

# Marginal maximum likelihood estimation of variance components in Poisson mixed models using Laplacian integration

RJ Tempelman<sup>1</sup>, D Gianola<sup>2</sup>

<sup>1</sup> *Louisiana State University, Department of Agricultural Statistics, 53 Agricultural Administration Building, Baton Rouge, LA 70803-5606*

<sup>2</sup> *University of Wisconsin, Department of Dairy Science, 266 Animal Sciences Building, Madison, WI 53706, USA*

(Received 10 November 1992; accepted 18 May 1993)

**Summary** – An algorithm for computing marginal maximum likelihood (MML) estimates of variance components in Poisson mixed models is presented. A Laplacian approximation is used to integrate fixed and random effects out of the joint posterior density of all parameters. This approximation is found to be identical to that invoked in the more commonly used expectation–maximization type algorithm for MML. Numerically, however, a different sequence of iterates is obtained, although the same variance component estimates should result. The Laplacian algorithm is precisely DFREML (derivative free REML) optimization when applied to normally distributed data, and could then be termed DFMML (derivative-free marginal maximum likelihood). Because DFMML is based on an approximation to the marginal likelihood of the variance components, it provides a mechanism for testing hypotheses about such components *via* posterior odds ratios or marginal likelihood ratio tests. Also, asymptotic posterior standard errors of the variance components can be computed with DFMML. A Tierney-Kadane procedure for computing the posterior mean of a variance component is also developed; however, it requires 2 joint maximizations, and consequently may not be expected to perform well in many linear and non-linear mixed models. An example of a Poisson model is presented in which the null estimate commonly found when jointly estimating variance components with fixed and random effects is observed; thus, the Tierney-Kadane procedure for computing the posterior mean failed. On the other hand, the Laplacian method succeeded in locating the mode of the marginal distribution of the variance component in a Bayesian model with flat priors for fixed effects and variance components; that is, the MML estimate.

generalized linear model / marginal maximum likelihood / variance component / mixed model / Laplacian estimation

**Résumé** – Estimation des composantes de variance par le maximum de vraisemblance marginale dans des modèles mixtes de Poisson à l'aide de la méthode d'intégration de Laplace. Un algorithme de calcul des estimées de composantes de variance par le maximum de vraisemblance marginale dans des modèles mixtes de Poisson est présenté. On utilise une approximation de Laplace pour éliminer par intégration les effets fixés et aléatoires de la densité conjointe a posteriori de tous les paramètres. Cette approximation se montre identique à celle à laquelle il est fait appel dans l'algorithme plus classique du type espérance-maximisation. Du point de vue numérique cependant, la séquence des valeurs obtenues par itération est différente, bien que les mêmes estimées de composantes doivent être obtenues. L'algorithme de Laplace est précisément l'optimisation de DFREML (maximum de vraisemblance restreinte sans dérivée) quand on l'applique à des données distribuées normalement, et pourrait être appelé DFML (maximum de vraisemblance marginale sans dérivée). Parce que DFML est basé sur une approximation de la vraisemblance marginale des composantes de la variance, il fournit un moyen de tester des hypothèses relatives à de telles composantes via des rapports de probabilités a posteriori ou des tests de rapport de vraisemblance. De plus, des valeurs asymptotiques a posteriori des composantes de variance peuvent être calculées au moyen de DFML. Une procédure de Tierney-Kadane pour calculer la moyenne a posteriori d'une composante de variance est également présentée; elle requiert cependant 2 maximisations conjointes et, en conséquence, on ne doit pas s'attendre à ce qu'elle donne de bons résultats dans beaucoup de modèles linéaires et non linéaires. Un exemple de modèle de Poisson est donné, dans lequel on obtient les valeurs nulles habituellement trouvées quand on estime conjointement des composantes de variance avec des effets fixés et aléatoires; ainsi, la procédure de Tierney-Kadane pour calculer la moyenne a posteriori échoue. En revanche, la méthode de Laplace réussit à localiser le mode de la distribution marginale des composantes de variance dans un modèle bayésien avec des a priori uniformes pour les effets fixés et les composantes de variance, ie l'estimée du maximum de vraisemblance marginale.

**composantes de variance / distribution de Poisson / modèle linéaires généralisé / maximum de vraisemblance marginale / intégration de Laplace**

## INTRODUCTION

Non-linear models for quantitative genetic analysis of categorically scored phenotypes have been developed in recent years (Gianola and Foulley, 1983; Harville and Mee, 1984). In these models, it is assumed that the observed polychotomies correspond to realizations of an underlying normal variate inside intervals of the real line that are delimited by fixed thresholds. The mathematical link between the underlying and the discrete scales is, thus, the probit function. Although threshold models have been used for analysis of different types of discrete data (*eg*, Meijering, 1985; Weller *et al*, 1988; Weller and Gianola, 1989; Manfredi *et al*, 1991) counted variates are probably better modelled using Poisson or related distributions such as the negative binomial distribution. Non-linear Poisson models for counted variates, *eg* litter size in swine and sheep, have been suggested by Foulley *et al* (1987), and an application to prolificacy in the Iberian pig is given by Pérez-Enciso *et al* (1993).

The model of Foulley *et al* (1987) requires knowledge of variance components, so these must be estimated somehow. Animal breeders have used restricted maximum likelihood (REML) to estimate genetic variances for a wide array of economically

important traits. This is not entirely satisfactory for discrete characters because REML relies on the assumption of multivariate normality; the degree of robustness of this method to departures from normality has not been sufficiently studied. Further, unless there is a large amount of statistical information in the data about the variance parameters, the sampling performance of REML when applied to discrete traits may be unsatisfactory, as suggested by the simulation study of Tempelman and Gianola (1991).

