Statistical Papers **48**,321-327 (2007)

Statistical Papers

A direct derivation of the REML likelihood function

Lynn Roy LaMotte

Biostatistics Program, Louisiana State University School of Public Health, 1600 Canal Street, New Orleans, LA 70112; (e-mail: llamot@lsuhsc.edu)

Received: January 18, 2005; revised version: May 30, 2005

Abstract The original derivation of the widely cited form of the REML likelihood function for mixed linear models is difficult and indirect. This paper derives it directly using familiar operations with matrices and determinants.

Key words REML likelihood, mixed models

1 Introduction

In accounts of restricted or residual maximum likelihood (REML) estimation of variance and covariance components in mixed linear models, the likelihood function is rarely shown. See Jennrich and Schluchter (1986), Searle, Casella, and McCulloch (1992), Littell et al. (1996), and Demidenko (2004), for example.

The REML likelihood is the density function for a statistic $A'y$, and so, with the n-vector y normally distributed with mean vector $X\beta$ and variance-covariance matrix $V(\theta)$, it is straightforward to write it down:

$$L_{\text{REML}} = (2\pi)^{-(n-p)/2}|A'VA|^{-1/2}\exp\{-y'A(A'VA)^{-1}A'y/2\}. \quad (1)$$

The rank of the $n \times p$ matrix X is assumed here to be p. The columns of the $n \times (n-p)$ matrix A form a basis for $\text{sp}(X)^{\perp}$, so $A'X = 0$, and any linear function $L'y$ with $L'X = 0$ can be expressed as a function of $A'y$. Elements of the $n \times n$ positive-definite variance-covariance matrix V are functions of parameters θ.

The logarithm of the REML likelihood function is commonly formulated so that the likelihood is implicitly represented as

$$L_{\text{REML}} = (\text{Const.})|V|^{-1/2}|X'V^{-1}X|^{-1/2} \times \exp\{-(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})/2\}, \quad (2)$$

with $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y$. In this form, it is clear that any effect of the choice of A on L_{REML} is through Const., and so it doesn't affect which value of θ maximizes L_{REML}.

Harville (1974) showed that, when $A'A = \text{I}$, the REML likelihood could be expressed as

$$L_{\text{REML}} = (2\pi)^{-(n-p)/2}|X'X|^{1/2}|V|^{-1/2}|X'V^{-1}X|^{-1/2} \times \exp\{-(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})/2\}. \quad (3)$$

Later, Harville (1977) noted that the likelihood for $A'y$ without the restriction that $A'A = \text{I}$ entails another factor that doesn't depend on θ. These two references appear to be the original sources of the form (2). Other works that discuss REML estimation cite one or both, or they cite other sources that cite them.

The argument that Harville (1974) used to show (3) is sophisticated, nuanced, and presented in a dense thirteen lines. Although it is edifying to work through it, it is indirect and somewhat difficult. The purpose here in this paper is to derive (3) from (1) using only familiar relations and operations on matrices and determinants. That derivation is presented in the next section, followed by a few related notes in the last.

2 Re-expressing the REML Likelihood Function

Notation used here is conventional. For a matrix M with m rows, its transpose is denoted M' and a generalized inverse by M^-. If M is square, its determinant is denoted $|M|$ and its trace by $\text{tr}(M)$. If in addition M is nonsingular, its inverse is denoted M^{-1}. The linear subspace spanned by the columns of M is denoted $\text{sp}(M)$, and its orthogonal complement $\{x \in \Re^m : M'x = 0\}$ by $\text{sp}(M)^\perp$, where $\Re^m$ denotes m-dimensional Euclidean space. The $m \times m$ identity matrix is denoted I, with its dimension understood from the context. For a symmetric, nonnegative-definite matrix S, there exists a symmetric, nonnegative-definite matrix M with the same rank as S such that $S = M^2$. Denote such a matrix M by $S^{1/2}$. In particular, if S is positive definite, then $S^{1/2}$ is positive definite.

