

Heterogeneous variances in Gaussian linear mixed models

JL Foulley¹, RL Quaas²

¹*Institut national de la recherche agronomique, station de génétique quantitative et appliquée, 78352 Jouy-en-Josas, France;*

²*Department of Animal Science, Cornell University, Ithaca, NY 14853, USA*

(Received 28 February 1994; accepted 29 November 1994)

Summary – This paper reviews some problems encountered in estimating heterogeneous variances in Gaussian linear mixed models. The one-way and multiple classification cases are considered. EM-REML algorithms and Bayesian procedures are derived. A structural mixed linear model on log-variance components is also presented, which allows identification of meaningful sources of variation of heterogeneous residual and genetic components of variance and assessment of their magnitude and mode of action.

heteroskedasticity / mixed linear model / restricted maximum likelihood / Bayesian statistics

Résumé – **Variances hétérogènes en modèle linéaire mixte gaussien.** *Cet article fait le point sur un certain nombre de problèmes qui surviennent lors de l'estimation de variances hétérogènes dans des modèles linéaires mixtes gaussiens. On considère le cas d'un ou plusieurs facteurs d'hétéroscédasticité. On développe des algorithmes EM-REML et bayésiens. On propose également un modèle linéaire mixte structurel des logarithmes des variances qui permet de mettre en évidence des sources significatives de variation des variances résiduelles et génétiques et d'appréhender leur importance et leur mode d'action.*

hétéroscédasticité / modèle linéaire mixte / maximum de vraisemblance résiduelle / statistique bayésienne

INTRODUCTION

Genetic evaluation procedures in animal breeding rely mainly on best linear unbiased prediction (BLUP) and restricted maximum likelihood (REML) estimation of parameters of Gaussian linear mixed models (Henderson, 1984). Although BLUP can accommodate heterogeneous variances (Gianola, 1986), most applications of

mixed-model methodology postulate homogeneity of variance components across subclasses involved in the stratification of data. However, there is now a great deal of experimental evidence of heterogeneity of variances for important production traits of livestock (*eg*, milk yield and growth in cattle) both at the genetic and environmental levels (see, for example, the reviews of Garrick *et al*, 1989, and Visscher *et al*, 1991).

As shown by Hill (1984), ignoring heterogeneity of variance decreases the efficiency of genetic evaluation procedures and consequently response to selection, the importance of this phenomenon depending on assumptions made about sources and magnitude of heteroskedasticity (Garrick and Van Vleck, 1987; Visscher and Hill, 1992). Thus, making correct inferences about heteroskedastic variances is critical. To that end, appropriate estimation and testing procedures for heterogeneous variances are needed. The purpose of this paper is an attempt to describe such procedures and their principles. For pedagogical reasons, the presentation is divided into 2 parts according to whether heteroskedasticity is related to a single or to a multiple classification of factors.

THE ONE-WAY CLASSIFICATION

Statistical model

The population is assumed to be stratified into several subpopulations (*eg*, herds, regions, *etc*) indexed by $i = 1, 2, \dots, I$, representing a potential source of heterogeneity of variances. For the sake of simplicity, we first consider a one-way random model for variances such as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \sigma_{u_i} \mathbf{Z}_i \mathbf{u}^* + \mathbf{e}_i \quad [1]$$

where \mathbf{y}_i is the $(n_i \times 1)$ data vector for subpopulation i , $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of fixed effects with incidence matrix \mathbf{X}_i , \mathbf{u}^* is a $(q \times 1)$ vector of standardized random effects with incidence matrix \mathbf{Z}_i and \mathbf{e}_i is the $(n_i \times 1)$ vector of residuals.

The usual assumptions of normality and independence are made for the distributions of the random variances \mathbf{u}^* and \mathbf{e}_i , *ie* $\mathbf{u}^* \sim N(\mathbf{0}, \mathbf{A})$ (\mathbf{A} positive definite matrix of coefficients of relationship) and $\mathbf{e}_i \sim \text{NID}(\mathbf{0}, \sigma_{e_i}^2 \mathbf{I}_{n_i})$ and $\text{Cov}(\mathbf{e}_i, \mathbf{u}^{*'}) = \mathbf{0}$ so that $\mathbf{y}_i \sim N(\mathbf{X}_i \boldsymbol{\beta}, \sigma_{u_i}^2 \mathbf{Z}_i' \mathbf{A} \mathbf{Z}_i + \sigma_{e_i}^2 \mathbf{I}_{n_i})$ where $\sigma_{e_i}^2$ and $\sigma_{u_i}^2$ are the residual and u-components of variance pertaining to subpopulation i . A simple example for [1] is a 2-way additive mixel model $y_{ijk} = \mu + h_i + \sigma_{s_i} s_j^* + e_{ijk}$ with fixed herd (h_i) and random sire ($\sigma_{s_i} s_j^*$) effects. Notice that model [1] includes the case of fixed effects nested within subpopulations as observed in many applications.

EM-REML estimation of heterogeneous variance components

To be consistent with common practice for estimation of variance components, we chose REML (Patterson and Thompson, 1971; Harville, 1977) as the basic estimation procedure for heterogeneous variance components (Foulley *et al*, 1990). A convenient algorithm to compute such REML estimates is the 'expectation-maximization' (EM) algorithm of Dempster *et al* (1977). The iterative scheme will

be based on the general definition of EM (see pages 5 and 6 and formula 2.17 in Dempster *et al*, 1977) which can be explained as follows.

Letting $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_i, \dots, \mathbf{y}'_I)'$, $\boldsymbol{\sigma}^2_e = \{\sigma^2_{e_i}\}$, $\boldsymbol{\sigma}^2_u = \{\sigma^2_{u_i}\}$ and $\boldsymbol{\sigma}^2 = (\boldsymbol{\sigma}^2_e', \boldsymbol{\sigma}^2_u')$, the derivation of the EM algorithm for REML stems from a complete data set defined by the vector $\mathbf{x} = (\mathbf{y}', \boldsymbol{\beta}', \mathbf{u}^*)'$ and the corresponding likelihood function $L(\boldsymbol{\sigma}^2; \mathbf{x}) = \ln p(\mathbf{x}|\boldsymbol{\sigma}^2)$. In this presentation, the vector $\boldsymbol{\beta}$ is treated in a Bayesian manner as a vector of random effects with variance fixed at infinity (Dempster *et al*, 1977; Foulley, 1993). A frequentist interpretation of this algorithm based on error contrasts can be found in De Stefano (1994). A similar derivation was given for the homoskedastic case by Cantet (1990). As usual, the EM algorithm is an iterative one consisting of an 'expectation' (E) and of a 'maximization' (M) step. Given the current estimate $\boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^{2[t]}$ at iteration $[t]$, the E step consists of computing the conditional expectation of $L(\boldsymbol{\sigma}^2; \mathbf{x})$, *ie*

$$Q(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]}) = E[L(\boldsymbol{\sigma}^2; \mathbf{x})|\mathbf{y}, \boldsymbol{\sigma}^{2[t]}] \tag{3}$$

given the data vector \mathbf{y} and $\boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^{2[t]}$.

