

ON A MINIMAX ESTIMATE FOR THE MEAN OF A NORMAL RANDOM VECTOR UNDER A GENERALIZED QUADRATIC LOSS FUNCTION

BY TAMER BASAR¹ AND MAX MINTZ²

Yale University

An admissible minimax estimate for the mean of a normal random vector with known covariance is derived for a generalized quadratic loss function. This loss function is quadratic in both the estimation error and the unknown mean. The estimate is derived using the method of least favorable prior distributions. The decision rule is linear, and the least favorable prior distribution for the unknown mean is normal with zero mean. The covariance of this least favorable normal distribution is determined by the solution of a certain nonlinear algebraic matrix equation.

1. Introduction. A minimax location parameter estimation problem is posed as follows:

Let z denote a single observation of a random vector, where $z \in N[\theta, \Sigma]$. It is assumed that Σ is known, where $\Sigma > \mathbf{0}$, and that θ is an unknown element of E^n . We wish to determine a minimax estimate $\delta(z)$ for θ with respect to the generalized quadratic loss function $L(\delta, \theta)$ defined by

$$(1) \quad L(\delta, \theta) = [\delta - \theta]'C[\delta - \theta] - \theta'D\theta,$$

where $C \geq \mathbf{0}$, and $D > \mathbf{0}$ (i.e. C is a symmetric nonnegative definite matrix and D is a symmetric positive definite matrix).

The second term in the generalized quadratic loss function defined by (1), places a quadratic cost on nature's choice of θ . This loss function, as compared with the usual quadratic loss function, forces nature to compromise between a gain due to estimation error and a loss which is equal to the square of the norm of θ weighted by D . This choice of loss function is motivated by the general problem of tracking an evasive target which is under the control of an intelligent adversary, who must balance a gain due to estimation error against a loss due to energy consumed in evasive maneuvers.

The conditional risk of $\delta(z)$ given θ will be denoted by $r[\delta, \theta]$ i.e.,

$$(2) \quad r[\delta, \theta] = E[L(\delta, \theta) | \theta].$$

2. A minimax estimate. Define $\Gamma(\Lambda)$ by

$$(3a) \quad \Gamma(\Lambda) = [\Lambda + \Sigma]^{-1}\Sigma C\Sigma[\Lambda + \Sigma]^{-1} - D,$$

with

$$(3b) \quad \Lambda \geq \mathbf{0}.$$

Received March 13, 1971; revised June 6, 1972.

¹ Now at Harvard University.

² Now at The University of Illinois.

Let Λ^* denote any solution to the matrix equation

$$(4a) \quad \Gamma(\Lambda^*)\Lambda^* = \mathbf{0},$$

such that

$$(4b) \quad \Gamma(\Lambda^*) \leq \mathbf{0}.$$

THEOREM 1.

(i) If $\Gamma(\Lambda)$ is defined by (3a) and (3b), then there exists at least one covariance matrix Λ^* , which satisfies (4a) and (4b).

(ii) If $\delta^*(z)$ is defined by

$$(5) \quad \delta^*(z) = \Lambda^*[\Lambda^* + \Sigma]^{-1}z,$$

then $\delta^*(z)$ is an admissible minimax estimate for θ .

A proof of part (i) of Theorem 1 is given in Appendix 1. A proof of part (ii) of Theorem 1 is developed by using Lemma 1, which is stated in Section 3. The admissibility of the estimate is established in Section 4.

3. A least favorable prior distribution for θ . The following lemma provides us with a useful relationship between a minimax estimate for θ and a corresponding least favorable prior distribution. This lemma will be used in the proof of part (ii) of Theorem 1.

Let λ denote a prior probability measure for θ . Further, let $\delta_\lambda(z)$ denote a Bayes estimate for θ with respect to λ , and the loss function defined by (1).

LEMMA 1. Assume that $\max_{\theta \in E^n} r[\delta_\lambda, \theta]$ exists, and denote its value by \bar{r}_λ . Let $\tau_\lambda = \{\theta : r[\delta_\lambda, \theta] = \bar{r}_\lambda\}$.

If $\lambda[\tau_\lambda] = 1$, then the Bayes estimate $\delta_\lambda(z)$ is a minimax estimate for θ , and λ is a least favorable prior probability measure for θ .

