Minimax Estimation of a Normal Mean Vector for Arbitrary Quadratic Loss and Unknown Covariance Matrix

by

J. Berger[1], M. E. Bock[2], L. D. Brown[3],
G. Casella[4], and L. Gleser[4]

Purdue University and Rutgers University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #461

July, 1976

[1]Research supported by National Science Foundation Grant # MCS 76-06627.
[2]Research supported by National Science Foundation Grant # MCS 76-10817.
[3]Research supported by National Science Foundation Grant #
[4]Research supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR-72-2350C at Purdue University.

ABSTRACT

Let X be an observation from a p-variate normal distribution ($p \geq 3$) with mean vector θ and unknown positive definite covariance matrix Σ. It is desired to estimate θ under the quadratic loss $L(\delta,\theta,\Sigma) = (\delta-\theta)^t Q(\delta-\theta)/\mathrm{tr}(Q\Sigma)$, where Q is a known positive definite matrix. Estimators of the following form are considered:

$$\delta^c(X,W) = (I - c\alpha Q^{-1}W^{-1}/(X^t W^{-1} X))\, X \quad ,$$

where W is a $p \times p$ random matrix with a Wishart (Σ,n) distribution (independent of X), α is the minimum characteristic root of $(QW)/(n-p-1)$ and c is a positive constant. For appropriate values of c, δ^c is shown to be minimax and better than the usual estimator $\delta^0(X) = X$.

1. Introduction

Assume $X = (X_1,\ldots,X_p)^t$ is a p-dimensional random vector ($p\geq 3$) which is normally distributed with mean vector $\theta = (\theta_1,\ldots,\theta_p)^t$ and positive definite covariance matrix Σ. It is desired to estimate θ by an estimator $\delta = (\delta_1,\ldots,\delta_p)^t$ under the quadratic loss

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

where Q is a positive definite ($p\times p$) matrix.

The usual minimax and best invariant estimator for θ is $\delta^0(X) = X$. Since Stein (1955) first showed that δ^0 could be improved upon for $Q=\Sigma=I$ (the identity matrix), a considerable effort by a number of authors (see the references) has gone into finding significant improvements upon δ^0. For the most part these efforts have been directed towards the problems where either Σ was known (or known up to a multiplicative constant) or where $Q=\Sigma^{-1}$ (a rather unrealistic assumption). For unknown Σ only a few special situations have been considered. Berger and Bock (1976a) and (1976b) found minimax estimators (better than δ^0) for problems in which Σ was an unknown diagonal matrix or could be reduced to one. Gleser (1976) found minimax estimators under the assumption that the characteristic roots of $Q\Sigma$ have a known lower bound.

In this paper the fundamental problem of completely unknown Σ will be considered. It will be assumed that an estimate W of Σ is available, where W has a Wishart distribution with parameter Σ and n degrees of freedom, and is independent of X. Let $ch_{min}(A)$ denote the minimum characteristic root of A, and define

$$\alpha = [(n-p-1)ch_{max}(Q^{-1}W^{-1})]^{-1} = ch_{min}(QW)/(n-p-1).$$

The estimators considered in this paper will be of the form

$$\delta^c(X,W) = \left(I - \frac{c\alpha Q^{-1}W^{-1}}{X^tW^{-1}X}\right)X \quad , \tag{1.1}$$

where c is a positive constant. For known Σ, estimators of this form (with $(n-p-1)W^{-1}$

replaced by $\ddagger^{-1}$) were shown to be minimax in Bock (1974) and Berger (1976b), providing $0 \le c \le 2(p-2)$. In this paper δ^c is shown to be minimax for

$$0 \le c \le c_{n,p} \quad ,$$

where the $c_{n,p}$ are solutions to equation (2.17), and are numerically calculated in Table 1 for certain values of n and p.