The M step consists of choosing the next value $\boldsymbol{\sigma}^{2[t+1]}$ of $\boldsymbol{\sigma}^2$ by maximizing $Q(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]})$ with respect to $\boldsymbol{\sigma}^2$

$$\boldsymbol{\sigma}^{2[t+1]} = \text{Arg max}_{\boldsymbol{\sigma}^2} Q(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]}) \tag{4}$$

Since $\ln p(\mathbf{x}|\boldsymbol{\sigma}^2) = \ln p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}^*, \boldsymbol{\sigma}^2) + \ln p(\boldsymbol{\beta}, \mathbf{u}^*|\boldsymbol{\sigma}^2)$ with $\ln p(\boldsymbol{\beta}, \mathbf{u}^*|\boldsymbol{\sigma}^2)$ providing no information about $\boldsymbol{\sigma}^2$, $Q(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]})$ can be replaced by

$$Q^*(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]}) = E[\ln p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{u}^*, \boldsymbol{\sigma}^2)|\mathbf{y}, \boldsymbol{\sigma}^{2[t]}]$$

Under model [1], the expression for $Q^*(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]})$ reduces to

$$Q^*(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]}) = \text{Const} - (1/2) \sum_{i=1}^I n_i \ln(\sigma^2_{e_i}) - (1/2) \sum_{i=1}^I \sigma^{-2}_{e_i} E_c^{[t]}(\mathbf{e}'_i \mathbf{e}_i)$$

where $E_c^{[t]}(\cdot)$ indicates a conditional expectation taken with respect to the distribution of $\boldsymbol{\beta}, \mathbf{u}^*|\mathbf{y}, \boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^{2[t]}$. This posterior distribution is multivariate normal with mean $E(\boldsymbol{\beta}|\mathbf{y}, \boldsymbol{\sigma}^2) = \text{BLUE}$ (best linear unbiased estimate) of $\boldsymbol{\beta}$, $E(\mathbf{u}|\mathbf{y}, \boldsymbol{\sigma}^2) = \text{BLUP}$ of \mathbf{u} , and $\text{Var}(\boldsymbol{\beta}, \mathbf{u}|\mathbf{y}, \boldsymbol{\sigma}^2) = \text{inverse of the mixed-model coefficient matrix}$.

The system of equations $\partial Q^*(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]})/\partial \boldsymbol{\sigma}^2 = \mathbf{0}$ can be written as follows: With respect to the *u*-component, we have

$$\frac{\partial}{\partial \sigma^2_{u_k}} Q^*(\boldsymbol{\sigma}^2|\boldsymbol{\sigma}^{2[t]}) = 0 \implies -\frac{1}{2\sigma^2_{e_i}} \left[\frac{\partial}{\partial \sigma^2_{u_i}} E_c^{[t]}(\mathbf{e}'_i \mathbf{e}_i) \right] = 0 \tag{5}$$

and

$$\frac{\partial}{\partial \sigma^2_{u_i}} E_c^{[t]}(\mathbf{e}'_i \mathbf{e}_i) = \sigma^{-1}_{u_i} \left\{ -E_c^{[t]} \left[\mathbf{u}^*{}' \mathbf{Z}'_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] + \sigma_{u_i} E_c^{[t]}[\mathbf{u}^*{}' \mathbf{Z}'_i \mathbf{Z}_i \mathbf{u}^*] \right\}$$

For the residual component,

$$\frac{\partial}{\partial \sigma_{e_i}^2} \mathbf{Q}^*(\sigma^2 | \sigma^{2[t]}) = 0 \implies -\frac{1}{2}(n_i/\sigma_{e_i}^2) + \frac{1}{2} \left[\mathbf{E}_c^{[t]}(\mathbf{e}'_i \mathbf{e}_i) / \sigma_{e_i}^4 \right] = 0 \quad [6]$$

Since $\mathbf{E}_c^{[t]}(\mathbf{e}'_i \mathbf{e}_i)$ is a function of the unknown σ_{u_i} only, equation [5] depends only on that unknown whereas equation [6] depends of both variance components. We then solve [5] first with respect to σ_{u_i} , and then solve [6] second substituting the solution $\sigma_{u_i}^{[t+1]}$ to σ_{u_i} back into $\mathbf{E}_c^{[t]}(\mathbf{e}'_i \mathbf{e}_i)$ of [6], *ie* with

$$\mathbf{e}_i^{[t+1]} = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \sigma_{u_i}^{[t+1]} \mathbf{Z}_i \mathbf{u}^*$$

Hence

$$\sigma_{u_i}^{[t+1]} = \frac{\mathbf{E}_c^{[t]}[\mathbf{u}^{*'} \mathbf{Z}'_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})]}{\mathbf{E}_c^{[t]}(\mathbf{u}^{*'} \mathbf{Z}'_i \mathbf{Z}_i \mathbf{u}^*)} \quad [7]$$

$$\sigma_{e_i}^{2[t+1]} = \mathbf{E}_c^{[t]}(\mathbf{e}'_i{}^{[t+1]} \mathbf{e}_i^{[t+1]}) / n_i \quad [8]$$

It is worth noticing that formula [7] gives the expression of the standard deviation of the *u*-component, and has the form of a regression coefficient estimator. Actually σ_{u_i} is the coefficient of regression of any element of \mathbf{y}_i on the corresponding element of $\mathbf{Z}_i \mathbf{u}^*$.

Let the system of mixed-model equations be written as

$$\begin{bmatrix} \sum_{i=1}^I \mathbf{X}'_i \mathbf{X}_i / \sigma_{e_i}^2 & \sum_{i=1}^I \mathbf{X}'_i \mathbf{Z}_i \sigma_{u_i} / \sigma_{e_i}^2 \\ \sum_{i=1}^I \mathbf{Z}'_i \mathbf{X}_i \sigma_{u_i} / \sigma_{e_i}^2 & \sum_{i=1}^I \mathbf{Z}'_i \mathbf{Z}_i \sigma_{u_i}^2 / \sigma_{e_i}^2 + \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}}^* \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^I \mathbf{X}'_i \mathbf{y}_i / \sigma_{e_i}^2 \\ \sum_{i=1}^I \mathbf{Z}'_i \mathbf{y}_i \sigma_{u_i} / \sigma_{e_i}^2 \end{bmatrix} \quad [9]$$

and

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{\beta\beta} & \mathbf{C}_{\beta u} \\ \mathbf{C}_{u\beta} & \mathbf{C}_{uu} \end{bmatrix} = \mathbf{g} \text{ inverse of the coefficient matrix.}$$

The elements of [7] and [8] can be expressed as functions of \mathbf{y} , $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$ and blocks of \mathbf{C} as follows

$$\begin{aligned} \mathbf{E}_c(\mathbf{e}'_i \mathbf{e}_i) &= \mathbf{E}_c[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] - 2\sigma_{u_i} \mathbf{E}_c[\mathbf{u}^{*'} \mathbf{Z}'_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] \\ &\quad + \sigma_{u_i}^2 \mathbf{E}_c(\mathbf{u}^{*'} \mathbf{Z}'_i \mathbf{Z}_i \mathbf{u}^*) \end{aligned}$$

$$\mathbf{E}_c[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] = (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}})' (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) + \text{tr}(\mathbf{X}'_i \mathbf{X}_i \mathbf{C}_{\beta\beta})$$

$$\mathbf{E}_c[\mathbf{u}^{*'} \mathbf{Z}'_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] = \hat{\mathbf{u}}^{*'} \mathbf{Z}'_i (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\beta}}) - \text{tr}(\mathbf{Z}'_i \mathbf{X}_i \mathbf{C}_{\beta u})$$

$$\mathbf{E}_c(\mathbf{u}^{*'} \mathbf{Z}'_i \mathbf{Z}_i \mathbf{u}^*) = \hat{\mathbf{u}}^{*'} \mathbf{Z}'_i \mathbf{Z}_i \hat{\mathbf{u}}^* + \text{tr}(\mathbf{Z}'_i \mathbf{Z}_i \mathbf{C}_{uu})$$

For readers interested in applying the above formulae, a small example is the presented in tables I and II for a (fixed) environment and (random) sire model. It

is worth noticing that formulae [7] and [8] can also be applied to the homoskedastic case by considering that there is just one subpopulation ($I = 1$). The resulting algorithm looks like a regression in contrast to the conventional EM whose formula ($\sigma_u^{2[t+1]} = E_c^{[t]}(\mathbf{u}'\mathbf{A}^{-1}\mathbf{u})/q$) where \mathbf{u} is not standardized ($\mathbf{u} = \sigma_u \mathbf{u}^*$) is in terms of a variance. Not only do the formulae look quite different, but they also perform quite differently in terms of rounds to convergence. The conventional EM tends to do quite poorly if $\sigma_e^2 \gg \sigma_u^2$ and (or) with little information, whereas the scaled EM is at its best in these situations. This can be demonstrated by examining a balanced paternal half-sib design (q families with progeny group size n each). This is convenient because in this case the EM algorithms can be written in terms of the between- and within-sire sums of squares and convergence performance checked for a variety of situations without simulating individual records. For this simple situation performance was fairly well predicted by the criterion $R^2 = n/(n + \alpha)$, where $\alpha = \sigma_e^2/\sigma_u^2$. Figure 1 is a plot of rounds to convergence for the scaled and usual EM algorithms for an arbitrary set of values of n and α . As noted by Thompson and Meyer (1986), the usual EM performs very poorly at low R^2 , *eg*, $n = 5$ and $h^2 = 4/(\alpha + 1) = 0.25$ or $n = 33$ and $h^2 = 0.04$, *ie* $R^2 = 0.25$, but very well at the other end of the spectrum: $n = 285$ and $h^2 = 0.25$ or $n = 1881$ and $h^2 = 0.04$, *ie* $R^2 = 0.95$. The performance of the scaled version is the exact opposite. Interestingly, both EM algorithms perform similarly for R^2 values typical of many animal breeding data sets ($n = 30$ and $h^2 = 0.25$, *ie* $R^2 = 2/3$).

