

Marginal inferences about variance components in a mixed linear model using Gibbs sampling

CS Wang*, JJ Rutledge, D Gianola

*University of Wisconsin-Madison,
Department of Meat and Animal Science, Madison, WI 53706-1284, USA*

(Received 9 March 1992; accepted 7 October 1992)

Summary – Arguing from a Bayesian viewpoint, Gianola and Foulley (1990) derived a new method for estimation of variance components in a mixed linear model: variance estimation from integrated likelihoods (VEIL). Inference is based on the marginal posterior distribution of each of the variance components. Exact analysis requires numerical integration. In this paper, the Gibbs sampler, a numerical procedure for generating marginal distributions from conditional distributions, is employed to obtain marginal inferences about variance components in a general univariate mixed linear model. All needed conditional posterior distributions are derived. Examples based on simulated data sets containing varying amounts of information are presented for a one-way sire model. Estimates of the marginal densities of the variance components and of functions thereof are obtained, and the corresponding distributions are plotted. Numerical results with a balanced sire model suggest that convergence to the marginal posterior distributions is achieved with a Gibbs sequence length of 20, and that Gibbs sample sizes ranging from 300 – 3 000 may be needed to appropriately characterize the marginal distributions.

variance components / linear models / Bayesian methods / marginalization / Gibbs sampler

Résumé – *Inférences marginales sur des composantes de variance dans un modèle linéaire mixte à l'aide de l'échantillonnage de Gibbs. Partant d'un point de vue bayésien, Gianola et Foulley (1990) ont établi une nouvelle méthode d'estimation des composantes de variance dans un modèle linéaire mixte: estimation de variance par les vraisemblances intégrées (VEIL). L'inférence est basée sur la distribution marginale a posteriori de chacune des composantes de variance, ce qui oblige à des intégrations numériques pour arriver aux solutions exactes. Dans cet article, l'échantillonnage de Gibbs, qui est une procédure numérique pour générer des distributions marginales à partir de distributions*

* Correspondence and reprints. Present address: Department of Animal Science, Cornell University, Ithaca, NY 14853, USA

conditionnelles, est employé pour obtenir des inférences marginales sur des composantes de variance dans un modèle linéaire mixte univarié général. Toutes les distributions conditionnelles a posteriori nécessaires sont établies. Des exemples basés sur des données simulées contenant plus ou moins d'information sont présentés pour un modèle paternel à un facteur. Des estimées des densités marginales des composantes de variance et de fonctions de celles-ci sont obtenues, et les distributions correspondantes sont tracées. Les résultats numériques avec un modèle paternel équilibré suggèrent que la convergence vers les distributions marginales a posteriori est atteinte avec une séquence de Gibbs longue de 20 unités, et que des tailles de l'échantillon de Gibbs allant de 300 à 3 000 peuvent être nécessaires pour caractériser convenablement les distributions marginales.

composante de variance / modèle linéaire / méthode bayésienne / marginalisation / échantillonnage de Gibbs

INTRODUCTION

Variance components and functions thereof are important in quantitative genetics and other areas of statistical inquiry. Henderson's method 3 (Henderson, 1953) for estimating variance components was widely used until the late 1970's. With rapid advances in computing technology, likelihood based methods gained favor in animal breeding. Especially favored has been restricted maximum likelihood under normality, known as REML (Thompson, 1962; Patterson and Thompson, 1971). This method accounts for the degrees of freedom used in estimating fixed effects, which full maximum likelihood (ML) does not do.

ML estimates are obtained by maximizing the full likelihood, including its location variant part, while REML estimation is based on maximizing the "restricted" likelihood, *ie*, that part of the likelihood function independent of fixed effects. From a Bayesian viewpoint, REML estimates are the elements of the mode of the joint posterior density of all variance components when flat priors are employed for all parameters in the model (Harville, 1974). In REML, fixed effects are viewed as nuisance parameters and are integrated out from the posterior density of fixed effects and variance components, which is proportional to the full likelihood in this case.

There are at least 2 potential shortcomings of REML (Gianola and Foulley, 1990). First, REML estimates are the elements of the modal vector of the joint posterior distribution of the variance components. From a decision theoretic point of view, the optimum Bayes decision rule under quadratic loss in the posterior mean rather than the posterior mode. The mode of the marginal distribution of each variance component should provide a better approximation to the mean than a component of the joint mode. Second, if inferences about a single variance component are desired, the marginal distribution of this component should be used instead of the joint distribution of all components.

Gianola and Foulley (1990) proposed a new method that attempts to satisfy these considerations from a Bayesian perspective. Given the prior distributions and the likelihood which generates the data, the joint posterior distribution of all parameters is constructed. The marginal distribution of an individual variance component is obtained by integrating out all other parameters contained in the model. Summary

statistics, such as the mean, mode, median and variance can then be obtained from the marginal posterior distribution. Probability statements about a parameter can be made, and Bayesian confidence intervals can be constructed, thus providing a full Bayesian solution to the variance component estimation problem. In practice, however, this integration cannot be done analytically, and one must resort to numerical methods. Approximations to the marginal distributions were proposed (Gianola and Foulley, 1990), but the conditions required are often not met in data sets of small to moderate size. Hence, exact inference by numerical means is highly desirable.