Table 1

Values of $c_{n,p}$

p \ n	8	10	12	14	16	18	20	25	30
3	.14	.41	.72	.88	1.03	1.10	1.23	1.51	1.53
4	.65	1.37	1.88	2.27	2.42	2.60	2.81	3.07	3.12
5		1.83	2.85	3.37	3.80	4.02	4.26	4.78	4.87
6		1.71	3.32	4.27	4.81	5.33	5.66	6.36	6.50
7			3.42	4.99	5.78	6.42	6.96	7.92	8.14
8			2.50	5.15	6.57	7.64	8.19	9.24	9.84
9				4.50	7.02	8.40	9.22	10.60	11.28
10				2.61	6.79	8.90	10.25	11.98	12.84
11					5.78	9.15	10.84	13.14	14.24
12					2.73	8.42	11.10	14.20	15.65
13						7.11	11.09	15.48	17.15
14						2.43	9.70	15.74	18.44
15							7.93	16.61	19.51
16							2.26	16.67	20.62
17								16.67	21.56
18								16.34	22.38
19									22.83
20									23.47

2. Minimaxity of δ^c

The notation E(Z) will be used for the expectation of Z. Subscripts on E will refer to parameter values, while superscripts on E will refer to the random variables with respect to which the expectation is to be taken. When obvious, subscripts and superscripts will be omitted.

For an estimator, δ, define the risk function

$$R(\delta,\theta,\ddagger) = E^{X,W}_{\theta,\ddagger}\,[L(\delta(X,W),\theta,\ddagger)] \quad .$$

For notational convenience define $n^* = (n-p-1)$ and

$$\Delta_c = \Delta_c(\theta,\ddagger) = tr(Q\ddagger)[R(\delta^c,\theta,\ddagger) - R(\delta^0,\theta,\ddagger)] \quad .$$

The estimator δ^c is clearly minimax (and as good as or better than δ^0) providing $\Delta_c(\theta,\ddagger) \le 0$ for all θ and $\ddagger$.

Expanding the quadratic loss L for δ^c verifies that

$$\text{(2.1)} \qquad \Delta_c = -2E\left[\frac{c\alpha(X-\theta)^t W^{-1} X}{X^t W^{-1} X} \right] + E\left[\frac{c^2\alpha^2 X^t W^{-1} Q^{-1} W^{-1} X}{(X^t W^{-1} X)^2} \right] \quad .$$

As in Berger (1976b) an integration by parts with respect to the X_i gives

$$E\left[\frac{(X-\theta)^t W^{-1} X}{X^t W^{-1} X} \right] = E\left[\frac{tr(\ddagger W^{-1})}{X^t W^{-1} X} - \frac{2X^t W^{-1} \ddagger W^{-1} X}{(X^t W^{-1} X)^2} \right] \quad .$$

Thus (2.1) becomes

$$\text{(2.2)} \qquad \Delta_c = -E\left[\frac{c\alpha}{(X^t W^{-1} X)} \left\{ 2tr(\ddagger W^{-1}) - \frac{4X^t W^{-1} \ddagger W^{-1} X}{X^t W^{-1} X} - \frac{c\alpha X^t W^{-1} Q^{-1} W^{-1} X}{X^t W^{-1} X} \right\} \right] \quad .$$

Note that

$$\frac{\alpha X^t W^{-1} Q^{-1} W^{-1} X}{X^t W^{-1} X} \le \frac{\alpha}{ch_{min}(QW)} = \frac{1}{n^*} \quad .$$

Using this in (2.2) gives

$$\text{(2.3)} \qquad \Delta_c \le -E\left[\frac{\alpha c}{(X^t W^{-1} X)} \left\{ 2tr(\ddagger W^{-1}) - \frac{4X^t W^{-1} \ddagger W^{-1} X}{X^t W^{-1} X} - \frac{c}{n^*} \right\} \right] \quad .$$

In this expression, perform the change of variables

$$Y = \ddagger^{-\frac{1}{2}}X \quad , \quad V = \ddagger^{-\frac{1}{2}}W\ddagger^{-\frac{1}{2}} \quad .$$

