The Annals of Statistics
1973, Vol. 1, No. 1, 142-145

GENERALIZED BAYES MINIMAX ESTIMATORS OF THE MULTIVARIATE NORMAL MEAN WITH UNKNOWN COVARIANCE MATRIX

BY PI-ERH LIN[1] AND HUI-LIANG TSAI

The Florida State University

Let $\mathbf{X}$ be a p-variate ($p \geqq 3$) vector normally distributed with mean $\boldsymbol{\theta}$ and covariance matrix Σ, positive definite but unknown. Let A be a $p \times p$ Wishart matrix with parameters (n, Σ), independent of $\mathbf{X}$. To estimate $\boldsymbol{\theta}$ relative to quadratic loss function $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\Sigma^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, we obtain a family of minimax estimators $\boldsymbol{\delta}(\mathbf{X}, A)$ based on $\mathbf{X}$ and A through $\mathbf{X}$ and $\mathbf{X}'A^{-1}\mathbf{X}$. It is shown that there are minimax estimators of the form $\boldsymbol{\delta}(\mathbf{X}, A)$ which are also generalized Bayes. A special case where $\Sigma = \sigma^2 I$ is also considered.

1. Introduction and summary. Let $\mathbf{X}$ be a p-variate random vector normally distributed with mean $\boldsymbol{\theta}$ and covariance matrix Σ, positive definite but unknown. Let A be a $p \times p$ Wishart random matrix with parameters (n, Σ), $n > p - 3$, and is independent of $\mathbf{X}$. Based on $\mathbf{X}$ and A, we estimate $\boldsymbol{\theta}$ by $\boldsymbol{\delta}(\mathbf{X}, A)$ relative to the quadratic loss function

$$L(\boldsymbol{\delta}(\mathbf{X}, A); \boldsymbol{\theta}, \Sigma) = (\boldsymbol{\delta}(\mathbf{X}, A) - \boldsymbol{\theta})'\Sigma^{-1}(\boldsymbol{\delta}(\mathbf{X}, A) - \boldsymbol{\theta}) . \tag{1.1}$$

In this paper we obtain a family of minimax estimators for $p \geqq 3$. We also produce a class of prior distributions for the parameters $\boldsymbol{\theta}$ and Σ, from which a family of generalized Bayes minimax estimators is derived for $p \geqq 3$.

Recently, Baranchik [1] has obtained a family of minimax estimators with the covariance matrix $\sigma^2 I$ and Strawderman [6] the Bayes minimax estimators for the case of known covariance matrix with $p \geqq 5$. We show how their results may be extended to include the case of unknown Σ and also, in Section 4, to the case where $\Sigma = \sigma^2 B$, B being a known positive definite matrix.

2. A family of minimax estimators. James and Stein [5] have obtained a minimax estimator $[1 - c/(\mathbf{X}'A^{-1}\mathbf{X})]\mathbf{X}$, where $c = (p - 2)/(n - p + 3)$. Let c be a function of $\mathbf{X}'A^{-1}\mathbf{X}$ satisfying certain conditions, we derive a family of minimax estimators.

THEOREM 2.1. *For $p \geqq 3$, an estimator of the form*

$$\boldsymbol{\delta}(\mathbf{X}, A) = [1 - r(y)/y]\mathbf{X} , \qquad \textit{where} \quad y = \mathbf{X}'A^{-1}\mathbf{X} , \tag{2.1}$$

is a minimax estimator of $\boldsymbol{\theta}$, relative to the loss function (1.1), if $r(y)$ is a nonnegative, non-decreasing function of y less than or equal to $2(p - 2)/(n - p + 3)$.

PROOF. Observe that the conditional distribution of $\mathbf{X}'A^{-1}\mathbf{X}$ given $\mathbf{X}$ is that of $\mathbf{X}'\mathbf{X}/S$, where S is chi-square distributed with $n - p + 1$ degrees of freedom and

Received October 6, 1971; revised May 22, 1972.

