# Minimax Estimation of a Multivariate Normal Mean with Unknown Covariance Matrix*

Leon Jay Gleser
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #460

July 1976

*Research supported in part by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR - 72-2350 C at Purdue University.

# Minimax Estimation of a Normal Mean

Leon Jay Gleser
Department of Statistics
Mathematical Sciences Bldg.
Purdue University
West Lafayette, Indiana 47907

# Minimax Estimation of a Multivariate Normal Mean with Unknown Covariance Matrix

by

Leon Jay Gleser
Purdue University

ABSTRACT

Let x be a p-variate ($p \geq 3$) vector, normally distributed with unknown mean $\theta$ and unknown covariance matrix $\Sigma$. Let $W: p \times p$ be distributed independently of x, and let W have a Wishart distribution with n degrees of freedom and parameter $\Sigma$. It is desired to estimate $\theta$ under the quadratic loss $(\delta-\theta)'Q(\delta-\theta)$, where Q is a known positive definite matrix. Under the condition that a lower bound for the smallest characteristic root of $Q\Sigma$ is known, a family of minimax estimators is developed.

AMS 1970 Subject classification: Primary 62 C 99; secondary 62 F 10, 62 H 99.

Key words and phrases: Multivariate normal distribution, unknown covariance matrix, estimation of mean vector, quadratic loss, minimax estimator.

# Minimax Estimation of a Multivariate Normal Mean with Unknown Covariance Matrix

by

Leon Jay Gleser
Purdue University

## 1. INTRODUCTION

Let $x:p\times 1$ be a normally distributed random vector with unknown mean $\theta$ and unknown covariance matrix $\Sigma$. Assume that we have an independent estimator $\hat{\Sigma} = n^{-1} W$ of $\Sigma$, where $W: p\times p$ has a Wishart distribution with n degrees of freedom and parameter $\Sigma = n^{-1}E(W)$. In the usual notation,

$$x \sim N(\theta,\Sigma) \;,\; W \sim \mathcal{W}_p(n,\Sigma). \tag{1}$$

We wish to estimate $\theta$ with an estimator $\delta(x,W)$ subject to the quadratic loss function

$$L(\delta,\theta,\Sigma) = (\delta-\theta)'Q(\delta-\theta)/tr(Q\Sigma) \tag{2}$$

Here, Q is a known $p\times p$ positive definite matrix, and tr(A) denotes the trace of the matrix A. Note that $tr(Q\Sigma)$ is just a normalizing constant, chosen to give the estimator $\delta_0(x,W) = x$ constant risk. It is well known that $\delta_0$ is a minimax estimator for this problem.

The limiting case of this problem where $\Sigma$ is completely known (corresponding here to $n = \infty$) has recently received a good deal of attention. [See Berger [1] for references.] The problem with $\Sigma$ unknown and $Q = \Sigma^{-1}$ (which is <u>not</u> a special case for our problem because $Q = \Sigma^{-1}$ cannot be known) has also been studied by James and Stein [5], Lin and Tsai [6], Bock [2], and Efron and Morris [3,4], among others. However, the assumption that $Q = \Sigma^{-1}$ is rather artificial (it seems to be motivated only by invariance arguments), and does not seem to be of practical importance. A possibly more reasonable assumption to make relating Q and $\Sigma$ is that something is known about the characteristic roots of $Q\Sigma$. [Note that if $Q = \Sigma^{-1}$, all of the characteristic roots of $Q\Sigma$ are equal to 1.] In the present paper, we assume that there exists a known constant $K > 0$ such that

$$ch_p(Q\Sigma) \geq K \text{ , all } \Sigma > 0, \tag{3}$$

where

$$ch_1(A) \geq ch_2(A) \geq \dots \geq ch_p(A)$$

denote the ordered characteristic roots of the $p\times p$ symmetric matrix $A$.

We consider estimators of the form

$$\delta_h(x,W) = (I_p - h(x'W^{-1}x)Q^{-1}W^{-1})x, \tag{4}$$

where $h(u)$ is an absolutely continuous function on $[0,\infty)$. Our main result, which is proven in Section 2, is the following.