Table I. Numerical example. Distribution of records.

<i>No</i>	<i>Environment</i>	<i>Sire</i>	<i>Record</i>	<i>No</i>	<i>Environment</i>	<i>Sire</i>	<i>Record</i>
1	1	1	470	19	2	1	385
2	1	1	510	20	2	2	450
3	1	1	345	21	2	2	605
4	1	1	395	22	2	3	575
5	1	2	450	23	2	4	530
6	1	2	345	24	2	4	310
7	1	2	495	25	2	4	415
8	1	3	410	26	2	4	370
9	1	3	335	27	3	2	805
10	1	3	365	28	3	2	475
11	1	3	480	29	3	3	875
12	1	4	410	30	3	3	850
13	1	4	330	31	3	3	510
14	1	4	300	32	3	3	310
15	1	4	330	33	3	4	565
16	2	1	530	34	3	4	330
17	2	1	880	35	3	4	410
18	2	1	575	36	3	4	480

Moreover, solutions given by the EM algorithm in [7] and [8] turn out to be within the parameter space in the homoskedastic case (see proof in the *Appendix*) but not necessarily in the heteroskedastic case as shown by a counter-example.

Table II. Numerical example. Estimates of variance components^a.

Situation	Method	Sire components			Residual components		
		1	2	3	1	2	3
Homoskedastic	REML	3 668	3 668	3 668	18 214	18 214	18 214
Heteroskedastic	REML	1 145	5 523	9 246	3 794	18 704	36 972
Heteroskedastic	Bayes ^b	2 612	3 112	3 108	8 502	17 005	25 912
Heteroskedastic	Bayes ^c	3 150	3 879	3 896	9 335	18 890	28 922

^a Model: environment 1 2 3 (fixed) + sire (random) + residual; ^b posterior mode of variance components (maximizing [11ab]): $\eta_e = \eta_u = 8$, $s_u^2 = 3\,668.42$, $s_e^2 = 18\,214.49$ as parameter values of the priors; ^c posterior mode of log-variance components (maximizing [26ab]); same parameter values of the priors.

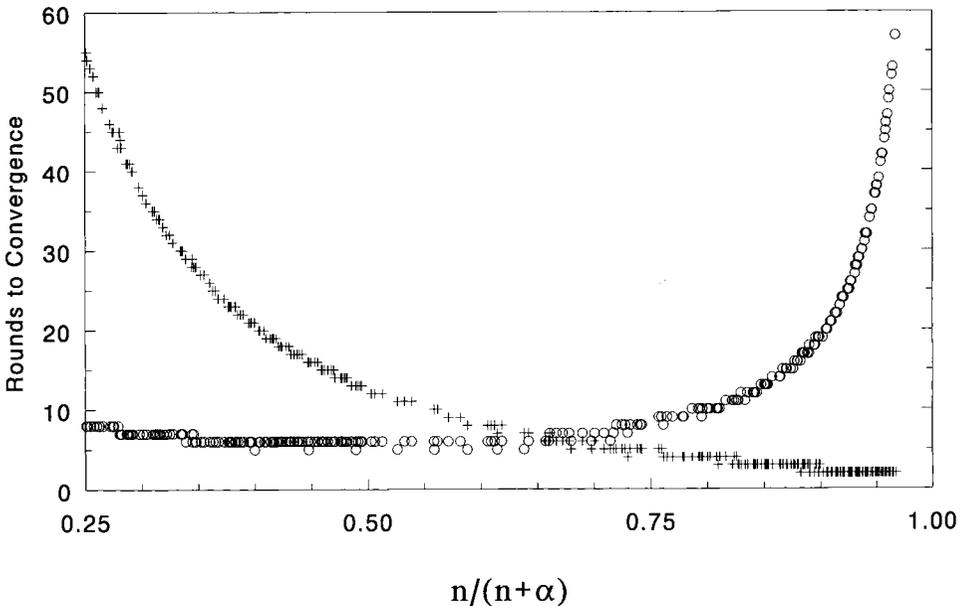


Fig 1. Rounds to convergence for scaled (O) and usual (+) EM algorithms with balanced designs of 100 families of n paternal half sibs as a function of $n/(n + \alpha)$ (α = ratio of the residual to the sire variance components).

Bayesian approach

When there is little information per subpopulation (*eg*, herd or herdx management unit), REML estimation of $\sigma_{e_i}^2$ and $\sigma_{u_i}^2$ can be unreliable. This led Hill (1984) and Gianola (1986) to suggest estimates shrunken towards some common mean variance. In this respect, Gianola *et al* (1992) proposed a Bayesian procedure to estimate heterogeneous variance components. Their approach can be viewed as a

natural extension of the EM-REML technique described previously. The parameters $\sigma_{e_i}^2$ and $\sigma_{u_i}^2$ are assumed to be independently and identically distributed random variables with scaled inverted chi-square density functions, the parameters of which are s_e^2, η_e and s_u^2, η_u respectively. The parameters s_e^2 and s_u^2 are location parameters of the prior distributions of variance components, and η_e and η_u (degrees of belief) are quantities related to the squared coefficient of variation (cv) of true variances by $\eta_e = (2/cv_e^2) + 4$ and $\eta_u = (2/cv_u^2) + 4$ respectively:

$$p(\sigma_{e_i}^2 | s_e^2, \eta_e) \propto (\sigma_{e_i}^2)^{-(\eta_e+2)/2} \exp(-\eta_e s_e^2 / 2\sigma_{e_i}^2) \quad [10a]$$

$$p(\sigma_{u_i}^2 | s_u^2, \eta_u) \propto (\sigma_{u_i}^2)^{-(\eta_u+2)/2} \exp(-\eta_u s_u^2 / 2\sigma_{u_i}^2) \quad [10b]$$

Moreover, let us assume as in Searle *et al* (1992, page 99), that the priors for residual and u -components are assumed independent so that $p(\sigma_{e_i}^2, \sigma_{u_i}^2) = p(\sigma_{e_i}^2)p(\sigma_{u_i}^2)$.

