theta)\,dx\,,$$

$$I_5 = -\tfrac{1}{2} \textstyle\sum_{i=1}^p \sum_{j=1}^p \int_{V^c} L^{(i,j)}(x-\theta)\gamma_i(x)\gamma_j(x) f(x-\theta)\,dx\,.$$

As before, it can be checked that if $x \in V^c$ and $|\theta| > T + 6\alpha + 1$, then $|\gamma(x)| \leqq 2\alpha/|\theta|$. Using this, together with Assumption 7, it is clear that

$$I_4 + I_5 = o(|\theta|^{-2})\,. \tag{2.5}$$

To handle I_3, define $V_1 = \{x : |x - \theta| > |\theta|/3\}$ and $V_2 = \{x : 3\alpha/|\theta| < |x - \theta| < |\theta|/3\}$. Clearly, if $x \in V_1$ and $|\theta| > 6\beta$, then $|\gamma(x)| < \beta < |\theta|/6 < |x - \theta|/2$. Assumption 6(b) thus implies that

$$|L^{(i,j,k)}(t\gamma(x) + x - \theta)| \leqq K_4 + K_5|L^{(i,j,k)}(x-\theta)|\,, \quad \text{if} \quad x \in V_1 \quad \text{and} \quad |\theta| > 6\beta\,. \tag{2.6}$$

Using (2.6), together with Assumption 5 and a simple Chebyshev argument, gives

$$\int_{V_1} |L^{(i,j,k)}(t\gamma(x) + x - \theta)|\,|\gamma_i(x)\gamma_j(x)\gamma_k(x)|\, f(x-\theta)\,dx \leqq K|\theta|^{-3} \int |x-\theta|^3[1 + |L^{(i,j,k)}(x-\theta)|] f(x-\theta)\,dx = o(|\theta|^{-2})\,. \tag{2.7}$$

Note next that if $x \in V_2$, then $|x| > 2|\theta|/3$. Assumption 9(a) and the definition of V_2 hence give

$$|\gamma(x)| \leqq \alpha/|x| < 3\alpha/(2|\theta|) < |x-\theta|/2\,, \quad \text{if} \quad x \in V_2 \quad \text{and} \quad |\theta| > 3T/2\,. \tag{2.8}$$

Using (2.8) and Assumptions 5 and 6(b), it is clear that

$$\int_{V_2} |L^{(i,j,k)}(t\gamma(x) + x - \theta)|\,|\gamma_i(x)\gamma_j(x)\gamma_k(x)|\, f(x-\theta)\,dx \leqq \left(\frac{3\alpha}{2|\theta|}\right)^3 \int_{V_2} [K_4 + K_5|L^{(i,j,k)}(x-\theta)|] f(x-\theta)\,dx = o(|\theta|^{-2})\,. \tag{2.9}$$

Combining (2.7) and (2.9) shows that

$$I_3 = o(|\theta|^{-2})\,. \tag{2.10}$$

Finally, the major terms I_1 and I_2 must be considered. It will be necessary to expand the $\gamma_i(x)$ in Taylor expansions about θ. For this purpose, define $W = \{x : |x - \theta| < |\theta|/2\}$. Assume $|\theta| > 2T$. Clearly, if $x \in W$ and $|\theta| > 2T$, then $|x| > |\theta|/2 > T$. Taylor expansions of the $\gamma_i(x)$ (up to third order terms) will thus be valid if $x \in W$. Using the expansions gives

$$\begin{aligned} I_1 = {} & \textstyle\sum_{i=1}^p \int_W L^{(i)}(x-\theta)[\gamma_i(\theta) + \sum_{j=1}^p \gamma_i^{(j)}(\theta)(x_j - \theta_j) \\ & + \tfrac{1}{2} \textstyle\sum_{j=1}^p \sum_{k=1}^p \gamma_i^{(j,k)}(tx + (1-t)\theta)(x_j - \theta_j)(x_k - \theta_k)] f(x-\theta)\,dx \\ & + \textstyle\sum_{i=1}^p \int_{W^c} L^{(i)}(x-\theta)\gamma_i(x) f(x-\theta)\,dx \\ = {} & \textstyle\sum_{i=1}^p \int_{R^p} L^{(i)}(x-\theta)[\gamma_i(\theta) + \sum_{j=1}^p \gamma_i^{(j)}(\theta)(x_j - \theta_j)] f(x-\theta)\,dx \qquad (2.11) \\ & + \textstyle\sum_{i=1}^p \int_{W^c} L^{(i)}(x-\theta)[\gamma_i(x) - \gamma_i(\theta) \end{aligned}$$

$$- \sum_{j=1}^{p} \gamma_i^{(j)}(\theta)(x_j - \theta_j)] f(x-\theta)\, dx$$
$$+ \tfrac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \int_W L^{(i)}(x-\theta)\gamma_i^{(j,k)}(tx + (1-t)\theta)$$
$$\times (x_j - \theta_j)(x_k - \theta_k) f(x-\theta)\, dx ,$$

where of course $0 \leqq t(x, \theta) \leqq 1$.

