The Annals of Statistics
1980, Vol. 8, No. 4, 811–819

EXPLICIT MAXIMUM LIKELIHOOD ESTIMATES FROM BALANCED DATA IN THE MIXED MODEL OF THE ANALYSIS OF VARIANCE

BY TED H. SZATROWSKI[1] AND JOHN J. MILLER

Rutgers University

A result of Szatrowski, giving necessary and sufficient conditions for the existence of explicit maximum likelihood estimates for multivariate normal means and covariances with linear structure, can be applied to the problem of obtaining explicit maximum likelihood estimates in the mixed model of the analysis of variance. This application yields a simple procedure for checking whether or not explicit maximum likelihood estimates exist for the parameters in the balanced mixed model of the analysis of variance. Examples of this procedure as well as a discussion on finding the explicit maximum likelihood estimates are given.

1. Introduction. Results in the literature on when the parameters of the mixed model of the analysis of variance have explicit maximum likelihood estimates are sparse and in general consist of solving the likelihood equations for specific balanced models. It has been shown that the balanced two-way nested ANOVA model has explicit maximum likelihood estimates, while the balanced two-way crossed random effects ANOVA model with interaction does not have explicit maximum likelihood estimators (e.g., Herbach (1959), Hartley and Rao (1967), Miller (1973, 1977)). Harville (1977) reviews maximum likelihood approaches to variance component estimation. Miller (1973) gives sufficient conditions for the existence of explicit maximum likelihood estimates for the balanced mixed model of the analysis of variance. Szatrowski (1980) gives necessary and sufficient conditions for the existence of explicit maximum likelihood estimates for multivariate normal means and covariances with linear structure. In the present study, the result of Szatrowski (1980) is applied to the problem of finding explicit maximum likelihood estimators in the mixed model of the analysis of variance. This application yields a simple procedure for determining whether or not explicit maximum likelihood estimates exist.

In Section 2, the basic model in the mixed model of the analysis of variance is reviewed along with the simplifications in the form of this model in the balanced case. In Section 3, Szatrowski's (1980) result is given along with a simple procedure for determining whether or not a balanced model has explicit maximum likelihood

Received October 1977; revised August 1978.

[1]Research partially supported by the Rutgers University Research Council and National Science Foundation Grant Number MCS77-28184.

AMS 1970 *subject classifications.* 62J10, 62F10.

Key words and phrases. Analysis of variance, explicit solutions, maximum likelihood estimate, mixed model, variance component estimation.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

estimates and examples of this procedure. The forms of the explicit estimators and several examples of explicit estimators are given in Section 4.

2. The basic model.

2.1 *The general form.* The basic model we shall use in the mixed model analysis of variance is that given by Hartley and Rao (1967) and Miller (1977). It can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}_1\mathbf{b}_1 + \mathbf{U}_2\mathbf{b}_2 + \cdots + \mathbf{U}_{p_1}\mathbf{b}_{p_1} + \mathbf{e} \tag{2.1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations; $\mathbf{X}$ is an $n \times p_0$ matrix of known constant (the design matrix for the fixed effects); $\boldsymbol{\alpha}$ is a $p_0 \times 1$ vector of unknown constants; $\mathbf{U}_i$ is an $n \times m_i$ matrix of known constants (a design matrix for a random effect), $i = 1, 2, \cdots, p_1$; $\mathbf{b}_i$ is an $m_i \times 1$ random vector, $\mathbf{b}_i \sim \mathcal{N}_{m_i}(\mathbf{0}, \sigma_i\mathbf{I})$, $i = 1, 2, \cdots, p_1$; $\mathbf{e}$ is an $n \times 1$ random vector, $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma_0\mathbf{I})$. Let $\mathbf{G}_i = \mathbf{U}_i\mathbf{U}_i'$, $i = 1, 2, \cdots, p_1$, and $\mathbf{G}_0 = \mathbf{I}_n$. Given the model (2.1) we make the following assumptions: $\mathbf{X}$ has full rank p_0; $n \geqslant p_0 + p_1 + 1$; $[\mathbf{X} : \mathbf{U}_i]$ has rank greater than p_0, $i = 1, 2, \cdots, p_1$, so that the fixed effects are not confounded with the random effects; the matrices $\mathbf{G}_0 \equiv \mathbf{I}$, $\mathbf{G}_i \equiv \mathbf{U}_i\mathbf{U}_i'$, $i = 1, 2, \cdots, p_1$ are linearly independent (i.e., $\Sigma_{i=0}^{p_1}\tau_i\mathbf{G}_i = \mathbf{0}$ implies that $\tau_i = 0$, $i = 0, 1, \cdots, p_1$.) so that the random effects are not confounded with each other; $\mathbf{U}_i$ consists of only zeroes and ones with exactly one 1 in each row and at least one 1 in each column, $i = 1, 2, \cdots, p_1$.