The $Q^{\textcircled{Q}}(\sigma^2 | \sigma^{2[t]})$ function to maximize in order to produce the posterior mode of σ^2 is now (Dempster *et al*, 1977, page 6):

$$Q^{\textcircled{Q}}(\sigma^2 | \sigma^{2[t]}) = Q^*(\sigma^2 | \sigma^{2[t]}) + G(\sigma^2) \quad [11a]$$

with

$$G(\sigma^2) = (-1/2) \left\{ (\eta_e + 2) \left(\sum_{i=1}^I \ln \sigma_{e_i}^2 \right) + \left(\sum_{i=1}^I \eta_e s_e^2 / \sigma_{e_i}^2 \right) + (\eta_u + 2) \left(\sum_{i=1}^I \ln \sigma_{u_i}^2 \right) + \left(\sum_{i=1}^I \eta_u s_u^2 / \sigma_{u_i}^2 \right) \right\} \quad [11b]$$

Equations based on first derivatives set to zero are:

$$E_c^{[t]}(\mathbf{e}'_i \mathbf{e}_i) \sigma_{e_i}^{-4} - n_i \sigma_{e_i}^{-2} + \eta_e s_e^2 \sigma_{e_i}^{-4} - (\eta_e + 2) \sigma_{e_i}^{-2} = 0 \quad [12a]$$

$$\sigma_{e_i}^{-2} \{ E_c^{[t]}[\mathbf{u}'^* \mathbf{Z}'_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] - E_c^{[t]}[\mathbf{u}'^* \mathbf{Z}'_i \mathbf{Z}_i \mathbf{u}^*] \sigma_{u_i} \} - (\eta_u + 2) \sigma_{u_i}^{-1} + \eta_u s_u^2 \sigma_{u_i}^{-3} = 0 \quad [12b]$$

Using [12ab], one can use the following iterative algorithm

$\sigma_{u_i}^{[t+1]}$ positive root of

$$\sigma_{u_i}^{2[t+1]} [E_c^{[t]}(\mathbf{u}'^* \mathbf{Z}'_i \mathbf{Z}_i \mathbf{u}^*) \delta_i^{[t]} + \eta_u + 2] - \sigma_{u_i}^{[t+1]} \{ E_c^{[t]}[\mathbf{u}'^* \mathbf{Z}'_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] \delta_i^{[t]} \} - \eta_u s_u^2 = 0 \quad [13a]$$

or, alternatively

$$\sigma_{u_i}^{2[t+1]} = \frac{E_c^{[t]}[\mathbf{u}'^* \mathbf{Z}'_i (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] \delta_i^{[t]} \sigma_{u_i}^{[t]} + \eta_u s_u^2}{E_c^{[t]}(\mathbf{u}'^* \mathbf{Z}'_i \mathbf{Z}_i \mathbf{u}^*) \delta_i^{[t]} + \eta_u + 2} \quad [13b]$$

and

$$\sigma_{e_i}^{2[t+1]} = \frac{E_c^{[t]}[\mathbf{e}_i^{[t+1]} \mathbf{e}_i^{[t+1]}] + \eta_e s_e^2}{n_i + \eta_e + 2} \quad [14]$$

where $\delta_i^{[t]} = \sigma_{u_i}^{2[t]} / \sigma_{e_i}^{2[t]}$.

Comparing [13b] and [14] with the EM-REML formulae [7] and [8] shows how prior information modifies data information (see also tables I and II). In particular when $\eta_e(\eta_u) = 0$ (absence of knowledge on prior variances), formulae [13b] and [14] are very similar to the EM-REML formulae. They would have been exactly the same if we had considered the posterior mode of log-variances instead of variances, η_e and η_u replacing $\eta_e + 2$ and $\eta_u + 2$ respectively in [11], and, consequently also in the denominator of [13b] and [14]. In contrast, if $\eta_e(\eta_u) \rightarrow \infty$ (no variation among variances), estimates tend to the location parameters $s_e^2(s_u^2)$.

Extension to several u -components

The EM-REML equations can easily be extended to the case of a linear mixed model including several independent u -components (\mathbf{u}_j^* ; $j = 1, 2, \dots, J$), *ie*

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \sum_{j=1}^J \sigma_{u_{ij}} \mathbf{Z}_{ij} \mathbf{u}_j^* + \mathbf{e}_i \quad [15]$$

In that case, it can be shown that formula [7] is replaced by the linear system

$$\sum_{k=1}^J \mathbf{E}_c^{[t]} \{ \mathbf{u}_j^{*'} \mathbf{Z}'_{ij} \mathbf{Z}_{ik} \mathbf{u}_k^* \} \sigma_{u_{jk}}^{[t+1]} = \mathbf{E}_c^{[t]} \{ \mathbf{u}_j^{*'} \mathbf{Z}'_{ij} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \}, \text{ for } j = 1, 2, \dots, J \quad [16]$$

The formula in [8] for the residual components of variance remains the same.

This algorithm can be extended to non-independent u -factors. As in a sire, maternal grand sire model, it is assumed that correlated factors j and k are such that $\text{Var}(\mathbf{u}_j^*) = \text{Var}(\mathbf{u}_k^*) = \mathbf{A}$, and $\text{Cov}(\mathbf{u}_j^*, \mathbf{u}_k^{*'}) = \rho_{jk} \mathbf{A}$ with $\dim(\mathbf{u}_j^*) = m$ for all j . Let $\tilde{\boldsymbol{\sigma}}^2 = (\boldsymbol{\sigma}_e^{2'}, \boldsymbol{\sigma}_u^{2'}, \boldsymbol{\rho}')$ with $\boldsymbol{\rho} = \text{vech}(\Omega)$, Ω being a $(m \times m)$ correlation matrix with ρ_{jk} as element jk . The $Q^\#(\tilde{\boldsymbol{\sigma}}^2 | \tilde{\boldsymbol{\sigma}}^{2[t]})$ function to maximize can be written here as

$$Q^\#(\tilde{\boldsymbol{\sigma}}^2 | \tilde{\boldsymbol{\sigma}}^{2[t]}) = \mathbf{E}_c^{[t]} [\ln p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}^*, \boldsymbol{\sigma}^2)] + \mathbf{E}_c^{[t]} [\ln p(\mathbf{u}^* | \tilde{\boldsymbol{\sigma}}^2)] \quad [17]$$

where $\mathbf{u}^* = (\mathbf{u}_1^{*'}, \mathbf{u}_2^{*'}, \dots, \mathbf{u}_j^{*'}, \dots, \mathbf{u}_j^{*'})'$

The first term $Q_1^\#(\boldsymbol{\sigma}^2 | \tilde{\boldsymbol{\sigma}}^{2[t]}) = \mathbf{E}_c^{[t]} [\ln p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}^*, \boldsymbol{\sigma}^2)]$ has the same form as with the case of independence except that the expectation must taken with respect to the distribution of $\boldsymbol{\beta}, \mathbf{u}^* | \mathbf{y}, \tilde{\boldsymbol{\sigma}}^2 = \tilde{\boldsymbol{\sigma}}^{2[t]}$. The second term $Q_2^\#(\boldsymbol{\rho} | \tilde{\boldsymbol{\sigma}}^{2[t]}) = \mathbf{E}_c^{[t]} [\ln p(\mathbf{u}^* | \tilde{\boldsymbol{\sigma}}^2)]$ can be expressed as

$$Q^\#(\boldsymbol{\rho} | \tilde{\boldsymbol{\sigma}}^{2[t]}) = \text{Const} - (m/2) \ln |\Omega| - \text{tr}[\Omega^{-1} \mathbf{E}_c^{[t]}(\mathbf{D})] \quad [18]$$

where $\mathbf{D} = \{ \mathbf{u}_j^{*'} \mathbf{A}^{-1} \mathbf{u}_k^* \}$ is a $(J \times J)$ symmetric matrix.

The maximization of $Q^\#(\tilde{\boldsymbol{\sigma}}^2 | \tilde{\boldsymbol{\sigma}}^{2[t]})$ with respect to $\tilde{\boldsymbol{\sigma}}^2$ can be carried out in 2 stages: i) maximization of $Q_1^\#(\tilde{\boldsymbol{\sigma}}^2 | \tilde{\boldsymbol{\sigma}}^{2[t]})$ with respect to the vector $\boldsymbol{\sigma}^2$ of variance components which can be solved as above; and ii) maximization of $Q_2^\#(\boldsymbol{\rho} | \tilde{\boldsymbol{\sigma}}^{2[t]})$ with respect to the vector of correlation coefficients $\boldsymbol{\rho}$ which can be performed *via* a Newton-Raphson algorithm.

THE MULTIPLE-WAY CLASSIFICATION

The structural model on log-variances

Let us assume as above that the σ_i^2 s (u and e types) are *a priori* independently distributed as inverted chi-square random variables with parameters s_i^2 (location) and η_i (degrees of belief), such that the density function can be written as:

$$p(\sigma_i^2 | s_i^2, \eta_i) = (\eta_i s_i^2 / 2)^{\eta_i / 2} (\sigma_i^2)^{-(\eta_i + 2) / 2} \exp(-\eta_i s_i^2 / 2 \sigma_i^2) / \Gamma(\eta_i / 2) \quad [19]$$

where $\Gamma(x)$ is the gamma function.

