Bu-875-M

ON A MATRIX IDENTITY USEFUL IN VARIANCE COMPONENT ESTIMATION

George Casella and Shayle R. Searle[1)]
Cornell University

January, 1985

BU-875-M April 1985

Abstract

Three straightforward proofs are given of the matrix identity that supplies the reason why restricted maximum likelihood estimation (REML) of variance components does not depend on which set of error contrasts are chosen as the basis of estimation.

---

1) This note was prepared while George Casella was Visiting Professor at the University of North-Carolina, Raleigh, N.C., and Shayle Searle was Gast-professor at Universität Augsburg supported by a U.S. Senior Scientist Award from the Alexander von Humboldt-Stiftung.

1. Introduction

The mixed linear model

$$y = X\beta + Zu + \varepsilon,$$

where $y_{n\times 1}$ is a vector of observations, $X_{n\times p}$ and $Z_{n\times q}$ are known matrices, β is a vector of unknown, fixed effects, and u and ε are vectors of random effects, has been widely used in applications, especially in animal breeding. Of particular interest is the estimation of the variance components due to the random effects, **u** and **e**, for which one quite popular method of estimation is restricted maximum likelihood (REML); see, for example, Patterson and Thompson, 1971, Corbeil and Searle, 1976, and Harville, 1977.

REML estimation is based on error contrasts (Harville, 1974), which are linear combinations of y that are orthogonal to the space spanned by the columns of X, and a well-known property of REML estimators is that they are invariant to the choice of error contrasts. Notwithstanding the importance of this fact, its proof is seldom seen, presumably because that proof relies on establishing a matrix identity that can be somewhat lengthy to verify (as given in Searle, 1979). This paper provides shorter and more insightful proofs of this matrix identity, and illustrates its use in REML estimation.

2. The Matrix Identity

Let $\rho(\mathbf{A})$, $\mathbf{A}'$ and $\mathbf{A}^-$ denote, respectively, the rank, transpose and a generalized inverse ($\mathbf{AA}^-\mathbf{A} = \mathbf{A}$) of any matrix **A**. We then propose the following theorem.

Theorem For $\rho(X_{n\times p}) = r \leq p$ and for any K having $\rho[K_{n\times(n-r)}] = n - r$ and satisfying $K'X = 0$, then for $V_{n\times n}$ being any positive definite matrix,

$$K(K'VK)^{-1}K' = V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}. \tag{1}$$

Proof of (1) is not immediately obvious and has, in at least one instance, been achieved only at great length. In contrast, we here give three proofs, all of them quite short. All three are based on observing that (1) can be established by verifying the simpler result

$$K(K'K)^{-1}K' = I - X(X'X)^{-}X' . \tag{2}$$

This is so, because with V being positive definite, define $X^* = V^{-\frac{1}{2}}X$ and $K^* = V^{\frac{1}{2}}K$, and observe that (1) reduces to (2) with K^* and X^* in place of K and X, respectively. We therefore proceed to establish (2).

3. Pre-requisites

Two pre-requisites are stated first.

Rohde's (1969) Lemma: $A^{-}AA^{-} = A^{-}$ if and only if $\rho(A^{-}) = \rho(A)$.

The Moore-Penrose (M-P) inverse, A^+ For given A, this is the unique matrix A^+ satisfying (e.g., Penrose, 1955)

$$\begin{array}{llll} \text{(i)} & AA^{+}A = A & \text{(ii)} & A^{+}AA^{+} = A^{+} \\ \text{(iii)} & (AA^{+})' = AA^{+} & \text{(iv)} & (A^{+}A)' = A^{+}A \end{array} \tag{3}$$

Using the result $A(A'A)^{-}A'A = A$ also yields $A^{+} = A'(AA')^{-}A(A'A)^{-}A'$ and thus $A(A'A)^{-}A' = AA^{+}$. We therefore rewrite (2) as

$$KK^{+} = I - XX^{+} \qquad (4)$$

and proceed to establish its validity. In passing, note from (3) that KK^{+} and XX^{+} are idempotent and symmetric; and from properties of X and K given in the theorem, $\rho(KK^{+}) = n - r$ and

$$KK^{+}(I - XX^{+}) = KK^{+} . \qquad (5)$$

4. Three Proofs of the Identity

Proof 1 Using (5) it is easy to show that conditions (i), (iii) and (iv) of (3) are satisfied by $I - XX^{+}$ as the possible M-P inverse of KK^{+}. Furthermore, since $\rho(I - XX^{+}) = n - r = \rho(KK^{+})$, Rohde's condition is satisfied and so $I - XX^{+}$ also satisfies condition (ii) of (3). Thus $I - XX^{+}$ is the M-P inverse of KK^{+}; but this is KK^{+} itself (due to its idempotency and symmetry). Hence, because the M-P inverse of a matrix is unique, (4) is established, and so (1) is true.

