The Annals of Statistics
1976, Vol. 4, No. 1, 11-21

FAMILIES OF MINIMAX ESTIMATORS OF THE MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION

BY BRADLEY EFRON AND CARL MORRIS

Stanford University and The Rand Corporation

Ever since Stein's result, that the sample mean vector $\mathbf{X}$ of a $k \geqq 3$ dimensional normal distribution is an inadmissible estimator of its expectation $\boldsymbol{\theta}$, statisticians have searched for uniformly better (minimax) estimators. Unbiased estimators are derived here of the risk of arbitrary orthogonally-invariant and scale-invariant estimators of $\boldsymbol{\theta}$ when the dispersion matrix $\boldsymbol{\Sigma}$ of $\mathbf{X}$ is unknown and must be estimated. Stein obtained this result earlier for known $\boldsymbol{\Sigma}$. Minimax conditions which are weaker than any yet published are derived by finding all estimators whose unbiased estimate of risk is bounded uniformly by k, the risk of $\mathbf{X}$. One sequence of risk functions and risk estimates applies simultaneously to the various assumptions about $\boldsymbol{\Sigma}$, resulting in a unified theory for these situations.

1. Introduction. Consider the problem of estimating the mean vector $\boldsymbol{\theta}$ of a $k \geqq 3$ dimensional multivariate normal distribution on the basis of the data vector $\mathbf{X}$,

$$\mathbf{X} \sim N_k(\boldsymbol{\theta}, D\mathbf{I}) \tag{1.1}$$

where for the moment the covariance matrix of $\mathbf{X}$ is assumed to be proportional to the identity so $\text{Var}(X_i) = D$ is the same and known for each component. For the squared error loss function $L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||^2/D$, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}^{(0)} = \mathbf{X}$ has risk $R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(0)}) = k$. James and Stein [9] showed that the estimator $\hat{\theta}^{(1)} = [1 - (k-2)D/S]\mathbf{X}$, $S \equiv ||\mathbf{X}||^2$ does substantially better. Its risk, which is a function only of $\lambda = ||\boldsymbol{\theta}||^2/2D$, increases from 2 at $\lambda = 0$ to the minimax value k as $\lambda \to \infty$.

A question of considerable theoretical interest is to characterize the class of minimax estimators $\hat{\boldsymbol{\theta}}(\mathbf{X})$, those having $\sup_{\boldsymbol{\theta}} R(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = k$. For orthogonally invariant estimators, those that can be written in the form

$$\hat{\boldsymbol{\theta}} = [1 - (k-2)\tau(F)/F]\mathbf{X}\,, \qquad F \equiv S/D\,, \tag{1.2}$$

Baranchik [2, 3] proved that the conditions

$$0 \leqq \tau(F) \leqq 2 \tag{1.3}$$

and

$$\tau(F) \quad \text{nondecreasing in} \quad F \tag{1.4}$$

are sufficient that $\hat{\boldsymbol{\theta}}$ be minimax.

Received February 1974; revised June 1975.

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

Key words and phrases. Estimation, minimax estimators, risk of invariant estimators, mean of a multivariate normal distribution, Stein's estimator.

Baranchik's condition (1.4) was recently widened somewhat by Alam [1]. A further extension is presented in Theorem 3: rules of the form (1.2) are minimax if they satisfy (1.3) and also

(1.5) $$F^{(k-2)/2}\tau(F)/[2-\tau(F)] \quad \text{is nondecreasing in} \quad F\,.$$

Similar results are shown to hold (Section 3) for more complicated cases when the covariance matrix $\mathbf{\Sigma}$ of $\mathbf{X}$ is unknown and must be estimated from supplementary data. Roughly speaking, the more data available for estimating $\mathbf{\Sigma}$, the weaker the substitute condition for (1.4) need be, (1.5) being the limiting case where $\mathbf{\Sigma}$ is completely known (Theorem 4).

Unlike (1.4), (1.5) allows $\tau(F)$ to decrease with increasing F, though not too quickly. Stein [16] recently has proposed some estimators satisfying (1.5) but not (1.4) (see Example 4, Section 3).

Our results are proved for the cases of known D, or when D must be estimated from an observed chi-square statistic having n degrees of freedom, by finding an unbiased estimator $\hat{R}_n(F)$ of the risk of an arbitrary invariant estimation rule (Theorem 1). This generalizes Stein's [16] unbiased estimator $\hat{R}_\infty(F)$ of the risk when D is known. The minimax proof characterizes all invariant estimators for which $\hat{R}_n(F) \leqq k$ uniformly, resulting in conditions (1.3), (1.5) for D known.

In addition to giving better minimax conditions, Stein's approach has the interesting property that $\hat{R}_n(F)$ is itself an observable statistic. Thus it may be used to estimate the risk of an estimator for a given situation and to suggest the best estimator from a class of estimators.