Note that V is now Wishart with parameter I and n degrees of freedom, and that $\alpha = ch_{min}(\ddagger^{\frac{1}{2}}Q\ddagger^{\frac{1}{2}}V)/n*$ Clearly (2.3) becomes

$$(2.4) \qquad \Delta_c \le -E[\frac{\alpha c}{(Y^tV^{-1}Y)}\{2tr(V^{-1}) - \frac{4Y^tV^{-2}Y}{Y^tV^{-1}Y} - \frac{c}{n*}\}] \quad .$$

For convenience, define

$$\beta = ch_{min}(Q\ddagger) \quad , \quad Z = Y/|Y| \quad , \quad \text{and } \ddagger^* = \ddagger^{\frac{1}{2}}Q\ddagger^{\frac{1}{2}}/\beta \quad .$$

Note that $ch_{min}(\ddagger^*) = 1$. Line (2.4) can then be rewritten

$$(2.5) \quad \Delta_c \le \frac{-\beta c}{n*} \; E^Y[\frac{1}{|Y|^2} E^V\{\frac{ch_{min}(\ddagger^*V)}{(Z^tV^{-1}Z)} (2tr(V^{-1}) - \frac{4Z^tV^{-2}Z}{Z^tV^{-1}Z} - \frac{c}{n*})\}] \quad .$$

To show that $\Delta_c \le 0$ it suffices to show for all $Z \in U_p$ (the unit p-sphere) and all $\ddagger^*$ with $ch_{min}(\ddagger^*) = 1$, that the following inequality holds:

$$(2.6) \qquad E^V\{\frac{ch_{min}(\ddagger^*V)}{(Z^tV^{-1}Z)}[2tr\,(V^{-1}) - \frac{4Z^tV^{-2}Z}{Z^tV^{-1}Z} - \frac{c}{n*})\} \ge 0 \quad .$$

(Note that the distribution of V does not depend on Z or on $\ddagger^*$.)

Let Γ be a p×p orthogonal matrix such that $\Gamma Z = (1,0,\ldots,0)^t$. Define $V^* = \Gamma V \Gamma^t$ and $\ddagger_Z = \Gamma\ddagger^*\Gamma^t$. Clearly V^* is also Wishart (I) and $ch_{min}(\ddagger_Z) = 1$. For convenience, let v_1 denote the (1,1) element of $(V^*)^{-1}$, v_2 denote the (1,1) element of $(V^*)^{-2}$, and let

$$\rho(V^*) = [2tr\{(V^*)^{-1}\} - 4v_2/v_1] \quad .$$

It is straightforward to verify that under the above change of variables for V, (2.6) becomes

$$(2.7) \qquad E^{V^*}\{\frac{ch_{min}(\ddagger_Z V^*)}{v_1}[\rho(V^*) - \frac{c}{n*}]\} \ge 0 \quad .$$

Since $ch_{min}(\ddagger_Z) = 1$, it is clear that

(2.8) $$ch_{min}(\ddagger_Z V^*) \geq ch_{min}(V^*) \quad .$$

Also if $a \in U_p$ (i.e. $|a| = 1$) then

$$ch_{min}(\ddagger_Z V^*) \leq a^t \ddagger_Z^{\frac{1}{2}} V^* \ddagger_Z^{\frac{1}{2}} a \quad .$$

Choosing a to be a^1, the characteristic vector of the root 1 of $\ddagger_Z^{\frac{1}{2}}$, it follows that

(2.9) $$ch_{min}(\ddagger_Z V^*) \leq (a^1)^t V^* a^1 \quad .$$

For convenience define

$$\Omega_c = \{V^*: \ \rho(V^*) < c/n^*\} \quad ,$$

let $\bar{\Omega}_c$ denote the complement of Ω_c, and let $I_A(V^*)$ denote the usual indicator function on A. Using (2.8) and (2.9) it then follows that (2.7) will hold (and δ^c will be minimax) if

(2.10) $$E^{V^*} \{\frac{(a^1)^t V^* a^1}{v_1} [\rho(V^*) - \frac{c}{n^*}] I_{\Omega_c}(V^*) + \frac{ch_{min}(V^*)}{v_1} [\rho(V^*) - \frac{c}{n^*}] I_{\bar{\Omega}_c}(V^*)\} \geq 0$$

for all $a^1 \in U_p$.