The procedure for estimating variance components suggested by Foulley *et al* (1987) in their Poisson model is marginal maximum likelihood (MML). In a Bayesian context with flat priors for variances and fixed effects, this method gives as point estimates the components of the mode of the marginal posterior distribution of all variance components (Foulley *et al*, 1990). With normal data, MML is identical to REML. With discrete traits, such as in the Poisson model of Foulley *et al* (1987), approximations to MML must be used, because the exact integration of nuisance parameters (fixed and random effects) out of the joint posterior distribution is onerous. In Foulley *et al* (1987), the posterior distribution of fixed and random effects, given the variance component, is approximated by a multivariate normal process when computing MML estimates.

The objective of this paper is to describe another approximation to marginal maximum likelihood estimation of variance components in a Poisson mixed model based on Laplace's method of integration, as suggested by Leonard (1982) for calculating posterior modes, and by Tierney and Kadane (1986) for computing posterior means. A model with a single variance component is considered in the present study, and the relationship of Laplacian integration to derivative-free methods for computing REML with normal data is highlighted. A numerical example is presented.

## THE POISSON MIXED MODEL

Foulley *et al* (1987) employ a Bayesian approach to make inferences in a Poisson mixed model. Given a location parameter vector,  $\theta$ , the conditional distribution  $f(\cdot)$  of a counted variate  $y_i$  is assumed to be Poisson.

$$f(y^i | \theta) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, y_i = 0, 1, \dots; \quad i = 1, \dots, n \quad [1]$$

where  $e$  denotes the natural exponent,  $n$  is the number of observations, and  $\lambda_i$  is the Poisson parameter for observation  $i$ . By definition, the Poisson parameter must be positive; however, the transformation  $\eta_i = \ln \lambda_i$ , defined as the canonical link function for Poisson variables (McCullagh and Nelder, 1989), can take any value on the real line. Foulley *et al* (1987) introduce the linear relationship

$$\eta_i = \mathbf{w}'_i \theta = \mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{u} \quad [2]$$

where  $\theta' = [\boldsymbol{\beta}', \mathbf{u}']$ , and  $\mathbf{w}'_i = [\mathbf{x}'_i, \mathbf{z}'_i]$  is the  $i$ th row of the  $n \times (p + q)$  incidence matrix  $\mathbf{W} = [\mathbf{X}, \mathbf{Z}]$ .  $\mathbf{X}$  and  $\mathbf{Z}$  are known incidence matrices of dimensions  $n \times p$  and  $n \times q$ , respectively, that associate the location vectors  $\boldsymbol{\beta}_{p \times 1}$  and  $\mathbf{u}_{q \times 1}$  to each

observation. Under the Poisson model, the mean and variance of an observation, given  $\theta$ , is equal to the Poisson parameter  $\lambda_i$ . Hence, the residual variance in this model is precisely  $\lambda_i$ .

The vectors  $\beta$  and  $u$  are distinct in the following sense. Typically, the elements of  $\beta$  pertain to levels of fixed effects such as herd, year and season, whereas those of the vector  $u$  pertain to "random" effects of the animals being recorded and of their known relatives. In a Bayesian context, a flat prior density is assigned to  $\beta$  and a multivariate normal prior distribution is assumed for  $u$  (Foulley *et al*, 1987). If  $u$  is a vector of breeding values,

$$u \mid \Sigma_u \sim N(0, \Sigma_u) \quad \text{with } \Sigma_u = A\sigma_u^2 \quad [3]$$

Above,  $A$  is a matrix of additive relationships, and  $\sigma_u^2$  is the additive genetic variance.

If the dispersion parameter  $\sigma_u^2$  is unknown, it can be estimated from its marginal posterior distribution so as to provide a parametric empirical Bayes approach to joint estimation of  $\beta$  and  $u$ . When the prior density assigned to  $\sigma_u^2$  is flat, then the mode of the marginal posterior distribution of  $\sigma_u^2$  is identical to the maximum of the marginal likelihood of  $\sigma_u^2$ .

The unknown parameters are thus  $\beta$ ,  $u$ , and  $\sigma_u^2$ . In animal breeding applications, often  $p + q > n$ . For example, in 'animal' models with a single observation per recorded individual, the dimension of  $u$  is often greater than the number of observations, that is,  $q \geq n$ . This leads to a highly parameterized model. When the elements of  $u$  are strongly intercorrelated, a potentially low degree of orthogonality can seriously slow down convergence of Monte Carlo Markov Chain methods, such as Gibbs sampling, as a means of estimating marginal densities, modes, or means (Smith, 1991). Under these conditions, approximating the marginal density of  $\sigma_u^2$  by Laplacian integration procedures may be attractive from a numerical point of view.

## ESTIMATION OF THE VARIANCE COMPONENT FROM THE MODE OF ITS MARGINAL POSTERIOR DISTRIBUTION

We first assume that, conditionally on  $\theta$ , the observations are independent, following a Poisson distribution as in [1]. Let  $\vartheta = [\theta', \sigma_u^2]' = [\beta', u', \sigma_u^2]'$  represent all parameters of interest. Assigning a flat prior to the variance component  $\sigma_u^2$  and to  $\beta$ , such that the joint prior density of  $\vartheta$  is proportional to that of  $u$ , we can write the log of the joint posterior density of  $\beta$ ,  $u$ , and  $\sigma_u^2$  as

$$L(\vartheta) = \sum_{i=1}^n \log [f(y_i \mid \theta)] + \log \pi(u \mid \Sigma_u) + \text{constant} \quad [4]$$

where  $\pi(u \mid \Sigma_u)$  is a multivariate normal density function. Further,

$$L(\vartheta) = \text{constant} + \sum_{i=1}^n (y_i \ln \lambda_i - \lambda_i) - \frac{1}{2} u' \Sigma_u^{-1} u - \frac{1}{2} \log |\Sigma_u| \quad [5]$$

