APPLYING THE **EM** ALGORITHM TO CALCULATING **ML** AND **REML** ESTIMATES OF VARIANCE COMPONENTS[1]

Shayle R. Searle
Biometrics Unit, Cornell University, Ithaca, N. Y.

BU-1213-M July 1993

ABSTRACT

Maximum likelihood (ML) is a firmly established estimation technique, and the estimation-maximization (EM) algorithm is a widely used iterative procedure for solving non-linear estimation equations. A confluence of these two methodologies is in the estimation of variance components. Because numerous descriptions of each method are often found difficult to understand, this paper is an attempt at providing a straightforward explanation of using the EM algorithm for estimating variance components.

INTRODUCTION

The development of ML (maximum likelihood) estimation began in a 1947 Ph.D. thesis by Crump at Iowa State University (Crump, 1947). That first attempt dealt with just the one-way classification, for both balanced and unbalanced (unequal subclass numbers) data, although for the latter it did not go much further than developing two non-linear equations in the two variance components. Herbach (1959) made the next step, dealing solely with balanced data but giving careful consideration to the need for maximum likelihood estimators of variance components to be non-negative (because ML estimation is designed to do its maximizing over the relevant parameter space which, for variance components, is non-negative).

The landmark paper is, without doubt, that of Hartley and J.N.K. Rao (1967). They develop general equations for the maximum likelihood estimation of fixed effects and variance components, equations that are applicable to *any* unbalanced data situation with any mixture of fixed and random effects. The secret of their success was innovativeness in handling a general mixed model in matrix-vector notation.

---

[1] Paper invited for the 1993 American Statistical Association meetings, San Francisco.

NOTATION FOR A GENERAL MIXED MODEL

To see what Hartley and Rao (1967) did, let us begin with the familiar model equation for y_{ijk}, the k'th observation on treatment i in block j of a randomized complete block experiment:

$$y_{ijk} = \mu + t_i + b_j + (tb)_{ij} + e_{ijk} \,. \tag{1}$$

On assembling all the y_{ijk}-values in a column vector $\mathbf{y}$, and likewise putting all the parameters, the μ, the t_is, b_js and $(tb)_{ij}$s in a vector $\boldsymbol{\beta}^*$, and the e_{ijk}s in a vector $\mathbf{e}$, we have the familiar matrix-vector form of the model equation

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{e} \,, \tag{2}$$

for some matrix $\mathbf{X}^*$ that has zeros and one as elements. Suppose (which is usually not so) that block effects are treated as fixed effects. Then from (2), the least squares estimator of $\mathbf{X}^*\boldsymbol{\beta}^*$ is $\mathbf{X}^*\mathbf{G}\mathbf{X}^{*'}\mathbf{y}$ where $\mathbf{G}$ is a generalized inverse of $\mathbf{X}^{*'}\mathbf{X}^*$.

Now suppose (as is usual) that block effects are random and treatment-by-block interaction effects are then random also. What Hartley and Rao (1967) did was to separate out those random effects into a vector that we will call $\mathbf{u}$, and so the model equation (2) could then be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \,, \tag{3}$$

for appropriate matrices $\mathbf{X}$ and $\mathbf{Z}$. And this formulation is suitable for any mixed linear model for all kinds of data, balanced or unbalanced, even with covariates as part of $\mathbf{X}$.

Facilitation of variance components models was then achieved by Hartley and Rao (1967) by their observing that $\mathbf{Z}\mathbf{u}$ could be expressed as

$$\mathbf{Z}\mathbf{u} = \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \cdots + \mathbf{Z}_i\mathbf{u}_i + \cdots + \mathbf{Z}_r\mathbf{u}_r \tag{4}$$

where each $\mathbf{u}_i$ is a vector of all the effects (q_i of them, say), of the i'th random factor occurring in the data, there being, in general, r random factors. For example, for $r = 2$ for (1), $\mathbf{u}_1$ could be the b_js, and $\mathbf{u}_2$ the $(tb)_{ij}$s. More than that, it was also possible to write $\mathbf{e}$ of (3) as

$$\mathbf{e} = \mathbf{Z}_0\mathbf{u}_0 \quad \text{with} \quad \mathbf{Z}_0 \equiv \mathbf{I}_N \quad \text{and} \quad \mathbf{u}_0 \equiv \mathbf{e} \,, \tag{5}$$

where N is the number of observations. In this way the model equation could be taken as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad \text{with} \quad \mathbf{Z}\mathbf{u} = \sum_{i=1}^{r} \mathbf{Z}_i\mathbf{u}_i \tag{6}$$

or, on using (5), it could be taken as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad \text{with} \quad \mathbf{Z}\mathbf{u} = \sum_{i=0}^{r} \mathbf{Z}_i\mathbf{u}_i \,. \tag{7}$$

