The Annals of Mathematical Statistics
1971, Vol. 42, No. 1, 385–388

PROPER BAYES MINIMAX ESTIMATORS OF THE MULTIVARIATE NORMAL MEAN

BY WILLIAM E. STRAWDERMAN

Rutgers, The State University

1. Introduction. Consider the problem of estimating the mean of a multivariate normal distribution with covariance matrix the identity and sum of squared errors loss. We show that there exist proper Bayes minimax estimators when the dimension is greater than or equal to 5. This answers, partially, a conjecture in the folklore attributed to Charles Stein to the effect that in 3 and 4 dimensions proper Bayes minimax estimators do not exist, while they do in dimensions greater than or equal to 5. In 1 and 2 dimensions, of course, $\bar{x}$ is unique minimax and is not proper Bayes. The 3- and 4-dimensional problems are studied by the author in a forthcoming paper [5].

In the next section we extend slightly the class of minimax estimators given by Baranchik [1]. In Section 3 we exhibit a subclass of proper Bayes estimators within the class of minimax estimators. Some remarks are given in Section 4.

2. A class of minimax estimators. Let X be a p-dimensional random vector distributed according to the multi-normal distribution with mean 0 and covariance matrix I. Baranchik [1] showed that if the dimension p is greater than or equal to 3, then any estimator of the form

$$(1)\qquad r(\tfrac{1}{2}\|x\|^2)\left(1-\frac{p-2}{\|x\|^2}\right)x+(1-r(\tfrac{1}{2}\|x\|^2))x=\left(1-r(\tfrac{1}{2}\|x\|^2)\frac{p-2}{\|x\|^2}\right)x$$

where

$$0 \leqq r(\tfrac{1}{2}\|x\|^2) \leqq 1$$

and $r(\cdot)$ is non-decreasing, is minimax.

We extend this class slightly so that in the above, $0 \leqq r(\frac{1}{2}\|x\|^2) \leqq 2$, and show that estimators in the resulting class are still minimax. It is clear from [2] that the improved bound is known to Baranchik. Our proof is based, however, on his earlier result in [1].

LEMMA. *Estimators of the form* (1) *with* $0 \leqq r(\frac{1}{2}\|x\|^2) \leqq 2$ *and* $r(\frac{1}{2}\|x\|^2)$ *non-decreasing are minimax.*

PROOF. As in Baranchik's proof, for any estimator of the form $h(\frac{1}{2}\|x\|^2)x$, the difference in risk between $h(\frac{1}{2}\|x\|^2)x$ and x is given by

$$(2)\quad \exp(-\tfrac{1}{2}\|\theta\|^2)\sum_{K=0}^{\infty}\frac{(\frac{1}{2}\|\theta\|^2)^K}{K!}[E\chi^2_{p+2K}h^2(\tfrac{1}{2}\chi^2_{p+2K})-4KEh(\tfrac{1}{2}\chi^2_{p+2K})-p+2K].$$

Received May 14, 1970.

To show an estimator is minimax it suffices, since x is itself minimax, to show the above difference is negative. To this end it suffices to show, for every integer $K \geqq 0$, that

$$E\chi^2_{p+2K}\, h^2(\tfrac{1}{2}\chi^2_{p+2K}) - 4KEh(\tfrac{1}{2}\chi^2_{p+2K}) - p + 2K \leqq 0. \tag{3}$$

For the estimators of interest to us

$$h(\tfrac{1}{2}\|x\|^2) = 1 - r(\tfrac{1}{2}\|x\|^2)\frac{p-2}{\|x\|^2}, \tag{4}$$

