Calcul des densités marginales a posteriori des paramètres génétiques d'un modèle animal multicaractère en utilisant une approximation laplacienne ou l'échantillonnage de Gibbs

A. Hofer, Vincent Ducrocq

Original article

Computing marginal posterior densities of genetic parameters of a multiple trait animal model using Laplace approximation or Gibbs sampling

A Hofer [*,**], V Ducrocq

Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, 78352 Jouy-en-Josas cedex, France

Summary – Two procedures for computing the marginal posterior density of heritabilities or genetic correlations, ie, Laplace's method to approximate integrals and Gibbs sampling, are compared. A multiple trait animal model is considered with one random effect, no missing observations and identical models for all traits. The Laplace approximation consists in computing the marginal posterior density for different values of the parameter of interest. This approximation requires the repeated evaluation of traces and determinants, which are easy to compute once the eigenvalues of a matrix of dimension equal to the number of animals are determined. These eigenvalues can be efficiently computed by the Lanczos algorithm. The Gibbs sampler generates samples from the joint posterior density. These samples are used to estimate the marginal posterior density, which is exact up to a Monte-Carlo error. Both procedures were applied to a data set with semen production traits of 1957 Normande bulls. The traits analyzed were volume of the ejaculate, motility score and spermatozoa concentration. The Laplace approximation yielded very accurate approximations of the marginal posterior density for all parameters with much lower computing costs.

marginal posterior density / Laplace approximation / Lanczos method / Gibbs sampling / genetic parameters

Résumé – **Calcul des densités marginales a posteriori des paramètres génétiques d'un modèle animal multicaractère en utilisant une approximation laplacienne ou l'échantillonnage de Gibbs.** *Deux procédures de calcul de la densité marginale a posteriori des héritabilités et des corrélations génétiques, à savoir la méthode de Laplace pour l'approximation des intégrales et l'échantillonnage de Gibbs, sont comparées. Pour cela, nous considérons un modèle animal multicaractère avec un effet aléatoire, sans observations manquantes et avec un modèle identique pour chaque caractère. L'approximation de*

---
[*] Correspondence and reprints
[**] On leave from and permanent address: Institute of Animal Science, Swiss Federal Institute of Technology (ETH), CLU, CH-8092 Zurich, Switzerland

Laplace conduit au calcul de la densité marginale a posteriori pour différentes valeurs du paramètre qui nous intéresse. Cela nécessite l'évaluation répétée de traces et de déterminants qui sont simples à calculer une fois que les valeurs propres d'une matrice de dimension égale au nombre d'animaux ont été déterminées. Ces valeurs propres peuvent être calculées de manière efficace à l'aide de l'algorithme de Lanczos. L'échantillonnage de Gibbs génère des échantillons de la densité conjointe a posteriori. Ces échantillons sont utilisés pour estimer la densité marginale a posteriori, qui est exacte à une erreur de Monte Carlo près. Les deux procédures ont été appliquées à un fichier de données comportant les caractères de production de semence recueillies sur 1957 taureaux Normands. Les caractères analysés étaient le volume de l'éjaculat, une note de motilité et la concentration en spermatozoïdes. L'approximation laplacienne a permis une approximation très précise de la densité marginale a posteriori de tous les paramètres avec un coût de calcul beaucoup plus réduit.

densité marginale a posteriori / approximation de Laplace / méthode de Lanczos / échantillonnage de Gibbs / paramètres génétiques

INTRODUCTION

The optimization of breeding programs relies on accurate estimates of genetic parameters, ie, heritabilities and genetic correlations. Genetic evaluations of selection candidates by best linear unbiased prediction (BLUP; Henderson, 1973) assumes known variance components which, in practice, are replaced by estimates. Thus, it seems desirable to have an idea about the accuracy of the estimates used. Restricted maximum likelihood (REML; Patterson and Thompson, 1971) is widely regarded as the method of choice for estimation of variance components. REML estimates have known large sample properties under the concept of repeated sampling of the data analyzed but unknown small sample distributions. In contrast, a Bayesian analysis yields an exact posterior probability density of the parameters of interest for the data set at hand (Gianola and Fernando, 1986). REML estimates correspond to the mode of the posterior distribution marginalized with respect to the location parameters in a Bayesian analysis with flat priors for fixed effects and variance components (Harville, 1977). Gianola and Foulley (1990) used two arguments for suggesting an alternative method for variance component estimation. First, REML estimates are joint modes of all (co)variance components rather than marginal modes. The latter are a better approximation to the posterior mean, which is the optimum Bayes estimator under a quadratic loss function. Second, if interest is only in a subset of the (co)variance components the remaining (co)variance components should be regarded as nuisance parameters and inferences should take into account the error incurred in estimating these nuisance variance components. Gianola and Foulley (1990) suggest basing inferences about the variance components of interest on their marginal posterior density.

Various methods exist for marginalizing the joint posterior density of variance components. Analytical integration is possible only for certain parameters in simple univariate models, eg, for obtaining the marginal posterior density of the variance ratio (or the heritability) in a single trait linear mixed model with one random effect (Gianola et al, 1990b). With more complex models, approximations or Monte-Carlo integration techniques are used to obtain marginal posterior densities. Tierney and Kadane (1986) have suggested computing marginal posterior densities by

Laplace's method to approximate integrals. In animal breeding the method was first considered by Cantet et al (1992) for a linear mixed model with maternal effects and has been applied by Tempelman and Gianola (1993) to a Poisson mixed model and recently by Ducrocq and Casella (1996) to a mixed survival model. Laplacian integration involves the repeated computation of second derivatives of the logarithm of the posterior density of all (co)variance components after integration of location parameters. For single trait animal models with one random effect or for multiple trait models where a canonical transformation is possible, the derivatives can be efficiently computed once all eigenvalues of a symmetric matrix of dimension of the number of animals are determined (Robert and Ducrocq, 1996). Robert and Ducrocq (1996) have shown that the Lanczos algorithm is well suited to computing these eigenvalues.

As an alternative, Monte-Carlo integration can be used to obtain the exact marginal posterior distribution up to the Monte-Carlo error. Gibbs sampling (Gelfand and Smith, 1990; Casella and George, 1992), a Markov chain Monte-Carlo procedure, is becoming an increasingly important tool in statistical analysis (Gilks et al, 1996). The Gibbs sampler generates samples from the joint (marginal) posterior distribution. The samples are used to derive the desired summary statistics of the target distribution. Applications in animal breeding have rapidly increased in recent years (eg, Wang al, 1994; Janss et al, 1995; Sorensen et al, 1995; Van Tassell and Van Vleck, 1996).

This paper considers the computation of the marginal posterior density of heritabilities or genetic correlations of a simple multiple trait animal model. Heritabilities and genetic correlations are nonlinear functions of (co)variance components. A second order Laplace approximation of marginal densities of nonlinear functions has been derived by Tierney et al (1989, 1991) and is presented in detail. The Laplace approximation is compared to Gibbs sampling by applying both procedures to a data set of semen production traits of Normande bulls.

METHODS

Model

Consider the multiple trait mixed model with only one random effect, equal incidence matrices for all t traits and no missing observations:

$$\mathbf{y} = (\mathbf{I}_t \otimes \mathbf{X})\mathbf{b} + (\mathbf{I}_t \otimes \mathbf{Z})\mathbf{a} + \mathbf{e} \qquad [1]$$

where $\mathbf{y}$ is the vector of observations of order $n \cdot t$, $\mathbf{I}_t$ is an identity matrix of order t by t, $\mathbf{X}$ and $\mathbf{Z}$ are known incidence matrices of order n by p and n by q, respectively, $\mathbf{b}$ is the vector of fixed effects of order $p \cdot t$ and $\mathbf{a}$ is the vector of random additive genetic effects of order $q \cdot t$, and $\mathbf{e}$ is the vector of residuals of order $n \cdot t$.

The conditional probability density function of the observations is:

$$p(\mathbf{y}|\mathbf{b}, \mathbf{a}, \mathbf{R}_0) \propto |\mathbf{R}_0|^{-\frac{n}{2}}$$

$$\exp\left\{ -\frac{1}{2}(\mathbf{y} - (\mathbf{I}_t \otimes \mathbf{X})\mathbf{b} - (\mathbf{I}_t \otimes \mathbf{Z})\mathbf{a})'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n)(\mathbf{y} - (\mathbf{I}_t \otimes \mathbf{X})\mathbf{b} - (\mathbf{I}_t \otimes \mathbf{Z})\mathbf{a})\right\}$$

with $\mathbf{R}_0$ the (co)variance matrix of residual effects of order t by t.

A Hofer, V Ducrocq

The prior distribution for the unknown location parameters are assumed to be:

$$p(\mathbf{b}) \propto \text{constant}$$

$$p(\mathbf{a}|\mathbf{G}_0) \propto |\mathbf{G}_0|^{-\frac{q}{2}} \exp\left\{-\frac{1}{2}\mathbf{a}'(\mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1})\mathbf{a}\right\}$$

with $\mathbf{G}_0$ the matrix of additive genetic (co)variances of order t by t and $\mathbf{A}$ the numerator relationship matrix of order q by q.

For the (co)variance matrices the prior distribution is assumed to be an inverse Wishart distribution (eg, Gelman et al, 1995):

$$p(\mathbf{G}_0|v_G, \mathbf{V}_G) \propto |\mathbf{G}_0|^{-\frac{1}{2}(v_G+t+1)} \cdot \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{V}_G\mathbf{G}_0^{-1})\right\}$$

$$p(\mathbf{R}_0|v_R, \mathbf{V}_R) \propto |\mathbf{R}_0|^{-\frac{1}{2}(v_R+t+1)} \cdot \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{V}_R\mathbf{R}_0^{-1})\right\}$$

with the known parameters degree of belief v_i (analogous to degrees of freedom) and scale matrix $\mathbf{V}_i$ of order t by t, where i stands either for G or R. The expected value of the inverse Wishart distribution is (for example for $\mathbf{G}_0$) $E(\mathbf{G}_0|v_G, \mathbf{V}_G) = (v_G - t - 1)^{-1}\mathbf{V}_G$, which may be used for choosing $\mathbf{V}_G$. The parameters $\mathbf{b}$, $\mathbf{a}$, $\mathbf{G}_0$ and $\mathbf{R}_0$ are assumed to be independent a priori.

The joint posterior probability density of all parameters is defined as the product of the probability density of the observations and the joint prior density of the parameters and can be written as:

$$p(\boldsymbol{\theta}, \mathbf{G}_0, \mathbf{R}_0 | \mathbf{y}, v_G, \mathbf{V}_G, v_R, \mathbf{V}_R) \propto$$

$$\exp\left\{-\frac{1}{2}[(\mathbf{y} - (\mathbf{I}_t \otimes \mathbf{W})\boldsymbol{\theta})'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n)(\mathbf{y} - (\mathbf{I}_t \otimes \mathbf{W})\boldsymbol{\theta}) + \boldsymbol{\theta}'(\mathbf{G}_0^{-1} \otimes \Sigma)\boldsymbol{\theta}]\right\}$$

$$\cdot |\mathbf{G}_0|^{-\frac{1}{2}(q+v_G+t+1)} \cdot \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{V}_G\mathbf{G}_0^{-1})\right\} \qquad\qquad [2]$$

$$\cdot |\mathbf{R}_0|^{-\frac{1}{2}(n+v_R+t+1)} \cdot \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{V}_R\mathbf{R}_0^{-1})\right\}$$

where $\mathbf{W} = [\mathbf{X}, \mathbf{Z}]$, $\boldsymbol{\theta}' = [\mathbf{b}', \mathbf{a}']$, and $\Sigma = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{A}^{-1} \end{bmatrix}$.