2. Unbiased estimates of the risk of invariant estimators. Only the canonical sampling situation of Section 1 is considered here. Certain more general cases which reduce to this one are treated in Section 4.

The multivariate distribution (1.1) of $\mathbf{X}$ and loss function $||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||^2/D$ are still assumed with $k \geqq 3$, but if D is unknown an estimate $\hat{D}$ is observed having distribution[1] $D\chi_n^2/(n+2)$ independent of $\mathbf{X}$. If D is known, notation is unified by setting $n = \infty$ and $\hat{D} = D$.

This problem of estimating $\boldsymbol{\theta}$ is invariant under orthogonal transformations and, if D is unknown, also under scale transformations: $\mathbf{X} \to \alpha\mathbf{\Gamma}\mathbf{X}$, $\boldsymbol{\theta} \to \alpha\mathbf{\Gamma}\boldsymbol{\theta}$, $\hat{D} \to \alpha^2\hat{D}$, $D \to \alpha^2 D$, $\mathbf{\Gamma}$ a $k \times k$ orthogonal matrix, $\alpha > 0$ a scalar. The class of estimation rules invariant under these transformations can be written in the form [14]

(2.1) $$\hat{\boldsymbol{\theta}} = (1-(k-2)\tau(F)/F)\mathbf{X}\,, \qquad F \equiv S/\hat{D}\,, \qquad S \equiv ||\mathbf{X}||^2$$

with $\tau(F)$ any real-valued function on $(0, \infty)$. The risk of these rules depends on $\boldsymbol{\theta}$ and D through the scalar parameter $\lambda \equiv ||\boldsymbol{\theta}||^2/2D$ only [9]. We denote the risk by $R_n(\lambda)$ or, if D is known, by $R_\infty(\lambda)$.

[1] The multiple $1/(n+2)$ which makes $\hat{D}$ biased is more convenient here than the unbiased choice $1/n$.

The conditions given in Section 3 for (2.1) to be minimax depend on the following estimate of $R_n(\lambda)$.

THEOREM 1. *Suppose in* (2.1) *that* $\tau(F)$ *is absolutely continuous with derivative* $\tau'(F)$. *Denote* $c_n \equiv (k-2)/(n+2)$. *If the risk* $R_n(\lambda)$ *of* (2.1) *is finite and if the expectation of each term in* (2.2) *exists, then a unique unbiased estimator of* $R_n(\lambda)$ *based on* F *exists and is*

$$\hat{R}_n(F) \equiv k - (k-2)\left[\frac{k-2}{F}\tau(F)(2-\tau(F)) + 4\tau'(F)(1+c_n\tau(F))\right]. \tag{2.2}$$

REMARK 1. If D is known, set $n = \infty$ and $c_n = 0$ in (2.2) for the estimator $\hat{R}_\infty(F)$ of the risk $R_\infty(\lambda)$.

REMARK 2. The distribution of $F = S/\hat{D}$ is a multiple of the noncentral F distribution with (k, n) degrees of freedom. The density of F is

$$f_{n,\lambda}(F) = \sum_{j=0}^{\infty} \frac{\lambda^j \exp(-\lambda)}{j!}\left\{\frac{1}{(n+2)\beta(j+k/2, n/2)}\frac{[F/(n+2)]^{j-1+k/2}}{[1+F/(n+2)]^{j+k/2+n/2}}\right\}. \tag{2.3}$$

If D is known then F has the noncentral chi-square distribution with k degrees of freedom, the density being

$$f_{\infty,\lambda}(F) \equiv \sum_{j=0}^{\infty} \frac{\lambda^j \exp(-\lambda)}{j!}\,\frac{(F/2)^{j-1+k/2}\exp(-F/2)}{2\Gamma(j+k/2)}. \tag{2.4}$$

REMARK 3. $\hat{R}_n(F)$ may have finite expectation even though some terms in (2.2) do not. For example, if $n < \infty$ and $\tau(F) \equiv 2 + F^{(n+2)/4}$ then $\hat{R}_n(F) = k - (k-2)(n+2)F^{(n-2)/4}$ has finite expectation. But $E\hat{R}_n(F) \neq R_n(\lambda)$ even though $R_n(\lambda) < \infty$. Theorem 1 does not apply because $E\tau^2(F)/F = \infty$.

Computation of (2.2) is illustrated in examples 1, 2 for the class of estimators suggested by James and Stein.

EXAMPLE 1. Suppose $\tau(F) = t$ is constant. Then

$$\hat{R}_n(F) = k - (k-2)^2 t(2-t)/F. \tag{2.5}$$

Expression (2.5) is minimized uniformly at $t = 1$, yielding the crude James–Stein estimator.