A proof of Lemma 1 can be found in Lehmann [4] and Wald [6].

Lemma 1 is used as follows:

Let λ_0 denote a prior probability measure for θ such that $\theta \in N[\mathbf{0}, \Lambda]$. For this choice of λ_0 , a Bayes estimate $\delta_{\lambda_0}(z)$ is

$$(6) \quad \delta_{\lambda_0}(z) = E[\theta | z] = \Lambda[\Lambda + \Sigma]^{-1}z,$$

and the corresponding conditional risk is

$$(7) \quad r[\delta_{\lambda_0}, \theta] = \theta^T \Gamma(\Lambda) \theta + \text{tr} [(\Lambda + \Sigma)^{-1} \Lambda C \Lambda (\Lambda + \Sigma)^{-1} \Sigma],$$

where $\Gamma(\Lambda)$ is defined as before by

$$(3a) \quad \Gamma(\Lambda) = [\Lambda + \Sigma]^{-1} \Sigma C \Sigma [\Lambda + \Sigma]^{-1} - D.$$

In the proof of part (i) of Theorem 1, it is shown that a covariance matrix Λ^* always exists such that relations (4a) and (4b) are satisfied. In what follows, we shall show that the pair

$$(8a) \quad \lambda^* = N[\mathbf{0}, \Lambda^*]$$

$$(8b) \quad \delta_{\lambda^*}(z) = \Lambda^*[\Lambda^* + \Sigma]^{-1}z$$

satisfies the sufficiency conditions of Lemma 1, and hence prove part (ii) of Theorem 1.

The determination of Λ^* is dependent on the signs of the eigenvalues of the matrix $[C - D]$. There are three cases to consider:

Case 1. $C - D \geq 0$.

Case 2. $C - D \leq 0$.

Case 3. $C - D$ has at least one positive and at least one negative eigenvalue.

Case 1. When $C - D \geq 0$, a suitable choice of Λ^* is

$$(9) \quad \Lambda^* = [D^\dagger]^{-1}[D^\dagger \Sigma C \Sigma D^\dagger]^\dagger [D^\dagger]^{-1} - \Sigma .$$

ASSERTION. The pair λ^* and $\delta_{\lambda^*}(z)$ defined by (8a), (8b), and (9) satisfies Lemma 1.

This assertion is verified as follows. If Λ^* is defined by (9) and $C - D \geq 0$, then $\Gamma(\Lambda^*) = 0$, and $\Lambda^* \geq 0$. [See Appendix 2 for further discussion of this point.] Now since $\Gamma(\Lambda^*) = 0$, it follows from (7) that the conditional risk $r[\delta_{\lambda^*}, \theta]$ is constant for all θ . Hence $\tau_{\lambda^*} = E^n$, and $\lambda^*[\tau_{\lambda^*}] = 1$. An estimate with constant conditional risk is often referred to as an equalizer decision rule.

Case 2. When $C - D \leq 0$, a suitable choice of Λ^* is

$$(10) \quad \Lambda^* = 0 .$$

ASSERTION. The pair λ^* and $\delta_{\lambda^*}(z)$ defined by (8a), (8b), and (10) satisfies Lemma 1.

This assertion is verified as follows. Note that

$$(11a) \quad \Gamma(0) = C - D \leq 0 ,$$

and further that

$$(11b) \quad r[\delta_{\lambda^*}, \theta] = \theta'[C - D]\theta .$$

The conditional risk function defined by (11b) attains its maximum value if and only if θ is in the null space of $[C - D]$. Hence the set τ_{λ^*} is defined by

$$\tau_{\lambda^*} = \{ \theta : \theta \in \text{null space of } C - D \} .$$

The one-point probability measure λ^* defined by (8a) and (10) assigns probability one to any set in E^n which contains the point $\theta = 0$. Now since $0 \in \tau_{\lambda^*}$, it follows that $\lambda^*[\tau_{\lambda^*}] = 1$.