Gibbs sampling (Geman and Geman, 1984) is a numerical integration method. It is based on all possible conditional posterior distributions, *ie*, the posterior distribution of each parameter given the data and all other parameters in the model. The method generates random drawings from the marginal posterior distributions through iteratively sampling from the conditional posterior distributions. Gelfand and Smith (1990) studied properties of the Gibbs sampler, and revealed its potential in statistics as a general numerical integration tool. In a subsequent paper (Gelfand *et al*, 1990), a number of applications of the Gibbs sampler were described, including a variance component problem for a one-way random effects model.

The objective of this paper is to extend the Gibbs sampling scheme to variance component estimation in a more general univariate mixed linear model. We first specify the Gibbs sampler in this setting and then use a sire model to illustrate the method in detail, employing 7 simulated data sets that encompass a range of parameter values. We also provide estimates of the posterior densities of variance components and of functions thereof, such as intraclass correlations and variance ratios.

SETTING

Model

Details of the model and definitions are found in Macedo and Gianola (1987), Gianola *et al* (1990a, b) and Gianola and Foulley (1990); only a summary is given here. Consider the univariate mixed linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i + \mathbf{e} \quad [1]$$

where: \mathbf{y} : data vector of order $n \times 1$; \mathbf{X} : known incidence matrix of order $n \times p$; \mathbf{Z}_i : known matrix of order $n \times q_i$; $\boldsymbol{\beta}$: $p \times 1$ vector of uniquely defined "fixed effects" (so that \mathbf{X} has full column rank); \mathbf{u}_i : $q_i \times 1$ "random" vector; and \mathbf{e}_i : $n \times 1$ vector of random residuals. The conditional distribution which generates the data is.

$$\mathbf{y} | \boldsymbol{\beta}, \mathbf{u}_1, \dots, \mathbf{u}_c, \sigma_e^2 \sim N(\mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i, \mathbf{R}\sigma_e^2) \quad [2]$$

where \mathbf{R} is an $n \times n$ known matrix, assumed to be an identity matrix here, and σ_e^2 is the variance of the random residuals.

Prior distributions

Prior distributions are needed to complete the Bayesian specification of the model. Usually, a “flat” prior distribution is assigned to $\boldsymbol{\beta}$, so as to represent lack of prior knowledge about this vector, so:

$$p(\boldsymbol{\beta}) \propto \text{constant} \quad [3]$$

Further

$$\mathbf{u}_i | \mathbf{G}_i, \sigma_{u_i}^2 \sim N_{q_i}(\mathbf{0}, \mathbf{G}_i \sigma_{u_i}^2), i = 1, 2, \dots, c \quad [4]$$

where \mathbf{G}_i is a known matrix and $\sigma_{u_i}^2$ is the variance of the prior distribution of \mathbf{u}_i . All \mathbf{u}_i 's are assumed to be mutually independent, *a priori*, as well as independent of $\boldsymbol{\beta}$.

Independent scaled inverted χ^2 distributions are used as priors for variance components, so that:

$$p(\sigma_e^2 | \nu_e, s_e^2) \propto (\sigma_e^2)^{-\nu_e/2-1} \exp\left(-\frac{1}{2}\nu_e s_e^2 / \sigma_e^2\right) \quad [5]$$

and

$$p(\sigma_{u_i}^2 | \nu_{u_i}, s_{u_i}^2) \propto (\sigma_{u_i}^2)^{-\nu_{u_i}/2-1} \exp\left(-\frac{1}{2}\nu_{u_i} s_{u_i}^2 / \sigma_{u_i}^2\right), i = 1, 2, \dots, c \quad [6]$$

Above $\nu_e(\nu_{u_i})$ is a “degree of belief” parameter, and $s_e^2(s_{u_i}^2)$ can be interpreted as a prior value of the appropriate variance. In this paper, as in Gelfand *et al* (1990) we assume the degree of belief parameters, ν_e and ν_{u_i} , to be zero to obtain the “naive” ignorance improper priors:

$$p(\sigma_e^2) \propto (\sigma_e^2)^{-1}; p(\sigma_{u_i}^2) \propto (\sigma_{u_i}^2)^{-1}, i = 1, 2, \dots, c \quad [7]$$

The joint prior density of $\boldsymbol{\beta}, \mathbf{u}_i (i = 1, 2, \dots, c), \sigma_{u_i}^2 (i = 1, 2, \dots, c)$ and σ_e^2 is the product of densities associated with [3–7], realizing that there are c random effects, each with their respective variances.

