Original article

Variance estimation from integrated likelihoods (VEIL)

D Gianola[1], JL Foulley[2]

[1] *University of Illinois, Department of Animal Sciences, Urbana, IL 61801 USA;*
[2] *Institut National de la Recherche Agronomique,*
Station de Génétique Quantitative et Appliquée,
Centre de Recherches de Jouy-en-Josas, 78352 Jouy-en-Josas Cedex, France

Summary – A method of variance component estimation in univariate mixed linear models based on the exact or approximate posterior distributions of the variance components is presented. From these distributions, posterior means, modes and variances can be calculated exactly, *via* numerical methods, or approximately, using inverted χ^2 distributions. Although particular attention is given to a Bayesian analysis with flat priors, informative prior distributions can be used without great additional difficulty. Implementation of the exact analysis can be taxing from a numerical point of view, but the computational requirements of the approximate method are not greater than those of REML.

variance components / maximum likelihood / Bayesian methods

Résumé – Estimation des composantes de la variance à partir de vraisemblances intégrées. *Cet article présente une méthode d'estimation des composantes de la variance, pour des modèles linéaires mixtes univariés, qui est fondée sur des expressions exactes et approchées de la distribution* a posteriori *de ces composantes. On peut calculer les espérances, modes et variances de ces distributions, de façon exacte par des méthodes numériques ou de façon approchée en utilisant des lois de* χ^2 *inverses. Bien que l'attention soit centrée sur une analyse bayésienne faisant intervenir des* a priori *uniformes, le recours à des lois* a priori *informatives est envisageable sans difficulté notoire additionnelle. La mise en œuvre de l'analyse exacte peut être contraignante d'un point de vue numérique, mais la méthode approchée ne demande pas plus de calculs que la résolution du REML.*

composantes de la variance / maximum de vraisemblance / méthodes bayésiennes.

INTRODUCTION

Harville and Callanan (1990) noted that likelihood-based methods of variance component estimation have gained favor among quantitative geneticists. In particular, the procedure known as "restricted maximum likelihood", or REML for short (Thompson, 1962; Patterson and Thompson, 1971), is now widely regarded in animal breeding as the method of choice. The approach is based on maximizing with

respect to the variances only the part of the likelihood function (normality is assumed) that does not depend on fixed effects. In so doing, Patterson and Thompson (1971) obtained an estimator that "accounts for the degrees of freedom lost in the estimation of fixed effects" which, according to their reasoning, is not accomplished by full maximum likelihood (ML). With balanced data, the REML estimating equations reduce to those used in estimation by analysis of variance (ANOVA) so that if the ANOVA estimates are within the parameter space, these are REML as well.

The basic idea underlying REML is obtaining a likelihood-based estimator (thus retaining the usual asymptotic properties) while reducing the bias of ML. However, REML estimators are biased as well and because these are constructed from a partial likelihood, one should expect larger variance of estimates than with ML. A more reasonable comparison is in terms of a loss function such as mean-squared error, but Monte Carlo studies (for example, Swallow and Monahan, 1984) have given ambiguous results. In fact, in a purely fixed model with a single location parameter and 1 variance component, the ML estimator has smaller mean-squared error than REML throughout the whole parameter space, ie, REML is inadmissible under quadratic loss for this model. Lacking evidence to the contrary, a similar sampling behavior can be expected at least in some other models. Searle (1988) puts it concisely : *"It is difficult to be anything but inconclusive about which of ML and REML is the preferred method".*

The 2 procedures have the same asymptotic properties (although their asymptotic variance is different) and unknown small sample distributions, a feature shared by all sampling theory estimators of variance components. Although ML and REML estimates are defined inside the appropriate parameter space, their asymptotic distribution (normal) can generate interval estimates that include negative values. This potentially embarrassing phenomenon is often overlooked in discussions of likelihood based methods.

Harville (1974) showed that REML corresponds to the mode of the marginal posterior density of the variance components in a Bayes analysis with flat priors for fixed effects and variance components. The fixed effects (β) are viewed in REML as nuisance parameters and, therefore are eliminated from the full likelihood by integration so that the restricted likelihood is a marginal likelihood (Dawid, 1980), seemingly a more appropriate term. This likelihood does take into account the error incurred in estimating β because it can be viewed as a weighted average of likelihoods of variance components evaluated at all possible values of β, the weight function being the posterior density of the fixed effects (proportional to the marginal likelihood of β in this context). Gianola *et al* (1986) gave a Bayesian justification of REML when estimating breeding values with unknown variance components. The same, by virtue of the symmetry of the Bayesian analysis, applies to estimation of fixed effects.