Because  $|\Sigma_{\mathbf{u}}| = |\mathbf{A}\sigma_u^2| = |\mathbf{A}|(\sigma_u^2)^q$ , and  $\mathbf{A}$  does not depend on the parameters, it follows that [5] is expressible as:

$$L(\boldsymbol{\theta}) = \text{constant} + \sum_{i=1}^n (y_i \ln \lambda_i - \lambda_i) - \frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}{2\sigma_u^2} - \frac{q}{2} \log(\sigma_u^2) \tag{6}$$

The joint posterior density of the full parameter set can be written as:

$$p(\boldsymbol{\theta} | \mathbf{y}) = p(\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}) \cdot p(\sigma_u^2 | \mathbf{y}) \tag{7}$$

where  $p(\boldsymbol{\theta} | \sigma_u^2, \mathbf{y})$  is the posterior density of  $\boldsymbol{\theta}$ , given that the variance components are known, and  $p(\sigma_u^2 | \mathbf{y})$  is the marginal density of the variance parameter.

Define:

$$\tilde{\boldsymbol{\theta}}_{\sigma} = \begin{bmatrix} \tilde{\boldsymbol{\beta}}_{\sigma} \\ \tilde{\mathbf{u}}_{\sigma} \end{bmatrix} = \text{Arg}_{\boldsymbol{\theta}} \max p(\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}) \tag{8a}$$

to be the mode of the joint posterior density of  $\boldsymbol{\theta}$ , given  $\sigma_u^2$ , and

$$\tilde{\mathbf{H}}_{\sigma} = \left[ \frac{-\partial^2 \log p(\boldsymbol{\theta} | \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_{\sigma}} = \left[ \frac{-\partial^2 \log p(\boldsymbol{\theta} | \sigma_u^2, \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]_{\boldsymbol{\theta}'=\tilde{\boldsymbol{\theta}}_{\sigma}} \tag{8b}$$

Ignoring third and higher order terms, the asymptotic approximation is then made that

$$\boldsymbol{\theta} | \sigma_u^2, \mathbf{y} \sim N(\tilde{\boldsymbol{\theta}}_{\sigma}, \tilde{\mathbf{H}}_{\sigma}^{-1})$$

which is also used by Foulley *et al* (1987) to obtain approximate MML. In order to compute [8a], these authors employ the Newton-Raphson algorithm, which can be shown to lead to the iteration:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}\mathbf{X} & \mathbf{X}'\mathbf{R}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}\mathbf{X} & \mathbf{Z}'\mathbf{R}\mathbf{Z} + \Sigma_{\mathbf{u}}^{-1} \end{bmatrix}_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_{\sigma}}^{[t]} \begin{bmatrix} \tilde{\boldsymbol{\beta}}_{\sigma} \\ \tilde{\mathbf{u}}_{\sigma} \end{bmatrix}^{[t+1]} = \begin{bmatrix} \mathbf{X}'\mathbf{R}\bar{\mathbf{y}} \\ \mathbf{Z}'\mathbf{R}\bar{\mathbf{y}} \end{bmatrix}_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_{\sigma}}^{[t]} \tag{9}$$

where  $[t]$  indicates iterate number,

$$\begin{aligned} \mathbf{R} &= \text{Diag} \{ \lambda_i \} \\ \bar{\mathbf{y}} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{R}^{-1}\mathbf{v} \quad \text{and} \\ \mathbf{v} &= \{ y_i - \lambda_i \} \end{aligned}$$

is a residual. Note that  $\mathbf{R}^{-1}\mathbf{v} = \{ (y_i - \lambda_i) / \lambda_i \}$  can be interpreted as a residual vector expressed in units of residual variance, or relative to the mean of the conditional distribution of the observations. It can also be shown that:

$$\tilde{\mathbf{H}}_{\sigma} = \begin{bmatrix} \mathbf{X}'\mathbf{R}\mathbf{X} & \mathbf{X}'\mathbf{R}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}\mathbf{X} & \mathbf{Z}'\mathbf{R}\mathbf{Z} + \Sigma_{\mathbf{u}}^{-1} \end{bmatrix}_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_{\sigma}} \tag{10}$$

Note that the solution to system [9], which resembles Henderson's mixed model equations, and the negative Hessian [10] are both a function of  $\sigma_u^2$ .

From [7]

$$p(\sigma_u^2 | \mathbf{y}) = \frac{p(\boldsymbol{\theta} | \mathbf{y})}{p(\boldsymbol{\theta} | \sigma_u^2, \mathbf{y})} \tag{11}$$

We wish to find the mode of the marginal distribution of  $\sigma_u^2$  (or maximum of the marginal likelihood of the data) by recourse to Laplacian integration, as in Leonard (1982). Now:

$$p(\boldsymbol{\theta} | \mathbf{y}) = \exp \{ \log p(\boldsymbol{\theta}, \sigma_u^2 | \mathbf{y}) \} \tag{12}$$