By Assumption 2 and the definition of the m_{ij}, it is clear that

$$\sum_{i=1}^{p} \int_{R^p} L^{(i)}(x-\theta)[\gamma_i(\theta) + \sum_{j=1}^{p} \gamma_i^{(j)}(\theta)(x_j-\theta_j)] f(x-\theta)\, dx = \sum_{i=1}^{p}\sum_{j=1}^{p} \gamma_i^{(j)}(\theta) m_{ij} = \operatorname{tr}[J_\gamma(\theta) M^t] . \tag{2.12}$$

Assumptions 5, 8, 9 and a simple Chebyshev argument show that

$$\sum_{i=1}^{p} \int_{W^c} L^{(i)}(x-\theta)[\gamma_i(x) - \gamma_i(\theta) - \sum_{j=1}^{p} \gamma_i^{(j)}(\theta)(x_j-\theta_j)] f(x-\theta)\, dx = o(|\theta|^{-2}) . \tag{2.13}$$

Finally, note that if $x \in W$ and $0 \leqq t \leqq 1$, then $tx + (1-t)\theta \in W$. It thus follows from Assumption 9(d) that

$$\sup_{x\in W} |\gamma_i^{(j,k)}(tx + (1-t)\theta)| \leqq \sup_{\xi:|\xi|>|\theta|/2} |\gamma_i^{(j,k)}(\xi)| = o(|\theta|^{-2}) . \tag{2.14}$$

Using (2.14) and Assumption 5 gives

$$\int_W L^{(i)}(x-\theta)\gamma_i^{(j,k)}(tx+(1-t)\theta)(x_j-\theta_j)(x_k-\theta_k) f(x-\theta)\, dx = o(|\theta|^{-2}) . \tag{2.15}$$

Combining (2.11), (2.12), (2.13) and (2.15), it is clear that

$$I_1 = \operatorname{tr}[J_\gamma(\theta) M^t] + o(|\theta|^{-2}) . \tag{2.16}$$

The term I_2 can be handled in an exactly analogous manner. In the Taylor expansion of $\gamma_i(x)\gamma_j(x)$, the dominant term will be $\gamma_i(\theta)\gamma_j(\theta)$ since its coefficient is $l_{ij} = E_0 L^{(i,j)}(X)$ and $\mathscr{L}$ is positive definite. The result is

$$I_2 = \tfrac{1}{2} \sum_{i=1}^{p}\sum_{j=1}^{p} \gamma_i(\theta)\gamma_j(\theta) l_{ij} + o(|\theta|^{-2}) = \tfrac{1}{2}\gamma^t(\theta)\mathscr{L}\gamma(\theta) + o(|\theta|^{-2}) . \tag{2.17}$$

Finally, combining (2.1), (2.2), (2.4), (2.5), (2.10), (2.16) and (2.17) gives the desired result. □

From Theorem 1 follow the following two important corollaries.

COROLLARY 1. *Assume Assumptions* 1 *through* 9 *hold, and that there exist* $\varepsilon > 0$ *and a sequence* $\{\theta^i \in R^p, i \geqq 1\}$ *such that* $|\theta^i| \to \infty$ *and such that*

$$\operatorname{tr}(J_\gamma(\theta^i) M^t) + \tfrac{1}{2}\gamma^t(\theta^i)\mathscr{L}\gamma(\theta^i) \geqq \varepsilon/|\theta^i|^2 , \qquad i \geqq 1 .$$

It can be concluded that $\delta(X) = X + \gamma(X)$ is not tail minimax, and hence cannot be minimax.

PROOF. Obvious. □

COROLLARY 2. *Assume Assumptions* 1 *through* 9 *hold, and that there exist* $\varepsilon > 0$

and $K > 1$ *such that if* $|\theta| > K$, *then*

$$\operatorname{tr}(J_\gamma(\theta)M^t) + \tfrac{1}{2}\gamma^t(\theta)\mathscr{L}\gamma(\theta) < -\varepsilon/|\theta|^2 . \tag{2.18}$$

It can be concluded that $\delta(X) = X + \gamma(X)$ is tail minimax.

PROOF. Obvious. □

Corollary 2 will be important in our search for tail minimax estimators. Corollary 1 is quite useful for demonstrating that a proposed estimator cannot indeed be minimax.

For the remainder of this paper, estimators of the following form will be considered:

$$\delta(X) = \left(I - \frac{r(X)B}{X^tCX}\right)X , \tag{2.19}$$

where I is the $p \times p$ identity matrix, B is a nonzero $p \times p$ matrix, C is a $p \times p$ positive definite matrix, and $r: R^p \to R^1$ is a measurable function. This class of estimators is relatively simple and yet includes most interesting minimax estimators so far discovered. Note that for the above estimator,

$$\gamma(X) = -r(X)BX/(X^tCX) .$$

LEMMA 1. *Assume* δ *is of the form* (2.19), *and that* $\nabla r(\theta)$ *exists for* $|\theta| > K$. *Then for* $|\theta| > K$,

$$\operatorname{tr}[J_\gamma(\theta)M^t] + \tfrac{1}{2}\gamma^t(\theta)\mathscr{L}\gamma(\theta)$$
$$= \frac{-\nabla r(\theta)M^tB\theta}{\theta^tC\theta} - \frac{r(\theta)}{\theta^tC\theta}\left[\operatorname{tr}(BM^t) - \frac{2\theta^tCM^tB\theta}{\theta^tC\theta} - \frac{r(\theta)\theta^tB^t\mathscr{L}B\theta}{2\theta^tC\theta}\right].$$

PROOF. Straightforward calculation. □

THEOREM 2. *Assume* L *and* f *satisfy Assumptions* 1 *through* 7. *Assume also that* δ *is of the form* (2.19), *where*

(i) $|\gamma(x)| < \beta < \infty$,
(ii) *there exist* $T > 0$, $\varepsilon_1 > 0$, *and* $\varepsilon_2 > 0$, *such that if* $|x| > T$, *then*
  (a) $r(x)$ *has all second order partial derivatives*,
  (b) $r^{(i)}(x) = o(1)$ *as* $|x| \to \infty$, $1 \leqq i \leqq p$,
  (c) $r^{(i,j)}(x) = o(|x|^{-1})$, $1 \leqq i \leqq p$, $1 \leqq j \leqq p$,
  (d) $\nabla r(x)M^tBx \geqq 0$,
  (e) $\varepsilon_1 \leqq r(x) \leqq (\operatorname{ch}_{\max}[C^{-1}B^t\mathscr{L}B])^{-1}\{2tr(BM^t) - 2\operatorname{ch}_{\max}[M^tB + C^{-1}B^tMC] - \varepsilon_2\}$.

