A NOTE ON A PREVIOUS LEMMA IN THE THEORY OF LEAST SQUARES AND SOME FURTHER RESULTS

By C. RADHAKRISHNA RAO

Indian Statistical Institute

SUMMARY. Let Y be a vector of random variables such that $E(Y) = X\beta$ where β is a vector of unknown parameters and Σ be the covariance matrix of Y. A linear function $L'Y$ is said to be best linear unbiased estimator (BLUE) of a parametric function $p'\beta$ with respect to Σ if $L'\Sigma L$ is a minimum subject to $p' = L'X$. The paper deals with necessary and sufficient conditions that, for every estimable parametric function or for a given subset, the BLUE with respect to Σ is the same as the BLUE with respect to $\Sigma = I$ (identity matrix) or the same as the BLUE with respect to $\Sigma = \Sigma_0$ (a given matrix).

Let Z be a matrix of maximum rank such that $X'Z = 0$. It is shown that when Σ_0 is non-singular, or rank $(X \vdots Z) =$ rank $(X \vdots \Sigma_0 Z)$, then a NAS condition for the equality of BLUE's of all estimable functions for Σ and Σ_0 is that Σ is of the form

$$\Sigma = X\Theta X' + \Sigma_0 Z \Gamma Z' \Sigma_0 + \Sigma_0$$

where Θ, Γ are arbitrary. The representations of Σ in other situations where Σ_0 is singular have also been obtained.

1. Statement of the previous lemma

In a previous paper (Rao, 1967), read at the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1965, the author proved the following result in the least square theory.

Let Y be a vector variable with $E(Y) = X\beta$ and covariance matrix Σ, where X is a known matrix, β is a vector of unknown parameters and the ranks of Σ and X may be arbitrary. The model is known as Gauss-Markoff set up and represented by $(Y, X\beta, \Sigma)$. We call a parametric function $p'\beta$ estimable if it is the expectation of a linear function of Y. A best linear unbiased estimator (BLUE) with respect to Σ is defined as $L'Y$ where L is chosen such that $V(L'Y) = L'\Sigma L$ is a minimum subject to the condition $E(L'Y) \equiv p'\beta$. A simple least square estimator (SLSE) of $p'\beta$ is defined to be the BLUE with respect to $\Sigma = I$, the identity matrix. *Then a necessary and sufficient (NAS) condition that for each estimable parametric function the BLUE with respect to Σ is the same as the SLSE is that Σ is of the form*

$$\Sigma = X\Theta X' + Z\Gamma Z' + \sigma^2 I \qquad \dots \quad (1.1)$$

where Θ, Γ are arbitrary matrices, σ^2 is an arbitrary scalar and Z is a matrix of maximum rank such that $Z'X = 0$.

In my previous paper (Rao, 1967), I gave details of the proof of (1.1) when Σ and $X'X$ are non-singular and mentioned that the same proof holds more generally for singular Σ and $X'X$. I omitted the details in the latter case as the extension was extremely simple, and not relevant to the main theme of the paper.

In a recent paper, Zyskind (1967) thought that there may be some difficulty in proving the NAS condition (1.1) in its widest generality when Σ is singular. Watson (1967) writes that, "Rao (1965) remarks that his result is true for Σ singular and rank $X = r < k$. Some skill with generalized inverses might show the proof is still valid." In view of these remarks and other statements it seems necessary to elaborate the earlier proof. It may be recalled that the basis of my earlier proof is the following : the NAS condition that a statistic is a minimum variance unbiased estimator is that it has zero covariance with statistics whose expectation is identically zero (Rao, 1965b, pp. 185, 257). Hence the proof given covers both the necessity and sufficiency of the condition. Further it is shown that the earlier proof carries over to the general case without any difficulty. No skill in generalized inverses seems to be necessary, although there is no harm in using new methods.

Incidentally some other results of interest are obtained. For instance, a representation for Σ has been obtained when SLSE and BLUE are the same only for a subset (subspace) of estimable parametric functions. Also, necessary and sufficient conditions for the BLUE's with respect to Σ and Σ_0 to be the same are also given, in addition to what is already stated in my earlier paper.