A second-order Taylor series expansion of the log joint posterior density about  $\tilde{\boldsymbol{\theta}}_\sigma$  at a fixed  $\sigma_u^2$  gives:

$$\begin{aligned} \log p(\boldsymbol{\theta}, \sigma_u^2 | \mathbf{y}) &\approx \log p(\tilde{\boldsymbol{\theta}}_\sigma, \sigma_u^2 | \mathbf{y}) \\ &+ \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \left[ \log p(\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}) + \log p(\sigma_u^2 | \mathbf{y}) \right] \right\}_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_\sigma} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_\sigma) \\ &+ \frac{1}{2} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_\sigma) \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \left[ \log p(\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}) + \log p(\sigma_u^2 | \mathbf{y}) \right] \right\}_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_\sigma} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_\sigma) \\ &\approx \log p(\tilde{\boldsymbol{\theta}}_\sigma, \sigma_u^2 | \mathbf{y}) - \frac{1}{2} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_\sigma)' \tilde{\mathbf{H}}_\sigma (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_\sigma) \end{aligned} \tag{13}$$

Employing [13] in [12] and letting  $p_A(\cdot)$  denote an approximate density,

$$p_A(\boldsymbol{\theta} | \mathbf{y}) = p(\tilde{\boldsymbol{\theta}}_\sigma, \sigma_u^2 | \mathbf{y}) \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_\sigma)' \tilde{\mathbf{H}}_\sigma (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_\sigma) \right\} \tag{14}$$

Using this in [11] and recalling that  $\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}$  is approximately normal,

$$p_A(\sigma_u^2 | \mathbf{y}) \propto \frac{p(\tilde{\boldsymbol{\theta}}_\sigma, \sigma_u^2 | \mathbf{y})}{|\tilde{\mathbf{H}}_\sigma|^{1/2}} \quad (0 < \sigma_u^2 < \infty) \tag{15}$$

Taking logs of [15], and using [6], we note that apart from a constant,

$$L_A(\sigma_u^2 | \mathbf{y}) = \sum_{i=1}^n (y_i \ln \tilde{\lambda}_i - \tilde{\lambda}_i) - \frac{1}{2} \frac{\tilde{\mathbf{u}}_\sigma' \mathbf{A}^{-1} \tilde{\mathbf{u}}_\sigma}{\sigma_u^2} - \frac{1}{2} q \log(\sigma_u^2) - \frac{1}{2} \log |\tilde{\mathbf{H}}_\sigma| \tag{16}$$

where  $\tilde{\lambda}_i = \exp \{ \mathbf{w}_i' \tilde{\boldsymbol{\theta}}_\sigma \}$  is computed from the mode of the joint posterior density of  $\boldsymbol{\beta}$  and  $\mathbf{u}$ , given  $\sigma_u^2$ . One can find the posterior marginal mode of  $\sigma_u^2$  by establishing a grid of points of  $\{ \sigma_u^2, L_A(\sigma_u^2 | \mathbf{y}) \}$  and then interpolating with a second order polynomial as in Smith and Graser (1986).

It is interesting to note that if the data were normally distributed, the algorithm just described reduces to that suggested by Graser *et al* (1987), or DFREML (Meyer, 1989). Hence, Laplacian integration provides a generalization of DFREML to a class of non-linear models that could be termed DFMML.

## VARIANCE COMPONENT ESTIMATION FROM THE POSTERIOR MEAN

### Theory

The posterior mean is an attractive point estimator; from a decision theory viewpoint, it can be shown to minimize expected posterior quadratic loss (Lee, 1989). The mean of the marginal posterior distribution of the variance component can be written as:

$$E(\sigma_u^2 | \mathbf{y}) = \int_{\mathfrak{R}_{\boldsymbol{\theta}}} \sigma_u^2 p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad [17]$$

where  $\mathfrak{R}_{\boldsymbol{\theta}}$  is the space of the entire parameter vector. In this section we consider developments for computing posterior means presented by Tierney and Kadane (1986) and derived in detail by Cantet *et al* (1992); these are extensions of Laplacian procedures introduced by Leonard (1982).

Let:

$$L = L(\boldsymbol{\theta}) = \log(p(\boldsymbol{\theta} | \mathbf{y})) \quad [18a]$$

$$L^* = L^*(\boldsymbol{\theta}) = \log(p(\boldsymbol{\theta} | \mathbf{y})) + \log(\sigma_u^2) \quad [18b]$$

The posterior mean of the variance component can then be represented as

$$E(\sigma_u^2 | \mathbf{y}) = \frac{\int_{\mathfrak{R}_{\boldsymbol{\theta}}} e^{L^*} d\boldsymbol{\theta}}{\int_{\mathfrak{R}_{\boldsymbol{\theta}}} e^L d\boldsymbol{\theta}} \quad [19]$$

Note that the denominator assures that the joint posterior integrates to one when the integration constant in the joint density is ignored. Define:

$$\hat{\boldsymbol{\theta}} = \text{Arg}_{\boldsymbol{\theta}} \text{Max}_{L(\boldsymbol{\theta})} \quad [20a]$$

$$\hat{\boldsymbol{\theta}}^* = \text{Arg}_{\boldsymbol{\theta}} \text{Max}_{L^*(\boldsymbol{\theta})} \quad [20a]$$

$$\mathbf{H} = \left[ -\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \quad [21a]$$

$$\mathbf{H}^* = \left[ -\frac{\partial^2 L^*}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^*} \quad [21b]$$

The negative joint Hessians above can be written as:

$$\mathbf{H} = -\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} -\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} & \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \sigma_u^2} \\ -\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \sigma_u^2 \partial \boldsymbol{\theta}'} & \frac{\partial^2 L(\boldsymbol{\theta})}{(\partial \sigma_u^2)^2} \end{bmatrix}$$

$$\mathbf{H}^* = -\frac{\partial^2 L^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} -\frac{\partial^2 L^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} & -\frac{\partial^2 L^*(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \sigma_u^2} \\ -\frac{\partial^2 L^*(\boldsymbol{\theta})}{\partial \sigma_u^2 \partial \boldsymbol{\theta}'} & -\frac{\partial^2 L^*(\boldsymbol{\theta})}{(\partial \sigma_u^2)^2} \end{bmatrix}$$

