The Annals of Statistics
1973, Vol. 1, No. 6, 1189–1194

PROPER BAYES MINIMAX ESTIMATORS OF THE MULTIVARIATE NORMAL MEAN VECTOR FOR THE CASE OF COMMON UNKNOWN VARIANCES[1]

BY WILLIAM E. STRAWDERMAN

Rutgers, The State University

We investigate the problem of estimating the mean vector $\boldsymbol{\theta}$ of a multivariate normal distribution with covariance matrix equal to $\sigma^2 I_p$, σ^2 unknown, and loss $||\delta - \boldsymbol{\theta}||^2/\sigma^2$. We first find a class of minimax estimators for this problem which enlarges a class given by Baranchik. This result is then used to show that for sufficiently large sample sizes (which never need exceed 4) proper Bayes minimax estimators exist for $\boldsymbol{\theta}$ if $p \geqq 5$.

1. Introduction. Given a random sample of size n $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$ from a p variate normal population with mean vector $\boldsymbol{\theta}$ and covariance matrix $\sigma^2 I_p$ (where σ^2 is unknown and I_p is the $p \times p$ identity matrix), we consider the problem of estimating $\boldsymbol{\theta}$ when the loss function is given by

$$L(\boldsymbol{\theta}, \delta) = ||\delta - \boldsymbol{\theta}||^2/\sigma^2 . \tag{1.1}$$

Stein [3] established that the usual estimator $\bar{X}$ is inadmissible for $p \geqq 3$ (for the case σ^2 known) and later with James [2] produced estimators which beat $\bar{X}$ for the cases of known or unknown common variance and for other cases as well. Baranchik [1] produced a class of minimax estimators which contained those of James and Stein. Strawderman [4] used the result of Baranchik in the case where σ^2 is known to produce for $p \geqq 5$ proper Bayes minimax estimators of $\boldsymbol{\theta}$.

In canonical form, $\mathbf{X} \sim N(\boldsymbol{\theta}, \sigma^2 I)$, s is independent of $\mathbf{X}$ and distributed as σ^2 times a χ^2 variable with m degrees of freedom. We extend the Baranchik class of minimax estimators to the following family

$$\delta(\mathbf{X}, s) = (1 - r(F, s)/F)\mathbf{X}$$

where $F = ||\mathbf{X}||^2/s$. The function $r(F, s)$ is assumed to be increasing in F for fixed s and decreasing in s for fixed F and bounded between 0 and $2(p-2)/(m+2)$. We then produce proper Bayes minimax admissible estimators for $p \geqq 5$ which are members of this extended class. The sample sizes required for our result are as follows: for $p = 5$ n must be at least 4; for $p = 6$, n must be at least 3; for $p \geqq 7$, n must be at least 2. To the author's knowledge these are the first known admissible minimax estimators for the case of unknown common variances.

Section 2 is devoted to the extension of Baranchik's result. The main result is proved in Section 3 and some remarks are given in Section 4.

2. A class of minimax estimators. We extend the result of Baranchik [1] as

Received February 1972; revised February 1973.
[1] Research partially supported by N.F.S. Grant No. GP-35018.

follows: Let $\mathbf{X}$ be a p dimensional ($p \geqq 3$) normal random vector with unknown mean vector $\boldsymbol{\theta}$ and covariance matrix of the form $\sigma^2 I$, and, let s be a statistic which is distributed as σ^2 times a χ^2 random variable on m degrees of freedom. Setting $F = \mathbf{X}'\mathbf{X}/s$ we establish the following generalization of Baranchik's result.

THEOREM. *Relative to the loss function* 1.1 *an estimator of the form* $\varphi(\mathbf{X}, s) = (1 - r(F, s)/F)\mathbf{X}$ *is minimax if*

(i) (a) *For each fixed* s, $r(\cdot, s)$ *is monotone non-decreasing.*
(b) *For each fixed* F, $r(F, \cdot)$ *is monotone non-increasing.*
(ii) $0 \leqq r(\cdot, \cdot) \leqq 2(p-2)/(m+2)$.