There are 2 aspects of REML that have not received sufficient discussion. First, the method produces joint modes rather than marginal modes. If the loss function is quadratic, the optimum Bayes estimator is the posterior mean, and marginal modes provide better approximations to this than joint modes (O'Hagan, 1976). Second, in some problems not all variance parameters have equal importance. For example, suppose that there is interest in making inferences about the amount of additive genetic variance in a population and that the statistical description of the problem

requires a model that, in addition to fixed effects, includes random herd, additive, dominance, permanent environmental and temporary effects. In this situation, the marginal or restricted likelihood of the variance components involves 5 dimensions, 1 for each of the variances, yet all components other than the additive one should be regarded as nuisance parameters. Carrying the logic of Patterson and Thompson (1971) 1 step further, REML would not take into account the error incurred in estimating the nuisance variances and, therefore, only the part of the likelihood that is a function of the additive genetic variance should be maximized. Construction of this likelihood employing classical statistical arguments seems impossible.

The objective of this note is to present a new method of variance component estimation in univariate mixed linear models based on the principle of finding the marginal posterior density of each of the variances. From this, means, modes, variances and higher order moments of the distribution of variance components can be calculated exactly, using numerical methods, or analytically via approximations. Particular attention is given to the Bayesian analysis with flat priors for fixed effects and variance components.

MODEL AND DEFINITIONS

Model

Consider the mixed linear model:

$$y = X\beta + \sum_{i=1}^{c} Z_i u_i \tag{1}$$

where :
y: data vector; X, Z_i: known incidence matrices. For $i = c$, $Z_i = I$, an identity matrix of appropriate order; β: $p \times 1$ vector of uniquely defined fixed effects (X has full-column rank); u_i: $q_i \times 1$ random vector distributed as $N(O, A_i \sigma_i^2)$, where A_i is a known square, positive-definite matrix and $\sigma_i^2 (i = 1, 2, \ldots, c)$ is a variance component. It is assumed that all u_i and u_j are mutually independent. For $i = c$, A_c is often taken to be an identity matrix of appropriate order; this will be assumed in the present paper.

The marginal distribution:

$$y \mid \beta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_c^2 \sim N(X\beta, \sum_{i=1}^{c} Z_i A_i Z_i' \sigma_i^2) \tag{2}$$

gives rise to the "full" likelihood function, from which ML estimators of β and of each of the variance components can be obtained.

Bayesian elements

In the Bayesian treatment of model (1), prior subjective uncertainty about β is often described via the "flat" prior density:

$$p(\beta) \propto \text{constant} \tag{3}$$

and the normal distributions assigned independently to the u_i's are viewed as prior probability distributions as well. To complete the model, assignment of prior probability distributions to the variance components is necessary. In this paper, and for the sole purpose of showing how a higher degree of marginalization can be achieved than in REML, independent "flat" prior distributions are assigned to each of the variances σ_i^2. With this formulation, the joint posterior density of $\boldsymbol{\beta}$ and of variances becomes:

$$p(\boldsymbol{\beta}, \sigma_i^2, \sigma_2^2, \ldots, \sigma_c^2 \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \ldots, \sigma_c^2) \qquad (4)$$

so that there is no information about fixed effects and variance parameters beyond that contained in the likelihood function. If (4) is maximized jointly with respect to $\beta, \sigma_1^2, \sigma_2^2, \ldots, \sigma_c^2$ one obtains the maximum likelihood estimates of all parameters. If (4) is integrated with respect to $\boldsymbol{\beta}$, the marginal (restricted) likelihood function of the variance components is obtained (Harville, 1974). This contains no information about β and maximization of this marginal likelihood with respect to the variances gives restricted maximum likelihood estimates of these parameters.

REPRESENTATION OF THE MARGINAL (RESTRICTED) LIKELIHOOD

As shown by Gianola *et al* (1990a, 1990b), integration of (4) with respect to β gives:

$$p(\sigma_1^2, \sigma_2^2, \ldots, \sigma_c^2 \mid \mathbf{y}) \propto (\sigma_c^2)^{-(n-p-q)/2} \exp[-(\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y})/2\sigma_c^2] \cdot \mid \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma} \mid^{-1/2}$$
$$\times \prod_{i=1}^{c-1} (\sigma_i^2)^{-q_i/2} \qquad (5)$$

where:

$$q = \sum_{i=1}^{c-1} q_i$$

$$\mathbf{W} = [\mathbf{X}\mathbf{Z}]$$

$$\mathbf{Z} = [\mathbf{Z}_1 \mathbf{Z}_2 \ldots \mathbf{Z}_{c-1}]$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{T}^{-1} \end{bmatrix}$$

$$\mathbf{T}^{-1} = \bigoplus_{i=1}^{c-1} \mathbf{A}_i^{-1} \sigma_c^2/\sigma_i^2$$

$$\bigoplus \text{ direct sum operator}$$

and

$$\widehat{\boldsymbol{\theta}} = [\widehat{\boldsymbol{\beta}}', \widehat{\mathbf{u}}_1', \widehat{\mathbf{u}}_2', \ldots, \widehat{\mathbf{u}}_{c-1}']' = (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})^{-1}\mathbf{W}'\mathbf{y}$$

is the solution to Henderson's mixed model equations. Although (5) is a posterior density, it is strictly proportional to the marginal likelihood of the variance components.