It may be noted that the form (1.1) of the covariance matrix arises in a natural way in the problem of estimation of growth curves (see Rao, 1965a). The phenomenon that for a certain class of covariance matrices, the BLUE is the same as that obtained on the assumption that the variables are uncorrelated and has common variance was noted in Lemma 3 of Rao (1965a).

2. Proof of the main lemma

We restate the previous lemma as Lemma 1 in an equivalent way.

Lemma 1 : *The following are equivalent NAS conditions that, in the Gauss-Markoff model* $(Y, X\beta, \Sigma)$, *for each estimable parametric function the BLUE with respect to* Σ *is the same as the SLSE.*

(i) $X'\Sigma Z = 0.$ ... (2.1)

(ii) $\Sigma = X\Lambda_1 X' + Z\Lambda_2 Z'.$ ... (2.2)

(iii) $\Sigma = X\,\Theta\, X' + Z\Gamma Z' + \sigma^2 I$ ... (2.3)

where $\Lambda_1, \Lambda_2, \Theta, \Gamma$ *are arbitrary symmetric matrices,* σ^2 *is an arbitrary scalar and* Z *is a matrix of maximum rank such that* $Z'X = 0$.

Corollary : *Let* X_r *be a matrix whose columns form a basis of* $\mathcal{M}(X)$ *the space generated by the columns of* X *and similarly let* Z_{n-r} *be a basis of* $\mathcal{M}(Z)$ *where* $r = \text{rank } X$ *and* $n-r = \text{rank } Z$. *Then the NAS conditions* (2.1), (2.2) *and* (2.3) *can also be stated as follows.*

(i)′ $X_r'\Sigma Z_{n-r} = 0$ ... (2.1)′

(ii)′ $\Sigma = X_r M_1 X_r' + Z_{n-r} M_2 Z_{n-r}'$ ... (2.2)′

(iii)′ $\Sigma = X_r N_1 X_r' + Z_{n-r} N_2 Z_{n-r}' + \sigma^2 I$ ... (2.3)′

where M_1, M_2, N_1, N_2 *are arbitrary symmetric matrices and* σ^2 *is an arbitrary scalar.*

Proof: Let N be a matrix of maximum rank such that $N'\Sigma = 0$, where N is a null matrix iff Σ is non-singular. Observe that

$$E(N'Y) = N'X\beta \quad \dots \quad (2.4)$$

and the dispersion matrix of $N'Y$ is null, so that $N'Y$ is a vector of constants say d and the equation $N'X\beta = d$ may be considered as a restriction on the unknown parameters β. Then a necessary and sufficient condition that a linear function, $G'Y$, has constant expectation independently of the unknown parameters but subject to $N'X\beta = d$ is that $X'G$ belongs to the space generated by the columns of $X'N$, or $X'G = X'NM$ for a suitable choice of M. Then $G'Y$ can be written

$$\begin{aligned} G'Y &= (G'Y - M'N'Y) + M'N'Y \\ &= H'Y + M'N'Y \end{aligned} \quad \dots \quad (2.5)$$

where $H'Y$ has zero expectation independently of the restrictions i.e., $X'H = 0$ or $H = ZB$ for some B.

Now a SLSE is of the form $\lambda'X'Y$ and conversely for any given λ, $\lambda'X'Y$ is the SLSE of its expected value (see, Rao, 1965b, pp. 181, 182). Let $G'Y$ be a function of the type considered in (2.5). Then the NAS condition that $\lambda'X'Y$ is a BLUE is (see Rao, 1965b, pp. 185 and 257)

$$\begin{aligned} \text{cov}(\lambda'X'Y, G'Y) &= \text{cov}(\lambda'X'Y, H'Y) \\ &= \lambda'X'\Sigma ZB = 0. \end{aligned} \quad \dots \quad (2.6)$$

In (2.6), $G'Y$ is replaced by $H'Y$ since the other part in $G'Y$ is a constant. If (2.6) is true for all λ and B, then

$$X'\Sigma Z = 0 \quad \dots \quad (2.7)$$

which proves the NAS condition (1) of Lemma 1 and which is exactly the condition obtained in my previous paper in the Fifth Berkeley Symposium, 1965, (see Rao, 1967, p. 364, equation, 66).