It follows that $\mathbf{y} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\alpha}, \boldsymbol{\Sigma}(\boldsymbol{\sigma}))$ where $\boldsymbol{\Sigma}(\boldsymbol{\sigma}) = \Sigma_{j=0}^{p_1}\sigma_j\mathbf{G}_j$. The parameter space is defined as follows: Let $p \equiv p_0 + p_1 + 1$ and let $\boldsymbol{\sigma} = (\sigma_0, \sigma_1, \cdots, \sigma_{p_1})'$. Then $\Theta \subset R^p$ is the parameter space, where

$$\Theta = \{\boldsymbol{\theta} \in R^p | \boldsymbol{\theta}' = (\boldsymbol{\alpha}', \boldsymbol{\sigma}'); \quad \boldsymbol{\alpha} \in R^{p_0}; \boldsymbol{\Sigma}(\boldsymbol{\sigma}) > \mathbf{0}; \sigma_0 > 0; \sigma_i \geqslant 0, i = 1, 2, \cdots, p_1\}. \tag{2.2}$$

The requirement $\sigma_i \geqslant 0$ is referred to as the "variance component constraint." The objective is to observe $\mathbf{y}$ and estimate $\boldsymbol{\alpha}, \sigma_0, \sigma_1, \cdots, \sigma_{p_1}$ by the method of maximum likelihood. (See Miller (1977) for additional details on the general model.)

2.2 *Simplifications for balanced models.* In this section we show that for balanced models, $\boldsymbol{\Sigma}$ is diagonalizable, and that the fixed effects have explicit maximum likelihood estimates that do not depend upon $\boldsymbol{\Sigma}$. To do this, we investigate properties of $\mathbf{X}$ and $\boldsymbol{\Sigma}$ in the general model (2.1). Consider the balanced model written out in the form

$$y_{j_1j_2\cdots j_r} = \mu + a_{j_1} + \cdots + e_{j_1j_2\cdots j_r}, j_k = 1, \cdots, J_k, k = 1, \cdots, r, \tag{2.3}$$

where $+ \cdots +$ may consist of crossed and nested, fixed, mixed or random effect terms. We can rewrite this model in the form

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\alpha}_1 + \cdots + \mathbf{X}_s\boldsymbol{\alpha}_s + \mathbf{U}_1\mathbf{b}_1 + \cdots + \mathbf{U}_{p_1}\mathbf{b}_{p_1} + \mathbf{e} \tag{2.4}$$

where $\mathbf{X}_i$ is the design matrix for the ith fixed term in the model and $\mathbf{U}_j$ is the design matrix for the jth random term. These $\mathbf{X}_i$ and $\mathbf{U}_j$ matrices are of the form

$[\mathbf{A}_1 \otimes \mathbf{A}_2 \otimes \cdots \otimes \mathbf{A}_r]$ where $\mathbf{A}_i$ is $\mathbf{I}_{J_i}$ if the term under consideration has the subscript **i**, $\mathbf{e}_{J_i}$ otherwise, $i = 1, 2, \cdots, r$. ($\mathbf{I}_n$ is the $n \times n$ identity matrix and $\mathbf{e}_n$ is an $n \times 1$ column vector of ones.) Note that we have used the Kronecker product, $\mathbf{A} \otimes \mathbf{B} \equiv (a_{ij}\mathbf{B})$ with well-known properties

$$(2.5) \qquad (\mathbf{A}_1 \otimes \mathbf{B}_1)(\mathbf{A}_2 \otimes \mathbf{B}_2) = \mathbf{A}_1\mathbf{A}_2 \otimes \mathbf{B}_1\mathbf{B}_2, (\mathbf{A} \otimes \mathbf{B})' = (\mathbf{A}' \otimes \mathbf{B}').$$

The $\mathbf{G}_i$ matrix corresponding to $\mathbf{U}_i$ is given by $\mathbf{G}_i = \mathbf{U}_i\mathbf{U}_i'$ which by (2.5) is of the form $[\mathbf{B}_1 \otimes \mathbf{B}_2 \otimes \cdots \otimes \mathbf{B}_r]$ where $\mathbf{B}_i$ is $\mathbf{I}_{J_i}$ if the term under consideration has the subscript i, $\mathbf{E}_{J_i}$ otherwise for $i = 1, \cdots, r$. ($\mathbf{E}_n$ is an $n \times n$ matrix of ones and $\mathbf{E}_n = \mathbf{e}_n\mathbf{e}_n'$.)