In Baranchik's result the function $r(\cdot)$ is allowed to be a function only of F, otherwise the results coincide.

PROOF. We follow Baranchik's proof and notation making the obvious changes in the definitions of $r(F)$ and $g(F)$ to $r(F, s)$ and $g(F, s)$. The computations up to (1.11) all being conditional on s, no changes are required to that point other than notational ones. Expression (1.11) in our notation then becomes

$$(1.11)' \qquad E[\chi^2_{p+2K}\, g^2(\chi^2_{p+2K}/\chi_m^2, \sigma^2\chi_m^2) - 4Kg(\chi^2_{p+2K}/\chi_m^2, \sigma^2\chi_m^2) - p + 2K]$$

and we must show for each $K = 0, 1, \cdots$ that this expression is not positive. Since the argument leading from (1.11) to (1.17) in Baranchik is conditional on χ_m^2 we have that our estimators will be minimax provided that

$$(1.17)' \qquad E\left\{r\left(\frac{2K}{\chi^2} + \frac{p-2}{m+2}, \sigma^2\chi_m^2\right)(\chi_m^2[-1 + \chi_m^2/(m+2)])\right\}$$

is less than or equal to zero. But by condition (i), (1.17)$'$ is bounded above by

$$\begin{aligned} &r\left(\frac{2K+p-2}{m+2}, (m+2)\sigma^2\right) E\{\chi_m^2[-1+\chi_m^2/(m+2)] \mid \chi_m^2 < m+2\} \\ &\quad \times P[\chi_m^2 < m+2] + r\left(\frac{2K+p-2}{m+2}, (m+2)\sigma^2\right) \\ &\qquad \times E\{\chi_m^2[-1+\chi_m^2/(m+2)] \mid \chi_m^2 \geqq m+2\} \cdot P[\chi_m^2 \geqq m+2] \\ &\qquad\quad = r\left(\frac{2K+p-2}{m+2}, (m+2)\sigma^2\right) E\{\chi_m^2[-1+\chi_m^2/(m+2)]\} = 0 \end{aligned}$$

which completes the proof.

3. The main result. We now apply the result of Section 2 to obtain a class of proper Bayes admissible minimax estimators of θ in the original formulation.

The following notation will be used:

$$\mathbf{X}_i = (X_{i1}, X_{i2}, \cdots, X_{ip})' \qquad \bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i = (\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_p)'$$

$$s^2 = \sum_{i=1}^p \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 \qquad F = n\|\bar{\mathbf{X}}\|^2/S^2 \qquad s = S^2/n\,.$$

Letting $\eta^2 = 1/\sigma^2$ we consider the following class of prior distributions: conditional on η and λ, $\boldsymbol{\theta}$ is normally distributed with mean 0 and covariance matrix

$n^{-1}\lambda^{-1}\eta^{-2}(1-\lambda)I_p$. The density of λ is given by

$$f(\lambda) = (1-a)\lambda^{-a}, \qquad 0 < \lambda < 1, a < 1,$$

independent of that of η, which has density

$$g(\eta) = C\eta^{-2K}, \qquad \gamma \leqq \eta < \infty, K > \tfrac{1}{2}, \gamma > 0.$$

The Bayes estimator of $\boldsymbol{\theta}$ with respect to the above prior for the loss function (1.1) is given by

