The Annals of Statistics
1976, Vol. 4, No. 3, 642–648

COMBINING INDEPENDENT NORMAL MEAN ESTIMATION PROBLEMS WITH UNKNOWN VARIANCES

BY JAMES O. BERGER[1] AND M. E. BOCK[2]

Purdue University

Let $X = (X_1, \cdots, X_p)^t$ be a p-variate normal random vector with unknown mean $\theta = (\theta_1, \cdots, \theta_p)^t$ and unknown positive definite diagonal covariance matrix A. Assume that estimates V_i of the variances A_i are available, and that V_i/A_i is $\chi^2_{n_i}$. Assume also that all X_i and V_i are independent. It is desired to estimate θ under the quadratic loss

$$[\sum_{i=1}^p q_i(\delta_i - \theta_i)^2]/[\sum_{i=1}^p q_i A_i]\,, \qquad \text{where} \quad q_i > 0,\, i = 1, \cdots, p\,.$$

Defining $W_i = V_i/(n_i - 2)$, $W = (W_1, \cdots, W_p)^t$, and $||X||_W{}^2 = \sum_{j=1}^p [X_j{}^2/(q_j W_j{}^2)]$, it is shown that under certain conditions on $r(X, W)$, the estimator given componentwise by

$$\delta_i(X, W) = (1 - r(X, W)/[||X||_W{}^2 q_i W_i])X_i$$

is a minimax estimator of θ. (The conditions on r require $p \geqq 3$.) A good practical version of this estimator is also given.

1. Introduction. Let $X = (X_1, \cdots, X_p)^t$ be a p-variate normal random vector with unknown mean $\theta = (\theta_1, \cdots, \theta_p)^t$ and positive definite covariance matrix Σ. Consider the problem of estimating θ, when the loss incurred in estimating θ by $\delta = (\delta_1, \cdots, \delta_p)^t$ is the quadratic loss

$$L(\delta, \theta, \Sigma) = (\delta - \theta)^t Q(\delta - \theta)/\text{tr}\,(Q\Sigma)\,.$$

Here Q is a $p \times p$ positive definite matrix and "tr" denotes the trace. Note that tr $(Q\Sigma)$ is just a normalizing constant.

The above problem has been of considerable interest since Stein (1955) demonstrated that if $Q = \Sigma = I$ (the $p \times p$ identity matrix) and if $p \geqq 3$, then the usual estimator $\delta_0(X) = X$ is inadmissible for estimating θ. Indeed he found minimax estimators which significantly improved upon the risk of δ_0. The generalization of these results to arbitrary Q and Σ was of obvious interest. For the case of known Σ, wide classes of minimax estimators have now been developed. (See Bhattacharya (1966), Hudson (1974), Berger (1976a), Bock (1975), and Berger (1975).) For unknown Σ, however, the results that have been obtained are very incomplete. For the special case $Q = \Sigma^{-1}$, James and

Received February 1975.

[1] The work of this author was partially supported by the National Science Foundation under Grant # MPS 75-07017.

[2] The work of this author was partially supported by the National Science Foundation under Grant # MPS 75-08554.

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

Key words and phrases. Minimax, unknown variances, independent normal means, quadratic loss, risk function.

Stein (1960) did obtain good minimax estimators better than δ_0. Bhattacharya (1966) obtained results for the situation $\Sigma = \sigma^2 B$, where B is a known $p \times p$ matrix and σ^2 is unknown. The subsequent literature considering unknown Σ has dealt with one or the other of the above two special situations. (See Strawderman (1973), Lin and Tsai (1973), and Efron and Morris (1976) among others.)

In this paper, a first step is made in dealing with arbitrary Q and unknown Σ. Results are obtained under the assumptions that the X_i are independent and that Q is diagonal with diagonal elements $q_i > 0$, $i = 1, \cdots, p$. Since the X_i are independent, it is clear that $\Sigma = A$, where A is an unknown $p \times p$ diagonal matrix with diagonal elements $A_i > 0$. It will be assumed that estimates V_i for A_i are available, where V_i/A_i has a chi-square distribution with n_i degrees of freedom. It will also be assumed that $n_i \geqq 3$, that all V_i are independent of V_j for $i \neq j$, and that the V_i are independent of the X_j.