Then, for the traditional variance components model $\mathrm{E}(\mathbf{u}_i) = \mathbf{0}$, and we attribute variances and covariances in the form

$$\mathrm{var}(\mathbf{u}_i) = \sigma_i^2\mathbf{I}_{q_i} \quad \text{and} \quad \mathrm{cov}(\mathbf{u}_i, \mathbf{u}'_{i'}) = \mathbf{0} \quad \forall\ i \neq i' \,. \tag{8}$$

In this way the dispersion matrix of $\mathbf{y}$, denoted by $\mathbf{V}$, is

$$\mathrm{var}(\mathbf{y}) = \mathbf{V} = \sum_{i=0}^{r} \mathbf{Z}_i\mathbf{Z}'_i\sigma_i^2 \,. \tag{9}$$

ML EQUATIONS

With the structure of (7), (8) and (9), and on assuming normality for all elements of $\mathbf{u}$, i.e., for all the random effects, Hartley and Rao developed, in a matrix-vector form, explicit ML equations for the fixed effects in $\boldsymbol{\beta}$ and for the variance components σ_i^2 for $i = 0, 1, \cdots, r$.

There are several forms of these equations; one of the more succinct of these, and easiest to comprehend, is the following for the variance components.

$$\left\{ \underset{m}{\mathrm{tr}}(\dot{\mathbf{V}}^{-1}\mathbf{Z}_i\mathbf{Z}'_i\dot{\mathbf{V}}^{-1}\mathbf{Z}_j\mathbf{Z}'_j) \right\}_{i,j=0}^{r} \dot{\boldsymbol{\sigma}}^2 = \left\{ \underset{c}{\mathbf{y}'\dot{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}'_i\dot{\mathbf{P}}\mathbf{y}} \right\}_{i=0}^{r} \tag{10}$$

where

$$\dot{\mathbf{V}} = \sum_{i=0}^{r} \mathbf{Z}_i\mathbf{Z}'_i\dot{\sigma}_i^2 \tag{11}$$

and

$$\dot{\mathbf{P}} = \dot{\mathbf{V}}^{-1} - \dot{\mathbf{V}}^{-1}\dot{\mathbf{X}}(\mathbf{X}'\dot{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\dot{\mathbf{V}}^{-1} \,. \tag{12}$$

The form of these $r + 1$ equations is that on the left-hand side of (10) the vector of variance components is pre-multiplied by a matrix whose elements are traces of matrix products; and the right-hand side is a vector of quadratic forms in the observation vector $\mathbf{y}$.

It is clear that equations (10) are non-linear in the elements $\dot{\sigma}_i^2$ of $\dot{\boldsymbol{\sigma}}^2$; they occur in $\dot{\mathbf{V}}$ and, through (12), in $\dot{\mathbf{P}}$ with $\dot{\mathbf{V}}$ and $\dot{\mathbf{P}}$ themselves occurring in (10). Moreover, solutions of these equations are not, as they stand, necessarily the ML estimators of the σ^2s. The ML estimators are solutions

that are non-negative; indeed, we must have $\tilde{\sigma}_0^2 > 0$ and $\tilde{\sigma}_i^2 \geq 0$ for $i = 1, \cdots, r$, for the ML estimators $\tilde{\boldsymbol{\sigma}}^2$ and $\tilde{\mathbf{V}} = \sum_{i=0}^{r} \mathbf{Z}_i \mathbf{Z}_i' \tilde{\sigma}_i^2$. Then the ML estimator of $\mathbf{X}\boldsymbol{\beta}$ is