Case 3. When $C - D$ has at least one positive and one negative eigenvalue, a suitable choice of Λ^* is made by picking any matrix $\Lambda^* \geq 0$ that satisfies the following relations:

$$(4a) \quad \Gamma(\Lambda^*)\Lambda^* = 0 ,$$

$$(4b) \quad \Gamma(\Lambda^*) \leq 0 .$$

The existence of at least one such Λ^* is proved in Appendix 1.

ASSERTION. The pair λ^* and $\delta_{\lambda^*}(z)$ defined by (8a), (8b), (4a) and (4b) satisfies Lemma 1.

This assertion is verified as follows. If (4a) is satisfied, then all of the eigenvectors of Λ^* with nonzero eigenvalues are in the null space of $\Gamma(\Lambda^*)$. Thus the null space of $\Gamma(\Lambda^*)$, which is a Borel set in E^n , receives probability one from the measure λ^* . Further, if (4b) is satisfied, it follows from (7) that τ_{λ^*} is the null space of $\Gamma(\Lambda^*)$. Hence $\lambda^*[\tau_{\lambda^*}] = 1$.

Since the relations (4a) and (4b) are satisfied in all three cases, the argument pertaining to Case 3 is sufficient to prove part (ii) of Theorem 1. In conclusion, we note that in Case 1 we obtained a minimax estimate which was an equalizer rule, and in Case 2 we obtained a minimax estimate which was degenerate.

4. Admissibility. The admissibility of the Bayes estimate $\delta_{\lambda^*}(z)$ is a consequence of the fact that all Bayes estimates for θ , with respect to the prior distribution $N[\mathbf{0}, \Lambda]$, have identical conditional risk functions. Hence, any Bayes estimate for θ is admissible. (See Ferguson [3] page 60.)

5. A wide sense property. The estimate $\delta^*(z)$ exhibits a wide sense property:

In particular, if $\mathcal{F}[\theta, \Sigma]$ denotes the class of all distributions F (for z) with mean θ and covariance Σ , then it follows from the linearity of $\delta_{\lambda^*}(z)$ and the quadratic nature of the loss function that $r[\delta_{\lambda^*}, \theta]$ is independent of $F \in \mathcal{F}[\theta, \Sigma]$. Thus if the distribution for z is an unknown element of $\mathcal{F}[\theta, \Sigma]$ then $\delta_{\lambda^*}(z)$ is an admissible minimax estimate for θ . In summary, the distribution $N[\theta, \Sigma]$ is least favorable for the statistician in the class $\mathcal{F}[\theta, \Sigma]$. This Gaussian result reminds us of the familiar maximum entropy result.

APPENDIX 1

The following observations and lemmas are used to prove part (i) of Theorem 1.

OBSERVATION 1. Let Ω denote the space of all $n \times n$ real symmetric matrices. Further define a metric $\rho(W, M): \Omega \times \Omega \rightarrow E^1$ by

$$(12) \quad \rho(W, M) = \left[\sum_{i=1}^n \sum_{j=1}^n (W_{ij} - M_{ij})^2 \right]^{\frac{1}{2}}.$$

With these definitions, the metric space (Ω, ρ) is isomorphic to $E^{\frac{1}{2}n(n+1)}$.

OBSERVATION 2. Define sets T_1 and $T_2(\alpha)$ as follows:

$$(13a) \quad T_1 =_{\Delta} \{ \Lambda \in \Omega : \Lambda \geq \mathbf{0} \},$$

$$(13b) \quad T_2(\alpha) =_{\Delta} \{ \Lambda \in \Omega : \rho(\Lambda, \mathbf{0}) \leq \alpha \}.$$

With these definitions the set $S(\alpha)$ defined by

$$(14) \quad S(\alpha) =_{\Delta} T_1 \cap T_2(\alpha) = \{ \Lambda \in \Omega : \Lambda \geq \mathbf{0}, \rho(\Lambda, \mathbf{0}) \leq \alpha \}$$

is both compact and convex in (Ω, ρ) .