Then δ *is tail minimax.*

PROOF. Corollary 2 will be applied. Note first that $\operatorname{ch}_{\max}[C^{-1}B^t\mathscr{L}B] > 0$ (since C and $\mathscr{L}$ are positive definite and B is nonzero). Straightforward calculation, together with condition (ii)(a), (b), and (c), verifies that γ satisfies Assumption 9. To apply Corollary 2, it remains only to verify (2.18).

Using Lemma 1 and condition (ii)(d), it is clear that if $|\theta| > K > T$, then

$$\text{(2.20)}\qquad \begin{aligned}&\operatorname{tr}(J_\gamma(\theta)M^t) + \tfrac{1}{2}\gamma^t(\theta)\mathscr{L}\gamma(\theta)\\ &\quad< \frac{-r(\theta)}{\theta^t C\theta}\left\{\operatorname{tr}(BM^t) - \frac{2\theta^t CM^t B\theta}{\theta^t C\theta} - \frac{r(\theta)\theta^t B^t\mathscr{L}B\theta}{2\theta^t C\theta}\right\}.\end{aligned}$$

Since C is positive definite, it is well known that

$$\frac{\theta^t B^t\mathscr{L}B\theta}{2\theta^t C\theta} \leqq \tfrac{1}{2}\operatorname{ch}_{\max}[C^{-1}B^t\mathscr{L}B],$$

$$\frac{2\theta^t CM^t B\theta}{\theta^t C\theta} = \frac{\theta^t(CM^tB + B^tMC)\theta}{\theta^t C\theta} \leqq \operatorname{ch}_{\max}[M^tB + C^{-1}B^tMC].$$

Together with (2.20) and the assumption that $r > 0$, these facts imply that

$$\text{(2.21)}\qquad \begin{aligned}&\operatorname{tr}(J_\gamma(\theta)M^t) + \tfrac{1}{2}\gamma^t(\theta)\mathscr{L}\gamma(\theta)\\ &\quad< \frac{-r(\theta)}{\theta^t C\theta}\Big\{\operatorname{tr}(BM^t) - \operatorname{ch}_{\max}[M^tB + C^{-1}B^tMC]\\ &\qquad - \frac{r(\theta)}{2}\operatorname{ch}_{\max}[C^{-1}B^t\mathscr{L}B]\Big\}.\end{aligned}$$

Condition (ii)(e) thus gives

$$\operatorname{tr}(J_\gamma(\theta)M^t) + \tfrac{1}{2}\gamma^t(\theta)\mathscr{L}\gamma(\theta) < -\varepsilon_1\varepsilon_2/(2\theta^t C\theta) \leqq -\varepsilon_1\varepsilon_2\operatorname{ch}_{\min}[C^{-1}]/(2|\theta|^2).$$

Since C is positive definite, $\operatorname{ch}_{\min}[C^{-1}] > 0$, and (2.18) is verified. Hence δ is tail minimax. □

For a given practical situation, the question of how to best choose C and B is very complex, involving not only f and L, but also the available prior information. The answers will of necessity be in the form of rough intuitive guidelines (an exact specification of the prior is assumed to be impossible for otherwise a Bayesian analysis should be run) rather than precise mathematical conditions, and hence seem best contained in a later manuscript.

For illustrative purposes, a particularly simple and attractive tail minimax estimator will be considered. The estimator will be of the form (2.19) with $B = (M^{-1})^t$ and $C = M^{-1}\mathscr{L}(M^{-1})^t$. For $p = 3$, this choice of B and C is the most attractive choice the author has yet encountered. For higher p, however, the estimator is good only if M has no "extreme" characteristic roots. Again, this issue will not be pursued here.

COROLLARY 3. *Assume L and f satisfy Assumptions* 1 *through* 7. *Let*

$$\text{(2.22)}\qquad \delta(X) = \left(I - \frac{r(X)(M^{-1})^t}{X^t M^{-1}\mathscr{L}(M^{-1})^t X}\right)X$$

where

$$\text{(i)}\qquad \frac{|r(x)(M^{-1})^t x|}{x^t M^{-1}\mathscr{L}(M^{-1})^t x} < \beta < \infty,$$

(ii) *there exist* $T > 0$, $\varepsilon_1 > 0$, *and* $\varepsilon_2 > 0$ *such that if* $|x| > T$, *then*

(a) $r(x)$ *has all second order derivatives*,

(b) $r^{(i)}(x) = o(1)$ *as* $|x| \to \infty$, $1 \leqq i \leqq p$,

(c) $r^{(i,j)}(x) = o(|x|^{-1})$, $1 \leqq i \leqq p$, $1 \leqq j \leqq p$,

(d) $r(x)$ *is nondecreasing in* $|x_i|$, $1 \leqq i \leqq p$,

(e) $\varepsilon_1 \leqq r(x) \leqq 2(p-2) - \varepsilon_2$.

Then δ *is tail minimax.*

PROOF. Noting that $r^{(i)}(x)x_i \geqq 0$ by condition (ii)(d), the corollary follows immediately from Theorem 2. □

The question of how to choose the function r arises next. The choice $r(x) \equiv c$ is attractive because of simplicity, but unfortunately the resulting estimator is such that $\gamma(x)$ has a singularity at zero. For quadratic loss, a relatively simple and good method of eliminating such singularities is given in Berger and Bock (1975). Indeed if $L(x) = x^tQx$, then the suggested choice of r is

$$r(x) = \min\{x^t(2Q)(M^{-1})^t x, c\} I_{(0,\infty)}\{x^t(2Q)(M^{-1})^t x\}, \tag{2.23}$$

where $I_{(0,\infty)}(\cdot)$ is the indicator function on $(0, \infty)$. The reader is referred to the above paper for a justification of this choice. The obvious analog of (2.23) for the situation of this paper is

$$r(x) = \min\{x^t \diagup (M^{-1})^t x, c\} I_{(0,\infty)}\{x^t \diagup (M^{-1})^t x\}. \tag{2.24}$$