From [19], one can alternatively consider the density of the log-variance $\ln \sigma_i^2$, or more interestingly that of $\nu_i = \ln(\sigma_i^2 / s_i^2)$. In addition, it can be assumed that $\eta_i = \eta$ for all i , and that $\ln s_i^2$ can be decomposed as a linear combination $\mathbf{p}'_i \boldsymbol{\delta}$ of some vector $\boldsymbol{\delta}$ or explanatory variables (\mathbf{p}'_i being a row vector of incidence), such that

$$\ln \sigma_i^2 = \mathbf{p}'_i \boldsymbol{\delta} + \nu_i \quad [20]$$

with

$$p(\nu_i | \eta) = (\eta / 2)^{\eta / 2} \exp\{-(\eta / 2)[\nu_i + \exp(-\nu_i)]\} / \Gamma(\eta / 2) \quad [21]$$

For $\nu_i \rightarrow 0$, the kernel of the distribution in [21] tends towards $\exp(-\eta \nu_i^2 / 4)$, thus leading to the following normal approximation

$$\nu_i | \eta \sim \text{NID}(0, 2 / \eta) \quad [22]$$

where the variance *a priori* (ξ) of true variances is inversely proportional to η ($\xi = 2 / \eta$), ξ also being interpretable as the square coefficient of variation of log-variances. This approximation turns out to be excellent for most situations encountered in practice ($cv \leq 0.50$).

Formulae [20] and [21] can be naturally extended to several independent classifications in $\mathbf{v} = (\mathbf{v}'_1, \mathbf{v}'_2, \dots, \mathbf{v}'_j, \dots, \mathbf{v}'_J)'$ such that

$$\ln \sigma_i^2 = \mathbf{p}'_i \boldsymbol{\delta} + \mathbf{q}'_i \mathbf{v} = \boldsymbol{\ell}'_i \boldsymbol{\lambda} \quad [23]$$

with

$$\mathbf{v} \sim \text{N}(\mathbf{0} \oplus_{j=1}^J \xi_j \mathbf{I}_{K_j}) \quad [24]$$

where $K_j = \dim(\mathbf{v}_j)$ and $\boldsymbol{\lambda} = (\boldsymbol{\delta}', \mathbf{v}')'$ is the vector of dispersion parameters and $\boldsymbol{\ell}'_i = (\mathbf{p}'_i, \mathbf{q}'_i)$ is the corresponding row vector of incidence.

This presentation allows us to mimick a mixed linear model structure with fixed $\boldsymbol{\delta}$ and random \mathbf{v} effects on log-variances, similar to what is done on cell means ($\boldsymbol{\mu}_i = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u} = \mathbf{t}'_i \boldsymbol{\theta}$), and thus justifies the choice of the log as the link function (Leonard, 1975; Denis, 1983; Aitkin, 1987; Nair and Pregibon, 1988) to use for this generalized linear mixed-model approach.

Equations [23] and [24] can be applied both to residual and u -components of variance *viz*,

$$\gamma_u = \mathbf{P}_u \boldsymbol{\delta}_u + \mathbf{Q}_u \mathbf{v}_u = \mathbf{L}_u \boldsymbol{\lambda}_u \quad [25a]$$

$$\boldsymbol{\gamma}_e = \mathbf{P}_e \boldsymbol{\delta}_e + \mathbf{Q}_e \mathbf{v}_e = \mathbf{L}_e \boldsymbol{\lambda}_e \quad [25b]$$

where $\boldsymbol{\gamma}_u = \{\ln \sigma_{u_i}^2\}$, $\boldsymbol{\gamma}_e = \{\ln \sigma_{e_i}^2\}$; \mathbf{P}_u , \mathbf{P}_e are incidence matrices pertaining to fixed effects $\boldsymbol{\delta}_u$, $\boldsymbol{\delta}_e$ respectively; \mathbf{Q}_u , \mathbf{Q}_e are incidence matrices pertaining to random effects $\mathbf{v}_u = (\mathbf{v}'_{u_1}, \mathbf{v}'_{u_2}, \dots, \mathbf{v}'_{u_j}, \dots)'$ and $\mathbf{v}_e = (\mathbf{v}'_{e_1}, \mathbf{v}'_{e_2}, \dots, \mathbf{v}'_{e_{j'}})'$ with $\mathbf{v}_{u_j} \sim \text{Nid}(\mathbf{0}, \xi_{u_j} \mathbf{I}_{K_{u_j}})$ and $\mathbf{v}_{e_{j'}} \sim \text{NID}(\mathbf{0}, \xi_{e_{j'}} \mathbf{I}_{K_{e_{j'}}})$ respectively.

Estimation

Let $\boldsymbol{\lambda} = (\boldsymbol{\lambda}'_u, \boldsymbol{\lambda}'_e)'$ and $\boldsymbol{\xi} = (\boldsymbol{\xi}'_u, \boldsymbol{\xi}'_e)'$ where $\xi_u = \{\xi_{u_j}\}$ and $\xi_e = \{\xi_{e_{j'}}\}$. Inference about $\boldsymbol{\lambda}$ is of an empirical Bayes type and is based on the mode $\widehat{\boldsymbol{\lambda}}$ of the posterior density $p(\boldsymbol{\lambda}|\mathbf{y}, \boldsymbol{\xi} = \widehat{\boldsymbol{\xi}})$ given $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}$ its marginal maximum likelihood estimator, *ie*

$$\widehat{\boldsymbol{\lambda}} = \text{Argmax} \ln p(\boldsymbol{\lambda}|\mathbf{y}, \boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}) \quad [26a]$$

$$\widehat{\boldsymbol{\xi}} = \text{Argmax} \ln p(\mathbf{y}|\boldsymbol{\xi}) \quad [26b]$$

Maximization in [26ab] can be carried out according to the procedure described by Foulley *et al* (1992) and San Cristobal *et al* (1993). The algorithm for computing $\widehat{\boldsymbol{\lambda}}$ can be written as (from iteration t to $t + 1$)

$$(\mathbf{L}'\mathbf{W}^{[t]}\mathbf{L} + \boldsymbol{\Gamma}^-)(\boldsymbol{\lambda}^{[t+1]} - \boldsymbol{\lambda}^{[t]}) = \mathbf{L}'\mathbf{z}^{[t]} - \boldsymbol{\Gamma}^-\boldsymbol{\lambda}^{[t]} \quad [27]$$

where

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_u & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_e \end{bmatrix}$$

$$\boldsymbol{\Gamma}^- = \bigotimes_{\alpha=u,e} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_\alpha^{-1} \end{bmatrix} \text{ with } \boldsymbol{\Lambda}_u = \bigotimes_j \xi_{u_j} \mathbf{I}_{K_{u_j}}, \boldsymbol{\Lambda}_e = \bigotimes_{j'} \xi_{e_{j'}} \mathbf{I}_{K_{e_{j'}}}$$

$\mathbf{z} = (\mathbf{z}'_u, \mathbf{z}'_e)$ are working variables updated at each iteration and such that

$$\mathbf{z}_u = (1/2)\text{E}(\sigma_{u_i} \sigma_{e_i}^{-2} \mathbf{u}^* \mathbf{Z}'_i \mathbf{e}_i | \mathbf{y}, \boldsymbol{\lambda})$$

$$\mathbf{z}_e = (1/2)\text{E}[(\sigma_{e_i}^{-2} \mathbf{e}'_i \mathbf{e}_i) - n_i | \mathbf{y}, \boldsymbol{\lambda}]$$

$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{uu} & \mathbf{W}_{ue} \\ \mathbf{W}_{eu} & \mathbf{W}_{ee} \end{bmatrix}$ is a (2I, 2I) matrix of weights described in Foulley *et al* (1990, 1992) for the environmental variance part, and in San Cristobal *et al* (1993) for the general case.