The joint posterior distribution resulting from priors [7] is improper mathematically, in the sense that it does not integrate to 1. The impropriety is due to [6], and it occurs at the tails. Numerical difficulties can arise when a variance component has a posterior distribution with appreciable density near 0. In this study, priors [7] were employed following Gelfand *et al* (1990), and difficulties were not encountered. However, informative or non informative priors other than [7] should be used in applications where it is postulated that at least one of the variance components is close to 0.

JOINT AND FULL CONDITIONAL POSTERIOR DISTRIBUTIONS

Denote $\mathbf{u}' = (\mathbf{u}'_1, \dots, \mathbf{u}'_c)$ and $\mathbf{v}' = (\sigma_{u_1}^2, \dots, \sigma_{u_c}^2)$. Let $\mathbf{u}'_{-i} = (\mathbf{u}'_1, \dots, \mathbf{u}'_{i-1}, \mathbf{u}'_{i+1}, \dots, \mathbf{u}'_c)$ and $\mathbf{v}'_{-i} = (\sigma_{u_1}^2, \dots, \sigma_{u_{i-1}}^2, \sigma_{u_{i+1}}^2, \dots, \sigma_{u_c}^2)$ be \mathbf{u}' and \mathbf{v}' with the i th element deleted from the set. The joint posterior distribution of the unknowns $(\boldsymbol{\beta}, \mathbf{u}, \mathbf{v}$ and $\sigma_e^2)$ is proportional to the product of the likelihood function and the joint prior distribution. As shown by Macedo and Gianola (1987) and Gianola *et al* (1990a, b), the joint posterior density is in the normal-gamma form:

$$p(\boldsymbol{\beta}, \mathbf{u}, \mathbf{v}, \sigma_e^2 | \mathbf{y}) \propto \frac{1}{(\sigma_e^2)^{n/2+1}} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i) \right\} \times \prod_{i=1}^c \left[\frac{1}{(\sigma_{u_i}^2)^{q_i/2+1}} \exp \left\{ -\frac{1}{2\sigma_{u_i}^2} \mathbf{u}'_i \mathbf{G}_i^{-1} \mathbf{u}_i \right\} \right]. \quad [8]$$

The full conditional density of each of the unknowns is obtained by regarding all other parameters in [8] as known. We then have:

$$p(\boldsymbol{\beta} | \mathbf{u}, \mathbf{v}, \sigma_e^2, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i - \mathbf{X}\boldsymbol{\beta}) \right\} \quad [9]$$

Manipulating [9] leads to

$$\boldsymbol{\beta} | \mathbf{u}, \mathbf{v}, \sigma_e^2, \mathbf{y} \sim N[\tilde{\boldsymbol{\beta}}, (\mathbf{X}'\mathbf{X})^{-1} \sigma_e^2],$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{y} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i)$. Note that this distribution does not depend on $\sigma_{u_i}^2$.

The full conditional distribution of each $\mathbf{u}_i (i = 1, 2, \dots, c)$ is multivariate normal:

$$\mathbf{u}_i | \boldsymbol{\beta}, \mathbf{u}_{-i}, \mathbf{v}, \sigma_e^2, \mathbf{y} \sim N \left[\tilde{\mathbf{u}}_i, \left(\mathbf{Z}'_i \mathbf{Z}_i + \mathbf{G}_i^{-1} \frac{\sigma_e^2}{\sigma_{u_i}^2} \right)^{-1} \sigma_e^2 \right], \quad [10]$$

where

$$\tilde{\mathbf{u}}_i = \left(\mathbf{Z}'_i \mathbf{Z}_i + \mathbf{G}_i^{-1} \frac{\sigma_e^2}{\sigma_{u_i}^2} \right)^{-1} \mathbf{Z}'_i (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{j=1, j \neq i}^c \mathbf{Z}_j \mathbf{u}_j), \quad i = 1, \dots, c$$

The full conditional density of σ_e^2 is in the scaled inverted χ^2 form:

$$p(\sigma_e^2 | \boldsymbol{\beta}, \mathbf{u}, \mathbf{v}, \mathbf{y}) \propto (\sigma_e^2)^{-n/2-1} \exp \left\{ -\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i) \right\}, \quad [11]$$

with parameters $\nu_e = n$ and $s_e^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i) / n$. Each full conditional density of $\sigma_{u_i}^2$ also is in the scaled inverted χ^2 form:

$$p(\sigma_{u_i}^2 | \boldsymbol{\beta}, \mathbf{u}, \mathbf{v}_{-i}, \sigma_e^2, \mathbf{y}) \propto (\sigma_{u_i}^2)^{-q_i/2-1} \exp \left(-\frac{1}{2\sigma_{u_i}^2} \mathbf{u}_i' \mathbf{G}_i^{-1} \mathbf{u}_i \right) \quad [12]$$

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

with parameters $\nu_{u_i} = q_i$ and $s_{u_i}^2 = \mathbf{u}_i' \mathbf{G}_i^{-1} \mathbf{u}_i / q_i$.