LEMMA 2. Consider the following abstract game:

Define

$$(15) \quad J(\Lambda_1, \Lambda_2) = {}_{\Delta} \text{tr}[CK\Sigma K'] + \text{tr}[(I - K)'C(I - K) - D]\Lambda_2,$$

where

$$(16) \quad K = {}_{\Delta} \Lambda_1[\Lambda_1 + \Sigma]^{-1};$$

then

$$\min_{\Lambda_1 \in S(\alpha)} \max_{\Lambda_2 \in S(\alpha)} J(\Lambda_1, \Lambda_2) = \max_{\Lambda_2 \in S(\alpha)} \min_{\Lambda_1 \in S(\alpha)} J(\Lambda_1, \Lambda_2).$$

PROOF. The proof of this lemma is in two parts:

(I) Since the kernel $J(\Lambda_1, \Lambda_2)$ is continuous in the pair (Λ_1, Λ_2) , and since $S(\alpha)$ is compact; there exist two probability measures on $S(\alpha)$ denoted by $\mu_1^*(\Lambda_1)$, $\mu_2^*(\Lambda_2)$ which are optimal mixed strategies for this game. (See Owen [5] page 88.)

(II) It will be demonstrated in what follows that the probability measures $\mu_1^*(\Lambda_1)$ and $\mu_2^*(\Lambda_2)$ referred to in (I) are one-point; that is, the game has a solution in pure strategies:

OBSERVATION 3. Let Λ_2 be any fixed element of T_1 which is defined by (13a). Then $J(\Lambda_1, \Lambda_2)$ is minimized when $\Lambda_1 = \Lambda_2$.

PROOF. The proof follows by identifying this minimization problem with a particular Bayesian estimation problem. Let θ be a normal random vector with zero mean and covariance Λ_2 . The Bayes estimate $\delta(z)$ for θ under the loss function $L(\delta, \theta)$ defined by (1) is given by

$$(17a) \quad \delta(z) = K_b z,$$

where

$$(17b) \quad K_b = {}_{\Delta} \Lambda_2[\Lambda_2 + \Sigma]^{-1}.$$

(Refer to Anderson [1] page 28.)

The Bayes risk is equal to $J(\Lambda_2, \Lambda_2)$. Comparing K_b with K defined by expression (16), it can then be concluded that $J(\Lambda_1, \Lambda_2)$ is minimized for fixed $\Lambda_2 \in T_1$ when $\Lambda_1 = \Lambda_2$.

Now consider the lower value of the game over (μ_1, μ_2) . That is, first fix μ_2 and then minimize with respect to μ_1 . Further, let

$$(18) \quad \hat{J}(\Lambda_1 | \mu_2) = {}_{\Delta} E_{\mu_2}[J(\Lambda_1, \Lambda_2)].^3$$

Since J defined by (15) is linear in Λ_2 ,

$$(19) \quad \hat{J}(\Lambda_1 | \mu_2) = J(\Lambda_1, E_{\mu_2}[\Lambda_2]).$$

A measure μ_1 that minimizes $E_{\mu_1}[\hat{J}(\Lambda_1 | \mu_2)]$ is the measure that assigns probability one to the element $\Lambda_1 = E_{\mu_2}[\Lambda_2]$. [This follows from Observation 3.] Hence for

³ “ $E_{\mu_i}[\cdot]$ ” is “expected value of $[\cdot]$ with respect to measure μ_i ”.

any μ_2 , the related minimizing measure μ_1 is one-point. To complete the evaluation of the lower value of the game with respect to the mixed strategies (μ_1, μ_2) , it is sufficient to compute

$$(20) \quad J_l = \Delta \max_{\mu_2} J(E_{\mu_2}[\Lambda_2], E_{\mu_2}[\Lambda_2]).$$

OBSERVATIONS.

(a) Since μ_2 is a probability measure on $S(\alpha)$, and since $S(\alpha)$ is convex, $E_{\mu_2}[\Lambda_2] \in S(\alpha)$.

(b) The $\max_{\Lambda \in S(\alpha)} J(\Lambda, \Lambda)$ exists since $J(\Lambda, \Lambda)$ is continuous in Λ and $S(\alpha)$ is compact.

Let Λ^* maximize $J(\Lambda, \Lambda)$ over $S(\alpha)$. A maximizing measure μ_2 defined by (20) is given by the measure which assigns probability one to the element Λ^* . Hence there exists a one-point maximizing measure $\underline{\mu}_2$ associated with the lower value of the game.

Let $(\bar{\mu}_1, \bar{\mu}_2)$ denote the measures associated with the upper value of the game.