THEOREM 1. *If (3) holds, then any estimator of the form (4) for which*

(i) *$u\,h(u)$ is nondecreasing in $u$,*

(ii) *$0 \leq h(u) \leq 2(p-2)(n-p)Ku/(n-1)$, all $u \geq 0$,*

*dominates $\delta_0(x,W) = x$ in risk, and hence is minimax.*

It is clearly of interest to determine what happens to estimators of the form (4) when the bound (3) can be violated. In Section 3 it is shown that when (3) does not hold, no estimator of the form (4) can be minimax. [Bock [2] has previously shown that for $Q = I_p$, no estimator of the form $h(x'W^{-1}x)x$ can be minimax.] It is conjectured that members of a certain family (see (36)) of estimators closely resembling the estimators (4) in form may be minimax, but no proof of this result is given.

## 2. PROOF OF THEOREM 1

Let

$$\Delta(\theta,\Sigma) = \mathrm{tr}(Q\Sigma)E[L(\delta_h,\theta,\Sigma) - L(\delta_0,\theta,\Sigma)]. \tag{5}$$

Clearly if $\Delta(\theta,\Sigma) \leq 0$, all $\theta$, all $\Sigma$ satisfying (3), then $\delta_h$ is minimax for our problem.

Using the fact that $a'Qa - b'Qb = (a-b)'Q(a+b)$, the fact that $\delta_0(x,W) = x$, and (4), we obtain

$$\Delta(\theta,\Sigma) = E[h^2(x'W^{-1}x)x'W^{-1}Q^{-1}W^{-1}x]-2E[h(x'W^{-1}x)x'W^{-1}(x-\theta)]. \tag{6}$$

Note that for any functions $g(x,W)$ for which $Eg(x,W)$ exists, we may write

$$E[g(x,W)] = E_W\left\{E_{x|W}[g(x,W)]\right\} = E_W\left\{E_x\ [g(x,W)]\right\}, \tag{7}$$

where $E_{x|W}[g(x,W)]$ denotes expectation over the conditional distribution of x given W, and $E_W$ and $E_x$ denote expectations over the marginal distributions of W and x respectively. The last equality in (7) holds since x and W are statistically independent. Further, using integration by parts term by term in the elements of x (with W treated as a fixed matrix), it can be shown (see Berger [1]) that

$$E_x[h(x'W^{-1}x)x'W^{-1}(x-\theta)] = E_x[h(x'W^{-1}x)\ \mathrm{tr}W^{-1}\ ] + 2E_x[h^{(1)}(x'W^{-1}x)\, x'W^{-1}\Sigma W^{-1}x], \tag{8}$$

where $h^{(1)}(u) = dh(u)/du$. [Note: We are assuming that $h(u)$ is differentiable; if not, a similar argument, using Riemann integration, produces a corresponding result; see Berger [1].]

From (6), (7), and (8), we have

$$\Delta(\theta,\Sigma) = E[h^2(x'W^{-1}x)x'W^{-1}Q^{-1}W^{-1}x-2h(x'W^{-1}x)\mathrm{tr}W^{-1}\Sigma-4h^{(1)}(x'W^{-1}x)\, x'W^{-1}\Sigma W^{-1}x]. \tag{9}$$

We now find a canonical representation for (9). Make the change of variables

$$y = \Sigma^{-1/2}x, \qquad V = \Sigma^{-1/2}W\Sigma^{-1/2}, \tag{10}$$

where $\Sigma^{1/2}$ is any square root of $\Sigma$. Then

$$y \sim N(\eta, I_p), \quad V \sim \mathcal{W}_p(n, I_p), \tag{11}$$

where $\eta = \Sigma^{-1/2}\theta$. Further, y and V are statistically independent. From (9) and (10), with

$$Q^* = \Sigma^{1/2} Q \Sigma^{1/2},$$

and using arguments and notation analagous to that used to obtain (7), we have

$$\Delta(\theta,\Sigma) = E_y E_V[h^2(y'V^{-1}y)\; y'V^{-1}(Q^*)^{-1}V^{-1}y - 2h(y'V^{-1}y)\operatorname{tr}V^{-1} \qquad (1)$$
$$-4h^{(1)}(y'V^{-1}y)y'V^{-2}y] \; . \tag{12}$$