Let $\boldsymbol{\Gamma}_n$ be the symmetric orthogonal matrix of dimension n with elements in the first row being all identically equal to $n^{-\frac{1}{2}}$,

(2.6)

$$\boldsymbol{\Gamma}_n = (\gamma_{ij}) = n^{-\frac{1}{2}}\left[\cos\{2\pi n^{-\frac{1}{2}}(i-1)(j-1)\} + \sin\{2\pi n^{-\frac{1}{2}}(i-1)(j-1)\}\right].$$

Let $\mathbf{P} = [\boldsymbol{\Gamma}_{J_1} \otimes \cdots \otimes \boldsymbol{\Gamma}_{J_r}]$. Then $\mathbf{G}_i^* = \mathbf{P}'\mathbf{G}_i\mathbf{p}$, $i = 1, \cdots, r$, and $\boldsymbol{\Sigma}^* = \mathbf{P}'\boldsymbol{\Sigma}\mathbf{P}$ are diagonal, since $\boldsymbol{\Gamma}_n'\mathbf{E}_n\boldsymbol{\Gamma}_n$ is diagonal. Thus for any balanced model, $\boldsymbol{\Sigma}$ can be diagonalized and the columns of $\mathbf{P}$ are eigenvectors of $\boldsymbol{\Sigma}$. Each design matrix $\mathbf{X}_i$ can be expressed as $\mathbf{X}_i = \mathbf{P}\mathbf{Z}_i$ where $\mathbf{Z}_i$ is of the form $\mathbf{Z}_i = [\mathbf{C}_1 \otimes \mathbf{C}_2 \otimes \cdots \otimes \mathbf{C}_r]$ where $\mathbf{C}_i$ is $\boldsymbol{\Gamma}_{J_i}$ if $\mathbf{A}_i = \mathbf{I}_{J_i}$, $(J_i)^{\frac{1}{2}}\mathbf{f}_{J_i}$ otherwise, $i = 1, \cdots, r$. ($\mathbf{f}_n$ is a $n \times 1$ column vector of zeroes with a one in the first position.) Each $\mathbf{X}_i$ is of full rank and can be expressed as a linear combination of rank $(\mathbf{X}_i)$ eigenvectors of $\boldsymbol{\Sigma}$. Thus, if we form the $\mathbf{X}$ matrix in (2.1) by deleting linearly dependent columns of the matrix $[\mathbf{X}_1 : \mathbf{X}_2 : \cdots : \mathbf{X}_s]$ until it is of full rank, we see that $\mathbf{X}$ can be expressed as a linear combination of rank $(\mathbf{X})$ eigenvectors of $\boldsymbol{\Sigma}$. Theorem 1 follows (e.g., Szatrowski (1980), Theorem 2) from this fact.

THEOREM 1. *The maximum likelihood estimates for the fixed effects in the balanced mixed model of the analysis of variance have explicit representations which do not depend upon* $\boldsymbol{\Sigma}$.

3. Explicit maximum likelihood estimates for balanced models. Szatrowski (1980, Theorem 5) gives the following necessary and sufficient conditions for the existence of explicit maximum likelihood estimates for the diagonal form of the balanced model without the variance component constraint. (The diagonal form is obtained by rotating the data so that $\boldsymbol{\Sigma}$ is diagonal, using the transformations in Section 2.2.)

THEOREM 2. *Assume that explicit maximum likelihood estimates exist for the mean and that* $\boldsymbol{\Sigma}$ *is diagonal. Under these assumptions* $\boldsymbol{\Sigma}$ *has explicit maximum likelihood estimates if and only if the diagonal elements of* $\boldsymbol{\Sigma}$ *consist of exactly* $p_1 + 1$ *linearly independent combinations of the* σ's.

REMARK. Note that by the results of Section 2.2 and Theorem 1, the assumptions of Theorem 2 are satisfied for balanced ANOVA models.