From (I) it follows that $\mu_1 = \bar{\mu}_1$ and $\underline{\mu}_2 = \bar{\mu}_2$. But since $(\underline{\mu}_1, \underline{\mu}_2)$ are one-point, this completes the proof of the lemma.

In addition to proving Lemma 2, it has been shown that the solution to the game is given by $\Lambda_1 = \Lambda_2 = \Lambda^*$ where Λ^* maximizes $J(\Lambda, \Lambda)$ over $S(\alpha)$.

DISCUSSION ON LEMMA 2. The solution $(\Lambda_1 = \Lambda^*(\alpha), \Lambda_2 = \Lambda^*(\alpha))$ to the game defined in Lemma 2 will depend in general on the value α which defines $S(\alpha)$. In what follows it will be shown that the solution $(\Lambda^*(\alpha), \Lambda^*(\alpha))$ will be independent of α for sufficiently large α . Hence, it will be possible to replace $S(\alpha)$ with T_1 and insure the existence of a solution to the game in pure strategies over T_1 .

When $\Lambda_1 = \Lambda_2 = \Lambda$, and the expression for K given by (16) is substituted into (15), the expression for $J(\Lambda_1, \Lambda_2)$ simplifies to:

$$(21) \quad J(\Lambda, \Lambda) = \text{tr}[C\Sigma] - \text{tr}[C\Sigma(\Lambda + \Sigma)^{-1}\Sigma] - \text{tr}[D\Lambda].$$

Since $\text{tr}[C\Sigma(\Lambda + \Sigma)^{-1}\Sigma] \geq 0$,

$$(22) \quad J(\Lambda, \Lambda) \leq \text{tr}[C\Sigma] - \text{tr}[D\Lambda] \quad \text{for all } \Lambda \in T_1.$$

It follows from inequality (22) that there exists a finite real number $\eta \geq 0$ such that $\|\Lambda\| > \eta$ implies $J(\Lambda, \Lambda) < 0$. A suitable choice of η is

$$(23) \quad \eta^* = \frac{\text{tr}[C\Sigma]}{d_{\min}}$$

where d_{\min} denotes the minimum eigenvalue of $D > 0$.

Since $J(\mathbf{0}, \mathbf{0}) = 0$, it follows that if $\alpha > \eta^*$ then $\Lambda^*(\alpha) \in S(\eta^*)$. Thus $\Lambda^*(\alpha)$ is independent of α when $\alpha > \eta^*$. Hence $S(\alpha)$ can be replaced by T_1 in the hypothesis of Lemma 2.

LEMMA 3. *The matrix,*

$$(24) \quad \Gamma(\Lambda^*) = {}_{\Delta}[\Lambda^* + \Sigma]^{-1}\Sigma C\Sigma[\Lambda^* + \Sigma]^{-1} - D$$

is nonpositive definite where Λ^ denotes a solution to the game defined in Lemma 2 with $S(\alpha)$ replaced by T_1 .*

PROOF BY CONTRADICTION. Write $J(\Lambda_1, \Lambda_2)$ as

$$(25) \quad J(\Lambda_1, \Lambda_2) = \text{tr}[CK\Sigma K'] + \text{tr}[\Gamma(\Lambda_1)\Lambda_2]$$

with K defined by (16) and $\Gamma(\Lambda_1^*)$ defined by (24). Let Λ_1^* be an optimal choice for Λ_1 and further assume that $\Gamma(\Lambda_1^*)$ has at least one positive eigenvalue. Then it is possible to find a $\Lambda_2^* \in T_1$ that will make $J(\Lambda_1^*, \Lambda_2)$ arbitrarily large. But this is a contradiction since the upper value of the game is bounded. Therefore $\Gamma(\Lambda_1^*)$ cannot have any positive eigenvalues.