$$\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y} \,. \tag{13}$$

Equations (12) and (13) are highly intractable analytically, as are other forms of the ML equations also and, except in a few cases of balanced data [but not all, e.g., the 2-way crossed classification, with interaction, random model – see Miller (1973)], they do not yield closed form solutions. Hence they have to be solved numerically – usually by iteration and, for deriving ML estimations the non-negativity requirement must also be taken into account. There are, of course, many ways of doing the iteration; e.g., gradient methods, Fisher scoring, Newton-Rhapson and so on, and nowadays the popular EM (expectation-maximization) algorithm. Because there are numerous ways of using this algorithm, and because descriptions of them are often difficult to understand, the purpose of this paper is to make an attempt at providing a succinct, straightforward description of applying the EM algorithm to estimation of variance components.

USING THE EM ALGORITHM

A starting point for using the EM algorithm [developed quite generally by Dempster, Laird and Rubin (1977)] is to distinguish between what are called complete data and what are called incomplete data. In many situations there can be more than one way of making this distinction, but for mixed models the distinction which provides a straightforward application of the algorithm is as follows. The actual data, the vector $\mathbf{y}$, is called the *incomplete data.* This is so because although $\mathbf{y}$ is known, the random effects in $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r$ are not known. If they were, we would know how to estimate $\boldsymbol{\beta}$ and the variance components. Therefore $\mathbf{y}$ and $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_r$ are called the *complete data.*

Suppose we could have the complete data. Then, on assuming normality, maximum likelihood estimation of $\boldsymbol{\beta}$ and the σ_i^2s leads to the equations (see, for example, Searle, Casella and McCulloch, 1992, Section 8.3b)

$$\hat{\sigma}_i^2 = \frac{\mathbf{u}_i'\mathbf{u}_i}{q_i} \quad \text{and} \quad \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\left(\mathbf{y} - \sum_{i=1}^{r} \mathbf{Z}_i\mathbf{u}_i\right). \tag{14}$$

But, in fact, we cannot use (14) because we do not know the $\mathbf{u}_i$s. So in their place we use some expected values: for $\hat{\sigma}_i^2$, the expected value of $\mathbf{u}_i'\mathbf{u}_i$, in place of $\mathbf{u}_i'\mathbf{u}_i$; and for $\mathbf{X}\hat{\boldsymbol{\beta}}$, the expected value of

$\mathbf{u}_i$, in place of $\mathbf{u}_i$. And, the expectation we use is the conditional expectation, conditional on the data being the vector $\mathbf{y}$ that has been observed. So in (14), for $\hat{\sigma}_i^2$, we use $\mathrm{E}(\mathbf{u}_i'\mathbf{u}_i \,|\, \mathbf{y})$ in place of $\mathbf{u}_i'\mathbf{u}_i$, and for $\mathbf{X}\hat{\boldsymbol{\beta}}$ we use $\mathrm{E}(\mathbf{u}_i \,|\, \mathbf{y})$ in place of $\mathbf{u}_i$. Expressions for these expected values are easily derived, because of the assumed normality.

Thus because

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_r \end{bmatrix} \sim \mathcal{N}\left\{ \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V} & \left\{ {}_r\, \sigma_i^2 \mathbf{Z}_i \right\}_{i=1}^{r} \\ \left\{ {}_c\, \sigma_i^2 \mathbf{Z}_i' \right\}_{i=1}^{r} & \left\{ {}_d\, \sigma_i^2 \mathbf{I}_{q_i} \right\}_{i=1}^{r} \end{bmatrix} \right\} \tag{15}$$

it is not difficult to show that

$$\mathbf{u}_i \,|\, \mathbf{y} \sim \mathcal{N}\left[\sigma_i^2 \mathbf{Z}_i \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad \sigma_i^2 \mathbf{I}_{q_i} - \sigma_i^4 \mathbf{Z}_i' \mathbf{V}^{-1} \mathbf{Z}_i \right]. \tag{16}$$