POSTERIOR DENSITY OF THE RESIDUAL VARIANCE AND OF RATIOS OF VARIANCE COMPONENTS

Consider the one-to-one reparameterization from $\sigma_1^2, \sigma_2^2, \ldots, \sigma_c^2$ to

$$\alpha_i = \sigma_c^2/\sigma_i^2 \qquad i = 1, 2, \ldots, c-1 \tag{6a}$$

and

$$\alpha_c = \sigma_c^2 \tag{6b}$$

Because maximum likelihood estimates are invariant under one-to-one reparameterizations, maximization of (5) with respect to the σ^2 or to the α-parameters gives the same estimates. However, and arguing from a likelihood veiwpoint, it is unclear how (5) can be factorized into separate components for each of the old or new parameters. Hence, estimates of a particular σ_i^2 or α_i from (5) are obtained in the presence of the remaining σ^2's or α's. To the extent that these are viewed as nuisance parameters, REML would not take into account the "degrees of freedom lost" in the estimation of such parameters.

In the Bayesian approach parameters are viewed as random variables, so a change from a σ^2 to an α-parameterization must be made using theory of transformation of random variables. The absolute value of the determinant of the Jacobian matrix of the transformations indicated in (6a) and (6b) is:

$$(\alpha_c)^{c-1} / \prod_{i=1}^{c-1} \alpha_i^2$$

Using this in (5) gives as joint posterior density of the α's:

$$p(\alpha_1, \alpha_2, \ldots, \alpha_c \mid \mathbf{y}) \propto (\alpha_c)^{-[n-p-2(c-1)]/2} \exp[-(\mathbf{y'y} - \widehat{\boldsymbol{\theta}}'\mathbf{W'y})/2\alpha_c]$$
$$\cdot \mid \mathbf{W'W} + \boldsymbol{\Sigma} \mid^{-1/2} \times \prod_{i=1}^{c-1} (\alpha_i)^{(q_i-4)/2} \tag{7}$$

The range of the α's depends on the model in question. The α_c parameter is always larger than 0. In an "animal" model with 2 variance components, the ratio between the environmental and the additive genetic variances can vary between 0 and infinity. On the other hand, in a "sire" model, the corresponding ratio varies between 3 and infinity.

POSTERIOR DENSITY OF THE RATIOS OF VARIANCE COMPONENTS

With the above parameterization, α_c can be integrated out analytically using inverted-gamma integrals (Box and Tiao, 1973) to obtain:

$$p(\alpha_1, \alpha_2, \ldots, \alpha_{c-1} \mid \mathbf{y}) = \int_0^{+\infty} p(\alpha_1, \alpha_2, \ldots, \alpha_c \mid y) d\alpha_c$$

$$\propto (\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y})^{-(n-p-2c)/2} \cdot \mid \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma} \mid^{-1/2} \prod_{i=1}^{c-1} (\alpha_i)^{(q_i-4)/2} \qquad (8)$$

Although the joint posterior density (8) is not in any obviously recognizable form, the components of the modal vector, or α-values maximizing (8) can be calculated. Let the logarithm of (8) be L, so that:

$$L = \text{Constant} - T \ln B(\alpha_1, \alpha_2, \ldots, \alpha_{c-1}) - (1/2) \ln \mid \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma} \mid + \sum_{i=1}^{c-1} E_i \ln \alpha_i \quad (9)$$

where:

$$T = (n - p - 2c)/2$$
$$B(\alpha_1, \alpha_2, \ldots, \alpha_{c-1}) = \mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y}$$
$$E_i = (q_i - 4)/2$$

The mixed model "residual" sum of squares $B(.)$ depends on the variance ratios through the solution vector $\widehat{\boldsymbol{\theta}}$. Following Macedo and Gianola (1987) and suppressing the dependence of $B(.)$ on the α's in the notation, we have for $i = 1, 2, \ldots, c - 1$:

$$\frac{\delta L}{\delta \alpha_i} = -(T/B)\frac{\delta B}{\delta \alpha_i} - \frac{1}{2}\frac{\delta}{\delta \alpha_i}(\ln \mid \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma} \mid) + E_i/\alpha_i \qquad (10)$$