$\widehat{\xi}_{u_j}$ and $\widehat{\xi}_{e_{j'}}$ can be computed as usual in Gaussian model methodology *via* the EM algorithm

$$\xi_{u_j}^{[n+1]} = \text{E}(\mathbf{v}'_{u_j} \mathbf{v}_{u_j} | \mathbf{y}, \boldsymbol{\xi} = \boldsymbol{\xi}^{[n]}) / K_{u_j} = [\widehat{\mathbf{v}}_{u_j}^{[n]'} \widehat{\mathbf{v}}_{u_j}^{[n]} + \text{tr}(\mathbf{H}_{v_j}^{[n]})] / K_{u_j} \quad [28a]$$

$$\xi_{e_{j'}}^{[n+1]} = \text{E}(\mathbf{v}'_{e_{j'}} \mathbf{v}_{e_{j'}} | \mathbf{y}, \boldsymbol{\xi} = \boldsymbol{\xi}^{[n]}) / K_{e_{j'}} = [\widehat{\mathbf{v}}_{e_{j'}}^{[n]'} \widehat{\mathbf{v}}_{e_{j'}}^{[n]} + \text{tr}(\mathbf{H}_{e_{j'}}^{[n]})] / K_{e_{j'}} \quad [28b]$$

where $\widehat{\mathbf{v}}_{u_j}^{[n]}$, $\widehat{\mathbf{v}}_{e_{j'}}^{[n]}$ are solutions of [26] for $\xi = \xi^{[n]}$, and $\mathbf{H}_{v_j v_j}^{[n]}$, $\mathbf{H}_{e_{j'} e_{j'}}^{[n]}$, are blocks of the inverse of the coefficient matrix of [27] pertaining to \mathbf{v}_{u_j} , $\mathbf{v}_{e_{j'}}$ respectively.

Testing procedure

The procedure described previously reduces to REML estimation of λ when flat priors are assumed for \mathbf{v}_u and \mathbf{v}_e , or equivalently when the structural model for log-variances is a purely fixed model. This property allows derivation of testing procedures to identify significant sources of heteroskedasticity. This can be achieved via the likelihood ratio test (LRT) as proposed by Foulley *et al* (1990, 1992) and San Cristobal *et al* (1993) but in using the residual (marginal) log-likelihood function $L(\lambda; \mathbf{y})$ instead of the full log-likelihood $L(\lambda, \beta; \mathbf{y})$.

Let $H_0: \lambda \in A_0$ be the null hypothesis and $H_1: \lambda \in A - A_0$ its alternative, where A refers to the complete parameter space and A_0 , a subset of it, pertains to H_0 . Under H_0 , the statistic

$$\phi(\mathbf{y}) = -2[\text{Max}_{A_0} L(\lambda; \mathbf{y}) - \text{Max}_A L(\lambda; \mathbf{y})] \tag{29}$$

has an asymptotic chi-square distribution with r degrees of freedom; r is the difference in the number of estimable parameters involved in A and A_0 , eg, $r = \text{rank}(\mathbf{K})$ when H_0 is defined by $\mathbf{K}'\lambda = \mathbf{0}$, \mathbf{K}' being a full row rank matrix.

Fortunately, for a Gaussian linear model, such as the one considered here $\mathbf{y}_i \sim \mathbf{N}(\mathbf{X}_i \beta, \sigma_{u_i}^2 \mathbf{Z}'_i \mathbf{A} \mathbf{Z}_i + \sigma_{e_i}^2 \mathbf{I}_{n_i})$, $L(\lambda; \mathbf{y})$, can be easily expressed as

$$-2L(\lambda; \mathbf{y}) = (N - p) \ln 2\pi + \ln |\mathbf{A}| + \ln \left| \sum_{i=1}^I \sigma_{e_i}^{-2} \mathbf{T}'_i \mathbf{T}_i + \Sigma^- \right| + \sum_{i=1}^I \sigma_{e_i}^{-2} \mathbf{y}'_i (\mathbf{y}_i - \mathbf{T}_i \widehat{\theta}) \tag{30}$$

where $N = \sum_{i=1}^I n_i$, $p = \text{rank}(\mathbf{X})$, $\mathbf{T}_i = (\mathbf{X}_i, \sigma_{u_i} \mathbf{Z}_i)$, $\Sigma^- = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{A}^{-1} \end{bmatrix}$, and

$\widehat{\theta} = (\widehat{\beta}', \widehat{u}^{*'})$ is a solution of the system of mixed-model equations in [9], ie

$$\left[\sum_{i=1}^I \sigma_{e_i}^{-2} \mathbf{T}'_i \mathbf{T}_i + \Sigma^- \right] \widehat{\theta} = \sum_{i=1}^I \sigma_{e_i}^{-2} \mathbf{T}'_i \mathbf{y}_i.$$

CONCLUSION

This paper is an attempt to synthesize the current state of research in the field of statistical analysis of heterogeneous variances arising in mixed-model methodology and in its application to animal breeding. For pedagogical reasons, the paper successively addressed the cases of one-way and multiple-way classification. In any case, the estimation procedures chosen were REML and its natural Bayesian extension, the posterior mode using inverted gamma prior distributions. For a single classification, simple formulae for computing REML and Bayes estimations were derived using the theory of EM. Emphasis was placed on this algorithm but other alternatives could be considered, eg, ECME (Liu and Rubin, 1994), AI-REML (Johnson and Thompson, 1994) and DF-REML (Meyer, 1989).

For multiple classifications, the key idea underlying the structural approach is to render models as parsimonious as possible, which is especially critical with the large data sets used in genetic evaluation having a large number of subclasses. Consequently, it was shown how heterogeneity of log-residual and u -components of variance can be described with a mixed linear model structure. This makes the corresponding estimation procedures a natural extension of what has been done for decades by BLUP techniques on subclass means.

An important feature of this procedure is its ability to assess *via* likelihood ratio tests the effects of potential sources of heterogeneity considered either marginally or jointly. Although computationally demanding, these procedures have begun to be applied. San Cristobal *et al* (1993) applied these procedures to the analysis of muscle development scores at weaning of 8 575 progeny of 142 sires in the Maine Anjou cattle breed where heroskedasticity was found both for the sire and residual components. Weigel *et al* (1993) and DeStefano (1994) also used the structural approach for sire and residual variances to assess sources of heterogeneity in within-herd variances of milk and fat records in Holstein. Herd size and within-herd means were associated with significant increases in residual variances as well as various management factors (*eg*, milking system). Approximations for the to estimation of within region-herd-year-parity phenotypic variances were also proposed for dairy cattle evaluation by Wiggans and Van Raden (1991) and Weigel and Gianola (1993). These techniques also open new prospects of research in different fields of quantitative genetics, *eg*, analyses of genotype \times environment interactions (Foulley *et al*, 1994; Robert *et al*, 1994), testing procedures for genetic parameters (Robert *et al*, 1995ab), crossbreeding experiments and QTL detection. However, research is still needed to improve the methodology and efficiency of algorithms.

ACKNOWLEDGMENTS

This paper draws from material presented at the 5th World Congress of Genetics Applied to Livestock Production, Guelph, August 7–12, 1994. Special thanks are expressed to AL DeStefano, V Ducrocq, RL Fernando, D Gianola, D Hébert, CR Henderson, S Im, C Robert, M Gaudy-San Cristobal and K Weigel for their valuable contribution to the discussion and to the publication of papers on this subject. The assistance of C Robert in the computations of the numerical example is also greatly acknowledged.