Some other basic results from linear algebra are required. An $n \times n$ matrix P such that, for every $y \in \Re^n$, Py is the orthogonal projection of y onto $\mathcal{S} = \text{sp}(P)$, is symmetric and idempotent, and it is unique for $\mathcal{S}$. For any matrix M such that $\text{sp}(M) = \mathcal{S}$, $P = M(M'M)^-M'$. If M and N are $m \times m$ matrices, then $|MN| = |M||N|$. There are several versions of the determinant of a square matrix in partitioned form. The only one we need here is for a nonsingular symmetric matrix M, partitioned as $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}' & M_{22} \end{pmatrix}$:

$$|M| = |M_{22}||M_{11} - M_{12}M_{22}^{-1}M_{12}'|. \tag{4}$$

This development depends on two propositions, one for the quadratic form in the exponent and the other for the determinant in (1). Proposition 1 is essentially the same as one shown in Searle, Casella, and McCulloch (1992, p. 452), proof of which they attribute to F. Pukelsheim. We provide two alternative statements and proofs leading to the same result. The first, due to an anonymous reviewer, relies on the fact that $(V^{-1}X, A)$ is nonsingular, which can be established readily.

Proposition 1. If V is an $n \times n$ positive-definite matrix and X and A are as described above, then

$$V^{-1} = V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} + A(A'VA)^{-1}A'.$$

Proof. Note that

$$\begin{aligned}(V^{-1}X, A)^{-1}V^{-1}(V^{-1}X, A)'^{-1} &= [(V^{-1}X, A)'V(V^{-1}X, A)]^{-1} \\ &= \begin{pmatrix} X'V^{-1}X & 0 \\ 0 & A'VA \end{pmatrix}^{-1} \\ &= \begin{pmatrix} (X'V^{-1}X)^{-1} & 0 \\ 0 & (A'VA)^{-1} \end{pmatrix}.\end{aligned}$$

Then

$$\begin{aligned}V^{-1} &= (V^{-1}X, A)[(V^{-1}X, A)'V(V^{-1}X, A)]^{-1}(V^{-1}X, A)' \\ &= (V^{-1}X, A)\begin{pmatrix} (X'V^{-1}X)^{-1} & 0 \\ 0 & (A'VA)^{-1} \end{pmatrix}(V^{-1}X, A)' \\ &= V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} + A(A'VA)^{-1}A',\end{aligned}$$

hence the result.

This version of Proposition 1 is all that we need for the development here. However, the same end can be achieved as an implication of a more general result that doesn't depend on X and A having full column rank, as shown next.

Proposition 1*. Let columns of the matrix (B_1, B_2) span $\Re^n$. Assume that both B_1 and B_2 have at least one column and that $B_1'B_2 = 0$. Then

$$B_1(B_1'B_1)^{-}B_1' + B_2(B_2'B_2)^{-}B_2' = \mathrm{I}. \tag{5}$$

Proof. The matrix on the left of the equal sign (call it P) is symmetric and idempotent, and so it is the unique matrix that performs orthogonal projection onto the linear subspace $\mathcal{S}$ spanned by its columns, which is $\mathrm{sp}(B_1, B_2) = \Re^n$. That is, $P = \mathrm{I}$.

Substituting $V^{-1/2}X$ and $V^{1/2}A$ for B_1 and B_2, it is straightforward to verify that $\mathrm{sp}(V^{-1/2}X, V^{1/2}A) = \Re^n$. Then using Proposition 1*, (5) becomes

$$V^{-1/2}X(X'V^{-1}X)^{-}X'V^{-1/2} + V^{1/2}A(A'VA)^{-}A'V^{1/2} = \mathrm{I}, \qquad (6)$$

hence

$$\begin{aligned} y'A(A'VA)^{-}A'y &= y'[V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}]y \\ &= (y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}), \end{aligned} \qquad (7)$$

and so the quadratic forms in the exponents of (1) and (3) are the same. In this paragraph, the only conditions required are that $\mathrm{sp}(X, A) = \Re^n$ and $A'X = 0$ and V be positive definite; the assumption that X and A have full column rank can be relaxed in showing equivalence of the quadratic forms in (1) and (3).