Throughout the paper, $E[\quad]$ will stand for the expectation of the argument. Subscripts on E (usually θ or A) will denote parameter values under which the expectation is taken. Superscripts on E will be used to clarify the random variable with respect to which the expectation is being taken. When obvious, no subscripts or superscripts will be given.

For notational convenience, let W be the $p \times p$ diagonal matrix with diagonal elements $W_i = V_i/(n_i - 2)$. Define

$$||X||_W^2 = X^t W^{-1} Q^{-1} W^{-1} X = \sum_{i=1}^p [X_i^2/(q_i W_i^2)] .$$

Let $|x|$ denote the usual Euclidean norm of x. Finally, let $\chi^2_{n_i}$, $i = 1, \cdots, p$, denote independent chi-square random variables with n_i degrees of freedom, and define

$$T = \min_{1 \leq i \leq p} [\chi^2_{n_i}/n_i], \qquad \text{and} \qquad \tau = \tau(n_1, \cdots, n_p) = E[T^{-1}] .$$

In Section 2, it is shown that under certain conditions, estimators of the form

$$\delta(X, W) = (I - r(X, W)||X||_W^{-2} Q^{-1} W^{-1}) X \tag{1.1}$$

are minimax and have risks smaller than 1 (the risk of δ_0). Thus, in combining p independent normal mean estimation problems with unknown variances it is often possible to improve upon the risk of the usual estimator. A simple, practically significant version of the above estimator is then suggested for application.

2. A class of minimax estimators.

THEOREM 1. *Assume δ is of the form* (1.1), *where*

(i) $0 \leqq r(X, W) \leqq 2(p - 2\tau)$,
(ii) *$r(X, W)$ is nondecreasing in $|X_i|$ for $i = 1, \cdots, p$,*
(iii) *$r(X, W)$ is nonincreasing in W_i for $i = 1, \cdots, p$,*
(iv) *$r(X, W)||X||_W^{-2}$ is nondecreasing in W_i for $i = 1, \cdots, p$.*

Then δ is a minimax estimator of θ.

PROOF. Throughout the proof it will be assumed that all first order partial derivatives of r exist. The generalization to r merely nondecreasing or nonincreasing in the various coordinates can be done analagously by treating all integrals as Riemann integrals.

The risk of δ, denoted $R(\delta, \theta, A)$, is given by

$$R(\delta, \theta, A) = E_{\theta,A}L(\delta, \theta, A) = E_{\theta,A}[(\delta - \theta)^t Q(\delta - \theta)/\operatorname{tr}(QA)] .$$

Writing $[\delta - \theta]$ as $[(X - \theta) - r\|X\|_W^{-2}Q^{-1}W^{-1}X]$, and expanding the above quadratic expression, gives

$$\begin{aligned} R(\delta, \theta, A) &= E_{\theta,A}[(X - \theta)^t Q(X - \theta)/\operatorname{tr}(QA)] \\ &\quad - E_{\theta,A}[2r\|X\|_W^{-2}(X - \theta)^t W^{-1}X/\operatorname{tr}(QA)] \\ &\quad + E_{\theta,A}[r^2\|X\|_W^{-4}X^t W^{-1}Q^{-1}QQ^{-1}W^{-1}X/\operatorname{tr}(QA)] \\ &= 1 - E_{\theta,A}[2r\|X\|_W^{-2}\{\textstyle\sum_{i=1}^p X_i(X_i - \theta_i)/W_i\}/\operatorname{tr}(QA)] \\ &\quad + E_{\theta,A}[r^2\|X\|_W^{-2}/\operatorname{tr}(QA)] . \end{aligned}$$

To show that δ is minimax, it is clearly only necessary to verify that

$$E_{\theta,A}[2r\|X\|_W^{-2}\{\textstyle\sum_{i=1}^p X_i(X_i - \theta_i)/W_i\}] - E_{\theta,A}[r^2\|X\|_W^{-2}] \geqq 0 . \tag{2.1}$$

A simple integration by parts with respect to X_i gives

$$\begin{aligned} &E_{\theta,A}[\{r\|X\|_W^{-2}X_i\}\{(X_i - \theta_i)/A_i\}] \\ &\quad = E_{\theta,A}\left[\frac{\partial}{\partial X_i}(r\|X\|_W^{-2}X_i)\right] \\ &\quad = E_{\theta,A}\left[\frac{r}{\|X\|_W^2} - \frac{2rX_i^2}{\|X\|_W^4 q_i W_i^2} + \frac{X_i}{\|X\|_W^2}\left\{\frac{\partial}{\partial X_i} r(X, W)\right\}\right] . \end{aligned}$$