We now develop a procedure for checking directly whether $\mathbf{\Sigma}$ has explicit maximum likelihood estimates when ignoring the variance component constraint. Suppose there are $p_1 + 1$ random effect terms in a model. Define the row vector $\mathbf{v}_j$ of dimension r as having a one in the ith position if the jth random effect term has the ith subscript, zero otherwise. Let $\mathbf{v}_0$ correspond to the $\mathbf{v}$ vector for the random effect $\mathbf{e}$, (i.e., $\mathbf{v}_0 = \mathbf{e}_r'$). Let $\mathbf{V}$ be the matrix with rows $\mathbf{v}_0, \mathbf{v}_1, \cdots, \mathbf{v}_{p_1}$. Let $\mathbf{w}_1, \cdots, \mathbf{w}_r$ be the column vectors of $\mathbf{V}$. We define the product of two column vectors $\mathbf{a} * \mathbf{b}$ as the column vector whose ith component is $a_i b_i$. Let $\mathbf{w}_0$ be a column vector of ones. Define $\mathscr{W}$ as the smallest set containing $\mathbf{w}_0, \mathbf{w}_1, \cdots, \mathbf{w}_r$ which is closed under $*$ multiplication and let $N(\mathscr{W})$ be the number of distinct vectors in $\mathscr{W}$.

THEOREM 3. *For the balanced mixed model in the analysis of variance, $\mathbf{\Sigma}$ has explicit maximum likelihood estimates for the model without the variance component constraint if and only if $N(\mathscr{W}) = p_1 + 1$.*

REMARK. Since the $\mathbf{G}$ matrices are linearly independent, $N(\mathscr{W}) \geqslant p_1 + 1$.

PROOF. By Theorem 2, it is sufficient to show that $N(\mathscr{W})$ is the number of distinct linear combinations of the σ's that are diagonal elements of $\mathbf{\Sigma}$. The diagonal matrix $\mathbf{G}_i^* = \mathbf{P}'\mathbf{G}_i\mathbf{P} = \text{diag}[\mathbf{R}_1 \otimes \cdots \otimes \mathbf{R}_r]$ where $\mathbf{R}_j$ is either $\mathbf{e}_{J_j}$ if the jth subscript is included in the term, $J_j \mathbf{f}_{J_j}$ otherwise, $i = 0, \cdots, p_1$. Thus $\mathbf{G}_i^*$ is of the form $c_i \, \text{diag}[\mathbf{S}_1 \otimes \cdots \otimes \mathbf{S}_r]$, $i = 0, \cdots, p_1$ where $\mathbf{S}_j$ is either $\mathbf{e}_{J_j}$ or $\mathbf{f}_{J_j}$. Without loss of generality, we may assume $J_1 = J_2 = \cdots = J_r = 2$ and can ignore c_i when counting the number of distinct linear combinations of σ's on the diagonal. Under these assumptions $\mathbf{G}_i^*$ is of the form

$$\mathbf{G}_i^* = \text{diag}\left[\begin{pmatrix} 1 \\ v_{i1} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ v_{i2} \end{pmatrix} \otimes \cdots \otimes \begin{pmatrix} 1 \\ v_{ir} \end{pmatrix}\right], \quad i = 0, \cdots, p_1.$$

The 2^r elements on the diagonal of G_i^* consist of products of the elements in the 2^r subsets of $\{v_{i1}, v_{i2}, \cdots, v_{ir}\}$ where the product of no elements is defined as one. If we let $\mathbf{G}$ be a $(p_1 + 1) \times 2^r$ dimensional matrix with rows the elements of the diagonals of $\mathbf{G}_0^*, \mathbf{G}_1^*, \cdots, \mathbf{G}_r^*$, i.e., $\mathbf{G}_{ij} = (\mathbf{G}_i^*)_{jj}$, we note that each column represents the $*$ multiplication product of a subset of vectors from $\mathbf{w}_1, \cdots, \mathbf{w}_r$. The number of distinct linear combinations of σ's is the same as the numbers of distinct column vectors of $\mathbf{G}$ which is the same as $N(\mathscr{W})$. □

Five examples follow applying the procedure of Theorem 3. The vectors of $\mathscr{W}$ appear as the columns of $\mathbf{W}$.