Let $\Gamma_y$ be p×p orthogonal with first row equal to $(y'y)^{-1/2}y'$. Let

$$U = \Gamma_y V \Gamma_y' \;, \quad Q_y^* = \Gamma_y Q^* \Gamma_y'. \tag{13}$$

Then, given y, $U \sim \mathcal{W}_p(n, I_p)$, so that U and y are statistically independent. Partition U as

$$U = \begin{pmatrix} u_{11} & u_{21}' \\ u_{21} & U_{22} \end{pmatrix}, \qquad u_{11}: 1\times 1,\; U_{22}: (p-1)\times(p-1),$$

and let

$$s = u_{11} - u_{21}'U_{22}^{-1}u_{21}, \quad t = U_{22}^{-1/2} \tag{14}$$

where $U_{22}{}^{1/2}$ is any square root of $U_{22}$. It is well known that s, t, and $U_{22}$ are statistically independent, with

$$s \sim \chi^2_{n-p+1},\; t \sim N(0, I_{p-1}),\; U_{22} \sim \mathcal{W}_{p-1}(n, I_{p-1}). \tag{15}$$

Further, $V^{-1} = \Gamma_y' U^{-1}\Gamma_y$ and

$$U^{-1} = s^{-1}\begin{pmatrix} 1 & -t'U_{22}^{-1/2} \\ -U_{22}^{-1/2}t & U_{22}^{-1/2}(sI_{p-1}+tt')U_{22}^{-1/2} \end{pmatrix}, \tag{16}$$

so that

$$y'V^{-1}y = s^{-1}y'y,\; y'V^{-2}y = s^{-2}y'y(1+t'U_{22}^{-1}t), \tag{17}$$

$$\mathrm{tr}V^{-1} = \mathrm{tr}U^{-1} = s^{-1}\ (1+t'U_{22}^{-1}t) + \mathrm{tr}U_{22}^{-1}, \tag{18}$$

and

$$y'V^{-1}(Q^{*})^{-1}V^{-1}y = s^{-2}y'y(1,-t'U_{22}^{-1/2})(Q_y^{*})^{-1}(1,-t'U_{22}^{-1/2})'. \tag{19}$$

Under the distributional assumptions given in (15), it is known that $E(U_{22}^{-1}) = (n-p)^{-1}I_{p-1}$, so that

$$E\mathrm{tr}U_{22}^{-1} = \mathrm{tr}\ EU_{22}^{-1} = (n-p)^{-1}(p-1). \tag{20}$$

For any constant matrix A,

$$\begin{aligned} &E[(1,-t'U_{22}^{-1/2})A(1,-t'U_{22}^{-1/2})'] \\ &= E_{U_{22}}E_t\left\{\mathrm{tr}\left[A\begin{pmatrix}1 & -t'U_{22}^{-1/2}\\ -U_{22}^{-1/2}t & U_{22}^{-1/2}tt'U_{22}^{-1/2}\end{pmatrix}\right]\right\} \\ &= E_{U_{22}}\mathrm{tr}\left[A\begin{pmatrix}1 & 0\\ 0 & U_{22}^{-1}\end{pmatrix}\right] \\ &= \mathrm{tr}\left[A\begin{pmatrix}1 & 0\\ 0 & (n-p)^{-1}I_{p-1}\end{pmatrix}\right]. \end{aligned} \tag{21}$$

Taking $A = I_p$, the result (21) allows us to verify that

$$E(1 + t'U_{22}^{-1}t) = (n-p)^{-1}(n-1). \tag{22}$$

Taking $A = (Q_y^{*})^{-1}$, the result (21) yields

$$\begin{aligned} &E[(1,-t'U_{22}^{-1/2})(Q_y^{*})^{-1}(1,-t'U_{22}^{-1/2})'] \\ &= \mathrm{tr}(Q_y^{*})^{-1}\begin{pmatrix}1 & 0\\ 0 & (n-p)^{-1}I_{p-1}\end{pmatrix}. \end{aligned} \tag{23}$$