Using the above equality in the first term of (2.1), and noting that $[X_i(\partial/\partial X_i)r(X, W)] \geqq 0$ by assumption (ii), it is clear that δ will be proven minimax if it can be shown that

$$E_{\theta,A}\left[\left(\sum_{i=1}^p \frac{2rA_i}{\|X\|_W^2 W_i}\right) - \frac{4r}{\|X\|_W^4}\left(\sum_{i=1}^p \frac{X_i^2 A_i}{q_i W_i^3}\right) - \frac{r^2}{\|X\|_W^2}\right] \geqq 0 . \tag{2.2}$$

At this point, the following equality is needed:

$$\begin{aligned} E_{\theta,A}\left[\frac{rA_i}{\|X\|_W^2 W_i}\right] = E_{\theta,A}\Bigg[&\frac{r}{\|X\|_W^2} - \frac{4A_i r X_i^2}{(n_i - 2)\|X\|_W^4 q_i W_i^3} \\ &- \frac{2A_i}{(n_i - 2)\|X\|_W^2}\left\{\frac{\partial}{\partial W_i} r(X, W)\right\}\Bigg] . \end{aligned} \tag{2.3}$$

PROOF OF (2.3). Let U be χ_n^2, $g: R^1 \to R^1$ be an absolutely continuous function, and g' denote the derivative of g (where it exists). Efron and Morris (1976) noted that an integration by parts will prove

$$E[Ug(U)] = nE[g(U)] + 2E[Ug'(U)] , \tag{2.4}$$

providing all integrals exist and are finite. In each of the integrals of (2.4) make the change of variables $Z = cU/(n-2)$ $(c > 0)$, and define $h(Z) = g([n-2]Z/c)$. Noting that $g'(U) = ch'(Z)/(n-2)$, (2.4) becomes

$$E[(n-2)Zh(Z)/c] = nE[h(Z)] + 2E[Zh'(Z)]\,.$$

Since $W_i(n_i - 2)/A_i = V_i/A_i$ is $\chi^2_{n_i}$, it follows that

$$E_{\theta,A}[(n_i-2)W_ih(W_i)/A_i] = E_{\theta,A}[n_ih(W_i)] + 2E_{\theta,A}[W_ih'(W_i)]\,. \tag{2.5}$$

Choose $h(W_i) = r(X, W)/(||X||_W^2 W_i)$, which under the assumptions on r is absolutely continuous unless $X_i = 0$ (which of course has measure 0). Noting that

$$h'(W_i) = -\frac{r}{||X||_W^2 W_i^2} + \frac{2rX_i^2}{||X||_W^4 q_i W_i^4} + \frac{1}{||X||_W^2 W_i}\left\{\frac{\partial}{\partial W_i} r(X, W)\right\},$$

the expression (2.5) reduces to (2.3).

Inserting the expression given by (2.3) for $E_{\theta,A}[rA_i/(||X||_W^2 W_i)]$ into (2.2), and collecting terms, gives as a sufficient condition for minimaxity

$$E_{\theta,A}\left[\frac{2pr - r^2}{||X||_W^2} - \frac{4r}{||X||_W^4}\left\{\sum_{i=1}^p\left[\frac{2A_i}{(n_i-2)} + A_i\right]\frac{X_i^2}{q_iW_i^3}\right\} - \frac{4}{||X||_W^2}\left\{\sum_{i=1}^p \frac{A_i}{(n_i-2)}\left(\frac{\partial}{\partial W_i} r\right)\right\}\right] \geqq 0\,. \tag{2.6}$$

Notice that $\{(\partial/\partial W_i)r(X, W)\} \leqq 0$ by assumption (iii). Hence (2.6) will be satisfied if

$$E_{\theta,A}\left[\frac{r}{||X||_W^2}\left\{2p - r - \frac{4}{||X||_W^2}\left(\sum_{i=1}^p \frac{n_iA_iX_i^2}{[n_i-2]q_iW_i^3}\right)\right\}\right] \geqq 0\,, \tag{2.7}$$

and hence if

$$E_{\theta,A}\left[\frac{r}{||X||_W^2}\left\{2p - r - \frac{4}{||X||_W^2} \times \left(\max_{1\leqq i\leqq p} \frac{n_iA_i}{[n_i-2]W_i}\right)\left(\sum_{i=1}^p \frac{X_i^2}{q_iW_i^2}\right)\right\}\right] \geqq 0\,. \tag{2.8}$$