[1] Research supported in part by the Florida State University Grant No. 011204-043 and in part by the National Institute of General Medical Sciences through Training Grant 5T1 GM-913.

is independent of $\mathbf{X}$ (see e.g. Wijsman [9]). The risk function of $\boldsymbol{\delta}(\mathbf{X}, A)$ is given by

$$
\begin{aligned}
&E\{\{[1 - r(y)/y]\mathbf{X} - \boldsymbol{\theta}\}'\Sigma^{-1}\{[1 - r(y)/y]\mathbf{X} - \boldsymbol{\theta}\} \mid \boldsymbol{\theta}, \Sigma\} \\
(2.2) \qquad &= E\{\{[1 - r(y)/y]\mathbf{X} - \boldsymbol{\theta}^*\}'\{[1 - r(y)/y]\mathbf{X} - \boldsymbol{\theta}^*\} \mid \boldsymbol{\theta}^*, I\} \\
&= E\{\{[1 - r(\mathbf{X}'\mathbf{X}/S)/(\mathbf{X}'\mathbf{X}/S)]\mathbf{X} - \theta^*\}' \\
&\quad \times \{[1 - r(\mathbf{X}'\mathbf{X}/S)/(\mathbf{X}'\mathbf{X}/S)]\mathbf{X} - \boldsymbol{\theta}^*\} \mid \boldsymbol{\theta}^*, I\},
\end{aligned}
$$

where $\boldsymbol{\theta}^* = [(\boldsymbol{\theta}'\Sigma^{-1}\boldsymbol{\theta})^{\frac{1}{2}}, 0, \cdots, 0]'$. The first equality is obtained by making the transformation $\mathbf{X} \to PD\mathbf{X}$, where D is a $p \times p$ nonsingular matrix such that $D\Sigma D' = I$ and P is a $p \times p$ orthogonal matrix with its first row proportional to $D\boldsymbol{\theta}$. The final expression of (2.2) is less than or equal to p, by an application of Baranchik's [1] result where n is replaced by $n - p + 1$. Since as is well known, relative to the loss function (1.1), $\mathbf{X}$ is minimax with constant risk p, the conclusion follows.

3. Generalized Bayes minimax estimators. A class of generalized prior distributions $\tau_\lambda(\boldsymbol{\theta}, \Sigma^{-1})$ of $\boldsymbol{\theta}$ and Σ^{-1}, conditional on λ is given by the densities

$$\tau_\lambda(\boldsymbol{\theta}, \Sigma^{-1}) = f_\lambda(\boldsymbol{\theta} \mid \Sigma^{-1}) \cdot g(\Sigma^{-1})$$

where

$$f_\lambda(\boldsymbol{\theta} \mid \Sigma^{-1}) = \left[\frac{\lambda}{2\pi(1 - \lambda)}\right]^{p/2} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2(1 - \lambda)}\, \boldsymbol{\theta}'\Sigma^{-1}\boldsymbol{\theta}\right\}, \qquad 0 < \lambda \leqq 1,$$

and

$$(3.1) \qquad g(\Sigma^{-1}) \propto |\Sigma|^{\frac{1}{2}\nu}, \qquad -\infty < \nu \leqq n, \quad \nu \text{ an integer.}$$

(Note that the prior (3.1) was considered by Geisser and Cornfield [4] for $-\infty < \nu \leqq n$, and it was used by Tiao and Zeller [7] and Geisser [3] for $\nu = p + 1$. Villegas [8] also gave a fiducial argument in support of this prior.) Since $E(\boldsymbol{\theta} \mid \mathbf{X}, A, \Sigma^{-1}, \lambda)$ does not depend on Σ^{-1}, it readily follows that

$$E(\boldsymbol{\theta} \mid \mathbf{X}, A, \lambda) = (1 - \lambda)\mathbf{X}.$$