The full conditional distributions [9–12] are essential for implementing the Gibbs sampling scheme.

OBTAINING THE MARGINAL DISTRIBUTIONS USING GIBBS SAMPLING

Gibbs sampling

In many Bayesian problems, marginal distributions are often needed to make appropriate inferences. However, due to the complexity of joint posterior distributions obtaining a high degree of marginalization of the joint posterior density is difficult or impossible by analytical means. This is so for many practical problems, *eg* inferences about variance components. Numerical integration techniques must be used to obtain the exact marginal distributions, from which functions of interest can be computed and inferences made.

A numerical integration scheme known as Gibbs sampling (Geman and Geman, 1984; Gelfand and Smith, 1990; Gelfand *et al*, 1990; Casella and George, 1990) circumvents the analytical problem. The Gibbs sampler generates a random sample from a marginal distribution by successively sampling from the full conditional distributions of the random variables involved in the model. The full conditional distribution presented in the previous sections are summarized below:

$$p(\sigma_e^2 | \boldsymbol{\beta}, \mathbf{u}, \mathbf{v}, \mathbf{y}) : \text{multiple of inverted } \chi^2 \quad [13]$$

$$p(\sigma_{u_i}^2 | \boldsymbol{\beta}, \mathbf{u}, \mathbf{v}_{-i}, \sigma_e^2, \mathbf{y}), \quad i = 1, \dots, c : \text{multiple of inverted } \chi^2; \quad [14]$$

$$p(\mathbf{u}_i | \boldsymbol{\beta}, \mathbf{u}_{-i}, \mathbf{v}, \sigma_e^2, \mathbf{y}), \quad i = 1, \dots, c : \text{multivariate normal}; \quad [15]$$

and

$$p(\boldsymbol{\beta} | \mathbf{u}, \mathbf{v}, \sigma_e^2, \mathbf{y}) : \text{multivariate normal.} \quad [16]$$

Although we are interested in the marginal distributions of σ_e^2 and $\sigma_{u_i}^2$ only, all full conditional distributions are needed to implement the sampler. The ordering placed above is arbitrary. Gibbs sampling proceeds as follows:

- (i) set arbitrary initial values for $\boldsymbol{\beta}$, \mathbf{u} , \mathbf{v} , σ_e^2 ;
- (ii) generate σ_e^2 from [13], and update σ_e^2 ;
- (iii) generate $\sigma_{u_i}^2$ from [14], and update $\sigma_{u_i}^2$;
- (iv) generate \mathbf{u}_i from [15], and update \mathbf{u}_i ;
- (v) generate $\boldsymbol{\beta}$ from [16], and update $\boldsymbol{\beta}$; and
- (vi) repeat (ii–v) k times, using the updated values.

We call k the length of the Gibbs sequence, $Ask \rightarrow \infty$, the points from the k th iteration are sample points from the appropriate marginal distributions. The convergence of the samples from the above iteration scheme to drawings from the marginal distributions was established by Geman and Geman (1984) and restated by Gelfand and Smith (1990) and Tierney (1991). It should be noted that there are no approximations involved. Let the sample points be:

$(\sigma_e^2)^{(k)}$, $(\sigma_{u_i}^2)^{(k)}$ ($i = 1, 2, \dots, c$), $(\mathbf{u}_i)^{(k)}$ ($i = 1, 2, \dots, c$) and $(\boldsymbol{\beta})^{(k)}$ respectively, where superscript (k) denotes the k th iteration. Then:

(vii) Repeat (i–vi) m times, to generate m Gibbs samples. At this point we have:

$$(\sigma_e^2)_1^{(k)}, (\sigma_e^2)_2^{(k)}, \dots, (\sigma_e^2)_m^{(k)} \sim p(\sigma_e^2 | \mathbf{y}), \quad [17]$$

$$(\sigma_{u_i}^2)_1^{(k)}, (\sigma_{u_i}^2)_2^{(k)}, \dots, (\sigma_{u_i}^2)_m^{(k)} \sim p(\sigma_{u_i}^2 | \mathbf{y}), \quad i = 1, 2, \dots, c \quad [18]$$

$$(\mathbf{u}_i)_1^{(k)}, (\mathbf{u}_i)_2^{(k)}, \dots, (\mathbf{u}_i)_m^{(k)} \sim p(\mathbf{u}_i | \mathbf{y}), \quad i = 1, 2, \dots, c \quad [19]$$

$$(\boldsymbol{\beta})_1^{(k)}, (\boldsymbol{\beta})_2^{(k)}, \dots, (\boldsymbol{\beta})_m^{(k)} \sim p(\boldsymbol{\beta} | \mathbf{y}). \quad [20]$$

Because our interest is in making inferences about σ_e^2 and $\sigma_{u_i}^2$, no attention will be paid hereafter to \mathbf{u}_i and $\boldsymbol{\beta}$. However, it is clear that the marginal distributions of \mathbf{u}_i and $\boldsymbol{\beta}$ are also obtained as a byproduct of Gibbs sampling.