For this study the parameters of interest are functions of the (co)variance components in $\mathbf{G}_0$ and $\mathbf{R}_0$, ie, variance ratios and correlations. Therefore, $\boldsymbol{\theta}$ is a nuisance parameter that should be integrated out of the joint posterior density. As shown by Gianola et al (1990a) this can be achieved analytically leading to the following marginal posterior probability density of all (co)variance components:

$$p(\mathbf{G}_0, \mathbf{R}_0 | \mathbf{y}, v_G, \mathbf{V}_G, v_R, \mathbf{V}_R) \propto$$

$$\exp\left\{-\frac{1}{2}[\mathbf{y}'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n)\mathbf{y} - \widehat{\boldsymbol{\theta}}'(\mathbf{R}_0^{-1} \otimes \mathbf{W})\mathbf{y}]\right\} \cdot \left|(\mathbf{R}_0^{-1} \otimes \mathbf{W}'\mathbf{W}) + (\mathbf{G}_0^{-1} \otimes \Sigma)\right|^{-\frac{1}{2}}$$

$$\cdot |\mathbf{G}_0|^{-\frac{1}{2}(q+v_G+t+1)} \cdot \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{V}_G\mathbf{G}_0^{-1})\right\} \cdot |\mathbf{R}_0|^{-\frac{1}{2}(n+v_R+t+1)} \cdot \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{V}_R\mathbf{R}_0^{-1})\right\}$$

$$[3]$$

where $\widehat{\boldsymbol{\theta}} = \left[(\mathbf{R}_0^{-1} \otimes \mathbf{W}'\mathbf{W}) + (\mathbf{G}_0^{-1} \otimes \Sigma) \right]^{-1} (\mathbf{R}_0^{-1} \otimes \mathbf{W}')\mathbf{y}$ is the solution to the mixed model equations (Henderson, 1973).

To draw inference about functions of a subset of the (co)variance components, a further marginalization of the posterior density is required. Analytical solutions are not available.

Laplace approximation

Approximation of marginal posterior probability densities using Laplace's method has been suggested by Tierney and Kadane (1986). They considered the case where the vector of parameters can be partitioned as $\boldsymbol{\sigma} = [\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2]$, where $\boldsymbol{\sigma}_1$ is the parameter of interest and $\boldsymbol{\sigma}_2$ is the nuisance parameter that is integrated out of the posterior distribution using Laplace's method to approximate integrals.

Genetic parameters such as heritabilities or genetic correlations are nonlinear functions of (co)variance components. Inferences about genetic parameters should therefore be based on marginal densities of nonlinear functions of (co)variance components. Tierney et al (1989, 1991) extended the work of Tierney and Kadane (1986) for the approximation of marginal posterior distributions of nonlinear functions.

Let $\tau = g(\boldsymbol{\sigma})$ be the nonlinear function of interest of the m parameters in $\boldsymbol{\sigma}$. We assume here that τ is a scalar (eg, heritability or genetic correlation) but this assumption is not required for the following derivation. A new parameterization is defined as $\boldsymbol{\phi}(\boldsymbol{\sigma}) = \{\boldsymbol{\phi}_1(\boldsymbol{\sigma}), \boldsymbol{\phi}_2(\boldsymbol{\sigma})\} = \{g(\boldsymbol{\sigma}), \boldsymbol{\phi}_2(\boldsymbol{\sigma})\}$ partitioned into the nonlinear function of interest and a function $\boldsymbol{\phi}_2$ that ensures that the transformation $\boldsymbol{\phi}$ is one to one, although it might be difficult to specify $\boldsymbol{\phi}_2$ explicitly. The approximation proposed by Tierney et al (1989, 1991) does not depend on an explicit reparameterization. The only requirement is the existence of a one to one transformation in a sufficiently small neighbourhood U of the joint mode $\widehat{\boldsymbol{\sigma}}$. In our application $\boldsymbol{\phi}_2$ is easy to specify and the approximation caused by the restriction to the neighbourhood U does not occur.

In order to obtain the marginal distribution of $\tau = g(\boldsymbol{\sigma})$, the nuisance parameters $\mathbf{u}_2 = \boldsymbol{\phi}_2(\boldsymbol{\sigma})$ in the new parameterization must be integrated out

$$p(\tau | \mathbf{y}) = \int \exp\{\log p(\boldsymbol{\phi}^{-1}(\tau, \mathbf{u}_2)|\mathbf{y})\} |\mathbf{J}(\boldsymbol{\phi}^{-1}(\tau, \mathbf{u}_2))|^{-1} d\mathbf{u}_2 \qquad [4]$$

where $\mathbf{J}(\boldsymbol{\phi}^{-1}(\tau, \mathbf{u}_2))$ is the Jacobian matrix of the transformation $\boldsymbol{\phi}^{-1}$.

Let $\widehat{\boldsymbol{\sigma}}_\tau$ be the argument on the original scale that maximizes $\log p(\boldsymbol{\sigma}|\mathbf{y})$ subject to the constraint $g(\boldsymbol{\sigma}) = \tau$. In the new parameterization $\boldsymbol{\phi}(\widehat{\boldsymbol{\sigma}}_\tau)$ maximizes $\log p(\boldsymbol{\phi}^{-1}(g(\boldsymbol{\sigma}), \mathbf{u}_2)|\mathbf{y})$ subject to the constraint $g(\boldsymbol{\sigma}) = \tau$. At the constrained mode $\boldsymbol{\phi}(\widehat{\boldsymbol{\sigma}}_\tau)$, the gradient of the Lagrangian $\log p(\boldsymbol{\phi}^{-1}(g(\boldsymbol{\sigma}), \mathbf{u}_2)|\mathbf{y}) + \lambda(g(\boldsymbol{\sigma}) - \tau)$ is zero, where λ is a Lagrange multiplier.

Now, $\log p(\boldsymbol{\phi}^{-1}(\tau, \mathbf{u}_2)|\mathbf{y})$ in [4] for fixed τ is approximated using a second order Taylor series expansion around $\boldsymbol{\phi}_2(\widehat{\boldsymbol{\sigma}}_\tau)$:

$$\log p(\boldsymbol{\phi}^{-1}(\tau, \mathbf{u}_2)|\mathbf{y}) \approx \log p(\boldsymbol{\phi}^{-1}(\tau, \boldsymbol{\phi}_2(\widehat{\boldsymbol{\sigma}}_\tau))|\mathbf{y})$$
$$-\frac{1}{2}(\mathbf{u}_2 - \boldsymbol{\phi}_2(\widehat{\boldsymbol{\sigma}}_\tau))' \boldsymbol{\Omega}_{\tau,22}^{-1}(\mathbf{u}_2 - \boldsymbol{\phi}_2(\widehat{\boldsymbol{\sigma}}_\tau)) \qquad [5]$$

where

$$\Omega_{\tau,22}^{-1} = -\frac{\partial^2 \, \log \, p(\boldsymbol{\phi}^{-1}(\tau, \mathbf{u}_2)|\mathbf{y})}{\partial \mathbf{u}_2 \, \partial \mathbf{u}_2'}\Bigg|_{\mathbf{u}_2 = \boldsymbol{\phi}_2(\widehat{\boldsymbol{\sigma}}_\tau)}$$

In expression [5] the part that depends on $\mathbf{u}_2$ is the kernel of a normal distribution with mean $\boldsymbol{\phi}_2(\widehat{\boldsymbol{\sigma}}_\tau)$ and variance $\Omega_{\tau,22}$. The normalizing constant of this kernel is $(2\pi)^{-\frac{m-1}{2}}|\Omega_{\tau,22}|^{-\frac{1}{2}}$, which leads to the following approximation of equation [4]:

$$p(\tau|\mathbf{y}) \approx (2\pi)^{\frac{m-1}{2}}|\Omega_{\tau,22}|^{\frac{1}{2}} \, p(\widehat{\boldsymbol{\sigma}}_\tau|\mathbf{y})|\mathbf{J}(\widehat{\boldsymbol{\sigma}}_\tau)|^{-1} \qquad [6]$$

In expression [6] $\Omega_{\tau,22}$ and $\mathbf{J}(\widehat{\boldsymbol{\sigma}}_\tau)$ depend on an explicit specification of $\boldsymbol{\phi}_2(\boldsymbol{\sigma})$. Following Tierney et al (1989, 1991), we now derive an alternative expression for [6] that does not require an explicit reparameterization.

Let Ω_τ^{-1} and $\mathbf{H}_\tau^{-1}$ be the matrices of second derivatives of the Lagrangian evaluated at the constrained mode in the new and the original parameterization, respectively, ie

$$\Omega_\tau^{-1} = -\frac{\partial^2 \left\{\log \, p(\boldsymbol{\phi}^{-1}(g(\boldsymbol{\sigma}), \mathbf{u}_2)|\mathbf{y}) + \lambda(g(\boldsymbol{\sigma}) - \tau)\right\}}{\partial \mathbf{u} \, \partial \mathbf{u}'}\Bigg|_{\mathbf{u} = \boldsymbol{\phi}(\widehat{\boldsymbol{\sigma}}_\tau)}$$

with $\mathbf{u}' = [g(\boldsymbol{\sigma}), \mathbf{u}_2']$, and

$$\mathbf{H}_\tau^{-1} = -\frac{\partial^2 \left\{\log \, p(\boldsymbol{\sigma}|\mathbf{y}) + \lambda(g(\boldsymbol{\sigma}) - \tau)\right\}}{\partial \boldsymbol{\sigma} \, \partial \boldsymbol{\sigma}'}\Bigg|_{\boldsymbol{\sigma} = \widehat{\boldsymbol{\sigma}}_\tau}$$

Note that $\Omega_{\tau,22}^{-1}$ in equation [5] is equal to the diagonal block of Ω_τ^{-1} pertaining to $\mathbf{u}_2 = \boldsymbol{\phi}_2(\boldsymbol{\sigma})$.