COROLLARY 4. *Assume L and f satisfy Assumptions* 1 *through* 7. *Assume in addition that* $x^t \diagup (M^{-1})^t x > 0$ *for all* $x \neq 0$. *Let*

$$\delta^c(X) = \left[I - \frac{\min\{X^t \diagup (M^{-1})^t X, c\}}{X^t M^{-1} \diagup (M^{-1})^t X} (M^{-1})^t\right] X, \tag{2.25}$$

where $0 < c < 2(p-2)$. *Then* δ^c *is tail minimax.*

PROOF. It is straightforward to check that condition (i) of Corollary 3 is satisfied for this estimator. Since $x^t \diagup (M^{-1})^t x$ is a positive definite quadratic form, it is clear that there exists $T > 0$ such that if $|x| > T$, then $x^t \diagup (M^{-1})^t x > c$. Hence if $|x| > T$, it follows that $r(x) = c$. Condition (ii) of Corollary 3 is thus trivially satisfied, and the conclusion follows. □

The additional condition, that $x^t \diagup (M^{-1})^t x > 0$, is true in most practical situations. See the discussion of Assumption 4 for three cases where the condition is clearly satisfied.

Note finally, that the estimator δ^c is probably not admissible since it is not even analytic. Nevertheless, comparisons of the risk functions of δ^c and similar admissible estimators have indicated that δ^c is "nearly admissible", in the sense that significant practical improvements upon δ^c seem unlikely to exist.

3. Applications. In this section only the tail minimax estimator δ^c, given in (2.25), will be considered. It will be seen that δ^c offers considerable improvement in risk over δ_0 at $\theta = 0$. It is clear that the estimator could be centered at any "likely" parameter value to take advantage of prior information. For simplicity, all graphs and calculations will just be for δ^c centered at zero.

Before proceeding, the question of choosing c arises. For $p = 3$, an examination of typical risk functions shows that choosing c as large as possible (recall $0 < c < 2(p-2)$) tends to result in the most attractive risks. For $p \geqq 4$ and the problem of estimating a multivariate normal mean under squared error loss, Efron and Morris (1973) suggest $c = p - .66$. This choice has worked well in other situations of our experience and so is the suggested choice of c if $p \geqq 4$.

The numerical results presented in this paper will be for $p = 3$. The difficulty in choosing c for $p = 3$ is that the obviously "largest" value, $c = 2 = 2(p-2)$, does not necessarily give rise to a tail minimax estimator. Indeed if

$$f(x-\theta) = K\exp[-|x-\theta|]\,, \qquad L(x) = |x|^2\,, \tag{3.1}$$

it can be shown that $\Delta_\delta c(\theta) > 0$ for large $|\theta|$ and $c = 2(p-2)$. Another problem with large c, as indicated by the numerical results, is that large values of c are more likely to give rise to estimators which are not strictly minimax. (It can actually be shown that if c is small enough, then the estimator of Corollary 4 is strictly minimax. Such a result would not, of course, provide a practically useful minimax estimator.) In spite of the two problems, the choice $c = 2$ will be considered since it often does give rise to a tail minimax and indeed minimax estimator, and since in any case it presents the "worst case" of tail minimaxity from the point of view of a tail minimax estimator being nonminimax. Numerical results will also be presented for the choice $c = 1$. Thus the estimators examined will be δ^1 and δ^2 when $p = 3$.

A. *Applications to the normal distribution.* Assume that X is a p-variate normal random variable with mean θ and known covariance matrix Σ.

LEMMA 2. *If L is such that Assumptions* 1 *through* 7 *of Section* 2 *are satisfied, then* $M = \mathscr{L}\Sigma$.

PROOF. Note first that $M^t = E_0[X\nabla L(X)]$, where the expectation is taken componentwise. Let A be a $p \times p$ nonsingular matrix such that $(A^{-1})^t\Sigma^{-1}A^{-1} = I$. Define $Y = AX$, $L^*(y) = L(A^{-1}y)$, $l_{ij}^* = E_0 L^{*(i,j)}(Y)$, $m_{ij}^* = E_0[L^{*(i)}(Y)Y_j]$, $\mathscr{L}^*$ as the $p \times p$ matrix with elements l_{ij}^*, and M^* as the $p \times p$ matrix with elements m_{ij}^*. It is easy to check that

$$\begin{aligned} M^{*t} &= E_0[Y\nabla L^*(Y)] = E_0[AX\nabla L(X)A^{-1}] = AM^tA^{-1}\,, \\ \mathscr{L}^* &= (A^{-1})^t\mathscr{L}A^{-1}\,. \end{aligned} \tag{3.2}$$

From the choice of A, it is clear that Y is a p-variate normal random variable with mean $A\theta$ and covariance matrix I. Defining $y^j = (y_1, \cdots, y_{j-1}, y_{j+1}, \cdots, y_p)$, it follows that

$$\begin{aligned} m_{ij}^* &= \textstyle\int_{R^p} L^{*(i)}(y)y_j(2\pi)^{-p/2}\exp[-|y|^2/2]\,dy \\ &= \textstyle\int_{R^{p-1}} (2\pi)^{-p/2}\exp[-|y^j|^2/2]\int_{-\infty}^{\infty} L^{*(i)}(y)y_j\exp[-y_j^2/2]\,dy_j\,dy^j\,. \end{aligned} \tag{3.3}$$

Except at the point $y^j = 0$, $L^{*(i)}(y)$ is everywhere differentiable as a function of y_j. Assumption 5(c) implies that for almost all y^j

$$\lim_{|y_j|\to\infty} L^{*(i)}(y)\exp[-y_j^2/2] = 0\,, \qquad 1 \leqq i \leqq p\,.$$