The upper left blocks in both negative Hessians pertaining to the vector of location parameters,  $\boldsymbol{\theta}$ , are as in [10]. The remaining terms are:

$$-\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma_u^2} = -\frac{\partial^2 L^*(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma_u^2} = 0$$

$$-\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \mathbf{u} \partial \sigma_u^2} = -\frac{\partial^2 L^*(\boldsymbol{\theta})}{\partial \mathbf{u} \partial \sigma_u^2} = -\frac{\mathbf{A}^{-1} \mathbf{u}}{(\sigma_u^2)^2}$$

$$-\frac{\partial^2 L(\boldsymbol{\theta})}{(\partial \sigma_u^2)^2} = \frac{\mathbf{u}' \mathbf{A}^{-1} \mathbf{u}}{(\sigma_u^2)^3} - \frac{q}{2(\sigma_u^2)^2}$$

Further:

$$\frac{\partial L^*(\boldsymbol{\theta})}{\partial \sigma_u^2} = \frac{\mathbf{u}' \mathbf{A}^{-1} \mathbf{u}}{2(\sigma_u^2)^2} - \frac{(\frac{q}{2} - 1)}{\sigma_u^2}$$

so that

$$-\frac{\partial^2 L^*(\boldsymbol{\theta})}{(\partial \sigma_u^2)^2} = \frac{\mathbf{u}' \mathbf{A}^{-1} \mathbf{u}}{(\sigma_u^2)^3} - \frac{(\frac{q}{2} - 1)}{(\sigma_u^2)^2}$$

Tierney and Kadane (1986) approximate the numerator and denominator in [19] via the second order Taylor series expansions

$$\int_{\mathbb{R}_{\boldsymbol{\theta}}} e^{L} d\boldsymbol{\theta} \approx \int_{\mathbb{R}_{\boldsymbol{\theta}}} \exp \left[ L(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{H}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] d\boldsymbol{\theta}$$

$$= (2\pi)^{(p+q+1)/2} |\mathbf{H}|^{-1/2} \exp \left[ L(\hat{\boldsymbol{\theta}}) \right] \quad [22a]$$

and

$$\int_{\mathbb{R}_{\boldsymbol{\theta}}} e^{L^*} d\boldsymbol{\theta} \approx \int_{\mathbb{R}_{\boldsymbol{\theta}}} \exp \left[ L^*(\hat{\boldsymbol{\theta}}^*) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^*)' \mathbf{H}^*(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^*) \right] d\boldsymbol{\theta}$$

$$= (2\pi)^{(p+q+1)/2} |\mathbf{H}^*|^{-1/2} \exp \left[ L^*(\hat{\boldsymbol{\theta}}^*) \right] \quad [22b]$$

Using [22a] and [22b] in [19], the posterior mean is approximately

$$\widehat{E}(\sigma_u^2 | \mathbf{y}) = \sqrt{\frac{|\mathbf{H}|}{|\mathbf{H}^*|}} \exp(L(\widehat{\boldsymbol{\theta}}^*) - L(\widehat{\boldsymbol{\theta}})) \quad [23]$$

This approximation has been deemed to be highly accurate. The errors of the approximations to the integrals in the denominator and the numerator in [19] are of order  $O(n^{-1})$ . This would also be the order of the error incurred when approximating the joint posterior by a normal distribution. However, the leading terms in the 2 errors are nearly identical and cancel when the ratio in [19] is taken (Tierney and Kadane, 1986), thereby leading to an error term that is proportional to  $O(n^{-2})$ .

### Computational considerations

Consider the ratio of integrals in [19], and the maximizers,  $\widehat{\boldsymbol{\theta}}$  and  $\widehat{\boldsymbol{\theta}}^*$  in [20a] and [20b]. Computing the denominator entails evaluating  $L(\widehat{\boldsymbol{\theta}})$ , which is equivalent to maximizing the joint log-posterior density in [6]. Likewise, computing the numerator would entail the same procedure, except that  $\log(\sigma_u^2)$  is added to the log joint posterior.

From [6], it follows that:

$$\frac{\partial L}{\partial \sigma_u^2} = \frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}{2(\sigma_u^2)^2} - \frac{q}{2\sigma_u^2}$$

and

$$\frac{\partial L^*}{\partial \sigma_u^2} = \frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}{2(\sigma_u^2)^2} - \frac{q}{2\sigma_u^2} + \frac{1}{\sigma_u^2}$$

Setting the first derivatives to zero leads to expressions:

$$\tilde{\sigma}_u^2 = \frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}{q} \quad [24a]$$

and

$$\tilde{\sigma}_u^{*2} = \frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}{q-2} \quad [24b]$$

which would be used in conjunction with system [9] (evaluated at the 'current' values of  $\sigma_u^2$ ) to obtain the joint posterior mode. We obtained estimates of  $\sigma_u^2$  equal to zero in several simulation tests of this algorithm, when applied to Poisson models. This implies that the joint density is maximum when  $\sigma_u^2 = 0$ . As noted by Lindley and Smith (1972), Harville (1977), Thompson (1980), and Gianola and Fernando (1986), joint maximization of a joint posterior density with respect to fixed and random effects and the variance components in a linear model, often leads to a sequence of iterates for the latter converging towards zero. Harville (1977) attributed the problem to 'severe dependencies' between  $\mathbf{u}$  and  $\sigma_u^2$  (clearly, the conditional distribution of  $\mathbf{u}|\Sigma_u$  depends on  $\sigma_u^2$ ). As noted by Gianola *et al* (1990),

the problem also arises when searching for the mode of  $p(\boldsymbol{\beta}, \mathbf{u}|\mathbf{y})$  or  $p(\mathbf{u}|\mathbf{y})$  where any 'dependency' would be eliminated by integration of  $\sigma_u^2$ . In general, the problem does not occur when informative priors are employed for  $\sigma_u^2$ . Höschele *et al* (1987) also found that many of their variance component iterations were drifting towards zero when using a first order algorithm for maximizing the joint posterior density in threshold models.