The other NAS conditions (2.2) and (2.3) are important in that they provide interesting representations of Σ. First we obtain a general representation of Σ. Since Σ is at least semi-positive definite, it can be written in the form $\Sigma = CC'$. By definition the columns of X and Z generate the whole space and hence there exist matrices U_1 and U_2 such that $C = XU_1 + ZU_2$ giving

$$\begin{aligned} \Sigma = CC' &= XU_1U_1'X' + ZU_2U_2'Z' + XU_1U_2'Z' + ZU_2U_1'X' \\ &= X\Lambda_1X' + Z\Lambda_2Z' + X\Lambda_3Z' + Z\Lambda_3'X' \end{aligned} \quad \dots \quad (2.8)$$

with the obvious notation, where Λ_1, Λ_2 are symmetric matrices.

Substituting in (2.7) the general expression (2.8) for Σ we find

$$X'X\Lambda_3 Z'Z = 0 \iff X\Lambda_3 Z' = 0 = Z\Lambda_3'X' \quad \dots \quad (2.9)$$

which shows that Σ is of the form

$$\Sigma = X\Lambda_1X' + Z\Lambda_2Z' \quad \dots \quad (2.10)$$

which is the condition (2.2).

It would be of interest to write Σ as the sum of $\sigma^2 I$ and another matrix which is easily done by using the identity

$$X(X'X)^{-}X'+Z(Z'Z)^{-}Z' = I \qquad \ldots \quad (2.11)$$

(see Rao, 1965b, p. 60). Multiplying (2.11) by an arbitrary scalar σ^2, subtracting from (2.10) and rearranging the terms we have the desired form

$$\Sigma = X\,\Theta\,X'+Z\,\Gamma\,Z'+\sigma^2 I \qquad \ldots \quad (2.12)$$

where Θ and Γ are also symmetric matrices, which is the condition (2.3).

The conditions (2.1)$'$, (2.2)$'$ and (2.3)$'$ are trivially equivalent to (2.1), (2.2) and (2.3).

Note 1 : The method employed in proving the lemma is the same as that used in the earlier paper. We note that in the equation (2.6), $G'Y$ is replaced by $H'Y$, which has zero expectation irrespective of whether Σ is singular or not, because the other part in $G'Y$ is constant. Zyskind apparently thought there would be some difficulty on this account. The main NAS condition (2.7) is deduced in a simple way without in any way depending on the knowledge and skill in the use of generalized inverses.

Note 2 : It is shown that BLUE and SLSE are the same when the covariance matrix has the structure given in (1.1). We may examine the expression for the variance of the estimator of $p'\beta$, which can be written as $p'(X'X)^{-}X'Y$, using a generalized inverse of $X'X$.

$$\begin{aligned} V(p'(X'X)^{-}X'Y) &= p'(X'X)^{-}X'\Sigma X[(X'X)^{-}]'p \\ &= p'(X'X)^{-}X'(X\Theta X'+\sigma^2 I)X[(X'X)^{-}]'p \\ &= p'\Theta p+\sigma^2 p'(X'X)^{-}p \end{aligned}$$

which is independent of Γ. The expression for the variance reduces to $\sigma^2 p'(X'X)^{-}p$ only when $\Theta = 0$. However, it is not possible to estimate the unknown parameter σ^2 unless Γ is known or is zero. The situation is different if an independent estimate of Σ is available. In such a case, the problem of inference on unknown parameters β admits an elegant solution even if Θ and Γ are not zero. I have considered the inference problem when $\Theta \neq 0$, $\Gamma = 0$ in a previous paper (Rao, 1965a). But it is easy to see that it makes no difference whether Γ is null or not and the entire discussion of Rao (1965a) when $\Sigma = X\Theta X'+\sigma^2 I$ remains the same for the more general structure (1.1) (see Rao, 1967).