Setting (10) to zero and rearranging:

$$\alpha_i = E_i[(T/B)\frac{\delta B}{\delta \alpha_i} + \frac{1}{2}\frac{\delta}{\delta \alpha_i}(\ln \mid \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma} \mid)]^{-1} \qquad (11)$$

with:

$$\frac{\delta B}{\delta \alpha_i} = -\mathbf{y}'\mathbf{W}\frac{\delta}{\delta \alpha_i}(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})^{-1}\mathbf{W}'\mathbf{y}$$
$$= \mathbf{y}'\mathbf{W}(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})^{-1}\mathbf{P}_i(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})^{-1}\mathbf{W}'\mathbf{y}$$
$$= \widehat{\boldsymbol{\theta}}'\mathbf{P}_i\widehat{\boldsymbol{\theta}} \qquad i = 1, 2, \ldots, c - 1$$

Above:

$$\mathbf{P}_i = \text{Diag}\{0, 0, \ldots, \mathbf{A}_i^{-1}, \ldots, 0, 0\}$$

is a block-diagonal matrix, and

$$\frac{\delta}{\delta \alpha_i}(\ln \mid \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma} \mid) = \text{tr}[(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})^{-1}\mathbf{P}_i]$$

Noting the special form of P_i one obtains:

$$\frac{\delta B}{\delta \alpha_i} = \widehat{\mathbf{u}}_i'\mathbf{A}_i^{-1}\widehat{\mathbf{u}}_i \qquad (12a)$$

$$\frac{\delta}{\delta \alpha_i}(\ln \mid \mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma} \mid) = \text{tr}(\mathbf{C}_{ii}\mathbf{A}_i^{-1}) \qquad (12b)$$

where $\mathbf{C}_{ii}$ is defined below. Using (12a) and (12b) in (11) and rearranging gives, for $i = 1, 2, \ldots, c-1$:

$$\alpha_i = s_c^2 / \left\{ [\widehat{\mathbf{u}}_i' \mathbf{A}_i^{-1} \widehat{\mathbf{u}}_i + s_c^2 \text{ tr}(\mathbf{C}_{ii} \mathbf{A}_i^{-1})] / (q_i - 4) \right\} \qquad (13)$$

Above,

$$s_c^2 = (\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}' \mathbf{W}' \mathbf{y}) / (n - p - 2c)$$

and $\mathbf{C}_{ii}$ is the partition of $(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})^{-1}$ corresponding to the $\mathbf{u}_i$ effects. Note that $q_i > 4$ and $n > p + 2c$ are required. The expression in (13) can be interpreted as the ratio between an "estimator" of σ_c^2 and an "estimator" of σ_i^2. Because (13) is not explicit in the α's, the expression must be iterated upon, thus generating a sequence of successive approximations. For example, starting with $\alpha_i^{[0]}$ $(i = 1, 2, \ldots, c-1)$ one can solve the mixed model equations, evaluate the right hand-side of (13), obtain a new set of α's and repeat the calculations until convergence.

In addition to first order algorithms such as the one suggested above, computing strategies based on second derivatives are possible. These are tedious and are not presented here. The important point is that the components of the mode of the joint posterior density of $\alpha_1, \alpha_2, \ldots, \alpha_{c-1}$ can be obtained without great conceptual difficulty. Hereafter, these components will be denoted as $\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_{c-1}$. Contrary to REML, account is being taken here of the "degree of freedom" lost in estimating α_c (α_c^2 in the original parameterization).

EXACT MARGINAL DISTRIBUTION OF EACH OF THE VARIANCE COMPONENTS

Residual component

Consider density (7) and write it as:

$$p(\alpha_1, \alpha_2, \ldots, \alpha_c \mid \mathbf{y}) = p(\alpha_c \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) p(\alpha_1, \alpha_2, \ldots, \alpha_{c-1} \mid \mathbf{y}) \qquad (14)$$

The first conditional density in (14) is obtained from (7) by regarding α_c as a variable and the remaining α's as constants. Hence:

$$p(\alpha_c \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, y) \propto (\alpha_c)^{-(n-p-2c+2)/2} \exp\left\{ -\frac{1}{2}(\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}' \mathbf{W}' \mathbf{y}) / \alpha_c \right\} \qquad (15)$$

This is exactly the kernel of the density of an inverted χ^2 random variable with parameters:

$$\nu = n - p - 2c$$

and

$$s_c^2 = (\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}' \mathbf{W}\mathbf{y}) / (n - p - 2c),$$

as before. The complete density (Zellner, 1971) is:

$$p(\alpha_c \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) = \frac{\left(\frac{1}{2}\nu s_c^2\right)^{-\frac{1}{2}\nu}}{\Gamma\left(\frac{1}{2}\nu\right)}(\alpha_c)^{\frac{1}{2}(\nu+2)} \exp\left\{-\frac{1}{2}\nu s_c^2/\alpha_c\right\} \qquad (16)$$

and the mean, mode and variance of this distribution can be shown by standard techniques to be:

$$E(\alpha_c \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) = s_c^2\nu/(\nu - 2) \qquad (17a)$$