It is instructive to contrast the log of the joint posterior density in [6],  $L(\boldsymbol{\theta})$ , with the approximate marginal density of  $\sigma_u^2$ ,  $L_A(\sigma_u^2|\mathbf{y})$ , in [16]. Apart from the constant terms, these 2 functions differ in that in [16], half the value of the log of the determinant of the negative Hessian matrix is subtracted. Ritter (1992) views this as an important 'width or variance adjustment' in the estimation of  $\sigma_u^2$  from its marginal distribution; this supports the claim made by O'Hagan (1976) that marginal modes are better estimators than joint modes.

Because the Tierney and Kadane (1986) approximation to the posterior mean fails whenever  $\sigma_u^2$  goes to zero in the joint maximization algorithm, alternative strategies must be sought. One possibility would be to evaluate the approximate marginal density of the variance component as in [16] and then compute the posterior mean by cubic spline fitting (deBoor, 1978) or by Gaussian quadrature involving 'strategic' evaluation points of  $\sigma_u^2$ .

## NUMERICAL EXAMPLE

Data on embryo yields within a nucleus scheme were simulated with a Poisson animal model according to procedures given in Tempelman and Gianola (1991). The underlying mean on the log scale was  $\log(4)$ . Two 'fixed' factors, one with 5 levels and the other with 20 levels were generated from a  $N(0, 0.10)$  distribution on the canonical log scale. Additive genetic effects were generated from a  $N(0, 0.05)$  distribution for a base population of 16 sires and 128 cows. Cows were superovulated and mated at random to outside sires also drawn at random from the population at large. The numbers of embryos produced per cow was a drawing from the Poisson distribution, with the value of its parameter depending on the fixed effects and the additive genetic value of the female in question. Sex ratios in the embryos was 50:50, and sexes were assigned at random, using the binomial distribution. Male embryos were discarded, and the genetic value of female embryos was obtained as:

$$a_0 = \frac{1}{2}a_s + \frac{1}{2}a_D + z_0\sqrt{\frac{1}{2}0.05}$$

where  $a_s$  is the breeding value of an outside sire,  $a_D$  is the breeding value of the donor cow, and  $z_0 \sim N(0, 1)$ . The female embryos were 'raised' (probability of survival to an embryo collection was 0.70), and mated at random to nucleus sires, to produce a new generation. Records on embryo yields obtained from these matings were simulated as before. Thus, information on embryo yields was available on foundation cows and their surviving female progeny. The simulation involved a 'natural selection' process because donor cows without embryos recovered left no progeny at all, whereas donor cows with higher embryo yields left more female progeny.

In the simulation,  $p = 24$  (25 levels of fixed effects minus 1 dependency) and  $q = 242$  (16 sires, 128 dams and 98 surviving progeny). The mode of the approximate marginal density of  $\sigma_u^2$  was located employing [16]. An iterative quadratic fit led to  $\sigma_u^2 = 0.0347$  as maximum, and the approximate log marginal density is depicted in figure 1, with a cubic spline fitted through the iterates. The EM-type algorithm of Foulley *et al* (1987) gave a modal value of  $\sigma_u^2 = 0.0343$ . Convergence criteria of the 2 algorithms were not directly comparable, thereby contributing to some of the discrepancy between the 2 estimates.

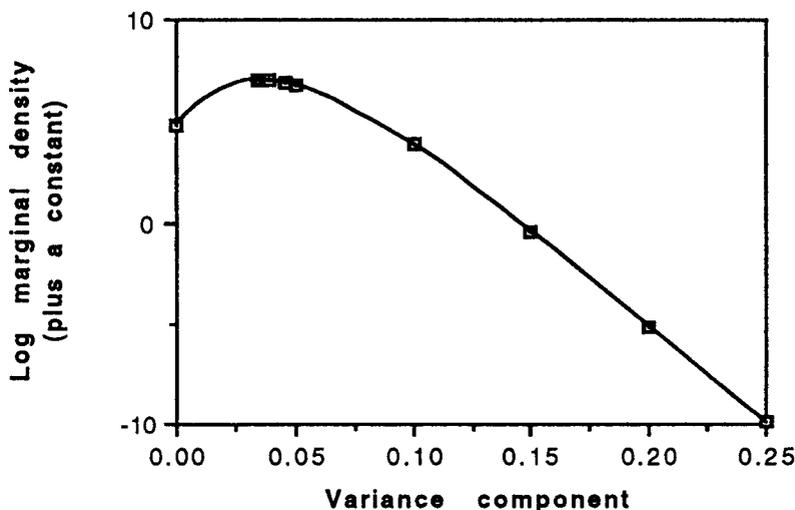


Fig 1. Approximate Laplacian log marginal density of  $\sigma_u^2$  (zero =  $5 \times 10^{-5}$ ).

As noted previously, the Tierney and Kadane (1986) approach gave  $\sigma_u^2 = 0$  as the value of the variance component that maximized the joint density. To illustrate, the log joint density [6] was evaluated at the 'current' value of  $\sigma_u^2$  (and of the resulting solution to system [9]) during iteration and the plot is shown in figure 2. Clearly,  $\sigma_u^2 = 0$  would be the maximizer, giving a density value of plus infinity. The degeneracy of this log density highlights the importance of the Hessian adjustment in [16].