3. Equivalence of lemma 1 with other results

It has been pointed out by the referee that the author's result of Lemma 1 which provides a representation of Σ unlike the results of other authors (Kruskal, 1968; Watson, 1967; Zyskind, 1967) could be used to derive their results in a simple way.

Zyskind (1967) : *There exists a subset of r (= rank X) orthogonal eigen vectors of Σ which form a basis of $\mathcal{M}(X)$ the space spanned by the columns of X.*

Kruskal (1968) : *The NAS condition is $\Sigma X = XB$ for some B.*

Let $\Sigma = \lambda_1 E_1 + \ldots + \lambda_k E_k$ be a spectral representation of Σ where $\lambda_1, \ldots, \lambda_k$ are eigen values and $E_1, \ldots, E_k$, are the corresponding eigen subspaces. Further, let E_{k+1} be the subspace corresponding to the zero eigen value when Σ is singular. By choice E_i are symmetric, idempotent and orthogonal to each other. Further, let

$$W_i = (X_r' X_r)^{-1} X_r' E_i, \quad i = 1, \ldots, k+1 \qquad \ldots \quad (3.1)$$

where X_r is a basis of $\mathcal{M}(X)$. Let us group W_i, $i = 1, \ldots, k+1$ into m (as large as possible) sets $S_1, \ldots, S_m$ such that $W_i' W_j = 0$ if W_i and W_j belong to different sets. We state Theorem 3 of Watson (1967), without assuming that $X'X$ and Σ are of full rank.

For the BLUE and SLSE to be the same for every estimable parametric function, a NAS condition is that the eigen values associated with W_i in any given set are equal.

We shall establish in a simple way the equivalence of the various results with that of Lemma 1.

First, the equivalence of Lemma 1 and Kruskal's result follows immediately by choosing the condition $X' \Sigma Z = 0$ of Lemma 1 since $(X'\Sigma)Z = 0 \Longleftrightarrow \Sigma X \subset \mathcal{M}(X)$ which is the same as $\Sigma X = XB$ for some B.

Second, let us consider the NAS condition (2.2)′ of Lemma 1, and the eigen values $\alpha_1, \ldots, \alpha_r$ with the eigen vectors $L_1, \ldots, L_r$ of $X_r' X_r M_1 X_r' X_r$ with respect to $X_r' X_r$.

$$X_r' X_r M_1 X_r' X_r L_i = \lambda_i X_r' X_r L_i$$

$$\Longrightarrow X_r M_1 X_r' X_r L_i = \lambda_i X_r L_i = \Sigma X_r L_i \qquad \ldots \quad (3.2)$$

using the expression (2.2)′ for Σ. The equation (3.2) shows that $X_r L_i$, $i = 1, \ldots, r$ are eigen vectors of Σ which is Zyskind's result. It is easy to see that Zyskind's result implies $X'\Sigma Z = 0$.

We prove the general version of Watson's theorem in a simple way using the representation (2.2)′ of Lemma 1.

$$\Sigma = X_r M_1 X_r' + Z_{n-r} M_2 Z_{n-r}'$$

$$\Longrightarrow X_r (X_r' X_r)^{-1} X_r' \Sigma = X_r M_1 X_r'. \qquad \ldots \quad (3.3)$$

Writing $G = X_r (X_r' X_r)^{-1} X_r'$ and $H = X_r M_1 X_r'$ which are symmetric, (3.3) is same as

$$G(\lambda_1 E_1 + \ldots + \lambda_k E_k) = H$$

$$\Longrightarrow \lambda_i G E_i = H E_i \Longrightarrow \lambda_i E_j G E_i = E_j H E_i, \quad i = 1, \ldots, k+1. \qquad \ldots \quad (3.4)$$

Noting that $W_j'W_i = E_jGE_i$, (3.4) yields

$$(\lambda_i - \lambda_j)(W_j'W_i + W_i'W_j) = 0 \Longrightarrow \lambda_i = \lambda_j \quad \dots \quad (3.5)$$

if $W_j'W_i + W_i'W_j \neq 0$ which is the same as $W_i'W_j \neq 0$ since. $W_iW_j' = 0$. This proves the necessity.