REFERENCES

- Aitkin M (1987) Modelling variance heterogeneity in normal regression using GLIM. *Appl Stat* 36, 332-339
- Cantet RJC (1990) Estimation and prediction problems in mixed linear models for maternal genetic effects. PhD thesis, University of Illinois, Urbana, IL, USA
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data *via* the EM algorithm. *J R Stat Soc B* 39, 1-38
- Denis JB (1983) Extension du modèle additif d'analyse de variance par modélisation multiplicative des variances. *Biometrics* 39, 849-856
- DeStefano AL (1994) Identifying and quantifying sources of heterogeneous residual and sire variances in dairy production data. PhD thesis, Cornell University, Ithaca, NY, USA

- Foulley JL (1993) A simple argument showing how to derive restricted maximum likelihood. *J Dairy Sci* 76, 2320-2324
- Foulley JL, Gianola D, San Cristobal M, Im S (1990) A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models. *J Dairy Sci* 73, 1612-1624
- Foulley JL, San Cristobal M, Gianola D, Im S (1992) Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Comput Stat Data Anal* 13, 291-305
- Foulley JL, Hébert D, Quaas RL (1994) Inferences on homogeneity of between-family components of variance and covariance among environments in balanced cross-classified designs. *Genet Sel Evol* 26, 117-136
- Garrick DJ, Van Vleck LD (1987) Aspects of selection for performance in several environments with heterogeneous variances. *J Anim Sci* 65, 409-421
- Garrick DJ, Pollak EJ, Quaas RL, Van Vleck LD (1989) Variance heterogeneity in direct and maternal weight traits by sex and percent purebred for Simmental sired calves. *J Anim Sci* 67, 2515-2528
- Gianola D (1986) On selection criteria and estimation of parameters when the variance is heterogeneous. *Theor Appl Genet* 72, 671-677
- Gianola D, Foulley JL, Fernando RL, Henderson CR, Weigel KA (1992) Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. *J Dairy Sci* 75, 2805-2823
- Graybill FA (1969) *Introduction to Matrices with Applications to Statistics*. Wadsworth Publishing Co, Belmont, CA, USA
- Harville DA (1977) Maximum likelihood approaches to variance component estimation and related problems. *J Am Stat Assoc* 72, 320-340
- Henderson CR (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph, Ontario, Canada.
- Hill WG (1984) On selection among groups with heterogeneous variance. *Anim Prod* 39, 473-477
- Johnson DL, Thompson R (1994) Restricted maximum likelihood estimation of variance component for univariate animal models using sparse matrix techniques and a quasi-Newton procedure. In: *5th World Congress on Genetics Applied to Livestock Production* (C Smith *et al*, eds), University of Guelph, Guelph, ON, Canada, vol 18, 410-413
- Leonard T (1975) A Bayesian approach to the linear model with unequal variances. *Technometrics* 17, 95-102
- Liu C, Rubin DB (1994) Application of the ECME algorithm and Gibbs sampler to general linear mixed model. In: *Proc XVIIth International Biometric Conference*, McMaster University, Hamilton, ON, Canada, vol 1, 97-107
- Meyer K (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet Sel Evol* 21, 317-340
- Nair VN, Pregon D (1988) Analysing dispersion effects from replicated factorial experiments. *Technometrics* 30, 247-257
- Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545-554
- Robert C, Foulley JL, Ducrocq V (1994) Estimation and testing of a constant genetic correlation among environments. In: *5th World Congress on Genetics Applied to Livestock Production* (C Smith *et al*, eds), University of Guelph, Guelph, ON, Canada, vol 18, 402-405
- Robert C, Foulley JL, Ducrocq V (1995a) Genetic variation of traits measured in several environments. I. Estimation and testing of homogeneous genetic and intra-class correlations between environments. *Genet Sel Evol* 27, 111-123

- Robert C, Foulley JL, Ducrocq V (1995b) Genetic variation of traits measured in several environments. II. Inference on between-environments homogeneity of intra-class correlations. *Genet Sel Evol* 27, 125-134
- San Cristobal M, Foulley JL, Manfredi E (1993) Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genet Sel Evol* 25, 3-30
- Searle SR, Casella G, McCulloch CE (1992) *Variance Components*. J Wiley and Sons, New York, USA
- Thompson R, Meyer K (1986) Estimation of variance components: what is missing in the EM algorithm? *J Stat Comput Simul* 24, 215-230
- Visscher PM, Hill WG (1992) Heterogeneity of variance and dairy cattle breeding. *Anim Prod* 55, 321-329
- Visscher PM, Thompson R, Hill WG (1991) Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. *Livest Prod Sci* 28, 273-290
- Weigel KA, Gianola D (1993) A computationally simple Bayesian method for estimation of heterogeneous within-herd phenotypic variances. *J Dairy Sci* 76, 1455-1465
- Weigel KA, Gianola D, Yandel BS, Keown JF (1993) Identification of factors causing heterogeneous within-herd variance components using a structural model for variances. *J Dairy Sci* 76, 1466-1478
- Wiggans GR, VanRaden PM (1991) Method and effect of adjustment for heterogeneous variance. *J Dairy Sci* 74, 4350-4357

APPENDIX

A proof that equation [7] is non-negative in the homoskedastic case is as follows. The numerator of [7] can be written as:

$$\begin{aligned} E(\boldsymbol{\varepsilon}'\mathbf{Zu}|\mathbf{K}'\mathbf{y}, \sigma_u, \sigma_e^2) &= \hat{\boldsymbol{\varepsilon}}'\mathbf{Z}\hat{\mathbf{u}} + \text{tr}\{\text{Cov}[\mathbf{Z}(\hat{\mathbf{u}} - \mathbf{u}), (\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})']\} \\ &= \hat{\boldsymbol{\varepsilon}}'\mathbf{Z}\hat{\mathbf{u}} + \text{tr}[\text{Cov}(\mathbf{Zu}, \boldsymbol{\varepsilon}') - \text{Cov}(\mathbf{Z}\hat{\mathbf{u}}, \hat{\boldsymbol{\varepsilon}}')] \end{aligned}$$

where $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{V}\mathbf{P}\mathbf{y}$, and $\hat{\mathbf{u}} = \sigma_u\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \sigma_u\mathbf{Z}'\mathbf{P}\mathbf{y}$ with $\mathbf{V} = \text{Var}(\mathbf{y})$, $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, and $\mathbf{K}'\mathbf{y}$, an $N - \text{rank}(\mathbf{X})$ vector of linear contrasts.

Now

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}'\mathbf{Z}\hat{\mathbf{u}} &= (\mathbf{y}'\mathbf{P}\mathbf{V})(\sigma_u\mathbf{Z}\mathbf{Z}'\mathbf{P}\mathbf{y}) \\ \text{Cov}(\mathbf{Zu}, \boldsymbol{\varepsilon}') &= \sigma_u\mathbf{Z}\mathbf{Z}' = \sigma_u\mathbf{Z}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{V} \\ \text{Cov}(\mathbf{Z}\hat{\mathbf{u}}, \hat{\boldsymbol{\varepsilon}}') &= (\sigma_u\mathbf{Z}\mathbf{Z}'\mathbf{P})\mathbf{V}(\mathbf{P}\mathbf{V}) = \sigma_u\mathbf{Z}\mathbf{Z}'\mathbf{P}\mathbf{V} \end{aligned}$$

so that

$$\text{tr}[\text{Cov}(\mathbf{Zu}, \boldsymbol{\varepsilon}') - \text{Cov}(\mathbf{Z}\hat{\mathbf{u}}, \hat{\boldsymbol{\varepsilon}}')] = \sigma_u \text{tr}[\mathbf{Z}\mathbf{Z}'(\mathbf{V}^{-1} - \mathbf{P})\mathbf{V}]$$

Moreover $\mathbf{V}^{-1} - \mathbf{P} = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$ is non-negative definite (nnd). For a one-way random factor

$$\begin{aligned} \mathbf{V} &= \sigma_u^2\mathbf{Z}\mathbf{Z}' + \sigma_e^2\mathbf{I} \\ \mathbf{V}\mathbf{Z}\mathbf{Z}' &= \sigma_u^2(\mathbf{Z}\mathbf{Z}')^2 + \sigma_e^2\mathbf{Z}\mathbf{Z}' \text{ nnd} \end{aligned}$$

then

$$\widehat{\boldsymbol{\varepsilon}}' \mathbf{Z} \widehat{\mathbf{u}} = \sigma_u (\mathbf{y}' \mathbf{P}) \mathbf{V} \mathbf{Z} \mathbf{Z}' (\mathbf{P} \mathbf{y}) \geq 0$$

$$\text{tr}[\mathbf{Z} \mathbf{Z}' (\mathbf{V}^{-1} - \mathbf{P}) \mathbf{V}] = \text{tr}[\mathbf{V} \mathbf{Z} \mathbf{Z}' (\mathbf{V}^{-1} - \mathbf{P})] \geq 0 \text{ because if } \mathbf{A}, \mathbf{B} \text{ are } (n \times n)$$

nnd matrices, then $\text{tr}(\mathbf{A}\mathbf{B}) \geq 0$ (Graybill, 1969).