LEMMA 4. *Let Λ^* denote a solution to the game defined in Lemma 2 with $S(\alpha)$ replaced by T_1 . Then,*

$$(26) \quad \Gamma(\Lambda^*)\Lambda^* = \mathbf{0}.$$

PROOF. $J(\Lambda_1, \Lambda_2) = \text{tr}[CK\Sigma K'] + \text{tr}[\Gamma(\Lambda_1)\Lambda_2]$ as defined by (25). Let Λ_1^* denote an optimal choice of Λ_1 , and Λ_2^* denote an optimal choice of Λ_2 . Assume $\Gamma(\Lambda_1^*)\Lambda_2^* \neq \mathbf{0}$. But $\Gamma(\Lambda_1^*) \leq \mathbf{0}$ by Lemma 3 and further $\Lambda_2^* \geq \mathbf{0}$ by definition. Note that $\text{tr}[AB] = 0$ if and only if $AB = \mathbf{0}$ when $A \leq \mathbf{0}$ and $B \geq \mathbf{0}$; hence it follows that $\text{tr}[\Gamma(\Lambda_1^*)\Lambda_2^*] < 0$. This contradicts the optimality of Λ_2^* since then

$$J(\Lambda_1^*, \mathbf{0}) > J(\Lambda_1^*, \Lambda_2^*).$$

Hence, $\Gamma(\Lambda_1^*)\Lambda_2^* = \mathbf{0}$. But $\Lambda_1^* = \Lambda_2^* = \Lambda^*$ by Lemma 2. Therefore $\Gamma(\Lambda^*)\Lambda^* = \mathbf{0}$.

PROOF OF PART (i) OF THEOREM 1. If $\Gamma(\Lambda)$ is defined by (3a) and (3b), then there exists at least one matrix $\Lambda^* \geq \mathbf{0}$, which satisfies (4a) and (4b).

Lemmas 3 and 4 prove this part of Theorem 1 by making use of the abstract game defined in Lemma 2 with $S(\alpha)$ replaced by T_1 .

APPENDIX 2

FURTHER DISCUSSION OF CASE 1.

LEMMA 5. *If $C - D \geq \mathbf{0}$, then $\Gamma(\Lambda^*) = \mathbf{0}$, and $\Lambda^* \geq \mathbf{0}$, when*

$$(27) \quad \Lambda^* = {}_{\Delta}[D^{\frac{1}{2}}]^{-1}[D^{\frac{1}{2}}\Sigma C\Sigma D^{\frac{1}{2}}]^{\frac{1}{2}}[D^{\frac{1}{2}}]^{-1} - \Sigma.^4$$

PROOF. We verify that $\Lambda^* \geq \mathbf{0}$ as follows: Note that, $C - D \geq \mathbf{0}$ implies

$$(28) \quad \Sigma(C - D)\Sigma \geq \mathbf{0},$$

⁴ $D^{\frac{1}{2}}$ denotes the unique symmetric nonnegative definite square root of the matrix D , where $D \geq \mathbf{0}$.

where

$$\Sigma > \mathbf{0}, \quad D > \mathbf{0}, \quad C > \mathbf{0}.$$

And further, (28) implies that:

$$(29) \quad D^{\frac{1}{2}} \Sigma C \Sigma D^{\frac{1}{2}} \geq D^{\frac{1}{2}} \Sigma D \Sigma D^{\frac{1}{2}} = (D^{\frac{1}{2}} \Sigma D^{\frac{1}{2}})^2.$$

But since it is well known that $A \geq B \geq \mathbf{0}$ implies $A^{\frac{1}{2}} \geq B^{\frac{1}{2}}$ (e.g. see Bellman [2]), it follows that (29) implies that Λ^* defined by (27) is nonnegative definite.

It can be shown by substitution that $\Gamma(\Lambda^*) = \mathbf{0}$.

REFERENCES

- [1] ANDERSON, T. W. (1959). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
- [2] BELLMAN, R. (1968). Some inequalities for the square root of a positive definite matrix, *Linear Algebra and its Applications* **1** 321-324.
- [3] FERGUSON, T. S. (1967). *Mathematical Statistics, A Decision Theoretic Approach*. Academic Press, New York.
- [4] LEHMANN, E. L. (1966). *Notes on the Theory of Estimation*. Univ. of California Press, Berkeley.
- [5] OWEN, G. (1968). *Game Theory*. Saunders, Philadelphia.
- [6] WALD, A. (1950). *Statistical Decision Functions*. Wiley, New York.

406 EAST MICHIGAN
APARTMENT 11
URBANA, ILLINOIS 61801