1. Two-factor random effects model. (a and b are random effects.)

$$y_{ijk} = \mu + a_i + b_j + ab_{ij} + e_{ijk},$$
$$i = 1, \cdots, I; j = 1, \cdots, J; k = 1, \cdots K.$$

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix} \qquad N(\mathscr{W}) = 5 > p_1 + 1, p_1 = 3.$$

Since $N(\mathcal{W}) = 5 > p_1 + 1 = 4$, there are no explicit maximum likelihood estimates. Note that eliminating the interaction term still does not yield a model with explicit maximum likelihood estimates.

2. Three-factor nested random effects model. (a, b and c are random effects.)

$$y_{ijkl} = \mu + a_i + b_{j(i)} + c_{k(ij)} + e_{ijkl}$$

$$i = 1, \cdots, I, \quad J = 1, \cdots, J$$
$$k = 1, \cdots, K, \quad l = 1, \cdots, L.$$

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}, \mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix},$$

$$N(\mathcal{W}) = 4 = p_1 + 1, p_1 = 3.$$

Since $N(\mathcal{W}) = 4 = p_1 + 1$, there are explicit maximum likelihood estimates when ignoring the variance component constraint. (See Theorem 4 below.)

3. Two-factor mixed effects model. (a is a fixed effect, b is a random effect.)

$$y_{ijk} = \mu + a_i + b_j + ab_{ij} + e_{ijk};$$

$$i = 1, \cdots, I; j = 1, \cdots, J; k = 1, \cdots, K.$$

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \mathbf{W} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix},$$

$$N(\mathcal{W}) = 3 = p_1 + 1, p_1 = 2.$$

Since $N(\mathcal{W}) = 3 = p_1 + 1$, there are explicit maximum likelihood estimates when ignoring the variance component constraint.

4. Three-factor model. (Two crossed random effects nested within a third random effect.)

$$y_{ijkl} = \mu + a_i + b_{j(i)} + c_{k(i)} + bc_{jk(i)} + e_{ijkl};$$

$$i = 1, \cdots, I; j = 1, \cdots, J; k = 1, \cdots, K; l = 1, \cdots, L.$$

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}, \mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix},$$

$$N(\mathcal{W}) = 5 = p_1 + 1, p_1 = 4.$$

Since $N(\mathcal{W}) = 5 = p_1 + 1$, there are explicit maximum likelihood estimates when ignoring the variance component constraint.

5. Three-factor model. (Two crossed random effects, one crossed fixed effect; a is a fixed effect, b and c are random effects.)

$$y_{ijkl} = \mu + a_i + b_j + ab_{ij} + c_k + ac_{ik} + bc_{jk} + abc_{ijk} + e_{ijkl};$$
$$i + 1, \cdots, I; j = 1, \cdots, J; k = 1, \cdots, K; l = 1, \cdots, L.$$

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}, \mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{bmatrix},$$

$$N(\mathcal{W}) = 9 > p_1 + 1, p_1 = 6.$$

Since $N(\mathcal{W}) = 9 > p_1 + 1 = 7$, there are no explicit maximum likelihood estimates. Note that this shows that $N(\mathcal{W})$ may exceed $p_1 + 1$ by more than one. If we introduce another fixed effect into this model which is crossed with all other terms, we find

$$N(\mathcal{W}) = 17 > p_1 + 1 = 13, p_1 = 12.$$

The result for example two is a special case of the following Theorem.

THEOREM 4. *Any balanced mixed model of the analysis of variance in which all effects, random or fixed, are nested has explicit maximum likelihood estimates with and without the variance component constraint.*

PROOF. It is sufficient to show this result for the problem without the variance component constraint since in the case where the unconstrained maximum likelihood estimates do not satisfy the variance component constraint, we drop the term(s) that violate the constraint from the model (setting them equal to zero) and solve the resulting model for the unconstrained maximum likelihood estimates. Note this new model is still a model in which all effects are nested.

Consider the completely nested random effects model with r subscripts. The $\mathbf{V}$ matrix is of the form given in example two where the first row consists of r ones, the second row one one followed by zeroes, the third row two ones followed by zeroes continuing in this fashion down to the last row consisting of $r - 1$ ones and one zero. The first column of $\mathbf{V}$ consists of all ones and $*$ multiplication of any two columns yields the column furthest to the right, i.e., if $i < j$, $\mathbf{w}_i * \mathbf{w}_j = \mathbf{w}_j$. Thus column multiplication yields no new elements and $N(\mathcal{W}) = p_1 + 1$.