Using the above two facts, an integration by parts gives that

$$\int_{-\infty}^{\infty} L^{*(i)}(y)y_j \exp[-y_j^2/2]\,dy_j = \int_{-\infty}^{\infty} L^{*(i,j)}(y)\exp[-y_j^2/2]\,dy_j\,,$$

for almost all y^j. Together with (3.2), this gives that

$$m_{ij}^* = \int_{R^p} L^{*(i,j)}(y)(2\pi)^{-p/2}\exp[-|y|^2/2]\,dy = l_{ij}^*\,.$$

Thus $M^* = \mathscr{L}^*$. Using (3.2), it follows that $M = \mathscr{L}A^{-1}(A^{-1})^t = \mathscr{L}\Sigma$. □

By the above lemma, it is clear that (2.25) becomes

$$\delta^c(X) = \left(I - \frac{\min(X^t\Sigma^{-1}X, c)\mathscr{L}^{-1}\Sigma^{-1}}{X^t\Sigma^{-1}\mathscr{L}^{-1}\Sigma^{-1}X}\right)X\,. \tag{3.4}$$

Corollary 4 shows that δ^c is tail minimax.

An interesting feature of this estimator becomes apparent when the case $\mathscr{L} = lI$, $l > 0$, is considered. δ^c then becomes

$$\delta^c(X) = \left(I - \frac{\min(X^t\Sigma^{-1}X, c)\Sigma^{-1}}{X^t\Sigma^{-2}X}\right)X\,, \tag{3.5}$$

which does not formally depend on the loss function. Thus, as long as the problem is symmetric in the sense that $\mathscr{L} = lI$, the actual shape of the loss function does not need to be known in order to determine δ^c. Of course the loss function will play an important role in introducing a possible bias term (a nuisance we avoided by Assumption 2), but the fact that it plays no further role in determining δ^c is extremely attractive from a practical point of view. At the risk of carrying the analogy too far, the following is a possible interpretation that could be given for the nonsymmetric situation. The matrix $\mathscr{L}$ intuitively represents the relative weightings or importance of the various coordinates of θ. From the point of view of tail minimaxity, this is the only feature of the loss that is important, and so the difficult task of accurately specifying the shape of the loss can be avoided.

We finally turn to an investigation of how good the tail minimax estimators actually are for estimating a normal mean. Consider first the case of quadratic loss. Thus assume $L(x) = x^tQx$, where Q is positive definite. Calculation shows that $\mathscr{L} = 2Q$. Hence

$$\delta^c(X) = \left(I - \frac{\min(X^t\Sigma^{-1}X, c)Q^{-1}\Sigma^{-1}}{X^t\Sigma^{-1}Q^{-1}\Sigma^{-1}X}\right)X\,. \tag{3.6}$$

In Berger (1974c) it was shown that δ^c is actually completely minimax for $0 \leqq c \leqq 2(p-2)$. Indeed, under similar conditions, the general class of tail minimax estimators given in Theorem 2 was shown in Berger (1975) to be completely minimax for estimating a normal mean under quadratic loss. These results lend considerable support to the idea that tail minimax estimators behave well.

The risk function of δ^2 (for $p = 3$) was numerically calculated in a variety of situations. The case $\Sigma = I$ and Q diagonal, with diagonal elements q_i, was considered for convenience. As could be expected, the more the q_i varied, the less the improvement in risk obtained (over the risk of δ_0). However, even with $(\max q_i)/(\min q_i) = 10$, δ^2 still had a risk 25% less than that of δ_0 at $\theta = 0$. For the improvement obtained at $q_1 = q_2 = q_3 = 1$, see Figure 3.

Of course, the major purpose of this paper was to provide a means of dealing with nonquadratic loss. Thus we will now evaluate the performance of δ^c for a variety of nonquadratic losses. For simplicity, assume that $\Sigma = I$ and $\mathscr{L} = lI$, where $l > 0$. The tail minimax estimator (3.5) then becomes

$$\delta^c(X) = (1 - c/|X|^2)^+ X. \tag{3.7}$$

The risk functions of δ^1 and δ^2 were evaluated for the following losses:

(i) $L(x) = |x|^a$, $a = \frac{1}{2}$, 1, 2, 3, and 4,
(ii) $L(x) = \sum_{i=1}^p x_i^4$,
(iii) $L(x) = |x|^2/(1 + |x|^2)$.

Note that these include a concave loss ($|x|^{\frac{1}{2}}$) which is "nasty" at zero, and a very reasonable bounded loss. For all the above losses, it is straightforward to verify that the assumptions of Section 2 hold. Note, in particular, that by Lemma 2, $l = l_{11} = m_{11} = E_0[X_1 L^{(1)}(X)] > 0$ (since the losses are increasing in $|X_1|$.)

The tail minimax estimator δ^1 appeared to be completely minimax for all the above losses, with improvements in risk (over that of δ_0) ranging from about 50%, at $|\theta| = 0$, to about 15% at $|\theta| = 2$. The flatter the loss, the less was the observed improvement.