Following Meyer and Smith (1996, equation 53) and using the fact that the gradient of $\log \, p(\boldsymbol{\phi}^{-1}(g(\boldsymbol{\sigma}), \mathbf{u}_2)|\mathbf{y}) + \lambda(g(\boldsymbol{\sigma}) - \tau)$ is zero at $\boldsymbol{\phi}(\widehat{\boldsymbol{\sigma}}_\tau)$, we have

$$\mathbf{H}_\tau^{-1} = \mathbf{J}'\Omega_\tau^{-1}\mathbf{J} \;\; \text{with} \;\; \mathbf{J}' = \mathbf{J}(\boldsymbol{\phi}^{-1}(\tau, \mathbf{u}_2))' = \left[\frac{\partial g(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}} \;\; \frac{\partial \boldsymbol{\phi}_2(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}}\right] = \left[\mathbf{d}_g \;\; \mathbf{D}_{\boldsymbol{\phi}_2}\right]$$

evaluated at $\boldsymbol{\sigma} = \widehat{\boldsymbol{\sigma}}_\tau$. This leads to:

$$\Omega_\tau = \mathbf{J}\mathbf{H}_\tau\mathbf{J}' = \begin{bmatrix} \mathbf{d}_g'\mathbf{H}_\tau\mathbf{d}_g & \mathbf{d}_g'\mathbf{H}_\tau\mathbf{D}_{\boldsymbol{\phi}_2} \\ \mathbf{D}_{\boldsymbol{\phi}_2}'\mathbf{H}_\tau\mathbf{d}_g & \mathbf{D}_{\boldsymbol{\phi}_2}'\mathbf{H}_\tau\mathbf{D}_{\boldsymbol{\phi}_2} \end{bmatrix} \qquad [7]$$

Using the formula for the determinant of the partitioned matrix in equation [7] (Searle, 1982) and recalling that $\Omega_{\tau,22}^{-1}$ is equal to the diagonal block of Ω_τ^{-1} pertaining to $\mathbf{u}_2 = \boldsymbol{\phi}_2(\boldsymbol{\sigma})$, the determinant of Ω_τ can be obtained as:

$$|\Omega_\tau| = |\mathbf{d}_g'\mathbf{H}_\tau\mathbf{d}_g||\mathbf{D}_{\boldsymbol{\phi}_2}'\mathbf{H}_\tau\mathbf{D}_{\boldsymbol{\phi}_2} - \mathbf{D}_{\boldsymbol{\phi}_2}'\mathbf{H}_\tau\mathbf{d}_g(\mathbf{d}_g'\mathbf{H}_\tau\mathbf{d}_g)^{-1}\mathbf{d}_g'\mathbf{H}_\tau\mathbf{D}_{\boldsymbol{\phi}_2}|$$

$$= |\mathbf{d}_g'\mathbf{H}_\tau\mathbf{d}_g||\Omega_{\tau,22}| \qquad [8]$$

Substituting $|\mathbf{\Omega_\tau}| = |\mathbf{H_\tau}||\mathbf{J}|^2$ which follows from equation [7] into equation [8] yields:

$$|\mathbf{\Omega}_{\tau,22}||\mathbf{J}|^{-2} = |\mathbf{H_\tau}||\mathbf{d}'_g\mathbf{H_\tau}\mathbf{d}_g|^{-1}$$

This equality can be used in equation [6] to obtain the following approximation of the marginal probability density of the nonlinear function $\tau = g(\mathbf{\sigma})$ (Tierney et al, 1989, 1991):

$$p(\tau|\mathbf{y}) \approx (2\pi)^{\frac{m-1}{2}} \cdot \left[\frac{|\mathbf{H_\tau}|}{\mathbf{d}'_g\mathbf{H_\tau}\mathbf{d}_g}\right]^{\frac{1}{2}} p(\widehat{\mathbf{\sigma}}_\tau|\mathbf{y}) \qquad [9]$$

Note that in contrast to equation [6] an explicit parameterization $\mathbf{\phi}_2$ is not needed for approximation [9]. To obtain the marginal posterior probability density of the genetic parameter $\tau = g(\mathbf{\sigma})$, equation [9] is repeatedly used for different values of τ.

In equation [9] observed second derivatives of log $p(\mathbf{\sigma}|\mathbf{y})$ are required. Model [1] allows a transformation to canonical scale, which facilitates the computation of derivatives. Colleau et al (1989) have given expressions on the canonical scale for the derivatives of the residual likelihood, which is one part of log $p(\mathbf{\sigma}|\mathbf{y})$. Derivatives of log $p(\mathbf{\sigma}|\mathbf{y})$ on the canonical scale and the transformation to the original scale are given in the *Appendix*.

Approximating determinant and traces

Evaluation of the marginal density [3], which is needed in equation [9], requires the computation of the determinant of the coefficient matrix of the mixed model equations for each canonical trait (*Appendix*). For the evaluation of derivatives of log $p(\mathbf{\sigma}|\mathbf{y})$ traces involving the inverse of the coefficient matrix of the mixed model equations on the canonical scale have to be computed (*Appendix*). The computation of these quantities becomes trivial once the eigenvalues, γ_i, of $\mathbf{L}'\mathbf{Z}'\mathbf{MZL}$ have been determined, where $\mathbf{L}$ is the Cholesky factor of $\mathbf{A}$, ie, $\mathbf{A} = \mathbf{LL}'$, and $\mathbf{M}$ is the absorption matrix, ie, $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. For the determinant we have (eg, Gianola et al, 1990a):

$$|\mathbf{W}'\mathbf{W} + \alpha_i\Sigma| \propto |\mathbf{L}'\mathbf{Z}'\mathbf{MZL} + \alpha_i\mathbf{I}_q| = \prod_{k=1}^{q}(\gamma_k + \alpha_i)$$

The traces can be obtained as (eg, Robert and Ducrocq, 1996):

$$\text{tr}\left[\mathbf{C}_i^{aa}\mathbf{A}^{-1}\right] = \text{tr}\left[(\mathbf{L}'\mathbf{Z}'\mathbf{MZL} + \alpha_i\mathbf{I}_q)^{-1}\right] = \sum_{k=1}^{q}\frac{1}{\gamma_k + \alpha_i}$$

$$\text{tr}\left[\mathbf{C}_i^{aa}\mathbf{A}^{-1}\mathbf{C}_j^{aa}\mathbf{A}^{-1}\right] = \text{tr}\left[(\mathbf{L}'\mathbf{Z}'\mathbf{MZL} + \alpha_i\mathbf{I}_q)^{-1}(\mathbf{L}'\mathbf{Z}'\mathbf{MZL} + \alpha_j\mathbf{J}_q)^{-1}\right]$$

$$= \sum_{k=1}^{q}\left(\frac{1}{\gamma_k + \alpha_i}\right)\left(\frac{1}{\gamma_k + \alpha_j}\right)$$

The procedure described by Robert and Ducrocq (1996) can be used to compute all eigenvalues of $\mathbf{L'Z'MZL}$. Here we only give a brief outline, for details see Robert and Ducrocq (1996) and the references cited therein. The procedure uses the Lanczos algorithm to transform the original matrix $\mathbf{B} = \mathbf{L'Z'MZL}$ of size q to a tridiagonal matrix $\mathbf{T}_k$ of a given size k, which in theory has the same eigenvalues as $\mathbf{B}$. A standard method can then be used to compute the eigenvalues of $\mathbf{T}_k$. The advantage of this algorithm is that memory requirements are minimal because matrix $\mathbf{B}$ is not altered. The matrix–vector product $\mathbf{Bv}$ needs to be computed repeatedly, which can be achieved very efficiently taking advantage of the special structure of $\mathbf{B}$ (Robert and Ducrocq, 1996).

Owing to rounding errors in the practical application, the resulting Lanczos matrix $\mathbf{T}_k$ is inaccurate. When the size k of $\mathbf{T}_k$ is increased, the eigenvalues of $\mathbf{T}_k$ provide increasingly accurate estimates of eigenvalues of $\mathbf{B}$. For $k > q$ $\mathbf{T}_k$ has 'good' eigenvalues which are true approximations of the distinct eigenvalues of $\mathbf{B}$ but also extra eigenvalues which are either copies of a good eigenvalue or spurious. A test is used to identify the spurious eigenvalues. A numerically multiple eigenvalue of $\mathbf{T}_k$ is accepted as an accurate approximation of an eigenvalue of $\mathbf{B}$. For the remaining good eigenvalues an error estimate is computed (Cullum and Willoughby, 1985).

The Lanczos algorithm is unable to determine the multiple eigenvalues and their multiplicities. The approach used by Robert and Ducrocq (1996) for identifying the multiple eigenvalues is based on the fact that if γ is a multiple eigenvalue of $\mathbf{B}$, then it will also be an eigenvalue of the matrix $\overline{\mathbf{B}}$ obtained from $\mathbf{B}$ by adding a symmetric rank one matrix, which is proportional to the product of the starting vector of the Lanczos algorithm and its transpose. Hence the Lanczos algorithm is applied to $\overline{\mathbf{B}}$ and the eigenvalues of the resulting Lanczos matrix $\overline{\mathbf{T}}_k$ that are also eigenvalues of $\mathbf{T}_k$ are identified as multiple eigenvalues of $\mathbf{B}$. Using the fact that the multiplicity m_1 of the zero eigenvalue is known to be $q - (n - \text{rank}(\mathbf{X}))$, the multiplicity m_i of the ith multiple eigenvalue γ_i can be obtained as a solution to the set of the following three equations:

$$\text{dim}(\mathbf{B}) = q = m_1 + \sum_{i=2}^{r_m} m_i + r_s \qquad [10]$$

$$\text{tr}(\mathbf{B}) = \sum_{i=2}^{r_m} m_i \gamma_i + \sum_{i=r_m+1}^{r_m+r_s} \gamma_i \qquad [11]$$

$$\text{tr}(\mathbf{B}^2) = \sum_{i=2}^{r_m} m_i \gamma_i^2 + \sum_{i=r_m+1}^{r_m+r_s} \gamma_i^2 \qquad [12]$$

subject to the constraint $m_i \geqslant 1$, where r_m is the number of multiple and r_s is the number of single eigenvalues of $\mathbf{B}$. Robert and Ducrocq (1996) successfully applied this method to a model with one fixed effect. Although the multiplicities determined were not consistent for different sizes of the Lanczos matrix, the traces were very well approximated using these multiplicities even for a Lanczos matrix of size $k = 2q$.

Gibbs sampling

The Gibbs sampler (Gelfand and Smith, 1990; Casella and George, 1992) generates samples from the joint posterior distribution by repeatedly sampling from the fully conditional distributions of the parameters. Fully conditional densities are derived from the joint posterior probability density given in equation [2] by collecting the terms that involve θ. These terms are proportional to the kernel of a multivariate normal density (eg, Gianola et al, 1990a):

$$p(\theta|\mathbf{y}, \mathbf{G}_0, \mathbf{R}_0) \propto \exp\left\{ -\frac{1}{2}(\theta - \widehat{\theta})'(\mathbf{R}_0^{-1} \otimes \mathbf{W}'\mathbf{W} + \mathbf{G}_0^{-1} \otimes \Sigma)(\theta - \widehat{\theta})\right\}$$

where, as before, $\widehat{\theta} = \left[(\mathbf{R}_0^{-1} \otimes \mathbf{W}'\mathbf{W}) + (\mathbf{G}_0^{-1} \otimes \Sigma)\right]^{-1}(\mathbf{R}_0^{-1} \otimes \mathbf{W}')\mathbf{y} = \mathbf{C}^{-1}\mathbf{r}$ is the solution to the mixed model equations. Therefore, obtaining a new realization of all location parameters jointly would require sampling from:

$$\theta|\mathbf{y}, \mathbf{G}_0, \mathbf{R}_0 \sim N(\widehat{\theta}, \mathbf{C}^{-1}) \tag{13}$$

which is impractical for larger problems. A computationally much less demanding alternative is to sample one location parameter at a time. Let θ_{-i} be θ without its ith component and $\mathbf{c}'_{-i}$ be the ith row of $\mathbf{C}$ without its ith element. Then, a new realization of the ith location parameter θ_i is obtained by sampling from (eg, Wang et al, 1994):

$$\theta_i|\mathbf{y}, \theta_{-i}, \mathbf{G}_0, \mathbf{R}_0 \sim N(\widetilde{\theta}_i, \widetilde{v}_i) \tag{14}$$

where $\widetilde{\theta}_i = (r_i - \mathbf{c}'_{-i}\theta_{-i})/c_{ii}$ and $\widetilde{v}_i = 1/c_{ii}$.