EXAMPLE 2. Using $\tau(F) = t$ but taking the positive part of $(1 - (k-2)t/F)$ uniformly decreases the risk (cf. Remark 6). This yields $\tau(F) = \min(t, F/(k-2))$. This estimator with $t = 1$ is the James–Stein estimator, but we prefer the less conservative choice $t = t^*$ which is based on the foreknowledge that the positive part will be taken[2] [7, page 124]

$$t^* = \min\left(2, \frac{1 + 1.34/(k-2)}{1 - 2.66/(n+2)}\right). \tag{2.6}$$

[2] More precisely, we prefer to choose $t \leqq 2$ so that $tn(k-2)/k(n+2)$ is the median of Snedecor's $F_{k,n}$ distribution. Formula (2.6) approximates this quite well.

The unbiased estimator of the risk of these rules is

$$\hat{R}_n(F) = k - (k-2)^2 t(2-t)/F \qquad \text{if} \quad F \geqq (k-2)t \tag{2.7}$$
$$= \frac{n-2}{n+2} F - k \qquad \text{if} \quad F < (k-2)t$$

which exceeds (2.5) for F near zero. This can be a poor estimate of $R_n(\lambda)$, being discontinuous at $(k-2)t$ and negative for $F < (k-2)t$ if $n \leqq 2$ or if $n \geqq 3$ and $F < k(n+2)/(n-2)$.

REMARK 4. Even if $\tau(F)$ has isolated simple discontinuities, $R_n(\lambda) = E\hat{R}_n(F)$ holds provided the [a.e.] derivatives of τ are used, provided τ' is interpreted as a delta function at points of discontinuity, and provided τ is given its average value across the discontinuity. This is illustrated by the following example.

EXAMPLE 3. Define $B(F) = (k-2)\tau(F)/F$, so (2.2) can be written

$$\hat{R}_n(F) = k - 2kB(F) - 4FB'(F) + FB^2(F) \tag{2.8}$$
$$- \frac{4}{n+2} FB(F)(B(F) + FB'(F)) \,.$$

The discontinuous function $B(F) \equiv I_{[0,c]}(F)$, the indicator function on $[0, c]$, defines an estimate $(1 - B(F))\mathbf{X}$ which corresponds to estimation following a preliminary test that $\boldsymbol{\theta} = \mathbf{0}$. Note that $B^2(F) = B(F)$ and $B'(F) = 0$ [a.e.]. At the point of discontinuity c, symbolically define $B'(c) = [B(c^+) - B(c^-)]\cdot\delta_c$ with δ_c the delta function at c, so $EB'(c)$ is replaced in (2.8) by $[B(c^+) - B(c^-)]\cdot f_{n,\lambda}(c) = -f_{n,\lambda}(c)$. Also redefine[3] $B(c) = [B(c^+) + B(c^-)]/2 = .5$. After slight simplification, (2.8) yields

$$R_n(\lambda) = k + 2cf_{n,\lambda}(c)(2 + c/(n+2)) + \int_0^c \left(\frac{n-2}{n+2} F - 2k\right) f_{n,\lambda}(F)\, dF \,. \tag{2.9}$$

Since $f_{n,\lambda}(c)$ depends on λ, (2.9) does not provide an observable unbiased estimator of $R_n(\lambda)$.

REMARK 5. There are other useful expressions for $R_n(\lambda)$ besides (2.2) which cannot be used to estimate $R_n(\lambda)$ because they involve λ. For example:

$$R_n(\lambda) \equiv E\|(1 - B(F))\mathbf{X} - \boldsymbol{\theta}\|^2/D = \sum_{j=0}^{\infty} \frac{\lambda^j \exp(-\lambda)}{j!} R_{n,j} \,, \tag{2.10}$$

$$R_{n,j} \equiv (n + k + 2j)E_j \frac{F}{n+2+F}(1 - B(F))^2 \tag{2.11}$$
$$- 4jE_j(1 - B(F)) + 2j \,.$$

The symbol E_j in (2.11) indicates that the integral is with F distributed conditionally as the density inside the braces of (2.3). This expression is similar to those appearing in [3, 13, 15, 17], and follows from [13], (3.3)—(3.4). It is useful for proving the following result.

---

[3] This is the only definition of $B(c)$ that will give the correct result when $n < \infty$; any definition suffices for $n = \infty$.

REMARK 6. If $B(F) > 1$ ever, then the "positive part rule" with

$$B^+(F) \equiv \min(1, B(F)) \tag{2.12}$$

reduces the risk (2.10) uniformly since both terms of (2.11) involving $1 - B(F)$ are thereby diminished. A similar proof of this fact was given in [15] for the case D known.

REMARK 7. The $R_\infty(\lambda)$ notation we have used is justified below by showing that $R_\infty(\lambda)$ is the limit of $R_n(\lambda)$. Suppose constants $0 \leqq a < (k-2)/4$, $1 \leqq b < \infty$, $c < \infty$ exist such that

$$|\tau(F)| \leqq c(F^{-a} + F^b) . \tag{2.13}$$

Suppose $R_{n_0}(\lambda_0)$ exists for some n_0, $\lambda_0 \neq 0$. Then $R_n(\lambda)$ exists for all $\lambda \geqq 0$ for all $n \geqq n_0$. For each λ, $\lim_{n\to\infty} R_n(\lambda) = R_\infty(\lambda)$. This is proved at the end of the section.