The results for δ^2 were more interesting and so will be considered in greater

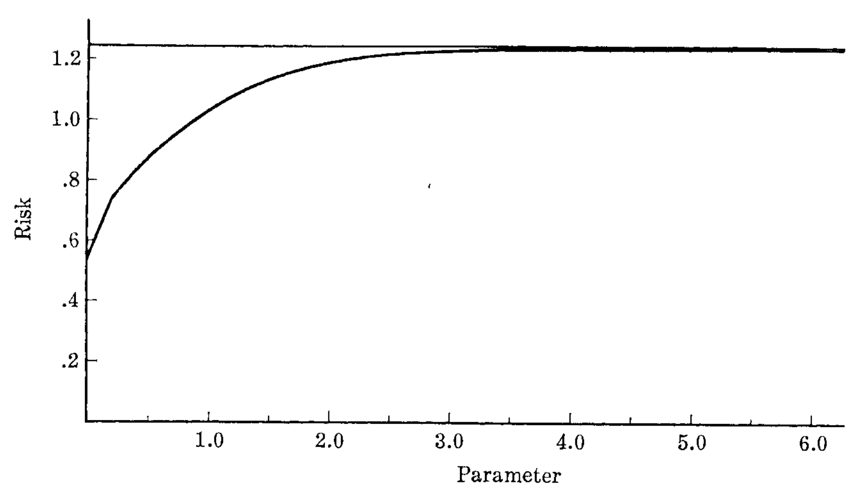

FIG. 1. $L(x) = |x|^{\frac{1}{2}}$.

detail. Figures 1–5 show the risk functions of δ^2 for some of the interesting cases. The scale labeled PARAMETER is $|\theta|$. The horizontal line is the risk of the best invariant estimator, δ_0. The following observations can be made from the graphs:

(i) δ^2 appears to be completely minimax if the loss is squared error or flatter.

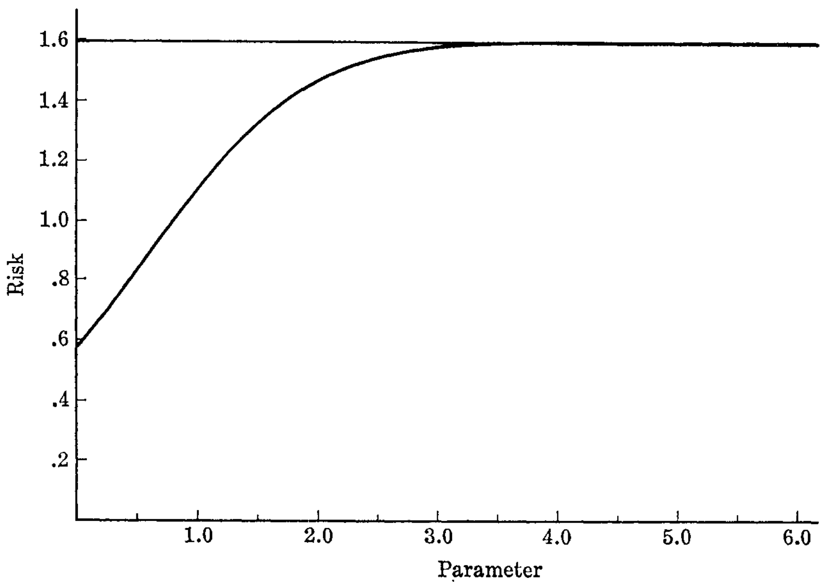

FIG. 2. $L(x) = |x|$.

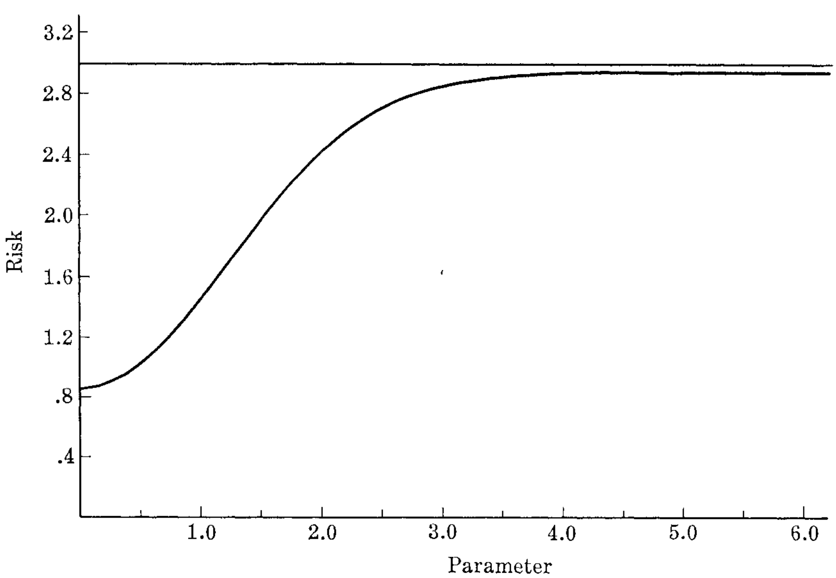

FIG. 3. $L(x) = |x|^2$.

Even when clearly not minimax, the risk function of δ^2 seems preferable to that of δ_0.

(ii) δ^2 gives improvements in risk of about 70% at $|\theta| = 0$. The region of significant improvements extends up to about 3 standard deviations from 0.

(iii) Again, the flatter the loss, the less is the improvement.

In conclusion, tail minimax estimators appear to be quite good for estimating a normal mean. They are very often completely minimax, can offer significant

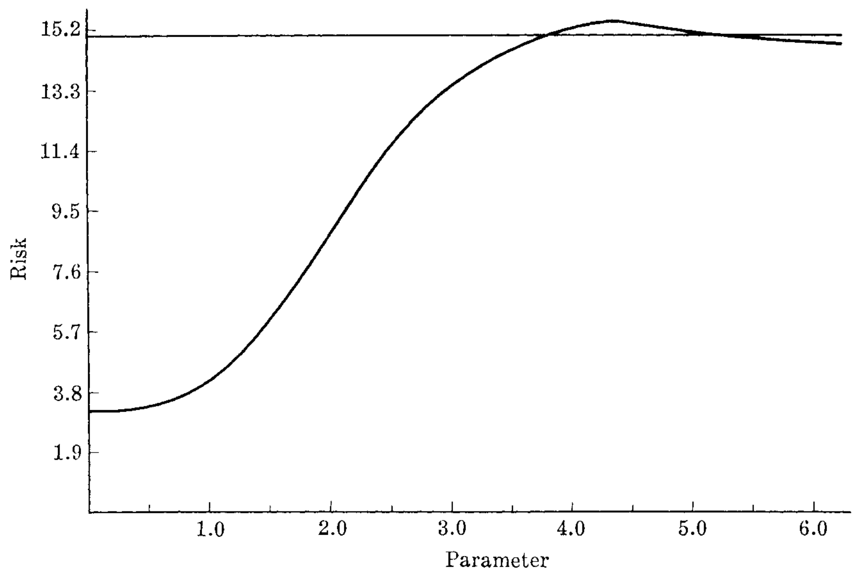

FIG. 4. $L(x) = |x|^4$.