$$\text{Mode}(\alpha_c \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) = s_c^2\nu/(\nu + 2) \qquad (17b)$$

$$\text{Var}(\alpha_c \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) = 2(\nu s_c^2)^2/[(\nu - 4)(\nu - 2)^2] \qquad (17c)$$

provided that $n - p - 2c > 4$. From (14), the marginal density of α_c is:

$$p(\alpha_c \mid \mathbf{y}) = \int_{R_\alpha} p(\alpha_c \mid \alpha_1, \alpha_2, \cdots, \alpha_{c-1}, y) \cdot p(\alpha_1, \alpha_2, \ldots, \alpha_{c-1} \mid \mathbf{y}) \qquad (18)$$
$$d\alpha_1 d\alpha_2 \ldots d\alpha_{c-1}$$

This is a weighted average of an infinite number of inverted χ^2 densities as in (16), the weight function being the posterior density (8). Whereas the marginal density of α_c cannot be written explicitly, its mean and variance can be calculated by numerical means. For example, note that:

$$E(\alpha_c \mid \mathbf{y}) = E[(17a)] = E(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y})/\nu - 2)$$
$$= \mathbf{y}'[\mathbf{y} - \mathbf{W}E(\hat{\boldsymbol{\theta}})]/(\nu - 2) \qquad (19)$$

because the expectation is calculated with respect to the distribution α_1, α_2, $\ldots$, $\alpha_{c-1} \mid y$. This is the Bayes estimator minimizing expected posterior quadratic loss. Evaluating the expectation in (19) requires solving Henderson's mixed model equations a large, technically an infinite, number of times. Monte-Carlo (Kloek and van Dijk, 1978) or numerical (Smith et al, 1987) integration procedures are needed to complete the calculation of (19).

Other components

Now consider the parameterization:

$$\sigma_i^2 = \alpha_c/\alpha_i \qquad (20a)$$

i different from c

$$\alpha_j = \alpha_j \qquad\qquad j = 1, 2, \ldots, c - 1 \qquad (20b)$$

The Jacobian matrix of the transformation from a parameterization in terms of all α-components to one in terms of σ_i^2 and the variance ratios as indicated above

is α_i. The joint posterior density of σ_i^2 and $\alpha_1, \alpha_2, \ldots, \alpha_{c-1}$ is then from (7):

$$p(\sigma_i^2, \alpha_1, \alpha_2, \ldots, \alpha_{c-1} \mid \mathbf{y}) \propto (\sigma_i^2)^{-\frac{1}{2}[n-p-2(c-1)]} \exp\left\{ -\frac{1}{2}(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y})/\sigma_i^2 \alpha_i \right\}$$

$$\times \alpha_i^{-\frac{1}{2}(n-p-2c)} \prod_{j=1}^{c-1} \alpha_i^{(q_i-4)/2} \tag{21}$$

As in the preceding section, given the ratios of variance components, and provided that each of the variances σ_i^2 takes values between 0 and ∞, one obtains:

$$p(\sigma_i^2 \mid \alpha_i, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) = \frac{\left(\frac{1}{2}\nu s_i^2\right)^{\frac{1}{2}\nu}}{\Gamma\left(\frac{1}{2}\nu\right)} (\sigma_i^2)^{\frac{1}{2}(\nu+2)} \exp\left\{ -\frac{1}{2}\nu s_i^2/\sigma_i^2 \right\} \tag{22}$$

$$i = 1, 2, \ldots, c-1$$

where:

$$s_i^2 = (\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y})/[(n-p-2c)\alpha_i]$$

Again, (22) is the density of an inverted χ^2 random variable with the features:

$$E(\sigma_i^2 \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) = s_i^2 \nu/(\nu-2) \tag{23a}$$

$$\text{Mode}(\sigma_i^2 \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) = s_i^2 \nu/(\nu+2) \tag{23b}$$

$$\text{Var}(\sigma_i^2 \mid \alpha_1, \alpha_2, \ldots, \alpha_{c-1}, \mathbf{y}) = 2(\nu s_c^2)^2/[(\nu-4)(\nu-2)^2] \tag{23c}$$

Using the argument of (18), the marginal density of σ_i^2 is a weighted average of an infinite number of inverted χ^2 densities, and the posterior density of the ratios of variance components, (8), is the weight function. The marginal posterior density cannot be written in explicit form, but its moments can be evaluated numerically. For example:

$$E(\sigma_i^2 \mid \mathbf{y}) = E[(23a)] = (\nu-2)^{-1}\mathbf{y}' E[\mathbf{y} - \mathbf{W}\hat{\boldsymbol{\theta}})/\alpha_i] \tag{24}$$

$$i = 1, 2, \ldots, c-1$$

where the expectation is taken with respect to the distribution $\alpha_1, \alpha_2, \ldots, \alpha_{c-1} \mid \mathbf{y}$. As discussed earlier, carrying out this calculation requires numerical integration or Monte Carlo techniques.