To prove sufficiency, let us consider the spectral representation

$$\Sigma = \nu_1 F_1 + \dots + \nu_m F_m \quad \dots \quad (3.6)$$

corresponding to the eigen values and spaces associated with $S_1, \dots, S_m$ and redefine

$$W_i = (X_r' X_r)^{-1} X_r' F_i. \quad \dots \quad (3.7)$$

Then, we show that $W_i'W_j = 0$, $i \neq j$ imply that $X_r'\Sigma Z_{n-r} = 0$. The condition $W_i'W_j = 0 \Longrightarrow F_iGF_j = 0$ and

$$F_1GF_1 + \dots + F_mGF_m = G \quad \dots \quad (3.8)$$

or $\lambda_i F_i G F_i Z_{n-r} = 0$, $i = 1, \dots, m$. Hence $\lambda_i X_r' F_i Z_{n-r} = 0$, $i = 1, \dots, m$ or $X_r'(\lambda_1 F_1 + \dots + \lambda_m F_m) Z_{n-r} = 0$ which is the condition (2.1) of Lemma 1.

Another version of Watson's theorem in the general case without assuming full rank for $X'X$ or Σ is given by the referee, which will be published elsewhere.

4. A variation of lemma 1

We shall investigate the form of Σ by demanding that the BLUE is the same as the SLSE only for a given subset and not for all estimable parametric functions. We refer to the equation (2.5) and write $\lambda' X' Y$ in the form $p'(X'X)^- X'Y$ to obtain

$$p'(X'X)^- X'\Sigma ZB = 0 \quad \dots \quad (4.1)$$

where $p'\beta$ is an estimable parametric function. If (4.1) is true for all B and $p \subset \mathcal{M}(P)$ where P is a given matrix, then

$$P'(X'X)^- X'\Sigma Z = 0. \quad \dots \quad (4.2)$$

Substituting the general expression (2.2) for Σ and observing that $P'(X'X)^- X'X = P'$ (see Rao, 1966, p. 266)

$$P'\Lambda_3 Z'Z = 0 \Longrightarrow P'\Lambda_3 Z' = 0$$

or $P'\Lambda_3 = 0$ if Z is chosen such that $Z'Z$ is of full rank. Thus we have proved the following lemma. It may be seen that the use of g-inverse was convenient but could have been avoided.

Lemma 2 : *A NAS condition that for any estimable function $p'\beta$ where $p \subset \mathcal{M}(P)$ the BLUE and the SLSE are the same is that*

$$\Sigma = X\Theta X' + Z\Gamma Z' + \sigma^2 I + X\Lambda_3 Z' + Z\Lambda_3' X' \quad \dots \quad (4.3)$$

where Θ, Γ, σ^2 are arbitrary and Λ_3 is such that $P'\Lambda_3 Z' = 0$ or $P'\Lambda_3 = 0$ when Z is chosen such that $Z'Z$ is of full rank.

6. Some further results

In this section we prove some results on the characterization of the covariance matrix Σ such that the BLUE with respect to Σ is the same as that with Σ_0. Results under very general conditions are contained in Lemmas 3 and 4, and under some restrictions in Lemma 5.

Lemma 3 : *A NAS condition that for each estimable parametric function the BLUE with respect to Σ is the same as that with Σ_0 is that the rows of the matrix $(0 \vdots 0 \vdots X)$ are contained in the space generated by the rows of the matrix $(\Sigma Z \vdots \Sigma_0 Z \vdots X)$, or the rows of the matrix $(\Sigma_0 Z \vdots \Sigma Z)$ are contained in the space generated by the rows of the matrix $(Z'\Sigma_0 Z \vdots Z'\Sigma Z)$.*