The disadvantage of using equation [14] is that the samples of subsequent cycles show higher correlations than if equation [13] were used. This leads to a slower mixing of the chain and therefore a lower effective number of independent samples (Sorensen et al, 1995), ie, a lower information content of a chain of given length (see below). Recently, García-Cortés and Sorensen (1996) have shown a way to sample from equation [13] that circumvents the need for the inverse of the coefficient matrix of the mixed model equations but requires a solution to the mixed model equations. With their simulated data for a single trait animal model, the effective number of independent samples for the additive genetic variance was only doubled using equation [13] instead of equation [14]. Because computing costs for equation [13] using the strategy proposed by García-Cortes and Sorensen (1996) are increased by a factor much larger than two it is concluded that equation [14] is the preferred way to sample the location parameters in single trait animal models. With the model considered here the canonical transformation can be used to transform the correlated traits to uncorrelated variables. This allows us to sample the location parameters on the canonical scale using equation [14], which is equivalent to jointly sampling the location parameters for all traits of one effect on the original scale. To improve mixing of the chain additive genetic values for sires and their final progeny can be sampled jointly as proposed by Janss et al (1995).

The fully conditional distribution for the genetic (co)variance matrix $\mathbf{G}_0$ is:

$$p(\mathbf{G}_0|\mathbf{y}, \boldsymbol{\theta}, v_G, \mathbf{V}_G) \propto |\mathbf{G}_0|^{-\frac{1}{2}(q+v_G+t+1)} \cdot \exp\left\{ -\frac{1}{2}\left[\boldsymbol{\theta}'(\mathbf{G}_0^{-1} \otimes \Sigma)\boldsymbol{\theta} + \mathrm{tr}(\mathbf{V}_G\mathbf{G}_0^{-1}) \right] \right\}$$

$$= |\mathbf{G}_0|^{-\frac{1}{2}(q+v_G+t+1)} \cdot \exp\left\{ -\frac{1}{2}\mathrm{tr}((\mathbf{V}_G + \mathbf{S}_a)\mathbf{G}_0^{-1}) \right\} \qquad [15]$$

using the fact that $\boldsymbol{\theta}'(\mathbf{G}_0^{-1} \otimes \Sigma)\boldsymbol{\theta} = \mathrm{tr}(\mathbf{S}_a\mathbf{G}_0^{-1})$, where $\mathbf{S}_a = \{\mathbf{a}_i'\mathbf{A}^{-1}\mathbf{a}_j\}$ for $i = 1, \ldots, t$ and $j = 1, \ldots, t$, and $\mathbf{a}_i$ is the vector of additive genetic effects for trait i. The conditional distribution is in the form of an inverse Wishart distribution with $q + v_G$ degrees of freedom and scale matrix $\mathbf{V}_G + \mathbf{S}_a$.

Similarly, the fully conditional distribution for the residual (co)variance matrix $\mathbf{R}_0$ is:

$$p(\mathbf{R}_0|\mathbf{y}, \boldsymbol{\theta}, v_R, \mathbf{V}_R) \propto |\mathbf{R}_0|^{-\frac{1}{2}(n+v_R+t+1)} \cdot \exp\left\{ -\frac{1}{2}\left[\mathbf{e}'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n)\mathbf{e} + \mathrm{tr}(\mathbf{V}_R\mathbf{R}_0^{-1}) \right] \right\}$$

$$= |\mathbf{R}_0|^{-\frac{1}{2}(n+v_R+t+1)} \cdot \exp\left\{ -\frac{1}{2}\mathrm{tr}\left[(\mathbf{V}_R + \mathbf{S}_e)\mathbf{R}_0^{-1} \right] \right\} \qquad [16]$$

where $\mathbf{e} = \mathbf{y} - (\mathbf{I}_t \otimes \mathbf{W})\boldsymbol{\theta}, \mathbf{S}_e = \{\mathbf{e}_i'\mathbf{e}_j\}$ for $i = 1, \ldots, t$ and $j = 1, \ldots, t$, and $\mathbf{e}_i$ is the subvector of $\mathbf{e}$ for trait i. This is an inverse Wishart distribution with $n+v_R$ degrees of freedom and scale matrix $\mathbf{V}_R + \mathbf{S}_e$. Sorensen (1996), among others, describes an algorithm for sampling from the inverse Wishart distribution.

One cycle of Gibbs sampling consists in drawing samples in turn from equation [14] for all location parameters followed by equations [15] and [16]. After convergence to the target distribution, the samples obtained are from the joint posterior distribution, or if only one parameter is considered, then samples are from the appropriate marginal distribution.

Samples from the marginal posterior density of the nonlinear function $g(\boldsymbol{\sigma})$ can be easily obtained by applying $g(\boldsymbol{\sigma})$ to the samples generated for $\boldsymbol{\sigma}$ in each cycle of the Gibbs sampler (eg, Wang et al, 1994).

APPLICATION

Data

The methodology was applied to data set DS2 of Ducrocq and Humblot (1995) with semen production traits of 1957 Normande bulls born between 1976 and 1986. The three traits considered were volume of the ejaculate (mL), motility score (on a scale from 0 to 4) and spermatozoa concentration(10^9/mL). The observations available were means of these traits for 11.4 ± 3.6 sperm collections. There were no missing observations. The means $\pm$ standard deviations were 3.08 ± 0.88, 2.99 ± 0.63 and 0.98 ± 0.26 for volume, motility and concentration, respectively. All known ancestors without records were added with the exception of ancestors that provided no additional information, ie, ancestors with both parents unknown and only one progeny. The total number of animals considered was 5 566. Further details about this data set are given by Ducrocq and Humblot (1995).

Model

The same model as considered by Ducrocq and Humblot (1995) was used for each of the three traits. The two fixed effects were birth year × birth trimester × station of performance test (65 classes) and the effect of total number of sperm collections (six classes: unknown, < 6, 7–9, 10–12, 13–15, > 15). The additive genetic effect was the only random effect. Because the number of sperm collections that made up a record did not vary much, and for simplicity, a constant residual variance was assumed for each of the three traits.

Proper inverse Wishart distributions were assumed as priors for the matrices of genetic and residual (co)variances. The degree of belief parameters were arbitrarily set to 20 for both prior distributions. The scale matrices were chosen such that the expected values correspond to the phenotypic standard deviations found in the raw data (see above) with heritabilities of 0.5 for all three traits and genetic and residual correlations of −0.2 between volume and motility as well as concentration and of 0.6 between motility and concentration. The expected prior heritability of 0.5 for means of about ten observations correspond to a heritability of around 0.1 for single observations. The chosen prior distribution reflects our prior belief based on literature values (Ouali, 1984; Taylor et al, 1985; Makulska et al, 1993).

Implementation of Laplace approximation

Eigenvalues of $\mathbf{L}'\mathbf{Z}'\mathbf{MZL}$ were computed using Lanczos matrices of size $2q$, $4q$, $8q$ and $12q$. Multiple eigenvalues and their multiplicities were determined using the procedure of Robert and Ducrocq (1996) described above. As observed already by Robert and Ducrocq (1996), eigenvalues were difficult to estimate in some intervals. With increasing size of the Lanczos matrix, new eigenvalues were detected in these intervals and the intervals with eigenvalues of low precision became smaller. These observations led to a new strategy where intervals with eigenvalues with high error estimates had the first eigenvalue with an error estimate $> 10^{-5}$ on each side of the interval arbitrarily declared as a multiple eigenvalue in order to compensate for the eigenvalues not yet detected in the interval. For these 'multiple' eigenvalues and the multiple eigenvalue of 0.5 found with the procedure of Robert and Ducrocq (1996) multiplicities were estimated with equations [10], [11] and [12], using a least squares procedure. Integer programming was also attempted but failed on all occasions.

Application of the formula for the Laplace approximation [9] requires maximization of $\log p(\boldsymbol{\sigma}|\mathbf{y})$ subject to the constraint $g(\boldsymbol{\sigma}) - \tau = 0$. The iterative algorithm used to solve this problem follows ideas developed by Bard (1974, ch 6-6). First, $\log p(\boldsymbol{\sigma}|\mathbf{y})$ is approximated by a second order Taylor series expansion around the current estimate $\widehat{\boldsymbol{\sigma}}^r$ at round r:

$$\widetilde{p} = \log p(\widehat{\boldsymbol{\theta}}^r|\mathbf{y}) + \left(\frac{\partial \log p(\boldsymbol{\sigma}|\mathbf{y})}{\partial \boldsymbol{\sigma}}\right)' \delta\boldsymbol{\sigma} + \frac{1}{2}\delta\boldsymbol{\sigma}'\frac{\partial^2 \log p(\boldsymbol{\sigma}|\mathbf{y})}{\partial \boldsymbol{\sigma}\,\partial \boldsymbol{\sigma}'}\delta\boldsymbol{\sigma}$$

where $\delta\boldsymbol{\sigma} = \boldsymbol{\sigma} - \widehat{\boldsymbol{\sigma}}^r$. Similarly, $g(\boldsymbol{\sigma}) - \tau$ is approximated by a first order Taylor series expansion:

$$\widetilde{g} = g(\widehat{\boldsymbol{\sigma}}^r) - \tau + \left(\frac{\partial g(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}}\right)' \delta\boldsymbol{\sigma}$$

Now, $\widetilde{p}$ is maximized subject to the constraint $\widetilde{g} = 0$, which is equivalent to finding the stationary point of the Lagrangian $\widetilde{p} + \lambda\widetilde{g}$. Setting the first derivatives of the Lagrangian with respect to $\delta\boldsymbol{\sigma}$ and λ to zero leads to the following system of equations:

$$-\begin{bmatrix} \dfrac{\partial^2 \log p(\boldsymbol{\sigma}|\mathbf{y})}{\partial\boldsymbol{\sigma}\,\partial\boldsymbol{\sigma}'} & \left(\dfrac{\partial g(\boldsymbol{\sigma})}{\partial\boldsymbol{\sigma}}\right)' \\[2mm] \dfrac{\partial g(\boldsymbol{\sigma})}{\partial\boldsymbol{\sigma}} & 0 \end{bmatrix} \begin{bmatrix} \delta\widehat{\boldsymbol{\sigma}}^{r+1} \\[2mm] \widehat{\lambda}^{r+1} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial \log p(\boldsymbol{\sigma}|\mathbf{y})}{\partial\boldsymbol{\sigma}} \\[2mm] g(\boldsymbol{\sigma}) - \tau \end{bmatrix} \qquad [17]$$

The derivatives are evaluated at the current value of $\widehat{\boldsymbol{\sigma}}^r$. The new solution is then found as $\widehat{\boldsymbol{\sigma}}^{r+1} = \widehat{\boldsymbol{\sigma}}^r + \delta\widehat{\boldsymbol{\sigma}}^{r+1}$. In our implementation observed second derivatives in equations [17] were replaced by expected second derivatives, which are faster to compute as will be shown below. Expected values for second derivatives on the canonical scale for the residual likelihood, which is one part of $\log p(\boldsymbol{\sigma}|\mathbf{y})$, were derived by the method of Meyer (1985). First derivatives and expected values of second derivatives of $\log p(\boldsymbol{\sigma}|\mathbf{y})$ on the canonical scale and the transformation to the original scale are given in the *Appendix*. Once the constraint was met $\log p(\boldsymbol{\sigma}|\mathbf{y})$ was computed for the proposed new solution and if necessary, step size was successively halved until the density was found to increase. Iterations were stopped when the maximum absolute change of heritability or genetic correlation estimates was less than 0.001. At convergence, observed second derivatives were computed for use in equation [9].