If, in addition, we assume λ has density

$$(3.2) \qquad h(\lambda) \propto \lambda^{-a}, \qquad -\infty < a < \tfrac{1}{2}p + 1,$$

it follows that the generalized Bayes estimator with respect to the generalized prior with density

$$(3.3) \qquad \tau(\lambda, \boldsymbol{\theta}, \Sigma^{-1}) = \tau_\lambda(\boldsymbol{\theta}, \Sigma^{-1}) \cdot h(\lambda),$$

relative to the loss in (1.1), is $\boldsymbol{\delta}(\mathbf{X}, A) = [1 - E(\lambda \mid \mathbf{X}, A)]\mathbf{X}$. Observe that $E(\lambda \mid \mathbf{X}, A)$ is a function of $y = \mathbf{X}'A^{-1}\mathbf{X}$ alone, say $s(y)$. In fact, if $y = \mathbf{X}'A^{-1}\mathbf{X}$,

$$(3.4) \qquad E(\lambda \mid \mathbf{X}, A) = \frac{\int_0^1 \lambda^{\frac{1}{2}p-a+1}(1 + \lambda y)^{-\frac{1}{2}(n-\nu+p+2)}\, d\lambda}{\int_0^1 \lambda^{\frac{1}{2}p-a}(1 + \lambda y)^{-\frac{1}{2}(n-\nu+p+2)}\, d\lambda}$$

or, as it is easily shown on integrating by parts the numerator of (3.4),

$$(3.5) \qquad E(\lambda \mid \mathbf{X}, A) = [(p - 2a + 2) - 2Q(y)]/[(n - \nu + 2a - 2)y],$$

where

$$[Q(y)]^{-1} = (1 + y)^{\frac{1}{2}(n-\nu+p)} \int_0^1 \lambda^{\frac{1}{2}p-a}(1 + \lambda y)^{-\frac{1}{2}(n-\nu+p+2)} \, d\lambda \, . \tag{3.6}$$

LEMMA 3.1. *Let $r(y) = ys(y)$. Then $r(y)$ is a nonnegative, non-decreasing function of $y \geqq 0$.*

PROOF. From (3.4), we have

$$r(y) = \frac{y \int_0^1 \lambda^{\frac{1}{2}p-a+1}(1 + \lambda y)^{-\frac{1}{2}(n-\nu+p+2)} \, d\lambda}{\int_0^1 \lambda^{\frac{1}{2}p-a}(1 + \lambda y)^{-\frac{1}{2}(n-\nu+p+2)} \, d\lambda} \, , \tag{3.7}$$

which is clearly nonnegative for all $y \geqq 0$. Let $t = \lambda y$ and take derivative of $r(y) + 1$ with respect to y, we have $r'(y) = d/k^2$, where k is the denominator of the right-hand side of (3.7) and

$$d = y^{-\frac{1}{2}p+a-2}(1 + y)^{-\frac{1}{2}(n-\nu+p+2)} \int_0^y t^{\frac{1}{2}p-a}(1 + t)^{-\frac{1}{2}(n-\nu+p+2)}(y - t) \, dt \, . \tag{3.8}$$

It is clear that $d \geqq 0$ for all $y \geqq 0$. Thus, $r'(y) \geqq 0$ for all $y \geqq 0$. This proves the lemma.

Using the above lemma and Theorem 2.1, the proof of Theorem 3.1 is immediate.

THEOREM 3.1. *For $p \geqq 3$ with*

(i) $-\infty < \nu < \min(n + 1, n + 2a - 2)$,
(ii) $2(p - 2)/(n - p + 3) \geqq (p - 2a + 2)/(n - \nu + 2a - 2)$,
(iii) $-\infty < a < \frac{1}{2}p + 1$, *and*
(iv) $n > p - 3$, *the estimators of the form*

$$\boldsymbol{\delta}(\mathbf{X}, A) = [1 - r(y)/y]\mathbf{X} \, , \qquad y = \mathbf{X}'A^{-1}\mathbf{X} \, , \tag{3.9}$$

are generalized Bayes minimax estimators with respect to the priors (3.3), *where*

$$r(y) = [(p - 2a + 2) - 2Q(y)]/(n - \nu + 2a - 2) \tag{3.10}$$

and $Q(y)$ is defined by (3.6).