The condition (2.13) is chosen because it includes the interesting functions $\tau(F)$ we know of and because it makes the result easy to verify. We conjecture that $R_n(\lambda) \to R_\infty(\lambda)$ with no restriction on $\tau(F)$ other than the existence of some $n_0 < \infty$ such that $R_{n_0}(\lambda) < \infty$.

PROOF OF THEOREM 1. Writing $B(F) = (k-2)\tau(F)/F$,

$$\begin{aligned} R_n(\lambda) &= E\|\mathbf{X} - \boldsymbol{\theta} - B(F)\mathbf{X}\|^2/D \\ &= k - 2E \sum B(F)X_i(X_i - \theta_i)/D + EB^2(F)S/D . \end{aligned} \tag{2.14}$$

Stein [16] observed that for any absolutely continuous function $h(x)$ with Lebesgue measurable derivative $h'(x)$ satisfying $E|h'(X_i)| < \infty$,

$$E(X_i - \theta_i)h(X_i) = DEh'(X_i) . \tag{2.15}$$

This formula is proved by integration by parts.
Using (2.15) with $h(X_i) = X_i B(F)$ and using

$$\frac{\partial F}{\partial X_i} = \frac{1}{\hat{D}} \frac{\partial S}{\partial X_i} = \frac{2X_i}{\hat{D}}$$

permits simplification of (2.14) to

$$R_n(\lambda) = k - 2kEB(F) - 4EB'(F)F + EB^2(F)S/D . \tag{2.16}$$

A formula like (2.15) for the chi-square distribution, provable by integration by parts, is

$$E(W - nb)h(W) = 2bEWh'(W) \tag{2.17}$$

for $W \sim b\chi_n^2$ and h such that all expectations in (2.17) exist. We take $W = \hat{D}$ in (2.17) so $b = D/(n+2)$ and apply (2.17) to $h(\hat{D}) = B^2(S/\hat{D})S/\hat{D}$ to obtain an expression for the last term in (2.16).

$$\begin{aligned} EB^2(F)S/D &= \frac{n}{n+2} EB^2(F)F + \frac{2}{n+2} E\hat{D} \frac{\partial h}{\partial F} \frac{\partial F}{\partial \hat{D}} \\ &= \frac{n-2}{n+2} EFB^2(F) + \frac{4}{n+2} EF^2B(F)B'(F) . \end{aligned} \tag{2.18}$$

Insertion of (2.18) into (2.16) yields (2.8). Substituting $B(F) = (k-2)\tau(F)/F$ into (2.8) gives $R_n(\lambda) = E\hat{R}_n(F)$ with $\hat{R}_n(F)$ defined by (2.2). This unbiased estimator is unique because the noncentral F distribution is complete.

Proof of Remark 7. Let δ satisfy $0 < \delta < (k-2-4a)/k$. From the moment convergence theorem ([11], page 184) it suffices to show that $\|\mathbf{X} - \boldsymbol{\theta} - [(k-2)\tau(F)/F]\mathbf{X}\|^{2+2\delta}$ has expectation bounded independently of n and since $E\|\mathbf{X} - \boldsymbol{\theta}\|^{2+2\delta} < \infty$ is independent of n we need only consider $E\{\tau^2(F)S/F^2\}^{1+\delta}$. Then

$$(2.19) \qquad \tau^2(F) \leqq F^{2b} = S^{2b}/\hat{D}^{2b} \qquad \text{so}$$
$$E\{\tau^2(F)S/F^2\}^{1+\delta} \leqq ES^{(2b-1)(1+\delta)}E\hat{D}^{(2-2b)(1+\delta)}.$$

If $b \geqq 1$ then $ES^{(2b-1)(1+\delta)}$ exists but the exponent of $\hat{D}$ in (2.19) is negative and $E\hat{D}^{(2-2b)(1+\delta)}$ will not exist until $n/2$ exceeds $(2b-2)(1+\delta)$ and then will exist for all larger n with expectation bounded by a constant independent of n. Setting $b = -a$ in (2.19), only the term $ES^{-(2a+1)(1+\delta)}$ can present a problem. But the conditions $0 \leqq a < (k-2)/4$ and $0 < \delta < (k-2-4a)/k$ guarantee that $(2a+1)(1+\delta) < k/2$ and $ES^{\varepsilon-k/2}$ exists for all $\varepsilon > 0$ and is independent of n. This completes the proof.