From this comes

$$\mathrm{E}(\mathbf{u}_i'\mathbf{u}_i \,|\, \mathbf{y}) = \sigma_i^4 (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} \mathbf{Z}_i \mathbf{Z}_i' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \mathrm{tr}\left(\sigma_i^2 \mathbf{I}_{q_i} - \sigma_i^4 \mathbf{Z}_i' \mathbf{V}^{-1} \mathbf{Z}_i \right) \tag{17a}$$

for use in $\hat{\sigma}_i^2$ of (14); and for $\mathbf{X}\hat{\boldsymbol{\beta}}$ we use

$$\mathrm{E}(\mathbf{u}_i \,|\, \mathbf{y}) = \sigma_i^2 \mathbf{Z}_i' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \,. \tag{17b}$$

Since (17a) and (17b) both involve $\boldsymbol{\beta}$ and the σ^2s, which we want to estimate using (14), we cannot just substitute (17) into (14) and be done. We have to use (17) and (14) alternately, in an iterative procedure: begin with some *à priori* starting values, $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\sigma}^{2(0)}$, say; use them in (17) to get computed values for $\mathrm{E}(\mathbf{u}_i'\mathbf{u}_i \,|\, \mathbf{y})$ and $\mathrm{E}(\mathbf{u}_i \,|\, \mathbf{y})$; these computed values are then used in (14) in place of, respectively, $\mathbf{u}_i'\mathbf{u}_i$ and $\mathbf{u}_i$. This gives computed values for $\hat{\sigma}_i^2$ and $\mathbf{X}\hat{\boldsymbol{\beta}}$ which are then used in (17) – and so the iterative process continues, (17), (14), (17), $\cdots$. Thus it is that the process is called the EM algorithm: (17) is the E-step, expectation and (14) is the M-step, maximization.

TWO VERSIONS OF THE EM ALGORITHM

The algorithm just described can be stated more explicitly as follows.

Version 1

Step 0 Choose starting values $\boldsymbol{\beta}_0$ and σ_0^2; set m = 0.

Step 1 (E-step) Calculate two terms that will be denoted $t_{i,m}$ and $\mathbf{s}_m$. First, from (17a)

$$\begin{aligned} t_{i,m} &= \mathrm{E}(\mathbf{u}_i'\mathbf{u}_i \,|\, \mathbf{y})\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_m \text{ and } \sigma^2=\sigma_m^2} \\ &= \sigma_{i,m}^4(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_m)'\mathbf{V}_m^{-1}\mathbf{Z}_i\mathbf{Z}_i'\mathbf{V}_m^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_m) \\ &\quad + \mathrm{tr}\Big[\sigma_m^2\mathbf{I}_{q_i} - \sigma_{i,m}^4\mathbf{Z}_i'\mathbf{V}_m^{-1}\mathbf{Z}_i\Big]. \end{aligned} \tag{18}$$

Second, based on (17b),

$$\mathbf{s}_m = \mathrm{E}\left[\left(\mathbf{y}-\sum_{i=1}^{r}\mathbf{Z}_i\mathbf{u}_i\right)|\,\mathbf{y}\right]\Bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_m \text{ and } \sigma^2=\sigma_m^2},$$

which easily reduces to

$$\mathbf{s}_m = \mathbf{X}\boldsymbol{\beta}_m + \sigma_{0,m}^2\mathbf{V}_m^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}_m)\,. \tag{19}$$

Step 2 (M-step)

$$\hat{\sigma}_{i,m+1}^2 = \frac{t_{i,m}}{q_i} \tag{20}$$

and

$$\mathbf{X}\boldsymbol{\beta}_{m+1} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{s}_m\,. \tag{21}$$

Step 3 If convergence is reached, set $\tilde{\sigma}^2 = \hat{\sigma}_{m+1}^2$ and $\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}\boldsymbol{\beta}_{m+1}$; otherwise increase m by unity and return to Step 1.