If in (12) we make the change of variables (13) and (14), and take account of the identities (17), (18), and (19), then by taking our expected values in the order $E_yE_sE_{t,U_{22}}$, and using (20), (22), and (23), we obtain

$$\Delta(\theta,\Sigma) = (n-p)^{-1}E_yE_s[h^2(s^{-1}y'y)s^{-2}y'y\ \tau(y,Q^*)$$
$$-2h(s^{-1}y'y)s^{-1}(n-1)-2h(s^{-1}y'y)(p-1)$$
$$-4h^{(1)}(s^{-1}y'y)s^{-2}y'y(n-1)], \qquad (24)$$

where

$$\tau(y,Q^*) = \mathrm{tr}(Q_y^*)^{-1}\begin{pmatrix} n-p & 0 \\ 0 & I_{p-1}\end{pmatrix}$$
$$= (n-p-1)(y'y)^{-1}y'(Q^*)^{-1}y + \mathrm{tr}(Q^*)^{-1}. \qquad (25)$$

Finally, integrating by parts in s, we can show that

$$E_sh(s^{-1}y'y) = (n-p-1)E_s[s^{-1}h(s^{-1}y'y)] - 2E_s[s^{-2}y'yh^{(1)}(s^{-1}y'y)], \qquad (26)$$

which, when substituted in (24), yields the expression

$$\Delta(\theta,\Sigma) = (n-p)^{-1}E_yE_s[h^2(s^{-1}y'y)s^{-2}y'y\tau(y,Q^*) - 2p(n-p)s^{-1}h(s^{-1}y'y)$$
$$-4(n-p)h^{(1)}(s^{-1}y'y)s^{-2}y'y], \qquad (27)$$

where

$$y \sim N(\eta, I_p),\ s \sim \chi^2_{n-p+1},$$

y and s are independent, $\eta = \Sigma^{-1/2}\theta$, $Q^* = \Sigma^{1/2}Q\Sigma^{1/2}$, and $\tau(y,Q^*)$ is given by (25). The expression (27) is the desired cononical form.

Now, we are ready to complete the proof of Theorem 1.

Let

$$r(u) = uh(u), \qquad (28)$$

and note that

$$h^{(1)}(u) = \frac{r^{(1)}(u)}{u} - \frac{r(u)}{u^2}, \qquad (29)$$

where $r^{(1)}(u) = dr(u)/du$. Substituting in (27), we obtain

$$\Delta(\theta,\Sigma) = (n-p)^{-1}E_y\{(y'y)^{-1}E_s[r^2(s^{-1}y'y)\tau(y,Q^*) - 2(p-2)(n-p)r(s^{-1}y'y)$$
$$-4(n-p)s^{-1}r^{(1)}(s^{-1}y'y)]\}$$
$$\leq (n-p)^{-1}E_yE_s\left[\frac{r(s^{-1}y'y)}{y'y}\ (\tau(y,Q^*)r(s^{-1}y'y) - 2(p-2)(n-p))\right], \qquad (30)$$

since, by assumption (i) of Theorem 1, $r(u)$ is nondecreasing in u. Note from (3) and (25) that

$$\tau(y,Q^*) \le (n-1)ch_1[(Q^*)^{-1}] \le (n-1)[ch_p(Q\Sigma)]^{-1} \le (n-1)\ K^{-1}. \tag{31}$$

Thus, applying assumption (ii) of Theorem 1, (30), and (31), we conclude that for all satisfying (3),

$$\Delta(\theta,\Sigma) \le 0, \quad \text{all } \theta.$$

This completes the proof of Theorem 1.1.

We remark that our proof actually demonstrates the following.

THEOREM 2. *Let an estimator $\delta_h(x,W)$ of the form (4) satisfy*

(i) *$u\,h(u)$ is nondecreasing in u,*

(ii) *$0 \le h(u) \le 2(p-2)(n-p)Lu$, all $u \ge 0$,*

*where $L > 0$ is a given constant. Then if $\Sigma$ satisfies*

$$(n-p-1)(ch_p(Q\Sigma))^{-1} + tr(Q\Sigma)^{-1} \le L^{-1}, \tag{32}$$

*we have*

$$\Delta(\theta,\Sigma) \le 0, \ \text{all } \theta,$$

*and $\delta_h(x,W)$ is minimax.*

Although Theorem 2 is more general than Theorem 1, the additional generality is unlikely to be of practical importance.