3. A more general minimax condition. Baranchik [2, 3] has shown that conditions (1.3), (1.4) lead to minimax estimators in the situation of Section 2. This is obvious from inspection of (2.2), for then $\hat{R}_n(F) \leqq k$ for all F, requiring its expectation $R_n(\lambda) \leqq k$. Increasing $\tau(F)$ is clearly not necessary from inspection of (2.2). The class of all functions τ making $\hat{R}_n(F) \leqq k$ uniformly in F is characterized in Theorem 2.

Theorem 2. *With $n \geqq 1$, $k \geqq 3$, assume that $\tau(F)$ is absolutely continuous and that $\hat{R}_n(F)$ exists*[4] *as an unbiased estimator of $R_n(\lambda)$. Necessary and sufficient conditions that*

$$(3.1) \qquad \hat{R}_n(F) \leqq k \qquad \text{for all} \quad F \geqq 0$$

are that

$$(3.2) \qquad 0 \leqq \tau(F) \leqq 2 \qquad \text{for all} \quad F,$$

that for all F with $\tau(F) < 2$

$$(3.3) \qquad \psi_n(F) \equiv F^{(k-2)/2}\tau(F)/(2-\tau(F))^{1+2c_n}$$

is nondecreasing, and that if F_1 exists such that $\tau(F_1) = 2$ then

$$(3.4) \qquad \tau(F) = 2 \qquad \text{for all} \quad F \geqq F_1.$$

Proof. Conditions (3.2), (3.3), (3.4) are sufficient, for with $l \equiv (k-2)/2$ we may write

$$(3.5) \qquad \hat{R}_n(F) = k - 4lF^{-l}(2-\tau(F))^{2+2c_n}\psi_n'(F)$$

when $\tau(F) < 2$ and $\hat{R}_n(F) = k$ when $\tau(F) = 2$.

---

[4] The reader is reminded of Remark 3 which includes an example for which Theorem 2 fails to hold.

To prove necessity of these conditions, suppose $\hat{R}_n(F) \leqq k$ for all $F \geqq 0$. If there exists F_0 such that $\tau(F_0) < 0$ then $\psi_n(F_0) < 0$. Note that $\psi_n(0) = \lim_{F\to 0} \psi_n(F) = 0$ whether $\tau(0) \equiv \lim_{F\to 0} \tau(F)$ is finite or not. Then there exists $0 < F < F_0$ such that $\psi_n'(F) < 0$ and $\hat{R}_n(F) > k$ from (3.5). Hence $\tau(F) \geqq 0$ for all F. If there exists F_2 such that $\tau(F_2) > 2$ then $\tau'(F_2) > 0$ from (2.2) and therefore $\tau(F)$ increases monotonically to $\tau(\infty) > 2$. Thus $\lim_{\lambda\to\infty} R_n(\lambda) > k$ ([7], page 121) implying $\hat{R}_n(F) > k$ for large F when $\hat{R}_n(F)$ exists. Condition (3.2) is established. Condition (3.3) now follows from (3.5). Condition (3.4) must hold because $\tau'(F) \geqq 0$ for all F such that $\tau(F) = 2$, from (2.2). But if $\tau'(F) > 0$ for some F, then there exists $F_2 > F$ such that $\tau(F_2) > 2$, violating (3.2). The proof is complete.

If $\tau(F)$ is not absolutely continuous then $\hat{R}_n(F)$ does not exist, but conditions (3.2)—(3.4) still are sufficient that the estimator corresponding to $\tau(F)$ is minimax.

THEOREM 3. *The estimator* $(1 - (k-2)\tau(F)/F)\mathbf{X}$ *is minimax for* $k \geqq 3$, $n \geqq 1$ *provided* (3.2), (3.3), *and* (3.4) *hold.*

PROOF. Since $\psi_n(F)$ is nondecreasing on the interval with $0 \leqq \tau(F) \leqq 2$, let $\psi_n(F) = \lim_{F\to\infty} \psi_{n,m}(F)$ with each function $\psi_{n,m}$ being nondecreasing and absolutely continuous. Define absolutely continuous functions $0 \leqq \tau_m(F) < 2$ by

$$\psi_{n,m}(F) = F^{(k-2)/2}\tau_m(F)/(2 - \tau_m(F))^{1+2c_n}\,. \tag{3.6}$$

Since τ_m satisfies (3.2)—(3.4), it yields a minimax estimator with risk $R_{n,m}(\lambda) \leqq k$ and $\tau_m \to \tau$ as $m \to \infty$. The dominated convergence theorem is easily applied because the $\{\tau_m\}$ are uniformly bounded, so as $m \to \infty$, $R_{n,m}(\lambda) \to R_n(\lambda) \leqq k$. Since points F where $\tau(F) = 2$ present no difficulty, the theorem is proved.