The computation of derivatives of $\log p(\boldsymbol{\sigma}|\mathbf{y})$ involves the solution to the mixed model equations on the canonical scale. Single trait mixed model equations were iteratively solved for each canonical trait by successive overrelaxation using a relaxation factor of 1.8 and solutions obtained in the previous round of the constrained maximization by equation [17] as starting values. The sum of squared differences between solutions in successive rounds of iterations divided by the sum of squared solutions in the current round was used as stopping criterion and was set to 10^{-12}.

The computation of observed second derivatives involves the quadratic form $\widehat{\mathbf{a}}_i' \mathbf{A}^{-1} \mathbf{C}_j^{aa} \mathbf{A}^{-1} \widehat{\mathbf{a}}_i$, where $\widehat{\mathbf{a}}_i$ is the solution to the mixed model equations for additive genetic effects for canonical trait i and $\mathbf{C}_j^{aa}$ is the submatrix of $(\mathbf{W}'\mathbf{W} + \alpha_j\boldsymbol{\Sigma})^{-1}$ (the inverse of the coefficient matrix of the mixed model equations for canonical trait j) pertaining to additive genetic effects. This quadratic form was evaluated by first computing $\widehat{\mathbf{u}}_i = \mathbf{A}^{-1}\widehat{\mathbf{a}}_i$, followed by iteratively solving the mixed model equations:

$$[\mathbf{W}'\mathbf{W} + \alpha_j\boldsymbol{\Sigma}]\begin{bmatrix} \widehat{\mathbf{w}}_b \\ \widehat{\mathbf{w}}_a \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \widehat{\mathbf{u}}_i \end{bmatrix}$$

then $\widehat{\mathbf{a}}_i' \mathbf{A}^{-1} \mathbf{C}_j^{aa} \mathbf{A}^{-1} \widehat{\mathbf{a}}_i = \widehat{\mathbf{u}}_i' \widehat{\mathbf{w}}_a$.

In addition to the Laplace approximation [9] a computationally less demanding approximation of the marginal posterior distribution was obtained by replacing the matrix of observed second derivatives in equation [9] by the matrix of expected second derivatives. This approximation will be called LaplaceE in the following.

For heritabilities (genetic correlations) both approximations were computed for 35 (55) values equally spaced by 0.02 that covered the expected range of non-negligible densities. For a more accurate location of the mode an additional three values spaced by 0.005 were evaluated in the intervals on each side of the value with the highest posterior probability density. The normalizing constant, the mean and quantiles of the marginal posterior densities were computed by simple numerical integration using the trapezoidal rule.

Implementation of Gibbs sampling

The Gibbs sampler was run in three parallel chains of 500 000 cycles each. At the start of each chain the (co)variance components were set to a random sample from their a priori distribution and the location parameters to the solution of the mixed model equations given the starting values for the (co)variance components. MZRAN of Marsaglia and Zaman (1994) was used as the basic random number generator, which has a period of the order of 2^{94}.

Because the starting point of the Gibbs sampler is chosen arbitrarily samples are not immediately from the desired target distribution. Therefore, the samples from the first cycles, the burn-in period, have to be discarded. In a preliminary analysis six parallel chains with extreme starting values were run for 10 000 cycles. Starting values were: all three heritabilities 0.05 or 0.95; within these two alternatives all three possible combinations of two genetic = residual correlations of +0.9 and one genetic = residual correlation of −0.9; variance of raw observations as phenotypic variance in all six chains. Visual inspection of the plots of samples for genetic parameters versus cycles indicated that after approximately 1 000 cycles convergence to the desired marginal posterior distributions was reached for all parameters. Based on these results the first 2 000 cycles of each of the three Gibbs chains were discarded.

Samples from subsequent Gibbs cycles are correlated. The higher the lag auto-covariances the lower the information content of the samples of a chain of given length. As suggested by Sorensen et al (1995) the 'initial positive sequence estimator' of Geyer (1992) was used to estimate the error variance of the marginal posterior mean and from this the effective chain length (ECL), ie, the number of independent samples that would result in the same error variance.

Marginal densities of heritabilities and genetic correlations were obtained by the normal kernel estimator (Silvermann, 1986, Wang et al, 1994). Let $\dot{\tau}_i = g(\dot{\sigma}_i)(i = 1, \ldots s)$ denote the samples of the parameter of interest $\tau = g(\sigma)$. Then the marginal posterior density can be estimated as:

$$\widehat{p}(\tau|\mathbf{y}) = \frac{1}{sh\sqrt{2\pi}} \sum_{i=1}^{s} \exp\left\{ -\frac{1}{2} \left(\frac{\tau - \dot{\tau}_i}{h} \right)^2 \right\}$$

where h is the window width. For all heritabilities and genetic correlations a window width of 0.01 was used and the posterior density computed for values equally spaced by 0.005 covering the whole range of the parameter space. Summary statistics of the marginal posterior distributions were computed by numerical integration using the estimated densities.

RESULTS

Approximation of determinant and traces

The eigenvalues determined as multiple and their estimated multiplicities are shown in table I. With the procedure of Robert and Ducrocq (1996) only the eigenvalue 0.5 was detected as multiple with all sizes of the Lanczos matrix. All other eigenvalues were determined as multiple with one size of the Lanczos matrix only. These results suggest that 0.5 is the only real multiple eigenvalue and that the others were falsely determined as multiple, probably due to a too loose criterion for comparing eigenvalues obtained from the two Lanczos matrices $\mathbf{T}$ and $\overline{\mathbf{T}}$. No attempt was made to change this criterion because it was also used for other comparisons within the Lanczos subroutine.

Table I. Multiple eigenvalues (γ) and their multiplicities (m) determined by the procedure of Robert and Ducrocq (1996) or by accepting only 0.5 as a multiple eigenvalue and choosing on each side of intervals with eigenvalues of low precision the first eigenvalue with an error estimate $> 10^{-5}$.

k^{a}	2q		4q		8q		12q	
	γ	m	γ	m	γ	m	γ	m
Robert and Ducrocq (1996)								
	0.500	885.0	0.500	7.5	0.500	276.2	0.500	267.9
			0.584	655.0	0.643	176.5	0.616	1.0
					0.695	1.0	0.655	169.0
					0.725	77.2		
0.5 & intervals of low precision								
	0.500	13.1	0.500	284.0	0.500	287.4	0.500	287.8
	0.583	738.4	0.652	181.9	0.666	140.1	0.667	94.3
	0.758	1.0	0.693	191.1	0.688	102.5	0.688	56.0
	0.816	135.5						
	0.846	1.0						

[a] Size of the Lanczos matrix defined as a multiple of the number of animals q.

In the alternative strategy, only 0.5 was accepted as a real multiple eigenvalue and the first eigenvalue with an error estimate $> 10^{-5}$ on each side of intervals with high error estimates were arbitrarily determined as 'multiple'. For the Lanczos matrix of size $2q$ two such intervals were found. For larger Lanczos matrices, only one of these intervals remained and became smaller with increasing size of the Lanczos matrix.

Table II shows the relative errors of the determinant and the traces obtained with increasing size of the Lanczos matrix for error to genetic variance ratios α of 99, 4 and 1/3 corresponding to heritabilities of 0.01, 0.20 and 0.75, respectively. For each strategy the values obtained with the Lanczos matrix of size $12q$ were regarded as exact. These values were very similar for the two procedures, the largest absolute relative difference being $< 2 \times 10^{-4}$ for the log-determinant and $\alpha = 1/3$. In general, absolute relative errors tended to increase with decreasing α. The new

Table II. Relative error for traces and determinant with different variance ratios α using eigenvalues obtained with Lanczos matrices of increasing size[a] and the multiplicities as shown in table I (relative error = deviation from value for $12q$ divided by absolute value for $12q$).

	α	$2q$	$4q$	$8q$	$12q$[b]
Robert and Ducrocq (1996)					
$\mathrm{tr}[B + \alpha I]$	99	2×10^{-4}	1×10^{-3}	1×10^{-5}	56.05
	4	4×10^{-3}	1×10^{-3}	3×10^{-5}	1315.33
	1/3	1×10^{-2}	1×10^{-3}	7×10^{-5}	12878.58
$\mathrm{tr}[(B + \alpha I)^2]$	99	4×10^{-4}	1×10^{-3}	1×10^{-5}	0.56
	4	7×10^{-3}	2×10^{-3}	4×10^{-5}	313.81
	1/3	8×10^{-3}	4×10^{-4}	7×10^{-5}	35012.51
$\log |B + \alpha I|$	99	-3×10^{-5}	1×10^{-3}	1×10^{-5}	25593.68
	4	-3×10^{-3}	1×10^{-3}	4×10^{-6}	8068.17
	1/3	-3×10^{-2}	-2×10^{-3}	-1×10^{-4}	-3901.86
0.5 & intervals of low precision					
$\mathrm{tr}[B + \alpha I]$	99	-4×10^{-10}	$<2 \times 10^{-10}$[c]	$<2 \times 10^{-10}$	56.05
	4	-2×10^{-6}	-3×10^{-8}	-3×10^{-9}	1315.36
	1/3	-1×10^{-4}	-2×10^{-6}	-2×10^{-7}	12879.59
$\mathrm{tr}[(B + \alpha I)^2]$	99	$<2 \times 10^{-8}$	$<2 \times 10^{-8}$	$<2 \times 10^{-8}$	0.56
	4	-7×10^{-6}	-1×10^{-7}	-1×10^{-8}	313.82
	1/3	-2×10^{-4}	-3×10^{-6}	-2×10^{-7}	35015.28
$\log |B + \alpha I|$	99	2×10^{-11}	$<4 \times 10^{-13}$	$<4 \times 10^{-13}$	25593.99
	4	5×10^{-7}	8×10^{-9}	7×10^{-10}	8068.19
	1/3	1×10^{-4}	2×10^{-6}	2×10^{-7}	-3902.46

[a] Size defined as a multiple of the number of animals q; [b] actual values obtained with a Lanczos matrix of size $12q$ regarded as exact; [c] no difference within eight decimal digits $\rightarrow$ absolute relative error $< 2 \times 10^{-10}$.

strategy resulted in more accurate approximations than the procedure of Robert and Ducrocq (1996). A Lanczos matrix of size $8q$ was required to obtain absolute relative errors $\leqslant 2 \times 10^{-4}$ with the procedure of Robert and Ducrocq (1996), whereas such accuracies were already obtained with a size of $2q$ using the new strategy. Eigenvalues computed with the new strategy and a Lanczos matrix of size $12q$ were used in the Laplace approximation.

Gibbs sampling

The lag-20 autocorrelations for samples of heritabilities and genetic correlations for the three Gibbs chains were all very high and ranged from 0.79 to 0.87 (table III). Lag correlations were highest for the heritability of motility and for the genetic correlation between volume and motility and lowest for the heritability of volume. Lag-1 autocorrelations were between 0.99 and 0.98 and lag-100 autocorrelations between 0.53 and 0.36 over all genetic parameters and all three chains (data not shown). The estimates of the standard deviation of the marginal posterior means

Table III. Lag-20 autocorrelations (lag20), estimates for the standard deviation of the marginal posterior mean (SDM) and effective chain length (ECL) for genetic parameters in three Gibbs sampling chains each of length 498 000.