APPROXIMATE MARGINAL DISTRIBUTION OF EACH OF THE VARIANCE COMPONENTS

Because of the numerical difficulties arising in the marginal analysis of the variance components, it would be desirable to develop approximations that permit an

analytical treatment of the resulting distributions. If the posterior density of the
ratios of variance components is peaked or symmetrical about its mode, one
can resort to the well known approximations (Box and Tiao, 1973; Gianola and
Fernando, 1986; Gianola et al, 1986):

$$p(\alpha_c \mid \mathbf{y}) \cong p(\alpha_c \mid \alpha_1 = \widehat{\alpha}_1, \alpha_2 = \widehat{\alpha}_2, \dots, \alpha_{c-1} = \widehat{\alpha}_{c-1}, \mathbf{y}) \qquad (25)$$

and

$$p(\sigma_i^2 \mid \mathbf{y}) \cong p(\sigma_i^2 \mid \alpha_1 = \widehat{\alpha}_1, \alpha_2 = \widehat{\alpha}_2, \dots, \alpha_{c-1} = \widehat{\alpha}_{c-1}, \mathbf{y}); \qquad (26)$$

$$i = 1, 2, \dots, c - 1$$

where the $\widehat{\alpha}'s$ are the components of the mode of the posterior distribution of
the variance ratios, with density as in (8). The marginal densities of the variance
components can be approximated using densities (16) and (22), but with the
parameters now being:

$$\nu = n - p - 2c$$

$$i = 1, 2, \dots, c$$

$$\widehat{s}_c^2 = (\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y})/(n - p - 2c)$$
$$\widehat{s}_i^2 = \widehat{s}_c^2/\widehat{\alpha}_i$$

$$i = 1, 2, \dots, c - 1$$

where $\widehat{\boldsymbol{\theta}}$ is the solution to Henderson's mixed model equations evaluated at
$\widehat{\alpha}_1, \widehat{\alpha}_2, \dots, \widehat{\alpha}_{c-1}$. Because the densities are in inverted χ^2 form, one can readily
obtain their moments and construct point estimates of the variance components in
a Bayesian decision-theoretical context. For example, for a quadratic loss function,
the estimator that minimizes expected quadratic loss is the posterior mean, or:

$$E(\sigma_c^2 \mid \mathbf{y}) \cong [\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y}]/[n - p - 2(c + 1)] \qquad (27a)$$

and

$$E(\sigma_i^2 \mid \mathbf{y}) \cong [\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y}]/\widehat{\alpha}_i[n - p - 2(c + 1)] = [27a]/\widehat{\alpha}_i \qquad (27b)$$

$$i = 1, 2, \dots, c - 1$$

Estimators minimizing "all or none" posterior loss are:

$$\text{Mode}(\sigma_c^2 \mid \mathbf{y}) \cong [\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y}]/[n - p - 2(c - 1)] \qquad (28a)$$

and

$$\text{Mode}(\sigma_i^2 \mid \mathbf{y}) \cong [\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y}]/\widehat{\alpha}_i[n - p - 2(c - 1)] = [28a]/\widehat{\alpha}_i \qquad (28b)$$

$$i = 1, 2, \dots, c - 1$$

The posterior variances can be approximated by evaluating expressions s_c^2 in
(17c) and s_i^2 in (23c) at the $\widehat{\alpha}$ values. In fact, all the approximate marginal

distributions can be depicted graphically by plotting densities (16) and (22), evaluated at the $\hat{\alpha}$ values, thus giving a full (approximate) solution to the problem of inferences regarding variance components in a univariate mixed linear model.

DISCUSSION

Harville (1990) stated: *"A more extensive use of Bayesian ideas by animal breeders and other practitioners is desirable and is more feasible from a computational standpoint than commonly thought... The Bayesian approach can be used to devise prediction procedures that are more sensible – from both a Bayesian and frequentist perspective – than those in current use".*

In this paper we have used the Bayesian mechanism of integrating nuisance parameters to construct marginal likelihoods from which marginal inferences about each of the variance components can be made. In this respect, the method advanced in this paper goes beyond REML in degree of marginalization. In REML, the fixed effects are viewed as nuisance parameters and are integrated out of the likelihood (Harville, 1974) so inferences about variance components are carried out jointly. In the present paper fixed effects and other variance components also regarded as nuisances are integrated out so that inferences about individual variances of interest can be completed after taking into account, exactly or approximately, the error due to not knowing all nuisance parameters. Hence, practitioners that accept REML in terms of the argument advanced by Patterson and Thompson (1971), *ie*, taking into account degrees of freedom "lost", should also feel comfortable with our approach because additional degrees of freedom (stemming from the nuisance variance components) are also taken into account.