REMARK 8. Alam [1] determined minimax conditions (1.3), (1.4) for the case $n = \infty$ (D known) which permit τ to decrease, his condition being that there exists $0 \leqq t < l \equiv (k-2)/2$ such that

$$0 \leqq \tau(F) < 2 - 2t/l \qquad \text{for all } F \tag{3.7}$$

$$F^t\tau(F) \quad \text{is nondecreasing.} \tag{3.8}$$

Theorem 3 is more general. This follows because ψ_∞ increases at any point of increase of τ, and at a point F where τ decreases, letting $\varepsilon \equiv l/t - 1 > 0$,

$$\psi_\infty(F) = \{F^t\tau(F)\}^{l/t}/\{\tau^\varepsilon(F)(2 - \tau(F))\} \tag{3.9}$$

still increases because $\tau^\varepsilon(2-\tau)$ is an increasing function of τ on $0 \leqq \tau \leqq 2\varepsilon/(1+\varepsilon) = 2 - 2t/l$.

Interesting minimax rules exist with τ sometimes decreasing although they have received little attention, perhaps because they could not be proved minimax easily. Several examples follow.

EXAMPLE 4. The James–Stein rules $\tau(F) \equiv \min(t, F/(k-2))$, $1 \leqq t \leqq 2$ are inadmissible, but can be dominated only by rules with $\tau(F)$ strictly decreasing at some points ([7], page 123). Stein [16] proposes the following class of estimators which may have members dominating the James-Stein rule with $t = 1$. With $n = \infty$ they take the form

$$\tau_d(F) = 1 + \frac{F/(k-2) - d - 1}{1 + d\varphi_k(F)}, \tag{3.10}$$

defining

$$\varphi_k(F) \equiv \Gamma(k/2) \textstyle\sum_{j=0}^{\infty} (F/2)^j/\Gamma(j + k/2) .$$

These rules are generated from the improper prior distributions on $\boldsymbol{\theta}$ which sets $\boldsymbol{\theta} = \mathbf{0}$ with probability $1 - p$ and with probability p requires $\boldsymbol{\theta}$ to have density $||\boldsymbol{\theta}||^{2-k}$ with respect to Lebesgue measure on k dimensional space. The constant d increases strictly from 0 at $p = 0$ to ∞ at $p = 1$ in which case $\tau_\infty(F)$ is interpreted as $1 - 1/\varphi_k(F)$. If $0 < d < \infty$ $(0 < p < 1)$ then $\tau_d(F)$ increases from zero to a value exceeding unity and decreases thereafter to unity.

Numerical study of the case $k = 4$ has revealed that the hypotheses of Theorem 2 are satisfied if and only if $d \geqq .542$, all smaller values of d leading to decreasing points of $\psi_\infty(F)$. While (3.10) cannot be minimax if d is too close to zero, there are values of $d < .542$ for which (3.10) is minimax, but the estimates of risk $\hat{R}_\infty(F)$ of these rules must exceed k at some points.

EXAMPLE 5. Other minimax estimators with $\tau(F)$ sometimes decreasing are Bayes estimators derived from priors suggested by Brown ([4], 16, Section 4) and by the authors ([7], page 125).

In addition to providing a condition that specific rules are minimax, Theorems 1 and 2 offer insight about the class of minimax rules with $\hat{R}_n(F) \leqq k$. This is the subject of the next two remarks and Theorem 4.

REMARK 9. If τ reaches the value $\tau(F_1) = 2$, τ must be constant at that value thereafter. Furthermore, having left zero, τ can never return (otherwise ψ_n would decrease from a positive value to zero). Stated differently, if $\tau(F_0) = 0$ then $\tau(0) = 0$ and $\tau(F) = 0$ at all points less than F_0. The rate of decrease of τ is also limited, according to

$$\tau'(F) \geqq -\frac{k-2}{4F}\tau(F)(2 - \tau(F))/(1 + c_n\tau(F)), \tag{3.11}$$

this being the condition that $\hat{R}_n(F) \leqq k$. Every point of discontinuity of τ must be a point of increase if Theorem 3 is to apply, for otherwise $\psi_n(F)$ would decrease at the discontinuity.

REMARK 10. Certain estimators, even if minimax, may be improved easily. Assume first that $n = \infty$. Estimators with $\tau(F)$ nonincreasing everywhere are uniformly dominated by the crude James-Stein rule, which has $\tau_1(F) = 1$, and also by the rule with $\tau_2(F) = 2 - \tau(F)$. Regardless of monotonicity, any estimator with $\tau(F) \leqq 1 - \varepsilon$ for all F, some $\varepsilon > 0$ fixed, is dominated by $\tau_3(F) = \tau(F) + \varepsilon$; or if $\tau(F) \geqq 1 + \varepsilon$ for all F then $\tau_4(F) = \tau(F) - \varepsilon$ is uniformly better.

Of course, from Remark 6, if $\tau(F) > F/(k-2)$ ever then it can be uniformly improved by $\tau_5(F) = \min(\tau(F), (k-2)/F)$.