Version 2

In Version 1, equation (21) gives

$$\mathbf{X}\boldsymbol{\beta}_m = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{s}_{m-1}\,. \tag{22}$$

This is not the same as

$$\mathbf{X}\boldsymbol{\beta}_{*,m} = \mathbf{X}\Big[\mathbf{X}'\mathbf{V}_m^{-1}\mathbf{X}\Big]^{-}\mathbf{X}'\mathbf{V}_m^{-1}\mathbf{y}\,, \tag{23}$$

which would be calculated from the ML equation (8). However, if we used (23) in place of (22) and

then substituted it into (18) where it occurs only in the form $\mathbf{V}_m^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_m)$ we would have

$$\mathbf{V}_m^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_m) = \left\{ \mathbf{V}_m^{-1} - \mathbf{V}_m^{-1}\mathbf{X}\left[\mathbf{X}'\mathbf{V}_m^{-1}\mathbf{X}\right]^{-}\mathbf{X}'\mathbf{V}_m^{-1} \right\}\mathbf{y} = \mathbf{P}_m\mathbf{y} ,$$

where $\mathbf{P}_m$ is $\dot{\mathbf{P}}$ of (12) with $\mathbf{V}_m$ in place of $\dot{\mathbf{V}}$. Then (18) would be

$$t_{i,m} = \sigma_{i,m}^4 \mathbf{y}'\mathbf{P}_m\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}_m\mathbf{y} + \mathrm{tr}\left[\sigma_{i,m}^2\mathbf{I}_{q_i} - \sigma_{i,m}^4\mathbf{Z}_i'\mathbf{V}_m^{-1}\mathbf{Z}_i\right] \tag{24}$$

and this second version of iterating would be as follows.

Step 0 Choose a starting value σ_0^2; set $m = 0$.

Step 1 (E-step): Calculate (24).

Step 2 (M-step): Calculate $\hat{\sigma}_{i,m+1}^2 = \mathrm{t}_{i,m} \Big/ \mathrm{q}_i$.

Step 3 If convergence is reached, set $\tilde{\sigma}^2 = \hat{\sigma}_{m+1}^2$ and calculate $\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\tilde{\mathbf{V}}^{-1}\mathbf{y}$; otherwise increase m by unity and return to Step 1.

Since the pure EM algorithm of Version 1 uses $\mathbf{X}\boldsymbol{\beta}_m$ of (22), which is not the same as $\mathbf{X}\boldsymbol{\beta}_{m,*}$ of (23), the use of the latter in (18) as shown in (24) is, to quote Searle, Casella and McCulloch (1992, p. 301), "strictly speaking, not EM. The differences are slight and probably do not affect convergence properties." The suggestion for using $\mathbf{P}_m\mathbf{y}$ was made by Laird (1983).

REML

The estimation idea $\hat{\sigma}_i^2 = \mathbf{u}_i'\mathbf{u}_i/\mathrm{q}_i$ was first published in Patterson and Thompson (1971) and it also appeared in Henderson (1973). In the Patterson and Thompson paper the innovation was to estimate the variance components of a mixed model in a manner that effectively removes the fixed effects from the model, particularly in the context of recovery of inter-block information in experiments that use blocking designs. This has come to be known as REML, restricted maximum likelihood, although in Europe it is more often called marginal maximum likelihood. Two descriptions of it alternative to that of removing fixed effects are that it is an estimation method (i) that maximizes that part of the likelihood which is location invariant (Thompson, 1962), i.e., invariant to the fixed effects, and (ii) that maximizes the likelihood function of the maximum number of linearly independent error contrasts (Harville, 1977), i.e., of linear combinations $\mathbf{k}'\mathbf{y}$ of the data such that $\mathbf{k}'\mathbf{X} = \mathbf{0}$. Then $\mathbf{k}'\mathbf{y}$ contains no $\boldsymbol{\beta}$. We represent a totality of such linearly independent combinations by $\mathbf{K}'\mathbf{y}$ where $\mathbf{K}'\mathbf{X}$ ($\mathbf{X}$ having rank $r_{\mathbf{X}}$) $= \mathbf{0}$ and $\mathbf{K}'$ has maximal row rank, $N - r_{\mathbf{X}}$ for $\mathbf{X}$ having rank $r_{\mathbf{X}}$.