Density estimation

After samples from the marginal distributions are generated, one can estimate the densities using these samples and the full conditional densities. As noted by Casella and George (1990) and Gelfand and Smith (1990), the marginal density of a random variable x can be written as:

$$\hat{p}(x) = \int p(x|y)p(y)dy = E[p(x|y)].$$

An estimator of $p(x)$ is:

$$\hat{p}(x) = \frac{1}{m} \sum_{i=1}^m p(x|y_i).$$

Thus, the estimator of the marginal density of σ_e^2 is:

$$\hat{p}(\sigma_e^2|\mathbf{y}) = \frac{1}{m} \sum_{j=1}^m p(\sigma_e^2 | (\boldsymbol{\beta})_j^{(k)}, (\mathbf{u})_j^{(k)}, (\mathbf{v})_j^{(k)}, \mathbf{y}), \quad [21]$$

where $(\mathbf{u}')_j^{(k)} = ((\mathbf{u}'_1)_j^{(k)}, \dots, (\mathbf{u}'_c)_j^{(k)})$, and
 $(\mathbf{v}')_j^{(k)} = ((\sigma_{u_1}^2)_j^{(k)}, \dots, (\sigma_{u_c}^2)_j^{(k)})$.

The estimated values of the density are thus obtained by fixing σ_e^2 (at a number of points over its space), and then evaluating [21] at each point. Similarly, the estimator of the marginal density of $\sigma_{u_i}^2$ is:

$$\hat{p}(\sigma_{u_i}^2|\mathbf{y}) = \frac{1}{m} \sum_{j=1}^m p(\sigma_{u_i}^2 | (\boldsymbol{\beta})_j^{(k)}, (\mathbf{u})_j^{(k)}, (\mathbf{v}_{-i})_j^{(k)}, \mathbf{y}), i = 1, 2, \dots, c \quad [22]$$

where $(\mathbf{v}'_{-i})_j^{(k)} = ((\sigma_{u_1}^2)_j^{(k)}, \dots, (\sigma_{u_{i-1}}^2)_j^{(k)}, (\sigma_{u_{i+1}}^2)_j^{(k)}, \dots, (\sigma_{u_c}^2)_j^{(k)})$.

Additional discussion about mixture density estimators is found in Gelfand and Smith (1990).

Estimation of the density of a function of the variance components is accomplished by applying theory of transformations of random variables to the estimated densities, with minimal additional calculations. Examples of estimating the densities of variance ratios and of an intraclass correlation are given later.

APPLICATION OF GIBBS SAMPLING TO THE ONE-WAY CLASSIFICATION

Model

We consider the one-way linear model:

$$y_{ij} = \beta + u_i + e_{ij}, \quad i = 1, 2, \dots, q; \quad j = 1, 2, \dots, n_i \quad \text{and} \quad N = \sum_{i=1}^q n_i, \quad [23]$$

where β is a "fixed" effect common to all observations, u_i could be, for example, sire effects, and e_{ij} is a residual associated with the record on the j th progeny of sire i . It is assumed that:

$$y_{ij} | \beta, u_i \sim NiD(\beta + u_i, \sigma_e^2) \quad [24]$$

and that

$$u_i | \sigma_u^2 \sim NiiD(0, \sigma_u^2) \quad [25]$$

where NiD and $NiiD$ stand for "normal, independently distributed" and "normal, independently and identically distributed", respectively.

Conditional distributions

For this model, the parameters of the normal conditional distribution of $\beta|\mathbf{u}, \sigma_u^2, \sigma_e^2, \mathbf{y}$ in [9] are:

$$\tilde{\beta} = \frac{1}{N} \left(\sum_{i=1}^q \sum_{j=1}^{n_i} y_{ij} - \sum_{i=1}^q n_i u_i \right). \quad [26]$$

and

$$(\mathbf{X}'\mathbf{X})^{-1}\sigma_e^2 = \frac{1}{N} \sigma_e^2 \quad [27]$$

Likewise, the parameters of the normal conditional distribution of $\mathbf{u}|\beta, \sigma_u^2, \sigma_e^2, \mathbf{y}$ in [10] are:

$$\{\tilde{u}_i\} = \left\{ \frac{n_i}{n_i + \alpha} (\bar{y}_i - \beta) \right\}, \quad i = 1, 2, \dots, q \quad [28]$$

with $\alpha = \sigma_e^2/\sigma_u^2$, and the covariance matrix is:

$$\text{Diag} \{c_{ii}\sigma_e^2\} \quad [29]$$

with $c_{ii} = 1/(n_i + \alpha)$. Because the covariance matrix is diagonal, each u_i can be generated independently as:

$$u_i|\beta, \sigma_u^2, \sigma_e^2, \mathbf{y} \sim N(\tilde{u}_i, c_{ii}\sigma_e^2) \quad [30]$$