To simplify this expression further, let

$$T = \begin{pmatrix} 1 & 0 \ldots 0 \\ 0 & \\ \vdots & S \\ 0 & \end{pmatrix} \quad ,$$

where S is a $(p-1)\times(p-1)$ orthogonal matrix such that

$$Ta^1 = (b, (1-b^2)^{\frac{1}{2}}, 0, \ldots, 0)^t \qquad (-1 \leq b \leq 1).$$

In (2.10), performing the change of variables $V = TV^*T^t$ (again Wishart (I)) then gives as the condition for minimaxity

(2.11) $$E^V \{ \frac{(Ta^1)^t V (Ta^1)}{v_1} [\rho(V) - \frac{c}{n^*}] I_{\Omega_c}(V) + \frac{ch_{min}(V)}{v_1} [\rho(V) - \frac{c}{n^*}] I_{\bar{\Omega}_c}(V) \} \geq 0$$

for all $a^1 \in U_p$.

(Note that $v_1 = (V^{*-1})_{11} = (T^t V^{-1} T)_{11} = (V^{-1})_{11}$ and likewise $v_2 = (V^{-2})_{11}$.)
The inequality (2.11) can be rewritten

$$(2.12) \qquad c \le \frac{n^* E^V \{ \rho(V) v_1^{-1} [(Ta^1)^t V (Ta^1) I_{\Omega_c}(V) + ch_{min}(V) I_{\overline{\Omega}_c}(V)]\}}{E^V \{v_1^{-1} [(Ta^1)^t V (Ta^1) I_{\Omega_c}(V) + ch_{min}(V) I_{\overline{\Omega}_c}(V)]\}} .$$

Note that

$$(Ta^1)^t V (Ta^1) = b^2 (V_{11} - V_{22}) + b(1-b^2)^{\frac{1}{2}} (V_{12} + V_{21}) + V_{22} .$$

Hence defining

$$\tau_0(c) = E^V\{\rho(V) v_1^{-1} [V_{22} I_{\Omega_c}(V) + ch_{min}(V) I_{\overline{\Omega}_c}(V)]\} ,$$

$$\tau_1(c) = E^V\{\rho(V) v_1^{-1} (V_{11} - V_{22}) I_{\Omega_c}(V)\} ,$$

$$\tau_2(c) = E^V\{\rho(V) v_1^{-1} (V_{12} + V_{21}) I_{\Omega_c}(V)\} ,$$

$$\tau_0'(c) = E^V\{v_1^{-1} [V_{22} I_{\Omega_c}(V) + ch_{min}(V) I_{\overline{\Omega}_c}(V)]\} ,$$

$$\tau_1'(c) = E^V\{v_1^{-1} (V_{11} - V_{22}) I_{\Omega_c}(V)\} , \text{ and}$$

$$\tau_2'(c) = E^V\{v_1^{-1} (V_{12} + V_{21}) I_{\Omega_c}(V)\} ,$$

it is clear that (2.12), the condition for minimaxity, can be rewritten

$$(2.13) \qquad c \le \frac{n^*[\tau_0(c) + \tau_1(c) b^2 + \tau_2(c) b (1-b^2)^{\frac{1}{2}}]}{\tau_0'(c) + \tau_1'(c) b^2 + \tau_2' b (1-b^2)^{\frac{1}{2}}}$$

for all $-1 \le b \le 1$. Finally, defining $\tilde{b} = (b, (1-b^2)^{\frac{1}{2}})$

$$A(c) = \begin{pmatrix} \tau_0(c) + \tau_1(c) & \tau_2(c)/2 \\ \tau_2(c)/2 & \tau_0(c) \end{pmatrix} , \text{ and } \quad B(c) = \begin{pmatrix} \tau_0'(c) + \tau_1'(c) & \tau_2'(c)/2 \\ \tau_2'(c)/2 & \tau_0'(c) \end{pmatrix}$$

line (2.13) becomes

$$(2.14) \qquad c \le \frac{n^* \tilde{b}^t A(c) \tilde{b}}{\tilde{b}^t B(c) \tilde{b}} .$$