and by the assumptions on $r(\cdot)$,

$$\begin{aligned}
&E\chi^2_{p+2K}\, h^2(\tfrac{1}{2}\chi^2_{p+2K}) - 4KEh(\tfrac{1}{2}\chi^2_{p+2K}) - p + 2K \\
&\quad = E\left\{\chi^2_{p+2K}\left[1 - \frac{p-2}{\chi^2_{p+2K}} r(\tfrac{1}{2}\chi^2_{p+2K})\right]^2 - 4K\left[1 - \frac{p-2}{\chi^2_{p+2K}} r(\tfrac{1}{2}\chi^2_{p+2K})\right] - p + 2K\right\} \\
&\quad = E\left\{\frac{(p-2)^2}{\chi^2_{p+2K}} r^2(\tfrac{1}{2}\chi^2_{p+2K}) - 2(p-2)r(\tfrac{1}{2}\chi^2_{p+2K}) + \frac{4K(p-2)}{\chi^2_{p+2K}} r(\tfrac{1}{2}\chi^2_{p+2K})\right\} \\
&\quad = E\left\{(p-2)r(\tfrac{1}{2}\chi^2_{p+2K})\left[\frac{p-2}{\chi^2_{p+2K}} r(\tfrac{1}{2}\chi^2_{p+2K}) + \frac{4K}{\chi^2_{p+2K}} - 2\right]\right\} \qquad (5) \\
&\quad \leqq E\left\{(p-2)r(\tfrac{1}{2}\chi^2_{p+2K})\left[\frac{2(p-2)+4K}{\chi_{p+2K}} - 2\right]\right\} \\
&\quad = (p-2)\operatorname{Cov}\left(r(\tfrac{1}{2}\chi^2_{p+2K}), \frac{2p-4+4K}{\chi^2_{p+2K}} - 2\right) \\
&\qquad + (p-2)E(r(\tfrac{1}{2}\chi^2_{p+2K}))E\left(\frac{2p-4+4K}{\chi^2_{p+2K}} - 2\right) \\
&\quad \leqq 0.
\end{aligned}$$

We used the fact that the covariance between an increasing function and a decreasing function is negative, and $E((2p-4+4K)/(\chi^2_{p+2K}) - 2) = 0$. This completes the proof of the lemma.

3. The main result. We now produce a class of proper prior distributions such that for $p \geqq 5$ the corresponding Bayes estimators satisfy the conditions of the lemma and are therefore minimax. Clearly such estimators are admissible.

Conditioned on λ, $0 < \lambda \leqq 1$, let the distribution of θ be normal with mean 0 and covariance matrix $\lambda^{-1}(1-\lambda)I$. The unconditional density of λ with respect to Lebesgue measure is given by $\lambda^{-a}/(1-a)$ for any a $0 \leqq a < 1$.

The (proper) Bayes estimator with respect to the above distribution on θ is given by

$$\begin{aligned}
E(\theta \mid x) &= E[E(\theta \mid x, \lambda) \mid x] \\
&= E[(1-\lambda)x \mid x] \qquad (6) \\
&= [1 - E(\lambda \mid x)]x.
\end{aligned}$$

The joint distribution of λ and x is given by

$$
\begin{aligned}
g(x,\lambda) &= \textstyle\int_{R^p} h(x,\lambda,\theta)\prod_{i=1}^{p} d\theta_i \\
&= K_1 \textstyle\int_{R^p} \lambda^{-a}\exp(-\frac{1}{2}||x-\theta||^2)(\lambda/(1-\lambda))^{p/2} \\
&\qquad \cdot\exp(-\frac{1}{2}(\lambda/(1-\lambda))||\theta||^2)\prod_{i=1}^{p} d\theta_i \\
&= K_2\,\lambda^{\frac{1}{2}p-a}\exp(-\frac{1}{2}\lambda\,||x||^2), \qquad 0<\lambda \leqq 1;\, ||x||^2 \geqq 0.
\end{aligned} \tag{7}
$$

Hence,

$$
\begin{aligned}
E(\lambda\,|\,x) &= \frac{\int_0^1 \lambda^{\frac{1}{2}p-a+1}\exp(-\frac{1}{2}\lambda\,||x||^2)\,d\lambda}{\int_0^1 \lambda^{\frac{1}{2}p-a}\exp(-\frac{1}{2}\lambda\,||x||^2)\,d\lambda} \\
&= \frac{p+2-2a}{||x||^2} - \frac{2\exp(-\frac{1}{2}||x||^2)}{||x||^2\int_0^1 \lambda^{\frac{1}{2}p-a}\exp(-\frac{1}{2}||x||^2\lambda)\,d\lambda},
\end{aligned} \tag{8}
$$

where integration by parts is used in the last equality. The Bayes estimator with respect to the above class of priors is then given by

$$
\left[1-\left(\frac{p+2-2a}{||x||^2} - \frac{2\exp(-\frac{1}{2}||x||^2)}{||x||^2\int_0^1 \lambda^{\frac{1}{2}p-a}\exp(-\frac{1}{2}\lambda\,||x||^2)}\right)\right]x. \tag{9}
$$

We have the following result

THEOREM. *The estimators given by* (9) *for* $0 \leqq a < 1$ *are proper Bayes minimax estimators if the dimension* p *is greater than or equal to* 6. *For* $p = 5$ *the estimators are minimax for* $\frac{1}{2} \leqq a < 1$.