The conditional density of σ_e^2 in [11] can be written as:

$$p(\sigma_e^2|\beta, \mathbf{u}, \sigma_u^2, \mathbf{y}) \propto (\sigma_e^2)^{-N/2-1} \exp \left\{ -\frac{1}{2\sigma_e^2} N s_e^2 \right\} \quad [31]$$

where:

$$s_e^2 = \sum_{i=1}^q \sum_{j=1}^{n_i} (y_{ij} - \beta - u_i)^2 / N \quad [32]$$

Because $N s_e^2 / \sigma_e^2 \sim \chi_N^2$, it follows that $\sigma_e^2 \sim N s_e^2 \chi_N^{-2}$, so [31] is the kernel of a multiple of an inverted χ^2 random variable.

Finally, the conditional density of σ_u^2 in [12] is expressible as:

$$p(\sigma_u^2|\beta, \mathbf{u}, \sigma_e^2, \mathbf{y}) \sim (\sigma_u^2)^{-q/2-1} \exp \left\{ -\frac{1}{2\sigma_u^2} q s_u^2 \right\} \quad [33]$$

with:

$$s_u^2 = \sum_{i=1}^q u_i^2 / q \quad [34]$$

Since $q s_u^2 / \sigma_u^2 \sim \chi_q^2$, then $\sigma_u^2 \sim q s_u^2 \chi_q^{-2}$, also a multiple of an inverted χ^2 variable.

Data sets and designs

Seven data sets (experiments) were simulated, so as to represent situations that differ in the amount of statistical information. Essentials of the experimental designs are in table I. Number of sire families (q) varied from 10 to 10 000, while number of progeny per family (n) ranged from 5 to 20. The smallest experiment was I, with 10 sires and 5 progeny per sire; the largest one was VII, with a total of 100 000 records. Only balanced designs ($n_i = n$, for $i = 1, 2, \dots, q$) were reported here, as similar results were found with unbalanced layouts. Data were randomly generated using parameter values of 0 and 1 for β and σ_u^2 , respectively. Parametric values for σ_e^2 were from 1 to 99, thus yielding intraclass correlations (ρ) ranging from 0.01 to 0.5. From a genetic point of view, an intraclass correlation of 0.5 (Data set I) is not possible in a sire model, but it is reported here for completeness.

Table I. Design of experiments simulated

| <i>Data set</i> | q | n | ρ | m | k | <i>CPU time</i> |
|-----------------|--------|-----|--------|-------|-----|-----------------|
| I | 10 | 5 | 0.5 | 600 | 30 | 2.39 |
| | | | | 300 | 20 | 0.80 |
| II | 50 | 10 | 0.2 | 1 000 | 20 | 10.41 |
| | | | | 300 | 20 | 3.13 |
| III | 500 | 20 | 0.1 | 1 000 | 10 | 43.80 |
| | | | | 500 | 50 | 104.67 |
| IV | 1 000 | 20 | 0.05 | 300 | 30 | 75.92 |
| V | 50 | 20 | 0.025 | 3 000 | 20 | 31.38 |
| | | | | 2 000 | 20 | 20.96 |
| VI | 50 | 20 | 0.01 | 3 000 | 20 | 31.41 |
| | | | | 300 | 20 | 3.14 |
| VII | 10 000 | 10 | 0.01 | 2 000 | 20 | 3 454.40 |

q = Number of families; n = number of observations per family; ρ = parametric value of intraclass correlation; m = Gibbs sample size; k = length of the Gibbs sequence; and CPU time = CPU time in s elapsed for generation of samples of [35–38], using the CRAY Y-MP8/864 at the San Diego Supercomputer Center.

The Gibbs sampler [13–16] was run at varying lengths of the Gibbs sequence ($k = 10$ to 100) and Gibbs sample sizes ($m = 300$ to 3 000), to assess the effects of k and m on the estimated marginal distributions. FORTRAN subroutines of the IMSL (IMSL Inc, 1989) were used to generate normal and inverted χ^2 random deviates. At the end of each run, the following quantities were retained:

$$(\sigma_e^2)_1^{(k)}, (\sigma_e^2)_2^{(k)}, \dots, (\sigma_e^2)_m^{(k)} \sim p(\sigma_e^2 | \mathbf{y}), \quad [35]$$

$$(\sigma_u^2)_1^{(k)}, (\sigma_u^2)_2^{(k)}, \dots, (\sigma_u^2)_m^{(k)} \sim p(\sigma_u^2 | \mathbf{y}), \quad [36]$$

$$(s_e^2)_1^{(k)}, (s_e^2)_2^{(k)}, \dots, (s_e^2)_m^{(k)}, \text{ and} \quad [37]$$

$$(s_u^2)_1^{(k)}, (s_u^2)_2^{(k)}, \dots, (s_u^2)_m^{(k)}. \quad [38]$$

Sample points in [37] and [38], are needed for density estimation, as noted below.