Now for fixed $\tilde{b}$, the nonnegative solutions to (2.14) lie in an interval $0 \le c \le c_{\tilde{b}}$. This can most easily be seen by looking at (2.11) (an expression equivalent to (2.14)) and noting that the left hand side is decreasing in c. Thus defining

$$c_{n,p} = \inf_{-1 \le \tilde{b} \le 1} c_{\tilde{b}} \quad ,$$

it follows that if

$$(2.15) \qquad 0 \le c \le c_{n,p}$$

then (2.14) will be satisfied for all $-1 \le b \le 1$, and hence δ^c will be minimax.

To get a more explicit equation for $c_{n,p}$, note from equation (2.12) (an equivalent expression to (2.14)) that B(c) is positive definite. Hence if (2.14) holds for all $-1 \le b \le 1$, then

$$(2.16) \qquad c \le n^* ch_{min}[B(c)^{-1}A(c)] \quad .$$

Thus (2.15)$\Rightarrow$ (2.14) for all $-1 \le b \le 1 \Rightarrow$ (2.16). It is also clear that the reverse implications hold, so that

$$\{c: \ 0 \le c \le c_{n,p}\} = \{c: \ c \le n^* ch_{min}[B(c)^{-1}A(c)]\} \quad .$$

It is also easy to check that

$$c_{n,p} = n^* ch_{min}[B(c_{n,p})^{-1}A(c_{n,p})] \quad ,$$

$$c < n^* ch_{min}[B(c)^{-1}A(c)] \qquad \text{if} \qquad 0 \le c < c_{n,p} \quad ,$$

and

$$c > n^* ch_{min}[B(c)^{-1}A(c)] \qquad \text{if} \qquad c > c_{n,p} \quad .$$

Hence $c_{n,p}$ is the unique solution to

$$(2.17) \qquad c = n^* ch_{min}(B(c)^{-1}A(c)) \quad .$$

As there appeared to be little hope of analytically obtaining solutions to (2.17), the computer was used to numerically compute the solutions. For a given n and p, the values of the $\tau_i(c)$ and $\tau_i'(c)$ (and hence A(c) and B(c)) were calculated by monte carlo methods using 4000 generations of V (for n=8) to 1000 generations of V (for n=30). (Unfortunately a larger number of generations could not be used due to the considerable expense of generating V and performing the calculations involving V^{-1}.) The resulting estimated solutions, $c_{n,p}$, to (2.17) were then found and are listed in Table 1. The standard deviations of these simulated solutions ranged from about .02 (for p=3) to about .1 (for n-p = 4).

3. Comments

1. The values $c_{n,p}$ are not the largest values of c for which δ^c is minimax. Approximations were made in the proof (lines (2.8) and (2.9)) which resulted in a smaller than necessary upper bound. If one could somehow determine the "least favorable" matrix $\ddagger_Z$ in (2.7), the approximations could be eliminated and the largest possible value of c obtained.

2. The estimators δ^c have a singularity as $X\to 0$. There are numerous ways of eliminating the singularity, one of the simplest being used in the following estimator:

$$\delta^{*c}(X,W) = \left(I - \frac{\min(n^* X^t W^{-1} X, c)\alpha Q^{-1} W^{-1}}{X^t W^{-1} X}\right)X \quad .$$

Through analogy with the known $\ddagger$ situation, it seems quite likely that δ^{*c} is itself minimax (for $0 \le c \le c_{n,p}$) and considerably better than δ^c.