Let $W = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$ and $V_i = \partial V/\theta_i$. Note that

$$\begin{aligned} \frac{\partial \log |A'VA|}{\partial \theta_i} &= \mathrm{tr}[(A'VA)^{-1}A'V_iA] \\ &= \mathrm{tr}[A(A'VA)^{-1}A'V_i] = \mathrm{tr}(WV_i), \end{aligned}$$

which is the same as

$$\begin{aligned} \frac{\partial}{\partial \theta_i}[\log |V| + \log |X'V^{-1}X|] &= \mathrm{tr}(V^{-1}V_i) - \mathrm{tr}[(X'V^{-1}X)^{-1}X'V^{-1}V_iV^{-1}X] \\ &= \mathrm{tr}(WV_i). \end{aligned}$$

This can be used to justify the relation

$$|A'VA| = (\mathrm{Const.})|V||X'V^{-1}X|.$$

Thus only Proposition 1 or 1* is required to get from (1) to (2).

Still, the relation between the determinants in (1) and (2) or (3) may be interesting in its own right. So far as I have found, this relation is never invoked in published descriptions of the REML likelihood, and my efforts to find it among known relations for determinants came up empty. After several attempts over several years, finally I came up with the argument shown below. A reviewer of the first version of this paper pointed out that the proposition and its proof differ only trivially from results shown in Searle (1979, pp. 26-27), a priority which I acknowledge gladly and with admiration. The proof is shown here because Searle's technical report is difficult to obtain.

Proposition 2. If V is an $n \times n$ symmetric, positive-definite matrix, and (X, A) is an $n \times n$ matrix with linearly independent columns, and $A'X = 0$, then

$$|V| = \frac{|A'VA||X'X|}{|A'A||X'V^{-1}X|}. \qquad (8)$$

Proof. Both (X, A) and its transpose are square, and so

$$\begin{aligned}
|A'A||X'X||V| &= \left|\begin{pmatrix} X' \\ A' \end{pmatrix} V(X, A)\right| \\
&= \begin{vmatrix} X'VX & X'VA \\ A'VX & A'VA \end{vmatrix} \\
&= |A'VA||X'VX - X'VA(A'VA)^{-1}A'VX| \\
&= |A'VA||X' \underbrace{[V - VA(A'VA)^{-1}A'V]}_{=X(X'V^{-1}X)^{-1}X' \text{ by } (6)} X| \\
&= |A'VA||X'X|^2/|X'V^{-1}X|,
\end{aligned}$$

and (8) follows.

In general, for any $n \times (n-p)$ matrix A with linearly independent columns such that $A'X = 0$, from (8),[1]

$$|A'VA| = |A'A||V||X'V^{-1}X|/|X'X|. \tag{9}$$

In general, then, these two results show that the REML likelihood function, the joint density function of $A'y$, can be expressed as

$$\begin{aligned}
L_{\text{REML}} = (2\pi)^{-(n-p)/2}|X'X|^{1/2}|V|^{-1/2}|X'V^{-1}X|^{-1/2}|A'A|^{-1/2} \\
\times \exp\{-(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta})/2\}. \quad (10)
\end{aligned}$$

When $A'A = \mathrm{I}$, the determinant $|A'VA|$ in (1) is, by (9), the same as the determinant $|V||X'V^{-1}X|/|X'X|$ in (3).

3 Discussion

The main reason for writing the REML likelihood function at all is to use its maximization as a rationale to follow to obtain estimators of variance and covariance components. Central to that objective are the REML equations. They may be obtained by setting to zero the partial derivatives of the logarithm of the REML likelihood.

The REML equations also can be derived without explicitly stating the REML likelihood function. In most applications, though, the parameter set is restricted by nonnegativity constraints, and the REML estimates may not be solutions to the REML equations. Recognizing this, there is some utility in deriving an explicit form of the REML likelihood function.