When $n < \infty$, we must assume also that $\tau(F) \geqq -1/c_n$ for $\tau_1(F)$ to dominate; that $\tau(F) \leqq 2 + 1/c_n$ for $\tau_2(F)$ to dominate; and that $\tau'(F) \geqq 0$, for $\tau_3(F)$ to dominate. The statement about $\tau_4(F)$ holds for $n < \infty$ if $\tau'(F) \leqq 0$, but this is an uninteresting condition. These assertions all are substantiated easily by inspecting $\hat{R}_n(F)$. Better conditions for $n < \infty$ can be given with slightly more effort, e.g., $\tau'(F) \leqq (k-2)\varepsilon/4c_n F$ is sufficient for $\tau_4(F)$ to dominate when $\tau(F) \geqq 1 + \varepsilon$.

The final theorem shows that more rules satisfy the minimax criteria of Theorems 2 and 3 as n increases.

THEOREM 4. *If* $0 \leqq \tau(F) \leqq 2$ *is given such that* $\psi_{n_0}(F)$ *is nondecreasing then* $\psi_n(F)$ *is nondecreasing for all* $n \geqq n_0$.

PROOF. Write

$$\psi_n(F) = \psi_{n_0}(F)(2 - \tau(F))^{2(c_{n_0} - c_n)} \tag{3.12}$$

with $c_{n_0} - c_n > 0$, so $\psi_n(F)$ is the product of increasing functions at any point of decrease of $\tau(F)$. At a point of increase of τ, ψ_n increases by its definition.

4. Risks for the case Σ unknown and other noncanonical situations. So far the covariance matrix $\mathbf{\Sigma}$ of $\mathbf{X}$ has been taken to be proportional to the identity. All of the results of Sections 2, 3 are shown here to apply to the more general assumptions about $\mathbf{\Sigma}$ described below because no new risk functions are encountered, provided the loss is redefined as the invariant loss

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{\Sigma}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{4.1}$$

and n and F are redefined properly. Three cases are considered.

(i) $\mathbf{\Sigma} > 0$ of general form and known.

Set $F = \mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X}$, $\lambda = \boldsymbol{\theta}'\mathbf{\Sigma}^{-1}\boldsymbol{\theta}/2$, $n = \infty$, $D = 1$.

(ii) $\mathbf{\Sigma} = D\mathbf{G}$, $\mathbf{G}$ known, D an unknown constant, $\hat{D} \sim D\chi_n^2/(n+2)$, independent of $\mathbf{X}$.

Set $F = \mathbf{X}'\mathbf{G}^{-1}\mathbf{X}/\hat{D}$, $\lambda = \boldsymbol{\theta}'G^{-1}\boldsymbol{\theta}/2D$, n and D as given.

(iii) $\mathbf{\Sigma}$ completely unknown, $\mathbf{W} \sim$ Wishart $(\mathbf{\Sigma}, k, n^*)$, independent of $\mathbf{X}$.

Set $F = (n^* - k + 3)\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}$, $\lambda = \boldsymbol{\theta}'\mathbf{\Sigma}^{-1}\boldsymbol{\theta}/2$, $n = n^* - k + 1$, $D = 1$.

Note that (ii) arises in multiple regression problems, $\mathbf{X}$ then being the usual Gauss–Markov estimator [12]. Cases (i) and (ii) also include the situation with $\mathbf{\Sigma}$ diagonal, having unequal diagonal elements. Although they are minimax, we think that these estimators are less appropriate for most applications than rules suggested specifically for the unequal variance situation, such as those of [6, 8].

In each case invariant estimators have the same form as estimators considered in the preceding sections,

$$\hat{\boldsymbol{\theta}} = (1 - (k-2)\tau(F)/F)\mathbf{X} \tag{4.2}$$

with F defined in (i), (ii), (iii). The risk $R_n(\cdot)$ of (4.2) is exactly the same function as that defined in Sections 2, 3 with $\boldsymbol{\Sigma}$ proportional to the identity. The unbiased estimate of $R_n(\lambda)$ therefore is given for these cases by $\hat{R}_n(F)$ (2.2). Consequently, the minimax theorems of Section 3 also apply to these covariance situations.

These assertions are easily established for cases (i), (ii) by making the usual transformations $\tilde{\mathbf{X}} = \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{X}$ and $\tilde{\mathbf{X}} = \mathbf{G}^{-\frac{1}{2}}\mathbf{X}$ to reduce (i), (ii) to the situation of Section 2. To deal with case (iii), it can be shown with $n = n^* - k + 1$ that $(n+2)\hat{D} \equiv \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}/\mathbf{X}'\mathbf{W}^{-1}\mathbf{X} \sim \chi_n^2$ independently of $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$. But then $F = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}/\hat{D}$, a situation identical to case (ii) with $D = 1$, $\mathbf{G} = \boldsymbol{\Sigma}$.

REMARK 11. The minimax theorems of Section 3 applied to case (iii) generalize the theorem of Lin–Tsai [10].