$$\begin{aligned}
&\delta_\iota(\mathbf{X}_1, \cdots, \mathbf{X}_n) \\
&\quad = E(\theta_\iota \eta^2) \mid \mathbf{X}_1, \cdots, \mathbf{X}_n)/E((\eta^2) \mid \mathbf{X}_1, \cdots, \mathbf{X}_n) \\
&\quad = [\textstyle\int_0^1 d\lambda \int_\gamma^\infty d\eta \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty d\boldsymbol{\theta}\, \theta_\iota(\eta^2)\lambda^{-a}(\eta^2)^{-K}\lambda^{p/2}(\eta^2)^{p/2}(1-\lambda)^{p/2} \\
&\qquad \times (\exp\{-\tfrac{1}{2}(\eta^2)\lambda(1-\lambda)^{-1}n\|\boldsymbol{\theta}\|^2\})(\eta^2)^{np/2} \\
&\qquad \times (\exp\{-\tfrac{1}{2}[n\|\bar{\mathbf{X}}-\boldsymbol{\theta}\|^2 + S^2][\eta^2]\})] \\
&\qquad \div [\textstyle\int_0^1 d\lambda \int_\gamma^\infty d\eta \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty d\boldsymbol{\theta}\, (\eta^2)\lambda^{-a}(\eta^2)^{-K}\lambda^{p/2}(\eta^2)^{p/2}(1-\lambda)^{-p/2} \\
&\qquad \times (\exp\{-\tfrac{1}{2}(\eta^2)\lambda(1-\lambda)^{-1}n\|\boldsymbol{\theta}\|^2\})(\eta^2)^{np/2} \\
&\qquad \times (\exp\{-\tfrac{1}{2}[n\|\bar{\mathbf{X}}-\boldsymbol{\theta}\|^2 + S^2][\eta^2]\})].
\end{aligned} \tag{3.1}$$

Completing the square in the exponent of the integrands of (3.1), integrating, and letting $A = np/2 - K + 1$ and $B = p/2 - a$ yields

$$\delta(\mathbf{X}_1, \cdots, \mathbf{X}_n) = [1 - r(F, s)/F]\bar{\mathbf{X}} \tag{3.2}$$

where

$$\begin{aligned}
r(F, s) &= [\textstyle\int_0^F du \int_\gamma^\infty d\eta\, (\eta^2)^A u^{B+1} \exp\{-\tfrac{1}{2}\eta^2(ns + nsu)\}] \\
&\quad \div [\textstyle\int_0^F du \int_\gamma^\infty d\eta\, (\eta^2)^A u^B \exp\{-\tfrac{1}{2}\eta^2(ns + nsu)\}].
\end{aligned} \tag{3.3}$$

LEMMA 1. *$r(F, s)$ is non-decreasing in F for fixed s ($F \geqq 0$, $s \geqq 0$).*

PROOF.

$$\begin{aligned}
\frac{\partial}{\partial F} r(F, s) &= \{[\textstyle\int_0^F du \int_\gamma^\infty d\eta\, (\eta^2)^A u^B \exp\{-\tfrac{1}{2}\eta^2(ns + nsu)\}] \\
&\quad \times [\textstyle\int_\gamma^\infty d\eta\, (\eta^2)^A F^{B+1} \exp\{-\tfrac{1}{2}\eta^2(ns + nsF)\}] \\
&\quad - [\textstyle\int_0^F du \int_\gamma^\infty d\eta\, (\eta^2)^A u^{B+1} \exp\{-\tfrac{1}{2}\eta^2(ns + nsu)\}] \\
&\quad \times [\textstyle\int_\gamma^\infty d\eta (\eta^2)^A F^B \exp\{-\tfrac{1}{2}\eta^2(ns + nsF)\}]\} \\
&\quad \div [\textstyle\int_0^F du \int_\gamma^\infty d\eta\, (\eta^2)^A u^B \exp\{-\tfrac{1}{2}\eta^2(ns + nsu)\}]^2.
\end{aligned}$$

The numerator of the above expression is equal to

$$\begin{aligned}
&[F^B \textstyle\int_\gamma^\infty d\eta\, (\eta^2)^A \exp\{-\tfrac{1}{2}\eta^2(ns + nsF)\}] \\
&\qquad \times [\textstyle\int_0^F du \int_\gamma^\infty d\eta\, (\eta^2)^A u^B (F-u) \exp\{-\tfrac{1}{2}\eta^2(ns + nsu)\}] \geqq 0.
\end{aligned}$$

This establishes the lemma.