When fixed effects are viewed as nuisance parameters, it is fortunate that the likelihood can be separated into a component that depends on the fixed effects plus an "error contrast" part (Patterson and Thompson, 1971). However, in likelihood inference it seems impossible to eliminate nuisance parameters other than by *ad hoc* methods or without recourse to Bayesian ideas. A possible method for dealing with nuisance variances would be treating all random effects other than that of interest as fixed, and then estimating the appropriate component by some translation invariant procedure such as REML. This suggests that the distribution of the resulting statistic would not depend on parameters of the distribution of the nuisance random effects. On the other hand, the variance of interest would be estimated jointly with the residual variance; in most instances this should cause little difficulty because of the large amount of information available about the residual component. It seems intriguing to examine the sampling properties of this method.

An interesting question is the extent to which the proposed method differs from REML. It is known that ML and REML can differ appreciably when the number of fixed effects is large relative to the number of observations. Likewise, the difference between estimates obtained with the new method and REML should be a function of the number of variance components in the model and of the amount of information about such components contained in the sample. For the sake of brevity, we will refer to the new method as "VEIL" (standing for "variance estimation from integrated likelihoods").

The finite sampling properties such as bias, variance and mean squared error of VEIL are unknown, but the same holds for REML and ML, except for some trivial models. Differences among methods depend on data structures and parameter values, and can only be assessed *via* simulation experiments. When the approximate analysis described in this paper is used, the goodness of VEIL will depend on the extent to which the density $\alpha_1, \alpha_2, \ldots, \alpha_{c-1} \mid y$ is peaked; this assumption surely does not hold when there is limited information in the data about the variance ratios. In this case, the exact approach is strongly recommended over the approximate one. With respect to large sample properties, the work of Cox (1975) on partial likelihoods suggests that the properties of maximum likelihood estimates apply to maximum marginal likelihood estimates as well. A good illustration is precisely REML relative to ML. Hence, the large sample properties of estimates obtained from the highly marginalized likelihoods presented here should be the same as those of REML although, of course, the parameters of the asymptotic distribution must differ. For example, the asymptotic variance-covariance matrix of ML estimates of variance components is different from that of REML estimates.

From a Bayesian viewpoint, the method described in this paper is a marked improvement over ML and REML provided the objective of the analysis is making marginal inferences about individual components of variance, as the other parameters are treated as nuisances. As stated earlier, the marginal densities of the variance components are approximately inverted χ^2. This allows to generate a complete posterior distribution from which moments can be calculated as needed, and probability statements can be obtained at least by numerical procedures. The inverted χ^2 distribution is defined only over the positive part of the real line so the embarrassing situations caused by encountering point estimates that fall outside of the parameter space, or negative values of the variances yielding a positive likelihood (Harville and Callanan, 1990) are avoided.

It is interesting to observe that the (approximate) marginal posterior densities of the variance components appear in the inverted χ^2 form. In a fixed model, the marginal posterior distribution of the residual variance is exactly inverted χ^2 and its modal value is the usual unbiased estimator (identical to REML in this case). This implies that this distribution is suitable for specifying marginal prior knowledge about variance parameters as done, for example, by Lindley and Smith (1972) and Höschele *et al* (1987). For instance, suppose that the prior distribution of σ_i^2 is taken as inverted χ^2 with parameters ν_0 and s_0^2. Then the marginal posterior density of σ_i^2 can be written as:

$$p(\sigma_i^2 \mid \mathbf{y}, \nu_0, s_0^2) \propto \text{integrated likelihood} \times p(\sigma_i^2 \mid \nu_0, s_0^2) \qquad (29)$$

Taking into account that the integrated likelihood is approximately proportional to (22) when the α's are replaced by the $\widehat{\alpha}$'s, (29) can be put as:

$$p(\sigma_i^2 \mid y, \nu_0, s_0^2) \propto (\sigma_i^2)^{-\frac{1}{2}(\nu+2)} \exp\left\{-\frac{1}{2}\nu s_i^2/\sigma_i^2\right\} (\sigma_i^2)^{-\frac{1}{2}(\nu_0+2)} \exp\left\{-\frac{1}{2}\nu_0 s_0^2/\sigma_i^2\right\}$$

$$\propto (\sigma_i^2)^{-\frac{1}{2}(\bar{\nu}+2)} \exp\left\{-\frac{1}{2}\bar{\nu}\,\bar{s}_i^2/\sigma_i^2\right\} \qquad (30)$$

where:

$$\bar{\nu} = \nu + \nu_0 + 2 \tag{31a}$$

and

$$\bar{s}_i^2 = \frac{\nu s_i^2 + \nu_0 s_0^2}{\nu + \nu_0 + 2} \tag{31b}$$

Hence, the marginal posterior density of component σ_i^2 remains in the inverted χ^2 form. This "conjugate" property of the inverted χ^2 density can be used to advantage when computing the marginal density of a variance component from a large data set. If the data can be partitioned into independent pieces, repeated use of (30) with parameters updated as in (31a) and (31b) will lead to the (approximate) marginal posterior density of the variance components. Each of these pieces of data should have a relatively large amount of information about the variance ratios so that replacing the variance ratios α_i by the modal values $\hat{\alpha}_i$ can be justified. Note in (31a) and (31b) that while the "degree of belief" parameter accumulates additively, the parameter s^2 is updated in the form of a weighted average which is a common feature in Bayesian posterior analysis.