## DISCUSSION

The Laplacian procedure for finding the mode of the marginal posterior distribution of a single variance component in a Poisson mixed model was found to be analogous to 'derivative-free' methods employed for computing REML estimates of variance components in a mixed linear model (Smith and Graser, 1986; Graser *et al*, 1987). In fact, if Laplacian marginalization is applied to variance estimation with normally distributed data in a Bayesian model with flat priors for fixed effects and variance components, this would yield precisely derivative-free REML. This is because the

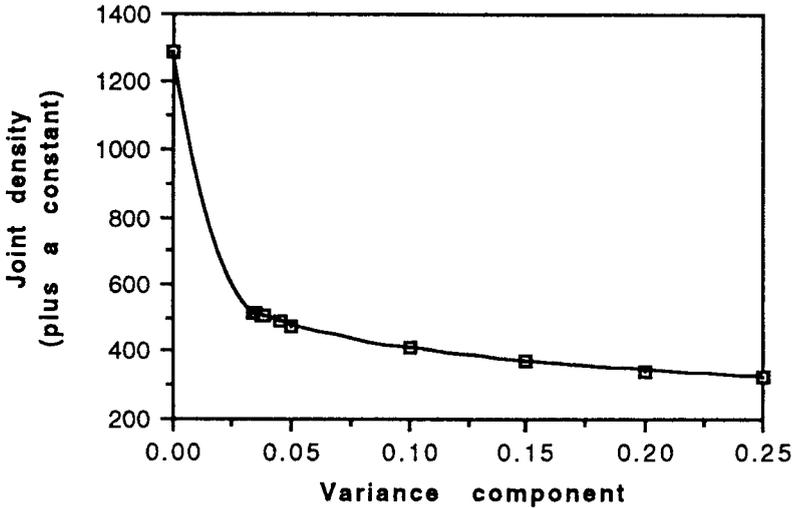


Fig 2. Log joint density as a function of  $\sigma_u^2$  («zero» =  $5 \times 10^{-5}$ ).

Laplacian integration is then exact. Although a single variance component was considered in this paper, the algorithm generalizes in a straightforward manner to a Poisson model with several variances, and one obtains MML estimates of variance components. Because of the analogy noted above, we suggest DFMMML (derivative-free marginal maximum likelihood) as a generic term for this algorithm, since the procedure extends beyond the class of mixed linear models.

The Laplacian technique used for finding the mode of the marginal posterior distribution of  $\sigma_u^2$  is theoretically, although not numerically, equivalent to the EM-type algorithm suggested by Foulley *et al* (1987). In order to obtain the mode of the marginal distribution of  $\sigma_u^2$  these authors employ the relationship:

$$\frac{\partial \log (p(\sigma_u^2 | \mathbf{y}))}{\partial \sigma_u^2} = \mathbf{E}_{\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}} \left[ \frac{\partial}{\partial \sigma_u^2} \log (p(\boldsymbol{\theta} | \mathbf{y})) \right] \tag{25}$$

Using the approximation [13] in [25], we obtain:

$$\begin{aligned} \frac{\partial \log (p(\sigma_u^2 | \mathbf{y}))}{\partial \sigma_u^2} &= \mathbf{E}_{\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}} \frac{\partial}{\partial \sigma_u^2} \left[ \log (p(\hat{\boldsymbol{\theta}}_\sigma, \sigma_u^2 | \mathbf{y})) - \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_\sigma)' \tilde{\mathbf{H}}_\sigma (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_\sigma) \right] \\ &= \mathbf{E}_{\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}} \left[ \frac{\partial}{\partial \sigma_u^2} \log (p(\hat{\boldsymbol{\theta}}_\sigma, \sigma_u^2 | \mathbf{y})) \right] - \mathbf{E}_{\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}} \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_\sigma)' \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{A}^{-1} \\ & & \sigma_u^4 \end{bmatrix} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_\sigma) \\ &= \mathbf{E}_{\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}} \left[ \frac{\partial}{\partial \sigma_u^2} \log (p(\hat{\boldsymbol{\theta}}_\sigma, \sigma_u^2 | \mathbf{y})) \right] - \mathbf{E}_{\boldsymbol{\theta} | \sigma_u^2, \mathbf{y}} \frac{1}{2} \text{tr} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{A}^{-1} \\ & & \sigma_u^4 \end{bmatrix} \tilde{\mathbf{H}}_\sigma^{-1} \end{aligned} \tag{26}$$

The first term of [26] is obtained by differentiating [6] with respect to  $\sigma_u^2$ , with  $\lambda_i$  and  $\mathbf{u}$  replaced by the numerical quantities,  $\tilde{\lambda}_i$  and  $\tilde{\mathbf{u}}$  respectively. Then

$$\frac{\partial \log (p(\sigma_u^2 | \mathbf{y}))}{\partial \sigma_u^2} = \frac{\tilde{\mathbf{u}}'_\sigma \mathbf{A}^{-1} \tilde{\mathbf{u}}_\sigma}{2\sigma_u^4} - \frac{q}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \text{tr}(\mathbf{A}^{-1} \mathbf{C}_{\mathbf{uu}}) \quad [27]$$

where  $\mathbf{C}_{\mathbf{uu}}$  is the random by random block of the inverse of the conditional negative Hessian [10]. Finally, setting [27] to zero and solving for  $\sigma_u^2$  gives the iteration.

$$\sigma_u^2 = \frac{\tilde{\mathbf{u}}' \mathbf{A}^{-1} \tilde{\mathbf{u}} + \text{tr}(\mathbf{A}^{-1} \mathbf{C}_{\mathbf{uu}})}{q}$$

which is precisely the algorithm of Foulley *et al* (1987). It is important to note that the algorithm developed in this paper in connection with [16] is numerically different from [8], *ie*, a different sequence of iterates is to be expected. However, within the limits of numerical precision, as determined by the local curvature of the marginal log likelihood, convergence to the same maximum should be attained, as verified in the numerical application discussed previously.