## 3. THE CASE WHERE Σ IS COMPLETELY UNRESTRICTED

When $\Sigma$ is unrestricted, and (3) need not hold, then $\delta_0(x,W)$ is essentially the only estimator of the form (4) that can be minimax.

THEOREM 3. *When $\Sigma$ is unrestricted, no estimator of the form $\delta_h(x,W) = (I_p - h(x'W^{-1}x)Q^{-1}W^{-1})x$ can be minimax unless $h(u) = 0$ for almost all $u \ge 0$.*

*Proof*. Note from (25) that

$$\tau(y,Q^*) \ge tr(Q^*)^{-1}, \qquad \text{for all } y. \tag{33}$$

Now from (33) and (27),

$$\Delta(\theta,\Sigma) \geq \mathrm{tr}(Q^*)^{-1}E[h^2(s^{-1}y'y)s^{-2}y'y] \\ -2(n-p)E[ps^{-1}h(s^{-1}y'y)-2h^{(1)}(s^{-1}y'y)s^{-2}y'y] \tag{34}$$

where the expected values in (34) are easily shown to depend only on $\theta'\Sigma^{-1}\theta$. Thus, if we choose a sequence $\{(\theta_i,\Sigma_i)\}$ of parameter values such that $\theta_i'\Sigma_i^{-1}\theta_i = c$, all i, and

$$\mathrm{tr}(Q^*)^{-1} = \mathrm{tr}(\Sigma_i)^{-1}Q^{-1}\to\infty, \text{ as } i\to\infty,$$

we see that unless

$$E[h^2(s^{-1}y'y)s^{-2}y'y] = 0, \text{ all } \theta'\Sigma^{-1}\theta = c, \tag{35}$$

we will have $\Delta(\theta_i,\Sigma_i)\to\infty$. Thus, for some parameter points $\Delta(\theta,\Sigma)$ will be positive (indeed, infinitely large), and hence $\delta_h(x,W)$ cannot be minimax. On the other hand, it is easy to show that (35) holds if and only if $h(u) = 0$ for almost all $u \geq 0$. This completes the proof.

Estimators of the form (4) do not perform well when any linear combination of the elements of x has low variability (implying that $ch_p(\Sigma)$ is small). To find a class of minimax estimators when $\Sigma$ is unrestricted, we might think of modifying members of the class (4) to produce new estimators of the form

$$\delta_h^*(x,W) = (I_p - ch_p(n^{-1}QW)h(x'W^{-1}x)Q^{-1}W^{-1})x. \tag{36}$$

Assuming that $ch_p(n^{-1}QW)$ and $ch_p(Q\Sigma)$ are close in value (which should be true at least when n is large), any member of the class (36) will behave like the minimax estimator x when $ch_p(\Sigma)$ is small, and will behave like $\delta_{ch_p(Q\Sigma)h}$ otherwise. Thus, we have good intuitive reasons for conjecturing that a member of the class (36) of estimators is minimax provided that (i) $uh(u)$ is nondecreasing in u, and (ii) $0 \leq h(u) \leq 2(p-2)u$, all $u \geq 0$. Unfortunately, we have not yet been able to prove this conjecture. One can follow the steps used in Section 2, but unlike the result (24) obtained for the class (4), integration over t and $U_{22}$ does not lead to any simplification. This lack of simplification is due to the fact that $ch_p(n^{-1}QW)$, after the change of variables from $(x,W)$ to $(y,s,t,U_{22})$, is a complicated and nonlinear function of $y$, $s$, $t$, and $U_{22}$.

## REFERENCES


[1] BERGER, J. (1976). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Multivariate Analysis **6**

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

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

[4] EFRON, B. AND MORRIS, C. (1976b). Multivariate empirical Bayes and estimation of covariance matrices. Ann. Statist. **4** 22-32.

[5] JAMES, W. AND STEIN, C. (1960). Estimation with quadratic loss. Proc. 4th Berkeley Symp. Math. Stat. Prob. 1, 361-379. Univ. California Press, Berkeley.

[6] LIN, PI-EHR AND TSAI, HUI-LIANG. (1973). Generalized Bayes minimax estimators of the multivariate normal mean with unknown covariance matrix. Ann. Statist. **1** 142-145.