To obtain the $\mathbf{V}$ matrix for a fully nested model with r subscripts, we start with the fully nested random model with r subscripts and cross out rows of $\mathbf{V}$ that correspond to fixed effect terms. Each time a row is crossed out $p_1 + 1$ decreases by one. However $N(\mathcal{W})$ also decreases by exactly one since crossing out a row yields two duplicate column vectors. Thus $N(\mathcal{W}) = p_1 + 1$. □

4. Solving for the explicit maximum likelihood estimates.

When $\boldsymbol{\Sigma} = \mathbf{I}$ is a possible value for $\boldsymbol{\Sigma}$ (as it is for the mixed model of the analysis of variance with

$\sigma_0 = 1, \sigma_i = 0, i = 1, \cdots, p_1$), and when the mean vector has explicit maximum likelihood estimates, then these estimates are given by

$$\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{5.1}$$

Define $\mathbf{C} = (\mathbf{y} - \mathbf{X}\hat{\alpha})(\mathbf{y} - \mathbf{X}\hat{\alpha})'$. When both the mean and covariance have explicit maximum likelihood estimates, the mean estimates are given by (5.1) and the covariance estimates for $\boldsymbol{\Sigma}(\boldsymbol{\sigma})$ are given by

$$\hat{\boldsymbol{\sigma}} = \left[\text{tr } \mathbf{G}_g\mathbf{G}_h\right]^{-1}(\text{tr } \mathbf{G}_g\mathbf{C}) \tag{5.2}$$

where $[\text{tr } \mathbf{G}_g\mathbf{G}_h]$ is a $(p_1 + 1) \times (p_1 + 1)$ matrix whose gh element is $\text{tr } \mathbf{G}_g\mathbf{G}_h$, $g, h = 0, \cdots, p_1$ and $(\text{tr } \mathbf{G}_g\mathbf{C})$ is a $(p_1 + 1) \times 1$ column vector whose gth element is $\text{tr } \mathbf{G}_g\mathbf{C}$, $g = 0, \cdots, p_1$ (Miller (1973), Szatrowski, (1980)). Anderson (1970) notes that the matrix $[\text{tr } \mathbf{G}_g\mathbf{G}_h]$ is positive definite.

For balanced ANOVA models, $\mathbf{G}_i^* = \mathbf{P}'\mathbf{G}_i\mathbf{P}$, $i = 0, \cdots, p_1$ are diagonal (see Section 2.2). Letting $\mathbf{V} = \mathbf{P}'\mathbf{C}\mathbf{P}$ we can rewrite (5.2) as

$$\hat{\boldsymbol{\sigma}} = \left[\text{tr } \mathbf{G}_g^*\mathbf{G}_h^*\right]^{-1}(\text{tr } \mathbf{G}_g^*\mathbf{V}). \tag{5.3}$$

$\boldsymbol{\Sigma}^* = \mathbf{P}'\boldsymbol{\Sigma}\mathbf{P} = \Sigma_{g=0}^{p_1}\sigma_g\mathbf{G}_g^*$ is diagonal and we can represent the $p_1 + 1$ distinct linear combinations on the diagonal by $\tau_0, \cdots, \tau_{p_1}$ and represent $\boldsymbol{\Sigma}^*$ by $\boldsymbol{\Sigma}^* = \Sigma_{g=0}^{p_1}\tau_g\boldsymbol{\Lambda}_g$ where the $\boldsymbol{\Lambda}_g$'s are diagonal matrices of zeroes and ones with the property that $\boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_h = 0$ and thus $\text{tr } \boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_h = 0$ for $g \neq h$. We can then rewrite (5.3) as (5.4) or (5.5) (noting $\text{tr } \Lambda_g^2 = \text{tr } \Lambda_g$),

$$\hat{\boldsymbol{\tau}} = \left[\text{tr } \boldsymbol{\Lambda}_g\boldsymbol{\Lambda}_h\right]^{-1}(\text{tr } \boldsymbol{\Lambda}_g\mathbf{V}). \tag{5.4}$$

$$\hat{\tau}_g = \frac{\text{tr } \boldsymbol{\Lambda}_g\mathbf{V}}{\text{tr } \boldsymbol{\Lambda}_g}, \qquad g = 0, 1, \cdots, p_1. \tag{5.5}$$

To find the maximum likelihood estimates for the σ's we need only solve the equations

$$\Sigma_{j=0}^{p_1}\sigma_j\mu_{kj} = \hat{\tau}_k \qquad k = 0, 1, \cdots, p_1, \tag{5.6}$$

where the μ's are known coefficients. Note that $\hat{\tau}_g$ is just the average of a subset of the diagonal components of $\mathbf{V}$. Several examples follow.