Chain	Lag20 1	2	3	SDM 1	2	3	ECL 1	2	3
h^2 (V[a])	0.79	0.79	0.79	0.0013	0.0013	0.0013	2377	2648	2648
h^2 (M)	0.87	0.87	0.86	0.0017	0.0016	0.0016	1569	1725	1603
h^2 (C)	0.81	0.81	0.81	0.0013	0.0013	0.0012	2093	2223	2374
r_g (V–M)	0.86	0.86	0.86	0.0032	0.0032	0.0032	1593	1535	1594
r_g (V–C)	0.84	0.84	0.84	0.0025	0.0023	0.0024	1753	2030	1904
r_g (M–C)	0.85	0.85	0.84	0.0016	0.0017	0.0016	2031	1930	2057

[a] V, volume; M, motility; C, concentration.

(SDM) ranged from 0.001 to 0.003. The effective chain length (ECL) ranged from 1535 to 2648, which is by a factor of 324 to 188 smaller than the actual chain length.

Marginal posterior probability density of genetic parameters

Figures 1 and 2 show an excellent agreement of the marginal posterior densities for heritabilities and genetic correlations, respectively, obtained with Gibbs sampling and the two Laplace approximations. Summary statistics of the marginal posterior densities obtained with the different procedures are shown in table IV for heritabilities and in table V for genetic correlations. The variation among the estimates of the marginal mean for the three Gibbs chains is in agreement with the estimated SDM in table III. Estimates obtained from the three chains combined are based on 4 722 to 7 673 effective number of independent samples (table III) and are regarded here as exact for comparison with estimates obtained with Laplace approximations. Both Laplace and LaplaceE approximations yielded very accurate estimates of summary statistics. In general, estimates based on approximation [9] were more accurate than estimates using LaplaceE. For marginal means, the largest absolute differences between Laplace (LaplaceE) and Gibbs sampling was 0.002 (0.004) for heritabilities and 0.006 (0.007) for genetic correlations. Absolute differences in mode estimates for heritabilities were never larger than the 0.005 spacing of the grid used in the region of the mode and only slightly larger for genetic correlations. For marginal quantiles, the largest absolute difference between Laplace (LaplaceE) and Gibbs sampling was 0.004 (0.008) for heritabilities and 0.007 (0.009) for genetic correlations.

Computing costs

Memory requirements for the Laplace approximation and Gibbs sampling were similar. Computing time for the determination of the eigenvalues of $\mathbf{L'Z'MZL}$ were 0.3 or 2.7 h for Lanczos matrices of size $4q$ or $12q$, respectively. As already noted by Robert and Ducrocq (1996) the computing time for the Lanczos algorithm to
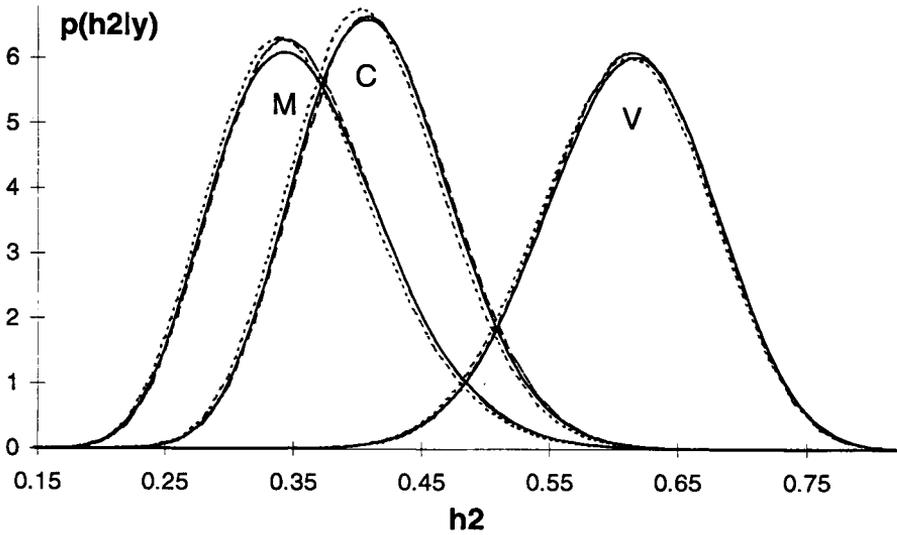
Fig 1. Marginal posterior densities of heritabilities for volume (V), motility (M) and concentration (C) obtained with Gibbs sampling (—), Laplace (- - -) or LaplaceE (···) approximation.
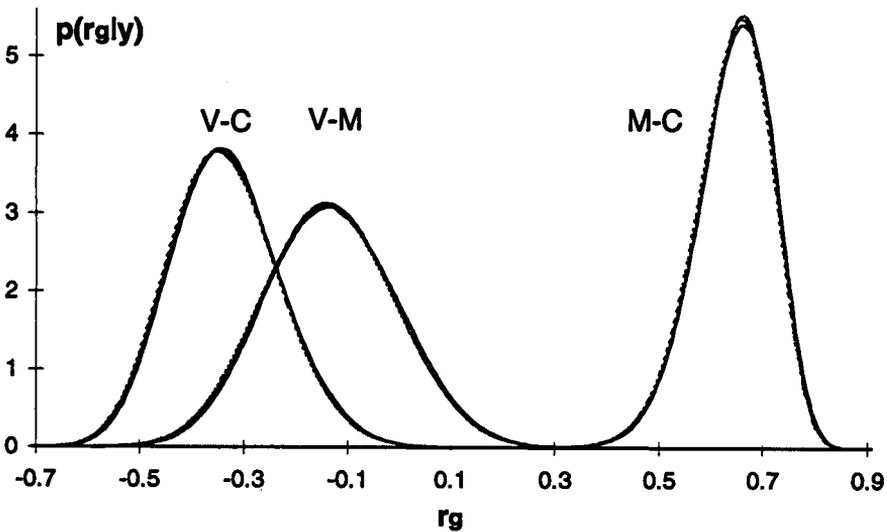
Fig 2. Marginal posterior densities of genetic correlations between the three traits volume (V), motility (M) and concentration (C) obtained with Gibbs sampling (—), Laplace (- - -) or LaplaceE (···) approximation.

Table IV. Summary statistics of marginal posterior distributions of heritabilities obtained with each of three chains of Gibbs sampling, the three chains combined, Laplace or LaplaceE approximation.

		Chain 1	Chain 2	Chain 3	Chains 1–3	Laplace	LaplaceE
h^2 (V[a])	mean	0.611	0.609	0.610	0.610	0.609	0.607
	mode[b]	0.615	0.615	0.620	0.615	0.615	0.610
	median	0.612	0.611	0.612	0.612	0.611	0.609
	5% quantile	0.500	0.498	0.497	0.497	0.500	0.496
	95% quantile	0.716	0.714	0.715	0.715	0.713	0.712
h^2 (M)	mean	0.355	0.359	0.360	0.358	0.359	0.355
	mode	0.340	0.340	0.345	0.345	0.345	0.340
	median	0.351	0.354	0.355	0.355	0.354	0.350
	5% quantile	0.256	0.260	0.261	0.261	0.261	0.257
	95% quantile	0.472	0.476	0.478	0.478	0.474	0.470
h^2 (C)	mean	0.412	0.415	0.415	0.414	0.416	0.410
	mode	0.405	0.405	0.415	0.405	0.410	0.400
	median	0.409	0.412	0.413	0.413	0.413	0.408
	5% quantile	0.318	0.321	0.319	0.319	0.321	0.317
	95% quantile	0.516	0.518	0.517	0.517	0.519	0.512

[a] V, volume; M, motility; C, concentration; [b] highest marginal posterior density on a 0.005 grid.

Table V. Summary statistics of marginal posterior distributions of genetic correlations obtained with each of three chains of Gibbs sampling, the three chains combined, Laplace or LaplaceE approximation.

		Chain 1	Chain 2	Chain 3	Chain 1–3	Laplace	LaplaceE
r_g (V–M[a])	mean	−0.130	−0.124	−0.128	−0.127	−0.133	−0.134
	mode[b]	−0.140	−0.130	−0.125	−0.130	− 0.140	−0.145
	median	−0.134	−0.128	−0.129	−0.130	−0.135	−0.136
	5% quantile	−0.333	−0.329	−0.332	−0.331	−0.336	−0.340
	95% quantile	0.083	0.089	0.083	0.085	0.078	0.079
r_g (V–C)	mean	−0.332	−0.328	−0.333	−0.331	−0.334	−0.337
	mode	−0.345	−0.335	−0.355	−0.345	−0.355	−0.355
	median	−0.337	−0.333	−0.337	−0.336	−0.338	−0.341
	5% quantile	−0.496	−0.492	−0.499	−0.496	−0.499	−0.502
	95% quantile	−0.149	−0.146	−0.152	−0.149	−0.155	−0.156
r_g (M–C)	mean	0.640	0.647	0.645	0.644	0.643	0.640
	mode	0.655	0.660	0.665	0.660	0.665	0.655
	median	0.645	0.652	0.650	0.649	0.649	0.646
	5% quantile	0.509	0.516	0.513	0.513	0.512	0.509
	95% quantile	0.755	0.758	0.755	0.756	0.753	0.752

[a] V, volume; M, motility; C, concentration; [b] highest marginal posterior density on a 0.005 grid.

determine $\mathbf{T}_k$ increases linearly but the time required for computing the eigenvalues of $\mathbf{T}_k$ increases quadratically with increasing k. The evaluation of the marginal posterior density at 41 values for heritabilities and 61 values for genetic correlations for the Laplace approximation required 0.4 and 0.5 h of CPU-time, respectively. Computing times for LaplaceE approximations were about 40% lower. Assuming that eigenvalues have to be determined by two (to identify real multiple eigenvalues) runs with a Lanczos matrix of size $4q$, total computing time for the approximation of the marginal densities of all genetic parameters amounts to 3.3 h. In contrast, generating the 500 000 samples of one Gibbs chain and post sampling analysis took 18 h of CPU-time, which adds up to 54 h for the three chains combined.

DISCUSSION AND CONCLUSIONS

Approximation of determinant and traces

Besides the zero multiple eigenvalue, only 0.5 was consistently detected as a multiple eigenvalue with different sizes of the Lanczos matrix in our application with two fixed effects. With a Lanczos matrix of size $2q$, 350 distinct eigenvalues were found in the interval [0.58, 0.76] with eigenvalues of low precision. In the same interval, additional 448 distinct eigenvalues were detected with a Lanczos matrix of size $12q$. Besides these 448 eigenvalues, only one eigenvalue in the other interval of low precision [0.81, 0.85] was additionally detected by increasing the size of the Lanczos matrix from $2q$ to $12q$. The adhoc declaration of two eigenvalues in the intervals of low precision as multiple with a Lanczos matrix of size $2q$ compensated well for the eigenvalues not yet detected leading to accurate approximations of traces and determinants. In contrast, much larger Lanczos matrices were required to achieve the same accuracy with the procedure used by Robert and Ducrocq (1996).