Proof: Let $p'\beta$ be an estimable parametric function and $L'Y$ the BLUE with respect to Σ. Then L satisfies the equation

$$\Sigma L - X\lambda = 0, \quad X'L = p$$

where λ is a vector of unknowns. Similarly if $L'Y$ is also the BLUE with respect to Σ_0, then

$$\Sigma_0 L - X\mu = 0, \; X'L = p$$

where μ is a vector of unknowns. Thus a NAS condition for the BLUE of $p'\beta$ to be common is that the equations

$$\Sigma L = X\lambda, \; \Sigma_0 L = X\mu, \; X'L = p \qquad \dots \quad (5.1)$$

in L, λ, μ are soluble. The NAS condition for the solubility of (5.1) is that for every set of vectors a_1, a_2, a_3 such that

$$a_1'\Sigma + a_2'\Sigma_0 + a_3'X' = 0, \; a_1'X = 0, \; a_2'X = 0$$

implies $a_3'p = 0$, or $a_3'X = 0$ if the BLUE's are the same for every estimable function. But $a_1'X = 0$ and $a_2 X = 0 \Longrightarrow a_1 = Zb_1$, $a_2 = Zb_2$ for some vectors b_1 and b_2. Then the condition for solubility of (5.1) reduces to

$$b_1'Z'\Sigma + b_2'Z'\Sigma_0 + a_3'X' = 0 \Longrightarrow a_3'X' = 0$$

which implies that the space generated by the rows of the matrix $(\Sigma Z \vdots \Sigma_0 Z \vdots X)$ contain the rows of the matrix $(0 \vdots 0 \vdots X)$. The equivalent result stated in Lemma 3 follows easily.

Lemma 4 : *A sufficient condition that for each estimable parametric function the BLUE with respect to Σ is the same as that with Σ_0 is that Σ is of the form*

$$\Sigma = X\Theta X' + \Sigma_0 Z\Gamma Z'\Sigma_0 + \Sigma_0. \qquad \dots \quad (5.2)$$

Proof: It is easy to verify that the conditions of Lemma 3 are satisfied if Σ is as in (5.2).

Lemma 5 : *Let any one of the conditions be satisfied.*

(i) *Σ_0 is of full rank.*

(ii) *The columns of X and $\Sigma_0 Z$ span the whole space.*

Then a NAS condition that for each estimable parametic function the BLUE with respect to Σ is the same as that with Σ_0 is that Σ is of the form

$$\Sigma = X\Theta X' + \Sigma_0 Z\Gamma Z'\Sigma_0 + \Sigma_0. \qquad \dots \quad (5.3)$$

Proof : It is clear that under any one of the conditions (i) and (ii) of Lemma 5, the columns of X and $\Sigma_0 Z$ span the whole space in which case Σ can be written as

$$\Sigma = X\Theta X' + \Sigma_0 Z\Gamma Z'\Sigma_0 + \Sigma_0 + X\Lambda Z'\Sigma_0 + \Sigma_0 Z\Lambda' X'. \qquad \dots \quad (5.4)$$

Applying the NAS conditions of Lemma 3, we find that $X\Lambda Z' = 0$, thus reducing (5.4) to (5.3).

References

Kruskal, William (1968) : When are Gauss-Markoff and least squares estimators identical ? A coordinate free approach. *Ann. Math. Stat.*, 39, 70-75.

Rao, C. Radhakrishna (1965a) : The theory of least squares when the parameters are stochastic and its application to analysis of growth curves. *Biometrika*, 52, 447-458.

——— (1965b) : *Linear Statistical Inference and its Applications*, John Wiley and Sons, New York.

——— (1966): Generalized inverse for matrices and its application in mathematical statistics. *Research Papers in Statistics* (Festchrift for J. Neyman), 263-280, John Wiley and Sons, New York.

——— (1967): Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, 1, 355-372, Univ. of California Press.

Watson, Geoffrey S. (1967): Linear least squares regression. *Ann. Math. Stat.*, 38, 1679-1699.

Zyskind, George (1967): On canonical forms, negative covariance matrices and best and simple least squares linear estimator in linear models. *Ann. Math. Stat.*, 38, 1092-1110.

Paper received : April, 1968.