4. A special case. Consider the case $\Sigma = \sigma^2 B$, where B is a $p \times p$ symmetric positive definite known matrix, and σ^2 is an unknown positive quantity. Since B is known, there exists a $p \times p$ nonsingular matrix C such that $CBC' = I$. If we let $\mathbf{Z} = C\mathbf{X}$, then $\mathbf{Z} \mid \lambda, \boldsymbol{\theta}, \sigma^2 \sim N(\boldsymbol{\mu}, \sigma^2 I)$ with $\boldsymbol{\mu} = C\boldsymbol{\theta}$, and the problem is reduced to that of estimating $\boldsymbol{\mu}$ relative to the quadratic loss function

$$L(\hat{\boldsymbol{\mu}}; \boldsymbol{\mu}, \sigma^2) = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})/\sigma^2 \, , \tag{4.1}$$

where $\hat{\boldsymbol{\mu}}$ is an estimator of $\boldsymbol{\mu}$. Without loss of generality, we may thus assume that $B = I$.

Consider that $\mathbf{X} \mid \lambda, \boldsymbol{\theta}, \sigma^2 \sim N(\boldsymbol{\theta}, \sigma^2 I)$ and $S \mid \lambda, \sigma^2 \sim \sigma^2 \chi_n^2$, independent of $\mathbf{X}$. If we assume further that the joint generalized prior density of $\boldsymbol{\theta}$, σ^{-2} and λ is

$$\tau(\lambda, \boldsymbol{\theta}, \sigma^{-2}) = \tau(\boldsymbol{\theta} \mid \lambda, \sigma^{-2}) \cdot g_1(\sigma^{-2}) \cdot h(\lambda) \, , \tag{4.2}$$

where $h(\lambda)$ is given by (3.2),

$$g_1(\sigma^{-2}) \propto (\sigma^2)^{\nu/2}, \qquad -\infty < \nu \leqq n, \quad \nu \text{ an integer}, \tag{4.3}$$

and

$$\tau(\boldsymbol{\theta} \mid \lambda, \sigma^{-2}) \propto \left(\frac{\lambda}{(1-\lambda)\sigma^2}\right)^{p/2} \exp\left\{-\frac{\lambda}{2(1-\lambda)\sigma^2}\boldsymbol{\theta}'\boldsymbol{\theta}\right\}, \tag{4.4}$$

then by the same argument as in the proof of Theorem 3.1 and the result of Baranchik [1], we obtain the following theorem, which is stated without proof.

THEOREM 4.1. *For* $p \geqq 3$, *with*

(i) $-\infty < \nu < \min(n+1, n+2a-2)$,
(ii) $2(p-2)/(n+2) \geqq (p-2a+2)/(n-\nu+2a-2)$, *and*
(iii) $-\infty < a < \frac{1}{2}p + 1$, *relative to the loss function* (4.1) *with* $\boldsymbol{\mu}$ *replaced by* $\boldsymbol{\theta}$, *the estimators of the form*

$$\boldsymbol{\delta}(\mathbf{X}, S) = [1 - r(F)/F]\mathbf{X} \tag{4.5}$$

are generalized Bayes minimax with respect to the priors (4.2), *where* $F = \mathbf{X}'\mathbf{X}/S$ *and* $r(F)$ *is defined by* (3.10) *with* $Q(F)$ *given by* (3.6).

Here, we note that the set of values of n, ν, a and p, which satisfies conditions (i)—(iii) of Theorem 4.1, is nonempty.

Acknowledgment. The authors wish to thank the referee and associate editor for their helpful suggestions.

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

[3] GEISSER, S. (1965). Bayes estimation in multivariate analysis. *Ann. Math. Statist.* **36** 150-159.

[4] GEISSER, S. and CORNFIELD, J. (1963). Posterior distributions for multivariate normal parameters. *J. Roy. Statist. Soc. Ser. B* **25** 368-376.

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

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

[7] TIAO, G. C. and ZELLNER, A. (1964). On the Bayesian estimation of multivariate regression. *J. Roy. Statist. Soc. Ser. B* **26** 277-285.

[8] VILLEGAS, C. (1969). On the a priori distribution of the covariance matrix. *Ann. Math. Statist.* **40** 1098-1099.

[9] WIJSMAN, R. A. (1957). Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis. *Ann. Math. Statist.* **28** 415-428.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306