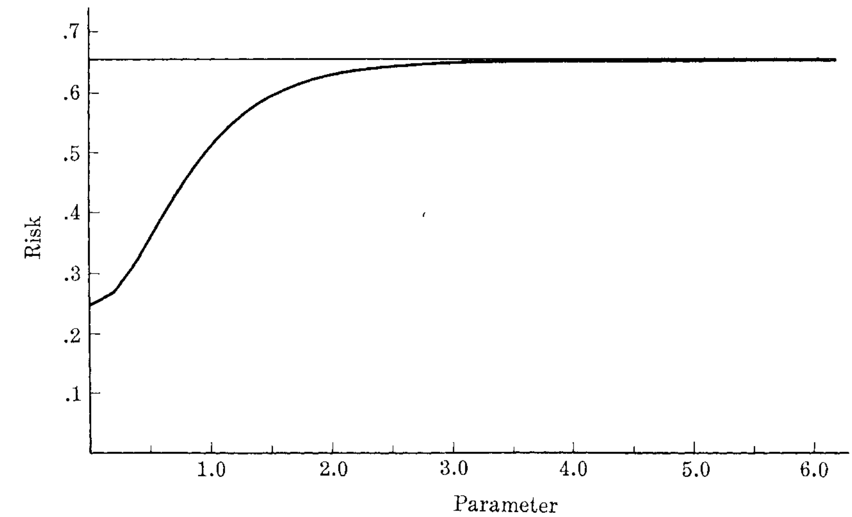

FIG. 5. $L(x) = |x|^2/(1 + |x|^2)$.

improvement in risk over δ_0, and in some sense are robust with respect to the loss function.

B. *Application to nonnormal densities.* For simplicity, only the situation $L(x) = |x|^2$ and $\Sigma = mI$ (where Σ is the covariance matrix of X) will be considered. It is easy to check that $\mathscr{L} = 2I$ and $M = 2mI$ for this situation. The tail minimax estimator (2.25) thus becomes

$$\delta^c(X) = (1 - mc/|X|^2)^+ X \,. \tag{3.8}$$

As mentioned at the beginning of this section, it is not always true that $\delta^{2(p-2)}$ is tail minimax. For estimating a normal mean, outside results indicated that $\delta^{2(p-2)}$ is often tail minimax. Because of the uncertainty for the nonnormal situation, however, we will deal with the more conservative estimator $\delta^{(p-2)}$ in this section.

For wide classes of densities of the form $f(|x - \theta|)$, Strawderman (1974) and Berger (1974d) showed that if an additional restriction is put on c, then δ^c in (3.8) is completely minimax. For example, it was shown that if f is a mixture of p-variate normal densities and if $c \leqq 2/(mE_0|X|^{-2})$, then δ^c is minimax. It is interesting to compare this bound on c, with the tail minimax bound of $2(p-2)$.

(i) If f is normal, then $mE_0|X|^{-2} = (p-2)^{-1}$. Thus the two bounds are the same.

(ii) If $f(|x - \theta|) = K\exp[-|x - \theta|]$, it can be calculated that $mE_0|X|^{-2} = (p+1)(p-1)^{-1}(p-2)^{-1}$. Thus the ratio of the tail minimax bound to the known minimax bound is $(p+1)/(p-1)$. Note further that $\delta^{(p-2)}$ is indeed strictly minimax for this problem if $p \geqq 3$.

(iii) Let $f(|x - \theta|) = K(1 + |x - \theta|^2)^{-a}$, where $a > (p+3)/2$ so that the assumptions of Section 2 are satisfied. It can be calculated that $mE_0|X|^{-2} = (2a - p)(p-2)^{-1}(2a - p - 2)^{-1}$, and hence that the ratio of the tail minimax bound to the known minimax bound is $(2a - p)/(2a - p - 2)$. For large a, the two bounds are nearly the same.

Numerical results will be presented for two commonly occuring densities, the double exponential and the exponential. Assume that the X_i are independent, with densities $f(x_i - \theta_i)$.

Consider first the double exponential density $f(x_i - \theta_i) = \exp[-|x_i - \theta_i|]/2$. It is easy to check that $m = E_0 X_1^2 = 2$. For $p = 3$, Figure 6 gives the risk along the θ_1 coordinate axis, of the estimator $\delta^1(X) = (1 - 2/|X|^2)^+ X$. (Thus PARAMETER is $|\theta_1|$.)

Finally, consider the exponential density

$$f(x_i - \theta_i) = I_{(0,\infty)}(x_i - \theta_i)\exp[-(x_i - \theta_i)] \,.$$

Before proceeding, note that $\delta_0(X) = X - \bar{1}$, where $\bar{1} = (1, 1, \cdots, 1)^t$. Thus the distribution is not properly centered for an application of the theory.