Marginal density estimation

The estimators of the marginal densities of σ_e^2 and σ_u^2 were:

$$\hat{p}(\sigma_e^2|\mathbf{y}) = \frac{(N/2)^{N/2}}{\Gamma(N/2)} (\sigma_e^2)^{-N/2-1} \frac{1}{m} \sum_{j=1}^m \left((s_e^2)_j^{(k)} \right)^{N/2} \exp \left(-\frac{1}{2} N (s_e^2)_j^{(k)} / \sigma_e^2 \right) \quad [39]$$

and

$$\hat{p}(\sigma_u^2|\mathbf{y}) = \frac{(q/2)^{q/2}}{\Gamma(q/2)} (\sigma_u^2)^{-q/2-1} \frac{1}{m} \sum_{j=1}^m \left((s_u^2)_j^{(k)} \right)^{q/2} \exp \left(-\frac{1}{2} q (s_u^2)_j^{(k)} / \sigma_u^2 \right). \quad [40]$$

Using the theory of transformation of random variables, we obtained the marginal density of the variance ratio $\gamma = \sigma_u^2/\sigma_e^2$ by fixing σ_e^2 , using [40] as:

$$\hat{p}(\gamma|\mathbf{y}) = \frac{(q/2)^{q/2}}{\Gamma(q/2)} \gamma^{-q/2-1} \frac{1}{m} \sum_{j=1}^m \left((\sigma_e^2)_j^{(k)} \right)^{-q/2} \left((s_u^2)_j^{(k)} \right)^{q/2} \exp \left(-\frac{1}{2\gamma} q (s_u^2)_j^{(k)} / (\sigma_e^2)_j^{(k)} \right). \quad [41]$$

If inferences are to be made about the intraclass correlation $\rho = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$, the Jacobian of the transformation from σ_u^2 to ρ is $J = \sigma_e^2/(1 - \rho)^2$, so from [40], considering σ_e^2 as fixed, the marginal density of the intraclass correlation is:

$$\hat{p}(\rho|\mathbf{y}) = \frac{(q/2)^{q/2}}{\Gamma(q/2)} \rho^{-q/2-1} (1 - \rho)^{q/2-1} \frac{1}{m} \sum_{j=1}^m \left((\sigma_e^2)_j^{(k)} \right)^{-q/2} \left((s_u^2)_j^{(k)} \right)^{-q/2} \exp \left(-\frac{1 - \rho}{2\rho} q (s_u^2)_j^{(k)} / (\sigma_e^2)_j^{(k)} \right). \quad [42]$$

Densities of any functions of the variance components can be estimated in a similar manner. Density [39] can also be used to make the transformations. Note that the Gibbs sampler [13–16] does not need to be run again to obtain the densities of the functions of variance components.

Plots were generated using densities [39–42] by selecting 50 to 100 equally spaced points in the “effective” range of the variables; the “effective” range of the variable covers at least 99.5% of the density mass. Summary statistics of the posterior distributions such as mean, mode, median and variance were calculated using the composite Simpson’s rules of numerical integration by dividing the effective range of the variables into 100 to 200 equally spaced intervals. Where appropriate, summary statistics were calculated using the densities estimated at higher values of m or k .

RESULTS

The estimated marginal densities of variance components and of their functions (γ and ρ) are depicted in figures 1 to 7, in which solid curves correspond to densities estimated at higher values of m or k and *vice versa* for the dotted ones. These figures correspond to the 7 designs given in table I. Figure 1 represents a situation where 50% of the total variance is “due to sires”, *ie*, a high intraclass correlation; as noted earlier, this is not possible genetically. All posterior distributions were unimodal, and convergence of the Gibbs sampling scheme to the appropriate marginals was achieved with $k = 20$ and $k = 300$, as it can be ascertained from direct inspection of the curves. Because of the limited information contained in data set I ($q = 10$, $n = 5$), posterior densities were not symmetric, so the mean, mode and median differ. The median was closer to the true values of the parameters than the mean and the mode; this was true for all 4 distributions considered.

For data sets II–IV, with heritabilities ranging from 20–80%, and number of sires from 50 to 1 000, the posterior densities (figs 2–4) were nearly symmetric, so the 3 location statistics were very similar to each other. In figure 4, corresponding to $\sigma_u^2 = 1$, $\sigma_e^2 = 10$ and to a data set with 1 000 sires and a total of 20 000 records, the posterior coefficients of variation were approximately 1% for σ_e^2 and 9% for the remaining parameters. This illustrates the well known result that σ_e^2 is less difficult to estimate than σ_u^2 or functions thereof. The plots suggest convergence of the Gibbs sampler at values of k as low as 10–20.