Considerable algebra has flowed from all these descriptions but one of the easiest ways of deriving the REML procedure is simply to adapt the ML results for

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta,\ \mathbf{V}) \quad \text{to} \quad \mathbf{K}'\mathbf{y} \sim \mathcal{N}(\mathbf{0},\ \mathbf{K}'\mathbf{V}\mathbf{K}) .$$

This entails making the following replacements in the ML results for $\mathbf{y}$: replace

$$\mathbf{y} \text{ by } \mathbf{K}'\mathbf{y} \qquad \mathbf{Z}_i \text{ by } \mathbf{K}'\mathbf{Z}_i \qquad N \text{ by } N - r_{\mathbf{X}}$$
$$\mathbf{X} \text{ by } \mathbf{K}'\mathbf{X} = \mathbf{0} \qquad \mathbf{V} \text{ by } \mathbf{K}'\mathbf{V}\mathbf{K}$$

The replacement of N by $N - r_{\mathbf{X}}$ is because $N - r_{\mathbf{X}}$ is the number of elements in $\mathbf{K}'\mathbf{y}$. These replacements, together with the important equality

$$\mathbf{K}(\mathbf{K}'\mathbf{V}\mathbf{K})^{-1}\mathbf{K}' = \mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}$$

(as detailed in Searle *et al.*, *loc. cit.*, p. 451), alter the ML equations (10) to be

$$\left\{ \underset{m}{\text{tr}}(\check{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i'\check{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i') \right\}_{i,j=0}^{r} \check{\sigma}^2 = \left\{ \underset{c}{\mathbf{y}'\check{\mathbf{P}}\mathbf{Z}_i\mathbf{Z}_i'\check{\mathbf{P}}\mathbf{y}} \right\}_{i=0}^{r} . \tag{25}$$

The only difference between these and the ML equations of (10) is that on the left-hand side of (25) there is $\mathbf{P}$ where in (10) there is $\mathbf{V}^{-1}$.

The effect of this on the EM algorithm is that it is exactly the same as Version 2 of EM for ML except that in (24) the $\mathbf{V}^{-1}$ in the trace term becomes $\mathbf{P}$. Hence the algorithm is as follows.

EM algorithm for REML

Step 0 Choose a starting value σ_0^2; set m = 0.

Step 1 Calculate

$$t_{i,m} = \sigma_{i,m}^4 \mathbf{y}'\mathbf{P}_m\mathbf{Z}_i\mathbf{Z}_i'\mathbf{P}_m\mathbf{y} + \text{tr}\left[\sigma_{i,m}^2 \mathbf{I}_{q_i} - \sigma_{i,m}^4 \mathbf{Z}_i'\mathbf{P}_m\mathbf{Z}_i\right].$$

Step 2 Calculate $\hat{\sigma}_{i,m}^2 = t_{i,m}/q_i$.

Step 3 At convergence, set $\check{\sigma}^2 = \hat{\sigma}_{m+1}^2$, otherwise increase m by unity and return to Step 1.

It is to be noted that REML provides no estimator for the fixed effects. Nevertheless, the practical procedure is to use

$$\mathbf{X}\check{\beta} = \mathbf{X}(\mathbf{X}'\check{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\check{\mathbf{V}}^{-1}\mathbf{y} ,$$

similar to the ML estimator of $\mathbf{X}\beta$ in (13). Kackar and Harville (1981, 1984) describe the unbiasedness and variance properties of this.

References

Crump, S.L. (1947). The estimation of variance in multiple classifications. Ph.D. Thesis, Iowa State University, Ames, Iowa.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc., Ser. B* **39**, 1-38.

Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93-108.

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

Henderson, C.R. (1973). Maximum likelihood estimation of variance components. Unpublished manuscript, Department of Animal Science, Cornell University, Ithaca, New York.

Herbach, L.H. (1959). Properties of Model II type analysis of variance tests, A: Optimum nature of the F-test for Model II in the balanced case. *Ann. Math. Statist.* **30**, 939-959.

Kackar, R.N. and Harville, D.A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Com. Statist. A: Theory & Methods* **10**, 1249-1261.

Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimation of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.* **79**, 853-861.

Laird, N.M. (1982). Computation of variance components using the EM algorithm. *J. Statist. Comp. & Simul.* **14**, 295-303.

Miller, J.J. (1973). Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. Technical Report No. 12, Department of Statistics, Stanford University, Stanford, California.

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

Searle, S.R., Casella, G. and McCulloch, C.E. (1992). *Variance Components*, Wiley & Sons, New York, New York.

Thompson, W.A., Jr. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33**, 273-289.