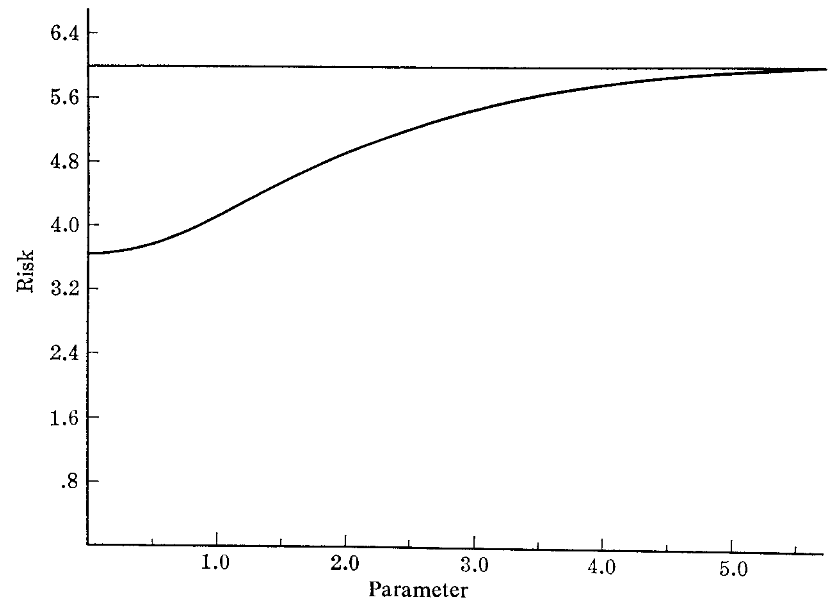

FIG. 6. Double exponential, $L(x) = |x|^2$.

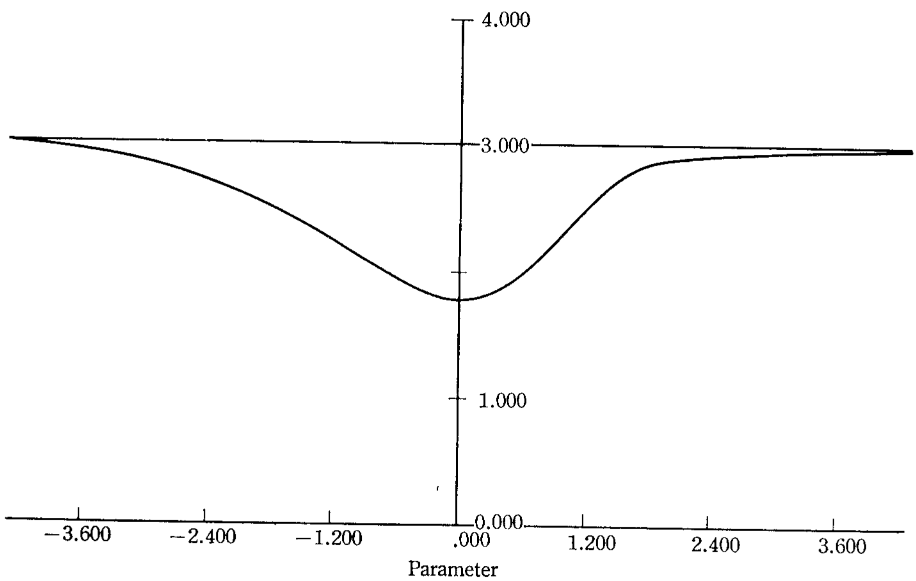

FIG. 7. Exponential, $L(x) = |x|^2$.

Reparametrize by defining $\eta_i = \theta_i + 1$. Then X_i has density $I_{(0,\infty)}(x_i + 1 - \eta_i) \exp[-(x_i + 1 - \eta_i)]$, and it is easy to check that X_i is the best invariant estimator of η_i. Calculation shows that $m = E_0 X_i^2 = 1$. The tail minimax estimator for $\eta = (\eta_1, \cdots, \eta_p)^t$ is thus $\delta^c(X) = (1 - c/|X|^2)^+ X$. The corresponding estimator for θ is $(\delta^c(X) - \bar{1})$. This is a tail minimax estimator centered at

$\theta = \bar{1}$. (See the comments at the beginning of this section.) The corresponding tail minimax estimator centered at zero is $\delta^c(X) = (1 - c/|X - \bar{1}|^2)^+(X - \bar{1})$. Figure 7 gives the risk of δ^1 along the θ_1 coordinate axis.

For both the above nonnormal densities, δ^1 performed quite well.

Acknowledgments. I would like to express my sincere thanks to Professor Roger Farrell and the referee for their many helpful comments and suggestions.

REFERENCES

[1] ALAM, KURSHEED (1973). A family of admissible, minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **1** 517–525.

[2] BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645.

[3] BERGER, J. (1974a). Inadmissibility results for generalized Bayes estimators of coordinates of a location vector. Technical Report 352, Purdue Univ.

[4] BERGER, J. (1974b). Admissibility results for generalized Bayes estimators of coordinates of a location vector. Technical Report 353, Purdue Univ.

[5] BERGER, J. (1974c). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226.

[6] BERGER, J. (1974d). Minimax estimation of location vectors for a wide class of densities. *Ann. Statist.* **3** 1318–1328.

[7] BERGER, J. (1975). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. Technical Report 413, Purdue Univ.

[8] BERGER, J. and BOCK, M. E. (1975). Eliminating singularities of Stein-type estimators of location vectors. Technical Report 414, Purdue Univ.

[9] BHATTACHARYA, P. K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. *Ann. Math. Statist.* **37** 1819–1824.

[10] BOCK, M. E. (1975). Minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **3** 209–218.

[11] BROWN, L. (1966). On the admissibility of estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1136.

[12] BROWN, L. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–904.

[13] BROWN, L. (1973). Estimation with incompletely specified loss functions. *J. Amer. Statist. Assoc.* **70** 417–427.

[14] BROWN, L. (1974). An heuristic method for determining admissibility of estimators—with applications. Technical Report, Rutgers Univ.

[15] EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130.

[16] FERGUSON, T. S. (1967). *Mathematical Statistics—A Decision Theoretic Approach.* Academic Press, New York.

[17] STEIN, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197–206. Univ. of California Press.

[18] STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385–388.

[19] STRAWDERMAN, W. E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *J. Multivariate Anal.* **4** 255–264.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907