Hence,

$$\begin{aligned} & \text{E}[\boldsymbol{\varepsilon}' \mathbf{Z} \mathbf{u} | \mathbf{K}' \mathbf{y}, \sigma_u, \sigma_e^2] \geq 0 \\ & \text{viz, if } \sigma_u^{[t]} \geq 0, \text{ and } \sigma_e^{[t]} > 0, \text{ then } \sigma_u^{[t+1]} \geq 0 \end{aligned}$$

COMMENT

Robin Thompson

Roslin Institute, Roslin, Midlothian EH25 9PS, UK

This paper is a synthesis of recent work of the authors and their co-workers in the area of heterogeneous variances. I think it is a valuable review giving a logical presentation showing how heterogeneous variance modelling can be carried out. The parallels between estimation of linear parameters and variance parameters is highlighted. My comments relate to transformation, convergence, simplicity and utility.

Transformation

Emphasis is on the genetic standard deviation as a parameter rather than the genetic variance. Such a parameterization has been used to allow estimation of binary variance components (Anderson and Aitkin, 1985). Using scaled variables can also allow reduction of the dimension of search in derivative-free methods of multivariate estimation (Thompson *et al*, 1994). It is also a special case of the Choleski transformation, which has been suggested to speed up estimation and keep parameters within bounds for repeated measures data (Lindstrom and Bates, 1988) and multivariate genetic data (Groeneveld, 1994).

Reverter *et al* (1994) have recently suggested regression type methods for estimation of variance parameters. As the authors point out, there is a natural regression interpretation to the similar equations [7] and [8]. However [7] and [8] include trace terms that essentially correct for attenuation or uncertainty in knowing the fixed or random effects.

Convergence

In the discussion of Dempster *et al* (1977) I pointed out that the rate of convergence for a balanced one-way analysis is (in the authors' notation) approximately $1 - R^2$, which, I think, explains one of the graphs in figure 1. In the time available, I have not been able to derive the rate of convergence for the scheme based on [7] and [8], but if σ_e^2 is known and [7] is used to estimate $\sigma_{u_i}^2$, which should be a good approximation, then the rate of convergence is $1 - 2R^2(1 - R^2)$. This is in good agreement with figure 1, suggesting symmetry about $R^2 = 0.5$ and equality of the speed of convergence when $R^2 = 2/3$.

As someone who has never understood the EM algorithm or its popularity, I would have thought schemes based on some form of second differentials would be more useful, especially as some of the authors' schemes allow negative standard deviations.

Simplicity

Whilst very elegant, I was wondering if there was a simple analysis, perhaps graphical, and based on score tests based on a homogeneous analysis that would

highlight the need for a heterogeneous analysis, both for fitting the data and measuring the possible loss of response.

Utility

I can see the use of these methods at a phenotypic level, but I am less clear if it is realistic to attempt to detect differences at a genetic level (Visscher, 1992). At the simplest level if the residuals are heterogeneous can one realistically discriminate between models with homogeneous genetic variances or homogeneous heritabilities?

Additional references

- Anderson DA, Aitkin MA (1985) Variance component models with binary response: interviewer variability. *J R Stat Soc B* 47, 203-210
- Groeneveld E (1994) A reparameterization to improve numerical optimization in multivariate REML (co)variance component estimation. *Genet Sel Evol* 26, 537-545
- Lindstrom MJ, Bates DM (1988) Newton-Raphson and EM algorithms for linear mixed-effects models for repeated measures data. *J Am Stat Assoc* 83, 1014-1022
- Thompson R, Crump RE, Juga J, Visscher PM (1995) Estimating variances and covariances for bivariate animal models using scaling and transformation. *Genet Sel Evol* 27, 33-42
- Reverter A, Golden BL, Bourdon RM, Brinks JS (1994) Method \mathfrak{R} variance components procedure: application on the simple breeding value model. *J Anim Sci* 72, 2247-2253
- Visscher PM (1992) Power of likelihood ratio tests for heterogeneity of intraclass correlation and variance in balanced half-sib designs. *J Dairy Sci* 75, 1320-1330

COMMENT

Daniel Gianola

Department of Meat and Animal Science, University of Wisconsin-Madison, USA

This paper reviews and extends previous work done by the senior author and collaborators in the area of inferences about heterogeneous variance components in animal breeding. The authors impose a model on the variance structure, and suggest estimation and testing procedures, so as to arrive to explanatory constructs that have as few parameters as possible. Their approach is systematic, in contrast to other suggestions available in the literature (some of which are referenced by Foulley and Quaas), where a model on the variances is adopted implicitly without reference to the amount of support that is provided by the data at hand, and often on the basis of mechanistic or *ad hoc* considerations. In this front, their developments are welcome.

They adopt either a likelihood or a Bayesian viewpoint, and restrict themselves to conducting a search of the appropriate maximizers (REML or posterior modes, respectively) *via* the EM algorithm, which is implemented in a clear, straightforward way. In so doing, they arrive at a formulation ('scaled' EM) which exhibits a different numerical behavior from that of the 'standard' EM in a simple model. In their Bayesian version, they employ scaled inverted chi-square priors, and arrive

at estimating equations that are similar to those of Gianola *et al* (1992), cited by the authors. It is now well known that setting all degree of belief parameters to zero leads to an improper posterior distribution; surprisingly, the authors (as well as Gianola *et al*, 1992) do not alert the readers about this pitfall.

The authors do not provide measures of the curvature of the likelihood or of the pertinent posterior distribution in the neighborhood of the maximum. It is useful to recover second derivatives either to implement fully the maximum likelihood analysis, or to approximate the joint (or marginal) posterior with a Gaussian distribution. Gianola *et al* (1992) gave expressions for second differentials for some simple heteroskedastic models, and an extension of these would have been an interesting contribution.

Their testing procedure relies on asymptotic properties of the (marginal) likelihood ratio test. I wonder how accurate this test is in situations where a heteroskedastic model of high dimensionality may be warranted. In this situation, the asymptotic distribution of the test criterion may differ drastically from the exact sampling distribution. It is somewhat surprising that the authors do not discuss Bayesian test and associated methods for assessing uncertainty; some readers may develop the false impression that there is a theoretical vacuum in this domain (see, for example, Robert, 1992).

There is a lot more information in a posterior (or normalized likelihood) than that contained in first and second differentials. In this respect, an implementation based on Monte-Carlo Markov chain (MCMC) methods such as the Gibbs sampler (*eg*, Tanner, 1993) can be used to estimate the whole set of posterior distributions in the presence of heterogeneous variances. In some simple heteroskedastic linear models it can be shown that the random walk involves simple chains of normal and inverted chi-square distributions. Further, it is possible to arrive at the exact (within the limits of the Monte-Carlo error) posterior distributions of linear and nonlinear functions of fixed and random effects. In the sampling theory framework one encounters immediately a Behrens–Fisher problem, even in a simple contrast between ‘treatments’. The Bayesian approach *via* MCMC would allow an exact analysis of, for example, breed comparison experiments, when the sources of germplasm involved have heterogeneous and unknown dispersion.

I have been intrigued for some time about the possible consequences of heterogeneous covariance matrices in animal evaluation in a multiple-trait analysis. If there is heterogeneous variance there must be heterogeneity in covariances as well! Perhaps the consequences on animal ranking are even more subtle than in the univariate case. It is not obvious how the structural model approach can be generalized here, and this is a challenge for future research.

Additional References

- Robert C (1992) *L'Analyse Statistique Bayésienne*. Economica, Paris, France
Tanner MA (1993) *Tools for Statistical Inference*. Springer-Verlag, New York, USA