PROOF. The estimators are proper Bayes for any p by the choice of the prior distribution on θ. To show they are minimax we show that the conditions of the lemma are satisfied for $p \geqq 5$.

Note that the estimators (9) are of the form (1) with

$$
r(\tfrac{1}{2}||x||^2) = \left(\frac{1}{p-2}\right)\left(p+2-2a-\frac{2\exp(-\frac{1}{2}||x||^2)}{\int_0^1 \lambda^{\frac{1}{2}p-a}\exp(-\frac{1}{2}\lambda\,||x||^2)\,d\lambda}\right). \tag{10}
$$

Since

$$
0 \leqq r(\tfrac{1}{2}||x||^2) \leqq \frac{p+2-2a}{p-2} \tag{11}
$$

we have that $r(\frac{1}{2}||x||^2) \leqq 2$, provided $p+2-2a \leqq 2p-4$ or $p \geqq 6-2a$. Hence $r(\frac{1}{2}||x||^2) \leqq 2$ if $p=5$ and $\frac{1}{2} \leqq a < 1$ and for any a in the interval $[0, 1)$ for $p > 5$.

It remains to show $r(\cdot)$ is non-decreasing. This is equivalent to showing

$$
\exp(-\tfrac{1}{2}||x||^2)/\textstyle\int_0^1 \lambda^{\frac{1}{2}p-a}\exp(-\frac{1}{2}\lambda\,||x||^2) = \left(\int_0^1 \lambda^{\frac{1}{2}p-a}\exp(\frac{1}{2}(1-\lambda)||x||^2)\,d\lambda\right)^{-1} \tag{12}
$$

is non-increasing, but this is clear from the fact that $0 < \lambda \leqq 1$. This completes the proof of the theorem.

4. Remarks. We have demonstrated that for $p \geqq 5$ there do exist proper Bayes minimax estimators. It is also clear from the proof of the theorem that $p = 5$ is the lowest dimension for which the above proof works. A result of Strawderman ([4] page 91) shows that no estimator in the class given by the lemma can be proper Bayes for $p \leqq 4$. This suggests that there do not exist proper Bayes minimax procedures in 3 and 4 dimensions. This is shown in a forthcoming paper by the author [5].

Work done by L. Brown [3] has suggested that prior distributions on θ which are approximately of the form

$$\begin{aligned} G(d\theta) &= d\theta, & ||\theta|| &\leqq 1; \\ &= d\theta/||\theta||^{2p-4}, & ||\theta|| &\geqq 1, \end{aligned}$$

should lead to minimax estimators. Note that for $p \geqq 5$ $G(d\theta)$ is a proper prior. It was with distributions of roughly this form that the author first tried to demonstrate minimaxity by using the Baranchik approach. However, attempts at proving $r(\cdot)$ to be non-decreasing were unsuccessful. We note that the family of priors which lead to the estimators (9) are such that for large $||\theta||$ the unconditional density of θ is approximately $1/||\theta||^{p+2-2a}$.[1]

5. Acknowledgments. The author wishes to thank L. Brown for bringing the problem to his attention and Bradley Efron for helpful discussions. The author wishes to thank Charles Stein for making many helpful suggestions, and in particular, for help in choosing a distribution on λ for which the calculations are tractable.

REFERENCES

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

[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] Brown, L. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. To appear *Ann. Math. Statist.* **42** No. 3.

[4] Strawderman, W. (1969). Generalized Bayes estimators and admissibility of estimators of the mean vector of a multivariate normal distribution with quadratic loss. Ph.D. thesis, Rutgers Univ.

[5] Strawderman, W. (1970). On the existence of proper Bayes minimax estimators of the mean of a multivariate normal distribution. To appear in *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*

[1] It has been pointed out by J. Kiefer and L. Brown that the above proof goes through provided only that $1 > a \geqq 3 - p/2$ which yields priors of precisely the order predicted by Brown.