Before the REML name was coined, LaMotte (1970) derived the same equations by equating quadratic forms $Q_i = y'WV_iWy$ that appear in the derivatives of the full log-likelihood function (substituting $y - X\hat{\beta} = VWy$ and noting that $WVW = W$) to expressions for their expected values in

[1] Apparently there is a typographical error in exercise E9.12(f) in Searle, Casella, and McCulloch (1992, p. 361), which states that $|A'VA| = |V|/|X'V^{-1}X|$.

terms of θ. That paper noted that the same equations could be obtained by maximizing a second-order approximation to the log-likelihood function, and it suggested that parameter space constraints (like non-negativity) could be imposed by solving an iterative sequence of constrained quadratic programming problems.

Searle, Casella, and McCulloch (1992) took a different approach, substituting $A'y$, 0, and $A'VA$ for y, $X\beta$, and V in the full likelihood equations. Other accounts of the REML equations start with the logarithm of (2) (see Jennrich and Schluchter 1986, Littell et al. 1996, and Demidenko 2004, for example).

The REML equations or the forms (2) and (3) are used to establish that REML estimates are invariant to the choice of A, apart from a constant that is independent of V (see Harville (1977, p. 325) and Searle, Casella, and McCulloch (1992, p. 252)). However, this kind of invariance is evident directly from (1). For any matrix A_* with linearly independent columns such that $\mathrm{sp}(A_*) = \mathrm{sp}(X)^{\perp}$, there exists a nonsingular matrix M such that $A_* = AM$. Then it can be shown directly that

$$A_*(A_*'VA_*)^{-1}A_*' = A(A'VA)^{-1}A'$$

and

$$|A_*'VA_*| = |M'M||A'VA|,$$

so that the REML likelihood functions in terms of $A'y$ and $A_*'y$ differ by the factor $|M'M|^{1/2}$. Once invariance is established, one may work with (1) or (2), whichever is more convenient for computations.

Laird (2004, pp. 45-46) develops the determinant of the REML likelihood from a one-to-one linear transformation $(A, G')'y$, which is similar to Harville's (1974) original derivation. The assertion (Laird 2004, p. 44) that the REML likelihood is not invariant to X is incorrect, as can be seen from (9).

The forms (3) or (2) may serve to make the relation between the full likelihood function and the REML likelihood function more transparent than (1). Using (3) illuminates relations to the Aitken equations and mixed-model equations (see Searle, Casella, and McCulloch (1992, p. 275)). Expressions for the derivatives of the log-likelihood function can be obtained from either form with about the same degree of difficulty. Although expressions for derivatives from (3) appear to be tedious, they are simplified considerably by noting the relation $\partial W/\partial\theta_i = -WV_iW$.

4 Acknowledgment

I am grateful for the careful reading and constructive suggestions from two reviewers and the editor. The result is a much better paper than I first submitted.

References

1. Demidenko, E. (2004). *Mixed Models: Theory and Applications.* John Wiley & Sons.
2. Harville, D.A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**: 383–385.
3. Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**: 320–338.
4. Jennrich, R.I., and Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**:805–820.
5. Laird, N. (2004). *Analysis of Longitudinal and Cluster-Correlated Data.* NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 8, Beachwood, Ohio: Institute of Mathematical Statistics.
6. LaMotte, L.R. (1970). A class of estimators of variance components. Technical Report No. 10, Department of Statistics, University of Kentucky, Lexington, KY.
7. Littell, R.C., Milliken, G.A., Stroup, W.W., and Wolfinger, R.D. (1996). *SAS System for Mixed Models*, Cary, NC: SAS Institute Inc.
8. Searle, S.R. (1979). Notes on variance component estimation. A detailed account of maximum likelihood and kindred methodology. Technical Report BU-673-M, Biometrics Unit, Cornell University, Ithaca, New York.
9. Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components.* John Wiley & Sons.