LEMMA 2. *$r(F, s)$ is non-increasing in s for fixed F, ($F \geqq 0$, $s \geqq 0$).*

PROOF.

$$\frac{\partial}{\partial F} r(F, s) = -\tfrac{1}{2}n \operatorname{Cov}(U, H^2(U+1)) = -\tfrac{1}{2}n \operatorname{Cov}(U, E(H^2(U+1) \mid U))$$

where the joint density of (U, H) is given by

$$g_{U,H}(u, \eta) = C(\eta^2)^A u^B \exp\{-\tfrac{1}{2}\eta^2(ns + nsu)\} \qquad 0 \leqq u < F, \gamma \leqq \eta < \infty .$$

To complete the proof of the lemma it suffices to show that $g(u) = E(H^2(1 + U) \mid U = u)$ is non-decreasing in u.

Write the integral expression for $E(H^2(1 + U) \mid U = u)$, make the substitution $v^2 = \eta^2(ns + nsu)$ in each of the integrals of this expression and differentiate with respect to u. The numerator of the resulting expression is nonnegative, which completes the proof.

The above lemmas imply that

$$\begin{aligned} 0 \leqq r(F, s) &\leqq r(\infty, 0) \\ &= \lim_{F\to\infty; s\to 0} r(F, s) \\ &= \lim_{F\to\infty; s\to 0} \{[\textstyle\int_0^F du \int_{\gamma(ns)^{\frac{1}{2}}} dv(v^2)^A u^{B+1} \exp\{-\frac{1}{2}v^2(1 + u)\}] \\ &\qquad \div [\textstyle\int_0^F du \int_{\gamma(ns)^{\frac{1}{2}}} dv(v^2)^A u^B \exp\{-\frac{1}{2}v^2(1 + u)\}]\} \\ &= [\textstyle\int_0^\infty du \int_0^\infty dv(v^2)^A u^{B+1} \exp\{-\frac{1}{2}v^2(1 + u)\}] \qquad (3.4) \\ &\qquad \div [\textstyle\int_0^\infty du \int_0^\infty dv(v^2)^A u^B \exp\{-\frac{1}{2}v^2(1 + u)\}] \\ &= \textstyle\int_0^\infty du(1 + u)^{-A-\frac{1}{2}} u^{B+1} / \int_0^\infty du(1 + u)^{-A-\frac{1}{2}} u^B \\ &= [\textstyle\int_0^1 dy\, y^{A+\frac{1}{2}-2}[(1 - y)/y]^{B+1}] / [\int_0^1 dy\, y^{A+\frac{1}{2}-2}[(1 - y)/y]^B] \\ &= \beta(A - B + \tfrac{1}{2} - 2, B + 2) / \beta(A - B + \tfrac{1}{2} - 1, B + 1) \\ &= (p - 2a + 2)/((n - 1)p - 2K + 2a - 1) . \end{aligned}$$

Since $\bar{\mathbf{X}} \sim N(\boldsymbol{\theta}(\sigma^2/n)I)$ and $s \sim (\sigma^2/n)\chi^2_{p(n-1)}$ the theorem will imply minimaxity of the estimators (3.2) provided we can find a and K such that

$$(3.5) \qquad (p - 2a + 2)/((n - 1)p - 2K - 1 + 2a) \leqq 2(p - 2)/(p(n - 1) + 2) .$$

We remark that the theorem implies minimaxity with respect to the loss function $\|\delta - \boldsymbol{\theta}\|^2/(\sigma^2/n)$ which is sufficient to claim minimaxity for the loss (1.1) as well.

Since we want our estimators to be proper Bayes as well, we want $K > \frac{1}{2}$ and $a < 1$. It is clear from (3.5) that it is "easier" for an estimator to be minimax if K is as close to $\frac{1}{2}$ as possible. Hence we take $K = \frac{1}{2}(1 + \varepsilon)$ for $\varepsilon > 0$. Condition (3.5) then becomes

$$(p - 2a + 2)/((n - 1)p - 2 + 2a - \varepsilon) \leqq 2(p - 2)/(p(n - 1) + 2)$$

which after a little algebra becomes $a \geqq a(p, n, \varepsilon)$ where

$$(3.6) \qquad a(p, n, \varepsilon) = [-p^2(n - 1) + 6np - 4 + 2\varepsilon(p - 2)]/[2(n + 1)p - 4] .$$

Hence we wil be able to find values of a and K which will give proper Bayes minimax estimators of $\boldsymbol{\theta}$ for those p, n for which there exists an $\varepsilon > 0$ such that $a(p, n, \varepsilon) < 1$. This is the case whenever $n(p - 4) > p - 2$ and $p \geqq 3$. For $p = 3, 4$, the inequality cannot hold for any n. For $p \geqq 5$, $n\ 1 + 2/(p - 4)$, so that for $p = 5$, $n > 3$, for $p = 6$, $n > 2$ and for $p \geqq 7$, $n \geqq 2$.