The computations required for completing the approximate analysis are similar to those of REML. As illustrated in (13), an iteration constructed on the basis of successive approximations requires at each round the solution vector and the inverse of the coefficient matrix of Henderson's equations. These are also the computing requirements arising in first and second order algorithms for REML (Harville and Callanan, 1990). Derivative free algorithms should be explored as well. On the other hand, carrying out the exact analysis requires numerical integration and the mixed model equations would need to be solved at each of the steps of integration. This may not be feasible in a large class of models. In summary, the exact method can be applied only in data sets of moderate size, but it is precisely in this situation where it would be most useful.

As stated by Searle (1988), the sampling distributions of analysis of variance and likelihood based estimates of variance components are unknown and will probably never be because of analytical intractability. It would be intriguing to study the extent to which the inverted χ^2 distributions with densitites as in (16) and (22) can provide a description of the variability encountered over repeated sampling among estimates of variance components obtained with the present method. In such a study the σ_i^2's would be estimates obtained in different (independent) samples and ν and s_i^2 would be parameters of the sampling distribution. In other words, perhaps the small sample distribution of point estimates obtained with VEIL can be suitably fitted by an inverted χ^2 distribution.

ACKNOWLEDGMENTS

REFERENCES

Box GEP, Tiao GC (1973) *Bayesian inference in statistical analysis.* Addison-Wesley, Reading

Cox DR (1975) Partial likelihood. *Biometrika* 62, 269-276

Dawid AP (1980) A Bayesian look at nuisance parameters. *In: Bayesian Statistics* (Bernardo JM, De Groot MH, Lindley DV, Smith AFM eds) Univ Press, Valencia, 167-175

Gianola D, Fernando RL (1986) Bayesian methods in animal breeding theory. *J Anim Sci* 63, 217-244

Gianola D, Foulley JL, Fernando RL (1986) Prediction of breeding values when variances are not known. *Genet Sel Evol* 18, 485-498

Gianola D, Im S, Fernando RL, Foulley JL (1990a) Mixed model methodology and the Box-Cox theory of transformations: a Bayesian approach. *In: Advances in statistical methods for genetic improvement of livestock* (Gianola D, Hammond K eds) Springer-Verlag, New York-Heidelberg-Berlin, 15-40

Gianola D, Im S, Macedo FWM (1990b) A framework for prediction of breeding value. *In: Advances in statistical methods for genetic improvement of livestock* (Gianola D, Hammond K eds) Springer-Verlag, New York-Heidelberg-Berlin, 210-239

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

Harville DA (1990) BLUP (Best linear unbiased prediction) and beyond. *In: Advances in statistical methods for genetic improvement of livestock* (Gianola D, Hammond K eds) Springer-Verlag, New York-Heidelberg-Berlin 239-276

Harville DA, Callanan TP (1990) Computational aspects of likelihood based inference for variance components. *In: Advances in statistical methods for genetic improvement of livestock* (Gianola D, Hammond K, eds) Springer-Verlag, New York-Heidelberg-Berlin 136-176

Höschele I, Gianola D, Foulley JL (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet* 104, 334-349

Kloek T, Van Dijk H (1978) Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. *Econometrica* 46, 1-19

Lindley DV, Smith AMF (1972) Bayes estimates for the linear model. *J Roy Stat Soc* 9, 45-58

Macedo FWM, Gianola D (1987) *Bayesian analysis of univariate mixed models with informative priors.* Eur Assoc Anim Prod, 38th Annu Meet, Lisbon, Portugal, 35 p

O'Hagan A (1976) On posterior joint and marginal modes. *Biometrika* 63, 329-333

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

Searle SR (1988) Mixed models and unbalanced data: wherefrom, whereat and whereto? *Comm Stat* 17, 935-968

Smith AFM, Skene AM, Shaw JEH, Naylor JC (1987) Progress with numerical and graphical methods for practical Bayesian statistics. *Statistician* 36, 75-82

Swallow W, Monahan J (1984) Monte Carlo comparison of ANOVA, MIVQUE, REML and ML estimators of variance components. *Technometrics* 26, 47-57

Thompson WA (1962) The problem of negative estimates of variance components. *Ann Math Stat* 26, 721-733
Zellner A (1971) An introduction to Bayesian inference in econometrics. Wiley, New York