Robert and Ducrocq (1996) successfully applied their strategy to a model with one fixed effect and found accurate approximations of traces already with a Lanczos matrix of size $2q$. In their application, Robert and Ducrocq (1996) consistently found three multiple eigenvalues (0.5, 0.6875, 0.75), which corresponded to animals with certain genetic relationships that had records in the same level of the fixed effect. They report that distinct eigenvalues not yet detected with a Lanczos matrix of $4q$ were in the intervals [0.72, 0.75] and [0.84, 0.87]. Therefore, the estimated multiplicities, especially for the multiple eigenvalue of 0.75, has compensated for the eigenvalues not yet detected with small sizes of the Lanczos matrix, leading to very accurate approximations of traces with a Lanczos matrix of $4q$.

Marginal posterior probability density of genetic parameters

The Laplace approximation yielded very accurate approximations of the marginal posterior densities. As evident from equations [5] and [6], the approximation assumes normality, which is true only asymptotically. Our results suggest that the 1 957 animals with observations were a large enough sample for normality to be a good approximation. The performance of the Laplace approximation with smaller samples is not known and will be the subject of another study.

The computing time for the Laplace approximation of the marginal densities of all six genetic parameters was by a factor of 16 smaller than for Gibbs sampling.

Computing time for the Laplace approximation could be reduced by considering more sophisticated numerical integration techniques, eg, iterative Gauss-Hermite quadrature as used by Ducrocq and Casella (1996), which might require evaluation of the density for less values. It should be kept in mind that in this study the three combined Gibbs sampling chains were used to obtain an exact reference with small Monte-Carlo error to investigate the accuracy of the Laplace approximation. The small variation of all summary statistics among the three chains suggests that a combined ECL $> 1\,500$ is enough for obtaining marginal densities sufficiently accurate for most practical applications. With increasing sample size computing time is expected to increase nearly linearly for both procedures, Gibbs sampling and Laplace approximation, with the exception of the determination of the eigenvalues of the Lanczos matrix. However, increasing the number of traits will lead to a quadratic increase in computing time for the Laplace approximation if marginal densities are desired for all genetic parameters. For Gibbs sampling, the increase in computing time is expected to be nearly linear because canonical transformation can be used.

A simple multiple trait animal model with one random effect, no missing observations and equal incidence matrices was considered in this study. Hence, canonical transformation was applied and traces and determinants for canonical traits were easily computed once the eigenvalues of $\mathbf{L'Z'MZL}$ had been determined. Without these computational advantages, the application of the Laplace approximation is feasible only for small problems, since for each value where the marginal density is computed, the traces need to be evaluated repeatedly to locate the constrained mode and the determinant has to be evaluated once. The application of the computational strategies used in this study to multiple trait models with missing observations or models with more than one random effect is not possible. For the Laplace approximation to be applied to such models new efficient computing strategies have to be developed. Recent advances in the application of the canonical transformation to situations with missing observations for the solution of multivariate mixed model equations (eg, Ducrocq and Chapuis, 1997) suggest that further research towards an extension to multiple trait models with one random effect but missing observations seems most promising.

ACKNOWLEDGMENTS

REFERENCES

Bard Y (1974) *Nonlinear Parameter Estimation*. Academic Press, New York
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
Casella G, George EI (1992) Explaining the Gibbs sampler. *Am Stat* 46, 167–174
Colleau JJ, Beaumont C, Regaldo D (1989) Restricted maximum likelihood (REML) estimation of genetic parameters for type traits in Normande cattle breed. *Livest Prod Sci* 23, 47–66

Cullum JK, Willoughby RA (1985) *Lanczos Algorithms for Large Symmetric Eigenvalue Computations. Vol I: Theory, Vol II: Programs.* Birkhauser Bosten Inc, Boston

Ducrocq V, Humblot P (1995) Genetic characteristics and evolution of semen production of young Normande bulls. *Livest Prod Sci* 41, 1–10

Ducrocq V, Casella G (1996) A Bayesian analysis of mixed survival models. *Genet Sel Evol* 28, 505–529

Ducrocq V, Chapuis H (1997) Generalizing the use of the canonical transformation for the solution of multivariate mixed model equations. *Genet Sel Evol* 29, 205–224

García-Cortés LA, Sorensen D (1996) On a multivariate implementation of the Gibbs sampler. *Genet Sel Evol* 28, 121–126

Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85, 398–409

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian Data Analysis.* Chapman & Hall, London

Geyer CJ (1992) Practical Markov chain Monte-Carlo (with discussion). *Stat Sci* 7, 467–511

Gianola D, Fernando RL (1986) Bayesian methods in animal breeding theory. *J Anim Sci* 63, 217–344

Gianola D, Foulley JL (1990) Variance estimation from integrated likelihoods (VEIL). *Genet Sel Evol* 22, 403–417

Gianola D, Im S, Fernando RL, Foulley JL (1990a) Mixed model methodology and the Box-Cox theory of transformations: a Bayesian approach. In: *Advances in Statistical Methods for the Genetic Improvement of Livestock* (Gianola D, Hammond K, eds), Springer-Verlag, Berlin, 15–40

Gianola D, Im S, Macedo FW (1990b) A framework for prediction of breeding values. In: *Advances in Statistical Methods for the Genetic Improvement of Livestock* (Gianola D, Hammond K, eds), Springer-Verlag, Berlin, 211–238

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Introducing Markov chain Monte-Carlo. In: *Markov Chain Monte-Carlo in Practice* (Gilks WR, Richardson S, Spiegelhalter DJ, eds), Chapman & Hall, London, 1–19

Harville DA (1977) Maximum likelihood approaches to variance component estimation and related problems. *J Am Stat Assoc* 72, 320–338

Henderson CR (1973) Sire evaluation and genetic trend. In: *Proc Animal Breeding and Genetics Symp in Honor of Dr JL Lush.* ASAS-ADSA, Champaign, IL, USA, 10–41

Janss LLG, Thompson R, van Arendonk JAM (1995) Application of Gibbs sampling for inference in a mixed major gene - polygenic inheritance model in animal populations. *Theor Appl Genet* 91, 1137–1147

Makulska J, Hagger C, Künzi N, Kupferschmid HU (1993) Genetic and environmental influences on semen traits in AI bulls. *Reprod Dom Anim* 28, 279–284

Marsaglia G, Zaman A (1994) Some portable very-long-period random number generators. *Comput Phys* 8, 117–121

Meyer K (1985) Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* 41, 153–165

Meyer K, Smith SP (1996) Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genet Sel Evol* 28, 23–49

Ouali F (1984) Composante génétique de la fonction sexuelle: héritabilité des caractères du spermogramme et de la morphologie testiculaire chez les jeunes taurillons de race Montbéliarde. Mémoire MSc Vét, ENVA-UNCEIA, Maisons-Alfort, France

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

Robert C, Ducrocq V (1996) Computation of all eigenvalues of matrices used in restricted maximum likelihood estimation of variance components using sparse matrix techniques. *Genet Sel Evol* 28, 51–65

Searle SR (1982) *Matrix Algebra Useful for Statistics.* John Wiley & Sons, New York

Silvermann BW (1986) *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London

Sorensen D, Andersen S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. *Genet Sel Evol* 27, 229–249

Sorensen D (1996) Gibbs sampling in quantitative genetics. Internal report 82, Danish Institute of Animal Science, Tjele, Denmark

Taylor JF, Bean B, Marshall CE, Sullivan JJ (1985) Genetic and environmental components of semen production traits of artificial insemination Holstein bulls. *J Dairy Sci* 68, 2703–2722

Tempelman RJ, Gianola D (1993) Marginal maximum likelihood estimation of variance components in Poisson mixed models using Laplacian integration. *Genet Sel Evol* 25, 305–319

Tierney L, Kadane JB (1986) Accurate approximations for posterior moments and marginal densities. *J Am Stat Assoc* 81, 82–86

Tierney L, Kass RE, Kadane JB (1989) Approximate marginal densities of non-linear functions. *Biometrika* 76, 425–433

Tierney L, Kass RE, Kadane JB (1991) Approximate marginal densities of non-linear functions: Amendments and corrections. *Biometrika* 78, 233–234

Van Tassell CP, Van Vleck LD (1996) Multiple-trait Gibbs sampler for animal models: Flexible programs for Bayesian and likelihood- based (co)variance component inference. *J Anim Sci* 74, 2586–2597

Wang CS, Rutledge JJ, Gianola D (1994) Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet Sel Evol* 26, 91–115

APPENDIX: DERIVATIVES OF THE LOG POSTERIOR DENSITY ON THE CANONICAL SCALE

The log of the posterior probability density [3] is

$$
\begin{aligned}
\log \pi = \log p(\mathbf{G}_0, \mathbf{R}_0 | \mathbf{y}, v_G, \mathbf{S}_G, v_R, \mathbf{S}_R) \propto \\
& -\frac{1}{2}\left[\mathbf{y}'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n)\mathbf{y} - \widehat{\boldsymbol{\theta}}'(\mathbf{R}_0^{-1} \otimes \mathbf{W})\mathbf{y}\right] \\
& -\frac{1}{2}\log\left|(\mathbf{R}_0^{-1} \otimes \mathbf{W}'\mathbf{W}) + (\mathbf{G}_0^{-1} \otimes \Sigma)\right| \\
& -\frac{1}{2}(v_G + t + 1)\log|\mathbf{G}_0| - \frac{1}{2}\operatorname{tr}(v_G \mathbf{S}_G \mathbf{G}_0^{-1}) \\
& -\frac{1}{2}(v_R + t + 1)\log|\mathbf{R}_0| - \frac{1}{2}\operatorname{tr}(v_R \mathbf{S}_R \mathbf{R}_0^{-1})
\end{aligned}
$$

The canonical transformation transforms the t correlated traits into t uncorrelated variables. The transformation matrix $\mathbf{Q} = [\mathbf{q}_1, \ldots \mathbf{q}_t]$ satisfies $\mathbf{Q}'\mathbf{R}_0\mathbf{Q} = \mathbf{R}_0^c = \mathbf{I}_0$, and $\mathbf{Q}'\mathbf{G}_0\mathbf{Q} = \mathbf{G}_0^c = \operatorname{diag}\{\alpha_i^{-1}\}$. Let $\mathbf{y}^c = (\mathbf{Q}' \otimes \mathbf{I}_n)\mathbf{y}$ and $\boldsymbol{\theta}^c = (\mathbf{Q}' \otimes \mathbf{I}_n)\boldsymbol{\theta}$ be the vector of observations and location parameters on the canonical scale. The sub-vectors $\mathbf{y}_i^c$ and $\boldsymbol{\theta}_i^{c'} = [\mathbf{b}_i^{c'}, \mathbf{a}_i^{c'}]$ contain the observations and location parameters for

canonical trait i. Elements of $\mathbf{R}_0^c$ and $\mathbf{G}_0^c$ will be denoted by r_{ij}^c and g_{ij}^c, respectively. $\mathbf{C}_i^{aa}$ is the submatrix of $(\mathbf{W}'\mathbf{W} + \alpha_i\Sigma)^{-1}$ pertaining to additive genetic effects.