Application of the Laplacian procedure to the threshold model of Gianola and Foulley (1983) and Harville and Mee (1984) is straightforward. One would simply replace the log-likelihood and the conditional Hessian given in this paper by the corresponding terms in the threshold model. For multidimensional problems, *ie*, more than one variance parameter, procedures suggested for DFREML by Meyer (1989) such as the simplex or quasi-Newton algorithms could be used. Optimization procedures that incorporate information on the vector of first derivatives and on the function to be optimized would be expected to be most useful.

Approximate posterior standard errors for the variance components may be computed using a quadratic fit of the log posterior density near the mode as in Graser and Smith (1987). The computation of the log-posterior densities allows also to construct marginal likelihood ratio tests, or posterior odds ratios, for assessing the importance of different sources of genetic and environmental variation.

Harville (1977) asserted that the posterior mode is an attractive estimator, being less sensitive than the mean to the tails of the posterior density. However, under a squared-error loss, the posterior mean is optimum. Unfortunately, the Laplacian procedure of Tierney and Kadane (1986) for computing the posterior mean would be expected to fail in many instances. Joint maximization should work well when there is a large amount of information on each random effect, *eg* sire models, but not in animal models. Hence, alternative numerical procedures should be sought for computing posterior means. Further enhancements to marginal estimation of parameters involving Laplacian integration are given by Kass and Steffey (1989) and Leonard *et al* (1989).

## REFERENCES

Cantet RJC, Fernando RL, Gianola D (1992) Bayesian inference about dispersion parameters of univariate mixed models with maternal effects: theoretical considerations. *Genet Sel Evol* 24, 107-135

- deBoor C (1978) *A Practical Guide to Spline Fitting*. Springer Verlag, New York
- Foulley JL, Gianola D, Im S (1987) Genetic evaluation of traits distributed as Poisson-binomial with reference to reproductive characters. *Theor Appl Genet* 73, 870-877
- Foulley JL, Gianola D, Im S (1990) Genetic evaluation for discrete polygenic traits in animal breeding. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer Verlag, Heidelberg, 361-409
- Gianola D, Fernando RL (1986) Bayesian methods in animal breeding theory. *J Anim Sic* 63, 217-244
- Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet Sel Evol* 15, 201-224
- Gianola D, Im S, Macedo FW (1990) A framework for prediction of breeding value. In: *Advances in Statistical Methods for Genetic Improvement of Livestock*. (Gianola D, Hammond K, eds) Springer Verlag, Heidelberg, 210-238
- Graser HU, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J Anim Sci* 64, 1362-1370
- Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* 72, 320-338
- Harville DA, Mee RW (1984) A mixed model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408
- 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
- Kass RE, Steffey D (1989) Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J Am Stat Assoc* 84, 717-726
- Lee PM (1989) *Bayesian Statistics: An Introduction*. Oxford University Press, NY
- Leonard T (1982) Comment on 'A simple predictive density function'. *J Am Stat Assoc* 77, 657-658
- Leonard T, Hsu JSJ, Tsui KW (1989) Bayesian marginal inference. *J Am Stat Assoc* 84, 1051-1058
- Lindley DV, Smith AFM (1972) Bayes estimates for the linear model. *J R Stat Soc B* 34, 1-18
- Manfredi ET, San Cristobal M, Foulley JL (1991) Some factors affecting the estimation of genetic parameters for cattle dystocia under a threshold model. *Anim Prod* 53, 151-156
- McCullagh P, Nelder JA (1989) *Generalized Linear Models*. Chapman and Hall, NY, 2nd edn
- Meijering A (1985) Sire evaluation for calving traits by best linear unbiased prediction and nonlinear methodology. *J Anim Breed Genet* 102, 95-105
- 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, 317-330
- O'Hagan AO (1976) On posterior joint and marginal modes. *Biometrika* 63, 329-333
- Pérez-Enciso M, Tempelman RJ, Gianola D (1993) A comparison between linear and Poisson mixed models for litter size in Iberian pigs. *Livest Prod Sci* (in press)

- Ritter C (1992) Modern inference in nonlinear least-squares regression. PhD thesis, Department of Statistics, University of Wisconsin-Madison, WI
- Smith AFM (1991) Bayesian computational methods. *Phil Trans R Soc Lond A* 337, 369-386
- Smith SP, Graser HU (1986) Estimating variance components in a class of mixed models by restricted maximum likelihood. *J Dairy Sci* 69, 1156-1165
- Tempelman RJ, Gianola D (1991) Evaluation of a Poisson animal model for the genetic evaluation of embryo yields. *J Dairy Sci* 74 (suppl. 1), 157
- Thompson R (1980) Maximum likelihood estimation of variance components. *Math Operationsforsch Stat* 11, 545-561
- Tierney L, Kadane JB (1986) Accurate approximations for posterior moments and marginal densities. *J Am Stat Assoc* 81, 82-86
- Weller JI, Misztal I, Gianola D (1988) Genetic analysis of dystocia and calf mortality in Israeli-Holsteins. *J Dairy Sci* 71, 2491-2501
- Weller JI, Gianola D (1989) Models of dystocia and calf mortality. *J Dairy Sci* 72, 2633-2643