1. Two-factor nested random effects model. (a and b are random effects.)

$$y_{ijk} = \mu + a_i + b_{j(i)} + e_{ijk},$$
$$i = 1 \quad I; j = 1, \cdots, J; k = 1, \cdots, K.$$

By Theorem 4, we know that this model has explicit maximum likelihood estimates. The $\mathbf{X}$ matrix is given by $\mathbf{X} = (\mathbf{e}_I \otimes \mathbf{e}_J \otimes \mathbf{e}_K)$ and thus $\hat{\alpha}$ and $\mathbf{C}$ are $\hat{\alpha} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (IJK)^{-1}\Sigma_{ijk}y_{ijk} \equiv \bar{y}_{...}$; $\mathbf{C} = (\mathbf{y} - \bar{y}_{...}\mathbf{e}_{IJK})(\mathbf{y} - \bar{y}_{...}\mathbf{e}_{IJK})'$. The diagonal $\mathbf{G}_i^*$ and $\boldsymbol{\Lambda}_i$ matrices are the τ's are $\mathbf{G}_0^* = [\mathbf{I}_I \otimes \mathbf{I}_J \otimes \mathbf{I}_K]$, $\mathbf{G}_1^* = K[\mathbf{I}_I \otimes \mathbf{I}_J \otimes \mathbf{F}_K]$, $\mathbf{G}_2^* = JK[\mathbf{I}_I \otimes \mathbf{F}_J \otimes \mathbf{F}_K]$, $\boldsymbol{\Lambda}_0 = \mathbf{G}_2^*/JK$, $\boldsymbol{\Lambda}_1 = (\mathbf{G}_1^* - \mathbf{G}_2^*/J)/K$, $\boldsymbol{\Lambda}_2 = \mathbf{G}_0^* - \mathbf{G}_1^*/K$, $\tau_0 = \sigma_0 + K\sigma_1 + JK\sigma_2$, $\tau_1 = \sigma_0 + K\sigma_1$, $\tau_2 = \sigma_0$, where $\mathbf{F}_n$ is an $n \times n$ matrix of zeroes, with a one in the (1, 1) position. Note $\boldsymbol{\Gamma}_n'\mathbf{F}_n\boldsymbol{\Gamma}_n = \boldsymbol{\Gamma}_n'\mathbf{f}_n\mathbf{f}_n'\boldsymbol{\Gamma}_n = \frac{1}{n}\mathbf{e}_n\mathbf{e}_n'$. We proceed to find the $\hat{\tau}$'s given in (5.5) in terms of the usual analysis of variance "means square" terms.

tr $\Lambda_0 \mathbf{V} = \text{tr}[\mathbf{I}_I \otimes \mathbf{F}_J \otimes \mathbf{F}_K]\mathbf{V} = \text{tr}[\mathbf{I}_I \otimes \mathbf{F}_J \otimes \mathbf{F}_K]\mathbf{P}'\mathbf{C}\mathbf{P} = \frac{1}{JK}\text{tr}[\mathbf{I}_I \otimes \mathbf{e}'_J \otimes \mathbf{e}'_K](\mathbf{y} - \bar{y}_{...}\mathbf{e}_{IJK})(\mathbf{y} - \bar{y}_{...}\mathbf{e}_{IJK})'[\mathbf{I}_I \otimes \mathbf{e}_J \otimes \mathbf{e}_K] = JK\Sigma_{i=1}^{I}(\bar{y}_{i..} - \bar{y}_{...})^2 = (I-1)\text{MS}_a$, where the last equality is obtained by noting $[\mathbf{I}_I \otimes \mathbf{e}'_J \otimes \mathbf{e}'_K]\mathbf{y}$ is an I component vector whose ith entry is $\Sigma_{j,k} y_{ijk} = JK\bar{y}_{i..}$. Similarly, we find tr $\Lambda_1 \mathbf{V} = K\Sigma_{i,j}(\bar{y}_{ij\cdot} - \bar{y}_{i..})^2 = I(J-1)\text{MS}_{b(a)}$ and tr $\Lambda_2 \mathbf{V} = \Sigma_{i,j,k}(y_{ijk} - \bar{y}_{ij\cdot})^2 = IJ(K-1)\text{MS}_e$. Dividing tr $\Lambda_i \mathbf{V}$ by tr Λ_i gives us $\hat{\tau}_i$ by (5.5) and the equations in (5.6) are $\sigma_e + K\sigma_{b(a)} + JK\sigma_a = \{(I-1)/I\}\text{MS}_a$; $\sigma_e + K\sigma_{b(a)} = \text{MS}_{b(a)}$; $\sigma_e = \text{MS}_e$; $\sigma_0 = \sigma_e$, $\sigma_1 = \sigma_{b(a)}$, $\sigma_2 = \sigma_a$. To get the explicit maximum likelihood estimates including the variance component constraint, we solve these equations for the σ's. If the constraint $\sigma_0 > 0$, $\sigma_i \geqslant 0$ $i = 1, \cdots, p_1$ is satisfied, we have the maximum likelihood estimates. Otherwise, we drop the σ's which do not satisfy the constraint from the model, and find the estimates for the new model.