Since $\sum_{i=1}^p X_i^2/(q_iW_i^2) = ||X||_W^2$, it can be concluded that δ is minimax if

$$E_{\theta,A}\left[r||X||_W^{-2}\left\{2p - r - 4\left(\max_{1\leqq i\leqq p} \frac{n_iA_i}{[n_i-2]W_i}\right)\right\}\right] \geqq 0\,. \tag{2.9}$$

For notational convenience, define

$$g(W) = 4\max_{1\leqq i\leqq p}\{n_iA_i/[(n_i-2)W_i]\}\,.$$

Note that $g(W)$ is nonincreasing in W_1, and by assumption (iii), $r(X, W)$ is nonincreasing in W_1. Hence $\{2p - r(X, W) - g(W)\}$ is nondecreasing in W_1. Assumption (iv) states that $\{r(X, W)||X||_W^{-2}\}$ is also nondecreasing in W_1. Hence

$$E_{A_1}^{W_1}[r||X||_W^{-2}\{2p - r - g(W)\}] \geqq (E_{A_1}^{W_1}[r||X||_W^{-2}])(E_{A_1}^{W_1}[2p - r - g(W)])\,.$$

Since the W_i are independent, it is again clear from assumptions (iii) and (iv) that $E_{A_1}^{W_1}[r\|X\|_W^{-2}]$ and $E_{A_1}^{W_1}[2p - r - g(W)]$ are nondecreasing in W_2. Hence

$$E_{A_2}^{W_2}\{(E_{A_1}^{W_1}[r\|X\|_W^{-2}])(E_{A_1}^{W_1}[2p - r - g(W)])\} \geqq (E_{A_1,A_2}^{W_1,W_2}[r\|X\|_W^{-2}])(E_{A_1,A_2}^{W_1,W_2}[2p - r - g(W)]) .$$

Continuing in the obvious manner verifies that

$$\begin{aligned} E_{\theta,A}^X\{E_A{}^W[r\|X\|_W^{-2}(2p - r - g(W))]\} & \\ \geqq E_{\theta,A}^X\{(E_A{}^W[r\|X\|_W^{-2}])(E_A{}^W[2p - r - g(W)])\} & \\ = E_{\theta,A}^X\{(E_A{}^W[r\|X\|_W^{-2}])(2p - E_A{}^W[r] - 4\tau)\} . & \end{aligned} \tag{2.10}$$

(The last step follows since $n_i A_i/[(n_i - 2)W_i] = [V_i/(A_i n_i)]^{-1} = [\chi^2_{n_i}/n_i]^{-1}$, and hence $E_A{}^W[g(W)] = 4E[T^{-1}] = 4\tau$.) Assumption (i) ensures that $r\|X\|_W^{-2} \geqq 0$ and that $(2p - E_A{}^W[r] - 4\tau) \geqq 0$. From (2.10), it is thus clear that (2.9) is satisfied, and hence that δ is minimax. □

Obviously, unless $(p - 2\tau) > 0$, assumption (i) and hence Theorem 1 is vacuous. To calculate τ, the following formula can be used if the n_i are even:

$$\tau = \sum_{k=1}^p \sum \left\{\left[\prod_{i\neq k} \frac{m_i{}^{(m_i - j(i))}}{(m_i - j(i))!}\right] \frac{(m - J(k) - 2)!\, m_k{}^{m_k}}{(m_k - 1)!\, m^{(m - J(k) - 1)}}\right\} ,$$

where $m_i = n_i/2$, $m = \sum_{i=1}^p m_i$, $J(k) = \sum_{i\neq k} j(i)$, and the inner summation is over all combinations $\{j(1), j(2), \cdots, j(k-1), j(k+1), \cdots, j(p)\}$ where the $j(l)$ are integers between 1 and m_l inclusive. The verification of this formula is a tedious but straightforward calculation. The following theorem does show that if $p \geqq 3$ and the n_i are large enough, then indeed $(p - 2\tau) > 0$.