3. If the linear restriction $R\theta = r^0$ is thought to hold, where R is an $(m\times p)$ matrix of rank m and r^0 is an $(m\times 1)$ vector, then the estimators δ^c and δ^{*c} can be modified so that their regions of significant risk improvement coincide with the linear restriction. Indeed, defining $Y = RX - r^0$, $W^* = RWR^t$, and $\alpha^* = ch_{min}[(RQ^{-1}R^t)^{-1}W^*]/(n-m-1)$, Theorem 2 of Berger and Bock (1976b) can be used to show that

$$\delta_R^c = X - c\alpha^* Q^{-1} R^t (W^*)^{-1} Y / [Y^t (W^*)^{-1} Y]$$

is minimax if $0 \le c \le c_{n,m}$. The appropriate modification of δ^{*c} is the above estimator with c replaced by $\min\{(n-m-1)Y^t(W^*)^{-1}Y, c\}$.

4. If $(Q\ddagger)$ has a characteristic root considerably smaller than the other characteristic roots, then $ch_{min}(Q\ddagger)$ will be small compared to $tr(Q\ddagger)$. From the definition of $\Delta_c(\theta,\ddagger)$ and line (2.2), it is apparent that the improvement obtained in using δ^c will be quite small. The estimator, δ^c, will therefore perform best when $(Q\ddagger)$ has no exceptionally small roots. (If it is suspected that a coordinate X_i might give rise to an exceptionally small root of $(Q\ddagger)$, it would probably pay to eliminate that coordinate in the construction of δ^c, providing of course that there are at least three coordinates left.)

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. Statist. 41, 642-645.

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

[4] Berger, J. (1976b). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Multivariate Anal. 6, No. 2, 1976.

[5] Berger, J. and Bock, M. E. (1976a). Combining independent normal mean estimation problems with unknown variances. Ann. Statist. 4, 642-648.

[6] Berger, J. and Bock, M. E. (1976b). Improved minimax estimators of normal mean vectors for certain types of covariance matrices. To appear in S. S. Gupta and D. S. Moore (Eds.) Statistical Decision Theory and Related Topics II, Academic Press.

[7] Berger, J. and Bock, M. E. (1976c). Eliminating singularities of Stein-type estimators of location vectors. J. Roy. Statist. Soc., B, 38, 166-170.

[8] Bhattacharya, P. K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. Ann. Math. Statist. 37, 1819-1927.

[9] Bock, M. E. (1974). Certain minimax estimators of the mean of a multivariate normal distribution. Ph.D. Thesis, Department of Mathematics, University of Illinois.

[10] Bock, M. E. (1975). Minimax estimators of the mean of a multivariate normal distribution. Ann. Statist. 3, 209-218.

[11] Efron, B. and Morris, C. (1973). Stein's estimation rule and its competitors - an empirical Bayes approach. J. Amer. Statist. Assoc. 68, 117-130.

[12] Efron, B. and Morris, C. (1976). Families of minimax estimators of the mean of a multivariate normal distribution. Ann. Statist. 4, 11-21.

[13] Gleser, Leon Jay. (1976). Minimax estimation of a multivariate normal mean with unknown covariance matrix. Technical Report #460, Purdue University.

[14] Hudson, M. (1974). Empirical Bayes estimation. Technical Report #58, Stanford University.

[15] James, W. and Stein, C. (1960). Estimation with quadratic loss. Proc. Fourth Berkeley Symposium Math. Stat. Prob. 1, 361-379. University of California Press.

[16] 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.

[17] Shinozaki, N. (1974). A note on estimating the mean vector of a multivariate normal distribution with general quadratic loss function. Keio Engineering Reports Vol. 27, No. 7, 105-112.

[18] 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. University of California Press.

[19] Stein, C. (1973). Estimation of the mean of a multivariate distribution Proc. Prague Symp. Asymptotic Statist. 345-381.

[20] 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.