Until now, "shrinking" of estimators like (4.2) has been toward the origin $\boldsymbol{\theta} = \mathbf{0}$, but it is well-known that shrinking may be toward any linear subspace or linear variety ([13], Section 4, [6], Section 7). When this is done, the same class of risk functions $R_n(\lambda)$ and unbiased estimates $\hat{R}_n(F)$ arise after redefinition of F and k.

EXAMPLE 6. Let $\boldsymbol{\Sigma} = \hat{D}\mathbf{I}$ and $\hat{D} \sim D\chi_n^2/(n+2)$. Suppose a d-dimensional ($d \leqq k-3$) linear variety is defined by a $k \times k$ orthogonal projection matrix $\mathbf{P}$ and a shift $\boldsymbol{\theta}^0$, each element of the variety being of the form $\mathbf{P}\boldsymbol{\eta} + \boldsymbol{\theta}^0$. Shrinking toward this variety is accomplished by decomposing $\mathbf{X} = (\mathbf{I}-\mathbf{P})(\mathbf{X}-\boldsymbol{\theta}^0) + [\mathbf{P}(\mathbf{X}-\boldsymbol{\theta}^0)+\boldsymbol{\theta}^0]$ and $\boldsymbol{\theta} = (\mathbf{I}-\mathbf{P})(\boldsymbol{\theta}-\boldsymbol{\theta}^0) + [\mathbf{P}(\boldsymbol{\theta}-\boldsymbol{\theta}^0)+\boldsymbol{\theta}^0]$ and applying (4.2) to $(\mathbf{I}-\mathbf{P})(\mathbf{X}-\boldsymbol{\theta}^0)$ only to estimate its mean $(\mathbf{I}-\mathbf{P})(\mathbf{X}-\boldsymbol{\theta}^0)$. With $F \equiv (\mathbf{X}-\boldsymbol{\theta}^0)'(\mathbf{I}-\mathbf{P})(\mathbf{X}-\boldsymbol{\theta}^0)/\hat{D}$, the resulting estimate is

$$\hat{\boldsymbol{\theta}} = [(1-(k-2)\tau(F)/F)(\mathbf{I}-\mathbf{P})(\mathbf{X}-\boldsymbol{\theta}^0)] + [\mathbf{P}(\mathbf{X}-\boldsymbol{\theta}^0)+\boldsymbol{\theta}^0]\,. \tag{4.3}$$

The risk of $\hat{\boldsymbol{\theta}}$ is $R_n(\lambda) + d$ with $\lambda \equiv (\boldsymbol{\theta}-\boldsymbol{\theta}^0)'(\mathbf{I}-\mathbf{P})(\boldsymbol{\theta}-\boldsymbol{\theta}^0)/2$ and $R_n(\lambda)$ computed as in Section 2, except k is reduced to $k-d$. The unbiased estimate of the risk of $\hat{\boldsymbol{\theta}}$ is therefore

$$\hat{R}_n(F) + d \tag{4.4}$$

and $\hat{\boldsymbol{\theta}}$ is minimax if τ satisfies the conditions of Theorem 3, k being replaced there by $k-d$.

REFERENCES

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

[2] BARANCHIK, A. J. (1964). Multiple regressions and estimation of the mean of a multivariate normal distribution. Technical Report No. 51, Department of Statistics, Stanford Univ.

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

[4] BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903.

[5] EFRON, B., and MORRIS, C. (1972). Limiting the risk of Bayes and empirical Bayes estimators—Part II: The empirical Bayes case. *J. Amer. Statist. Assoc.* **67** 130–139.

[6] EFRON, B., and MORRIS, C. (1972). Empirical Bayes on vector observations—An extension of Stein's method. *Biometrika* **59** 335–347.

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

[8] EFRON, B., and MORRIS, C. (1975). Data analysis using Stein's estimator and its generalizations. *J. Amer. Statist. Assoc.* **70** 311–319.

[9] JAMES, W., and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379. Univ. of California Press.

[10] LIN, PI-ERH, and TSAI, HUI-LIANG (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. *Ann. Statist.* **1** 142–145.

[11] LOÈVE, MICHEL (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

[12] SCLOVE, S. L. (1968). Improved estimators for coefficients in linear regression. *J. Amer. Statist. Assoc.* **63** 596–606.

[13] SCLOVE, S. L., MORRIS, C., and RADHAKRISHNAN, R. (1972). Nonoptimality of preliminary-test estimators for the mean of a multivariate normal distribution. *Ann. Math. Statist.* **43** 1481–1490.

[14] 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.

[15] STEIN, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs, in *Festschrift for J. Neyman*, 351–366. Wiley, New York.

[16] STEIN, C. (1973). Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symp. Asymptotic Statist.* 345–381.

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

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

THE RAND CORPORATION
1700 MAIN STREET
SANTA MONICA, CALIFORNIA 90406