We summarize the results in the following.

THEOREM. *The estimator given in* (3.2) *is a proper Bayes minimax admissible estimator of* $\boldsymbol{\theta}$ *subject to the loss* (1.1) *provided that* p, n, a *and* ε *satisfy*

(a) $p \geqq 5$

(b) $1 > a \geqq a(p, n, \varepsilon)$ *when* $a(p, n, \varepsilon)$ *is defined as in* (3.6). *In particular proper Bayes minimax estimators exist for* $p \geqq 5$ *provided the sample size* n *is at least* 4 *for* $p = 5$, *at least* 3 *for* $p = 6$, *and at least* 2 *for* $p \geqq 7$.

4. Remarks. We note first that if we do not insist that $a < 1$ we can construct generalized Bayes minimax estimators of $\boldsymbol{\theta}$ for $p = 3$ and 4 and $n \geqq 2$ as well as for those values of n not covered by the theorem for $p \geqq 5$. The basic requirement in order for all expressions to make sense is that $a < p/2 + 1$ and it can be shown that such an a can be chosen to satisfy (3.5) provided $p \geqq 3$ for any $n \geqq 2$. Of course if we choose $a \geqq 1$ the estimator is (probably) no longer proper Bayes and hence no claim can be made as to its admissibility.

We also note that we can obtain estimators which are "conditionally" proper Bayes minimax in the sense that the prior distribution conditional on σ^2 is proper for each σ^2 but the prior on σ^2 is not proper. To do this one considers the following classs of "priors." Conditional on λ and σ^2 let $\boldsymbol{\theta}$ be normally distributed with mean 0 and covariance matrix $n\lambda^{-1}(1 - \lambda)\sigma^{-2}Ip$. The density of λ is proportional to λ^{-a} as before and the "density" for σ is proportional to σ^{2K} on $0 < \sigma < \infty$. The estimators thus generated are easily seen to be

$$\delta(\mathbf{X}_1, \cdots, \mathbf{X}_n) = \{1 - r(F)/F\}\bar{\mathbf{X}}$$

where

$$r(F) = [\textstyle\int_0^F du\, u^{p/2-a+1}/(1 + u)^{np/2-K+\frac{1}{2}}][\int_0^F du\, u^{p/2-a}/(1 + u)^{np/2-K+\frac{1}{2}}]$$

which can be shown to be of the form (3.3) with $\gamma = 0$.

These estimators, then, provided care is taken in choosing a and K, will be in Baranchik's class and hence minimax. Again, of course, no claim for admissibility can be made.

We note also, that the analysis in Section 3 can be done for the problem in canonical form. In this case, n^{-1} would not be used in the covariance matrix of the prior distribution. The Bayes estimator would be given by (3.2) and (3.3) where $A = (p + m)/2 - k + 1$, $B = p/2 - a$ and when $\mathbf{X}$ replaces $\bar{\mathbf{X}}$ and s replaces ns. The lemmas remain true with the obvious notational changes and we find that the resulting proper Bayes éstimator is minimax provided $1 > a > a(p, m, \varepsilon)$ where

$$a(p, m, \varepsilon) = [-mp + 6(m + p) - 4 + 2\varepsilon(p - 2)]/[2(m + 2p) - 4]\,.$$

We find that no such a and ε exist if $p = 3$ or 4, but that values can be found if $p \geqq 5$ and $m > 2p/(p - 4)$.

REFERENCES

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

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

[3] 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.

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

STATISTICS CENTER
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903