Proof 2 Consider $T = I - XX^{+} - KK^{+}$; that it is symmetric and idempotent is easily shown. Hence, for tr(A) being the trace of A,

$$\begin{aligned} tr(TT') &= tr(T^2) = tr(T) \\ &= tr(I) - tr(XX^{+}) - tr(KK^{+}) \\ &= n - \rho(X) - \rho(K) \\ &= n - r - (n - r) \\ &= 0. \end{aligned}$$

But $\mathbf{T}$ is real, and since for any real matrix $\mathbf{A}$, $\mathrm{tr}(\mathbf{AA}') = 0$ implies $\mathbf{A} = 0$, we have $\mathbf{T} = \mathbf{0}$; i.e., $KK^+ = I - XX^+$.

Proof 3 Because $I - XX^+$ of order n is symmetric and idempotent, it is a projection matrix, and it is the projection matrix on to the space orthogonal to the column space of $\mathbf{X}$. But KK^+ is also a projection matrix, and because K has order $n \times (n - r)$ and is of rank $n - r$ with $\mathbf{K'X} = 0$ it too is the projection matrix on to the same space; i.e., $KK^+ = I - XX^+$.

5. Application to REML

REML is based on the distribution of $\mathbf{K'y}$ for $\mathbf{K'X} = 0$ when y is assumed to have a multivariate normal distribution $\mathbf{N(X\beta, V)}$. Then $K'y \sim N(0, K'VK)$, and so the log likelihood function of $\mathbf{K'y}$ is

$$\ell(V|K'y) = -\tfrac{1}{2}(n-p)\log 2\pi - \tfrac{1}{2}\log|K'VK| - \tfrac{1}{2}y'K(K'VK)^{-1}K'y, \qquad (6)$$

where $|\mathbf{A}|$ denotes the determinant of the matrix $\mathbf{A}$.

On defining $\tau(AB)$ as the product of non-zero eigenvalues of AB, then when AB and BA are both square, $\tau(AB) = \tau(BA)$. Therefore

$$\tau[K(K'VK)^{-1}K'] = \tau[(K'VK)^{-1}K'K] = |K'K|/|K'VK|.$$

Hence for the second term of (6)

$$\log|K'VK| = \log|K'K| - \log\{\tau[K(K'VK)^{-1}K']\}$$

and by (1) this is

$$\log|K'VK| = \log|K'K| - \log\{\tau[V^{-1} - V^{-1}X(X'V^{-1}X)X'V^{-1}]\}.$$

Since $\log|K'K|$ does not involve V, it plays no role in the maximization of (6) with respect to variance components that occur in V. And a second use of (1), in the last term of (6) shows that that term is not a function of K. Hence, since REML is derived from maximizing (6) with respect to variance components, REML does not depend on K. Thus REML estimation is not dependent on which particular set of n - r error contrasts are used as the basis of estimation.

Acknowledgement

Thanks to to Friedrich Pukelsheim for suggesting proofs 2 and 3.

References

Corbeil, R.R. and Searle, S.R. (1976). Restriced maximum likelihood (REML) estimation of variance components in the mixed model. Technometrics 18, 31-38.

Harville, D.A. (1974). Bayesian inference for variance components using only error contrasts. Biometrika 61, 383-385.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. J. Amer. Statist. Assoc. 72, 320-340.

Patterson, H.D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. Biometrika 58, 545-554.

Penrose, R.A. (1955). A generalized inverse for matrices. Proc. Cambridge Philos. Soc. 51, 406-413.

Rohde, C.A. (1966). Some results on generalized inverses. SIAM Review 8, 201-205.

Searle, S.R. (1979). Notes on variance component estimation. Paper BU-673-M, Biometrics Unit, Cornell University.