Note that

$$\log|(\mathbf{R}_0^{-1}\otimes\mathbf{W}'\mathbf{W} + (\mathbf{G}_0^{-1}\otimes\Sigma)| = 2(p+q)\cdot\log|\mathbf{Q}| + \log(\mathbf{I}_0\otimes\mathbf{W}'\mathbf{W} + (\mathbf{G}_0^{c^{-1}}\otimes\Sigma)|$$

$$= 2(p+q)\cdot\log|\mathbf{Q}| + \sum_{i=1}^{t}\log|\mathbf{W}'\mathbf{W} + \alpha_i\Sigma|$$

Following Colleau et al (1989) and adding contributions from the prior distribution, first derivatives on the canonical scale given the transformation matrix $\mathbf{Q}$ are as follows:

$$\frac{\partial\log\pi}{\partial g_{ii}^c} = \frac{1}{2}\alpha_i[\alpha_i\widehat{\mathbf{a}}_i^{c'}\mathbf{A}^{-1}\widehat{\mathbf{a}}_i^c + \alpha_i\mathrm{tr}(\mathbf{C}_i^{aa}\mathbf{A}^{-1}) - q + \alpha_i v_G\mathrm{tr}(\mathbf{S}_G\mathbf{q}_i\mathbf{q}_i') - (v_G + t + 1)]$$

$$\frac{\partial\log\pi}{\partial g_{ij}^c} = \alpha_i\alpha_j[\widehat{\mathbf{a}}_i^{c'}\mathbf{A}^{-1}\widehat{\mathbf{a}}_j^c + \frac{1}{2}v_G\mathrm{tr}(\mathbf{S}_G(\mathbf{q}_i\mathbf{q}_j' + \mathbf{q}_j\mathbf{q}_i'))]$$

$$\frac{\partial\log\pi}{\partial r_{ii}^c} = \frac{1}{2}[\mathbf{y}_i^{c'}\mathbf{y}_i^c - \mathbf{y}_i^{c'}\mathbf{W}\widehat{\theta}_i^c - \alpha_i\widehat{\mathbf{a}}_i^{c'}\mathbf{A}^{-1}\widehat{\mathbf{a}}_i^c - \alpha_i\mathrm{tr}(\mathbf{C}_i^{aa}\mathbf{A}^{-1})$$

$$+ p + q - n + v_R\mathrm{tr}(\mathbf{S}_R\mathbf{q}_i\mathbf{q}_i') - (v_R + t + 1)]$$

$$\frac{\partial\log\pi}{\partial r_{ij}^c} = \mathbf{y}_i^{c'}\mathbf{y}_j^c - \mathbf{y}_j^{c'}\mathbf{W}\widehat{\theta}_i^c - \alpha_i\widehat{\mathbf{a}}_i^{c'}\mathbf{A}^{-1}\widehat{\mathbf{a}}_j^c + \frac{1}{2}v_R\mathrm{tr}(\mathbf{S}_R(\mathbf{q}_i\mathbf{q}_j' + \mathbf{q}_j\mathbf{q}_i'))$$

Again following Colleau et al (1989) and adding contributions from the prior distribution, the non-zero observed second derivatives on the canonical scale are:

$$\frac{\partial^2\log\pi}{\partial^2 g_{ij}^c} = (1-\delta_{ij}/2)\alpha_i\alpha_j[q - \alpha_i\{\widehat{\mathbf{a}}_i^{c'}\mathbf{A}^{-1}\widehat{\mathbf{a}}_i^c + \mathrm{tr}(\mathbf{C}_i^{aa}\mathbf{A}^{-1})\}$$

$$- \alpha_j\{\widehat{\mathbf{a}}_j^{c'}\mathbf{A}^{-1}\widehat{\mathbf{a}}_j^c + \mathrm{tr}(\mathbf{C}_j^{aa}\mathbf{A}^{-1})\}$$

$$+ \alpha_i\alpha_j\{\widehat{\mathbf{a}}_i^{c'}\mathbf{A}^{-1}\mathbf{C}_j^{aa}\mathbf{A}^{-1}\widehat{\mathbf{a}}_i^c + \widehat{\mathbf{a}}_j^{c'}\mathbf{A}^{-1}\mathbf{C}_i^{aa}\mathbf{A}^{-1}\widehat{\mathbf{a}}_j^c$$

$$+ \mathrm{tr}(\mathbf{C}_i^{aa}\mathbf{A}^{-1}\mathbf{C}_j^{aa}\mathbf{A}^{-1})\}$$

$$- v_G\mathrm{tr}(\mathbf{S}_G(\alpha_i\mathbf{q}_i\mathbf{q}_i' + \alpha_j\mathbf{q}_j\mathbf{q}_j')) + (v_G + t + 1)]$$

where $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise.

$$\frac{\partial^2\log\pi}{\partial g_{ij}^c\partial g_{ik}^c} = -\frac{1}{2}\alpha_i\alpha_j\alpha_k v_G\,\mathrm{tr}(\mathbf{S}_G(\mathbf{q}_j\mathbf{q}_k' + \mathbf{q}_k\mathbf{q}_j'))$$

$$\frac{\partial^2\log\pi}{\partial g_{ij}^c\partial r_{ij}^c} = (1 - \delta_{ij}/2)\alpha_i\alpha_j[\frac{1}{2}\{\mathrm{tr}(\mathbf{C}_i^{aa}\mathbf{A}^{-1}) + \mathrm{tr}(\mathbf{C}_j^{aa}\mathbf{A}^{-1})$$

$$- (\alpha_i + \alpha_j)\mathrm{tr}(\mathbf{C}_i^{aa}\mathbf{A}^{-1}\mathbf{C}_j^{aa}\mathbf{A}^{-1})\} - \alpha_i\widehat{\mathbf{a}}_i^{c'}\mathbf{A}^{-1}\mathbf{C}_j^{aa}\mathbf{A}^{-1}\widehat{\mathbf{a}}_i^c$$

$$- \alpha_j\widehat{\mathbf{a}}_j^{c'}\mathbf{A}^{-1}\mathbf{C}_i^{aa}\mathbf{A}^{-1}\widehat{\mathbf{a}}_j^c]$$

$$\frac{\partial^2 \log \pi}{\partial^2 r_{ij}^c} = (1 - \delta_{ij}/2)[n - p - q + \alpha_i \alpha_j \mathrm{tr}(\mathbf{C}_i^{aa} \mathbf{A}^{-1} \mathbf{C}_j^{aa} \mathbf{A}^{-1})$$

$$- \mathbf{y}_i^{c'} \mathbf{y}_i^c - \mathbf{y}_j^{c'} \mathbf{y}_j^c + \mathbf{y}_i^{c'} \mathbf{W} \widehat{\boldsymbol{\theta}}_i^c + \mathbf{y}_j^{c'} \mathbf{W} \widehat{\boldsymbol{\theta}}_j^c$$

$$+ \alpha_i \widehat{\mathbf{a}}_i^{c'} \mathbf{A}^{-1} \widehat{\mathbf{a}}_i^c + \alpha_j \widehat{\mathbf{a}}_j^{c'} \mathbf{A}^{-1} \widehat{\mathbf{a}}_j^c + \alpha_i^2 \widehat{\mathbf{a}}_i^{c'} \mathbf{A}^{-1} \mathbf{C}_j^{aa} \mathbf{A}^{-1} \widehat{\mathbf{a}}_i^c$$

$$+ \alpha_j^2 \widehat{\mathbf{a}}_j^{c'} \mathbf{A}^{-1} \mathbf{C}_i^{aa} \mathbf{A}^{-1} \widehat{\mathbf{a}}_j^c - v_R \mathrm{tr}(\mathbf{S}_R(\mathbf{q}_i \mathbf{q}_i' + \mathbf{q}_j \mathbf{q}_j')) + (v_R + t + 1)]$$

$$\frac{\partial^2 \log \pi}{\partial r_{ij}^c \partial r_{ik}^c} = -\frac{1}{2} v_R \mathrm{tr}(\mathbf{S}_R(\mathbf{q}_j \mathbf{q}_k' + \mathbf{q}_k \mathbf{q}_j'))$$

Following Meyer (1985) and adding contributions from the prior distribution, non-zero expected values of second derivatives on the canonical scale are:

$$E \frac{\partial^2 \log \pi}{\partial^2 g_{ij}^c} = -(1 - \delta_{ij}/2)\alpha_i \alpha_j [q - \alpha_i \mathrm{tr}(\mathbf{C}_i^{aa} \mathbf{A}^{-1}) - \alpha_j \mathrm{tr}(\mathbf{C}_j^{aa} \mathbf{A}^{-1})$$

$$+ \alpha_i \alpha_j \mathrm{tr}(\mathbf{C}_i^{aa} \mathbf{A}^{-1} \mathbf{C}_j^{aa} \mathbf{A}^{-1}) + v_G \mathrm{tr}(\mathbf{S}_G(\alpha_i \mathbf{q}_i \mathbf{q}_i' + \alpha_j \mathbf{q}_j \mathbf{q}_j') - (v_G + t + 1)]$$

$$E \frac{\partial^2 \log \pi}{\partial g_{ij}^c \partial g_{ik}^c} = -\frac{1}{2} \alpha_i \alpha_j \alpha_k v_G \, \mathrm{tr}(\mathbf{S}_G(\mathbf{q}_j \mathbf{q}_k' + \mathbf{q}_k \mathbf{q}_j'))$$

$$E \frac{\partial^2 \log \pi}{\partial g_{ij}^c \partial r_{ij}^c} = -\frac{1}{2}(1 - \delta_{ij}/2)\alpha_i \alpha_j [\mathrm{tr}(\mathbf{C}_i^{aa} \mathbf{A}^{-1}) + \mathrm{tr}(\mathbf{C}_j^{aa} \mathbf{A}^{-1})$$

$$- (\alpha_i + \alpha_j)\mathrm{tr}(\mathbf{C}_i^{aa} \mathbf{A}^{-1} \mathbf{C}_j^{aa} \mathbf{A}^{-1})]$$

$$E \frac{\partial^2 \log \pi}{\partial^2 r_{ij}^c} = -(1 - \delta_{ij}/2)[n - p - q + \alpha_i \alpha_j \mathrm{tr}(\mathbf{C}_i^{aa} \mathbf{A}^{-1} \mathbf{C}_j^{aa} \mathbf{A}^{-1})$$

$$+ v_R \mathrm{tr}(\mathbf{S}_R(\mathbf{q}_i \mathbf{q}_i' + \mathbf{q}_j \mathbf{q}_j')) - (v_R + t + 1)]$$

$$E \frac{\partial^2 \log \pi}{\partial r_{ij}^c \partial r_{ik}^c} = -\frac{1}{2} v_R \mathrm{tr}(\mathbf{S}_R(\mathbf{q}_j \mathbf{q}_k' + \mathbf{q}_k \mathbf{q}_j'))$$

First and second derivatives on the original scale can be obtained by appropriate transformation of the first and second derivatives on the canonical scale, ie,

$$\frac{\partial \log \pi}{\partial \boldsymbol{\sigma}} = \mathbf{P}' \frac{\partial \log \pi}{\partial \boldsymbol{\sigma}^c}, \text{ and } \frac{\partial^2 \log \pi}{\partial \boldsymbol{\sigma} \partial \boldsymbol{\sigma}'} = \mathbf{P}' \frac{\partial^2 \log \pi}{\partial \boldsymbol{\sigma}^c \partial \boldsymbol{\sigma}^{c'}} \mathbf{P}, \text{ where the element of } \mathbf{P} \text{ in row } i$$
and column j is $\dfrac{\partial \sigma_i^c}{\partial \sigma_j}$.