2. Two-factor mixed effects model. (See Example 3 in Section 3.) Here we have two fixed effects terms, μ with design matrix $\mathbf{X}_1 = [\mathbf{e}_I \otimes \mathbf{e}_J \otimes \mathbf{e}_K]$ and a_i with design matrix $\mathbf{X}_2 = [\mathbf{I}_I \otimes \mathbf{e}_J \otimes \mathbf{e}_K]$. The matrix $[\mathbf{X}_1 : \mathbf{X}_2]$ is not of full rank, but the matrix with any one column deleted is of full rank. Let $\mathbf{X} = \mathbf{X}_2$. Then $\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (1/JK)(\mathbf{I}_I \otimes \mathbf{e}'_J \otimes \mathbf{e}'_K)\mathbf{y} = (\bar{y}_{1..}, \bar{y}_{2..}, \cdots, \bar{y}_{I..})'$. Solving for the σ's using the techniques of Example 1 yields the equations (5.6) in the form $\sigma_e + K\sigma_{ab} + IK\sigma_b = \{(J-1)/J\}\text{MS}_b$; $\sigma_e + K\sigma_{ab} = \text{MS}_{ab}$; $\sigma_e = \text{MS}_e$; $(J-1)\text{MS}_b = IK\Sigma_j(\bar{y}_{\cdot j \cdot} - \bar{y}_{...})^2$; $(I-1)(J-1)\text{MS}_{ab} = K\Sigma_{i,j}(\bar{y}_{ij\cdot} - \bar{y}_{i..} - \bar{y}_{\cdot j\cdot} + \bar{y}_{...})^2$; $(I-1)(J-1)(K-1)\text{MS}_e = \Sigma_{i,j,k}(y_{ijk} - \bar{y}_{ij\cdot})^2$.

3. Three-factor model. (Two-crossed random effects, nested within a third random effect.) (See Example 4, Section 3.) Solving for the σ's using the techniques of example 1 yields the equations (5.6) in the form

$$\sigma_e + L\sigma_{bc(a)} + JL\sigma_{c(a)} + KL\sigma_{b(a)} + JKL\sigma_a = \{(I-1)/I\}\text{MS}_a;$$
$$\sigma_e + L\sigma_{bc(a)} + JL\sigma_{c(a)} = \text{MS}_{c(a)};\ \sigma_e + L\sigma_{bc(a)} + KL\sigma_{b(a)} = \text{MS}_{b(a)};$$
$$\sigma_e + L\sigma_{bc(a)} = \text{MS}_{bc(a)};\ \sigma_e = \text{MS}_e.$$

REFERENCES

[1] Anderson, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. *Essays in Probability and Statistics*, pages 1–24, Univ. North Carolina Press, Chapel Hill.

[2] Anderson, T. W. (1973). Asymptotic efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

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

[4] Hartley, H. O. and W. K. Vaughn (1972). A computer program for the mixed analysis of variance model based on maximum likelihood. *Statistical Papers in Honor of George W. Snedecor* (T. A. Bancroft, ed.) Iowa State Univ. Press.

[5] Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–338.

[6] 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.

[7] MILLER, J. (1973). Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. Technical Report No. 12, Stanford University.

[8] MILLER, J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Ann. Statist.* **5** 746–762.

[9] MILLER, J. (1979). Maximum likelihood estimation of variance components, a Monte Carlo study, *J. Statist. Comp. and Simulation.* **8** 175–190.

[10] SZATROWSKI, T. H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *Ann. Statist.* **8** 802–810.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903