THEOREM 2. *Assume $p \geqq 3$. There exists an N such that if $n_i \geqq N$, $i = 1, \cdots, p$, then $(p - 2\tau) > 0$.*

PROOF. Since $p \geqq 3$, it clearly suffices to show that $\limsup_{N\to\infty} \tau(n_1, \cdots, n_p) \leqq 1$. From the definition of T, it is clear that $T^{-q} \leqq \sum_{i=1}^p (\chi^2_{n_i}/n_i)^{-q}$. For $q > 1$, Jensen's inequality thus gives

$$\tau^q = (E[T^{-1}])^q \leqq E[T^{-q}] \leqq \sum_{i=1}^p E[(\chi^2_{n_i}/n_i)^{-q}] . \tag{2.11}$$

An easy calculation shows that

$$E[(\chi^2_{n_i}/n_i)^{-q}] = (n_i/2)^q \Gamma\left(\frac{n_i}{2} - q\right)\Big/\Gamma\left(\frac{n_i}{2}\right) .$$

Together with (2.11) this gives

$$\tau \leqq \left\{\sum_{i=1}^p \left[(n_i/2)^q \Gamma\left(\frac{n_i}{2} - q\right)\Big/\Gamma\left(\frac{n_i}{2}\right)\right]\right\}^{1/q} . \tag{2.12}$$

For fixed $\varepsilon > 0$, q can be chosen large enough so that $(2p)^{1/q} < 1 + \varepsilon$. For fixed q, it is straightforward to verify by Stirling's approximation that

$$(n_i/2)^q \Gamma\left(\frac{n_i}{2} - q\right)\Big/\Gamma\left(\frac{n_i}{2}\right) \to 1 \qquad \text{as} \quad n_i \to \infty .$$

Combining these two observations with (2.12) gives the desired result. □

At this point it should be mentioned that condition (i) of Theorem 1 is undoubtedly stronger than necessary. An examination of the proof of Theorem 1, specifically the passage from (2.7) to (2.8), leads one to think that τ could be replaced by something much closer to one. Indeed, one would guess that τ could be replaced by

$$\max_{1\leq i\leq p} E(n_i/\chi^2_{n_i}) = \max_{1\leq i\leq p} [n_i/(n_i - 2)] .$$

Unfortunately, we were unable to verify any such better condition. The proof of Theorem 2 does indicate, in any case, that if the n_i are large (relative to p), then little is lost by the rougher bound.

When it comes to suggesting an estimator to use in practice, the choice determined by $r(X, W) \equiv c$, $0 \leqq c \leqq 2(p - 2\tau)$, is attractive because of its simplicity. In Berger and Bock (1975), it is shown that this simple choice can be considerably improved upon by using the "positive part" version, given componentwise by

$$\delta_i^c(X, W) = [1 - c/(||X||_W{}^2 q_i W_i)]^+ X_i .$$

(Here "+" stands for the usual positive part.) Choosing c as close as possible to $(p - 1)$ (while still preserving minimaxity) has given very attractive results in numerical studies, with the resulting estimator having a risk considerably better than the risk of the usual estimator $\delta_0(X) = X$. Typically, the improvement in risk is about 50 % at $\theta = 0$ (depending, of course, on p, Q, A, and the n_i), with the amount of improvement decreasing as $|\theta|$ gets large. The estimator should, of course, be centered at what is a priori considered to be the "most likely" parameter value, so that the major improvement in risk is obtained at this point.

Acknowledgments. The authors are grateful to the referee for greatly simplifying the proof of Theorem 2.

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. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645.

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

[4] BERGER, J. (1976a). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. *Ann. Statist.* **4** 223–226.

[5] BERGER, J. (1976b). Tail minimaxity in location vector problems and its applications. *Ann. Statist.* **4** 33–50.

[6] BERGER, J. and BOCK, M. E. (1976). Eliminating singularities of Stein-type estimators of location vectors. *J. Roy. Statist. Soc. Ser. B.*

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

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

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

[10] EFRON, B. and MORRIS, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. *Ann. Statist.* **4** 11–21.

[11] HUDSON, M. (1974). Empirical Bayes estimation. Technical Report 58, Stanford Univ.

[12] JAMES, W. and STEIN, C. (1960). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–379. Univ. of California Press.

[13] 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.

[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] STRAWDERMAN, W. E. (1973). Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. *Ann. Statist.* **1** 1189–1194.

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907