Some difficulties were encountered with the Gibbs sampling schemes in designs V and VI (fig 5 and 6, respectively). These designs correspond to situations of low heritability and of mild information about parameters contained in the data. While there was no problem in general with the posterior distribution of σ_e^2 , this was not so for the remaining 3 parameters. Typical problems were bi-modality or lack of smoothness in the left tail of the estimated densities. These were found to be related to insufficient Gibbs sample size, and were corrected by increasing the Gibbs sampling size (m). Compare, for example, the dotted ($m = 300$) with the solid ($m = 3 000$) curves in figure 6. A more awkward distribution requires more samples to be characterized accurately. Recall that each of the density estimators [40]–[42] was obtained by averaging m conditional densities. In the case of badly behaved marginal distributions, which are associated with low heritability and small data sets, it is possible to obtain “outlier conditional densities”. The outliers can be influential and distort the values of the estimated density unless m is large enough. It is clear from the figures that smooth posterior estimated densities for awkward distributions, *eg*, for heritability as low as 4%, can be obtained by increasing Gibbs sample size, albeit at computational expense. At this level of heritability, when the number of sire families was increased to 10 000, posterior distributions were nearly symmetric (fig 7), and there was little variability about the parametric values.

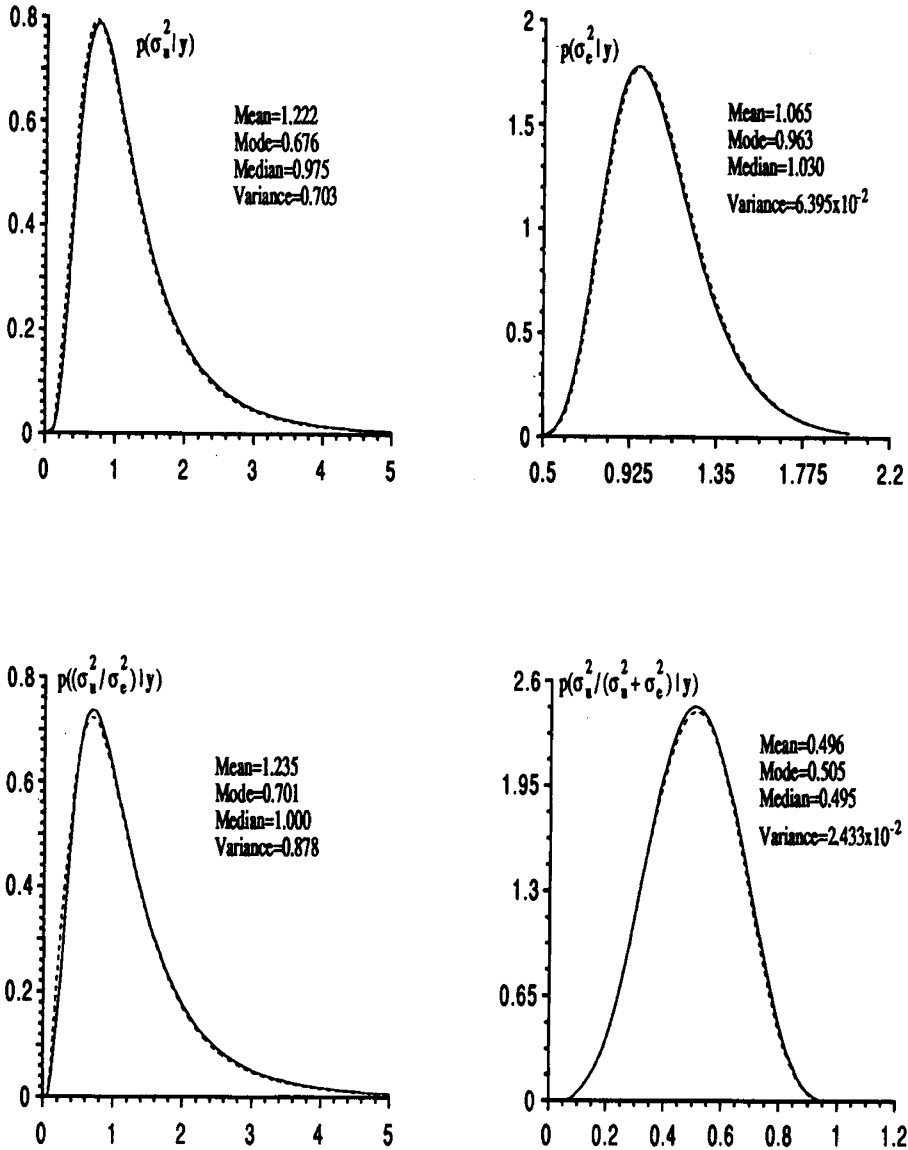


Fig 1. Marginal densities of variance components from data set I. The dotted curve was generated with $k = 20$ and $m = 300$, while the solid curve was with $k = 30$ and $m = 600$. See footnotes to table I for definitions of symbols.

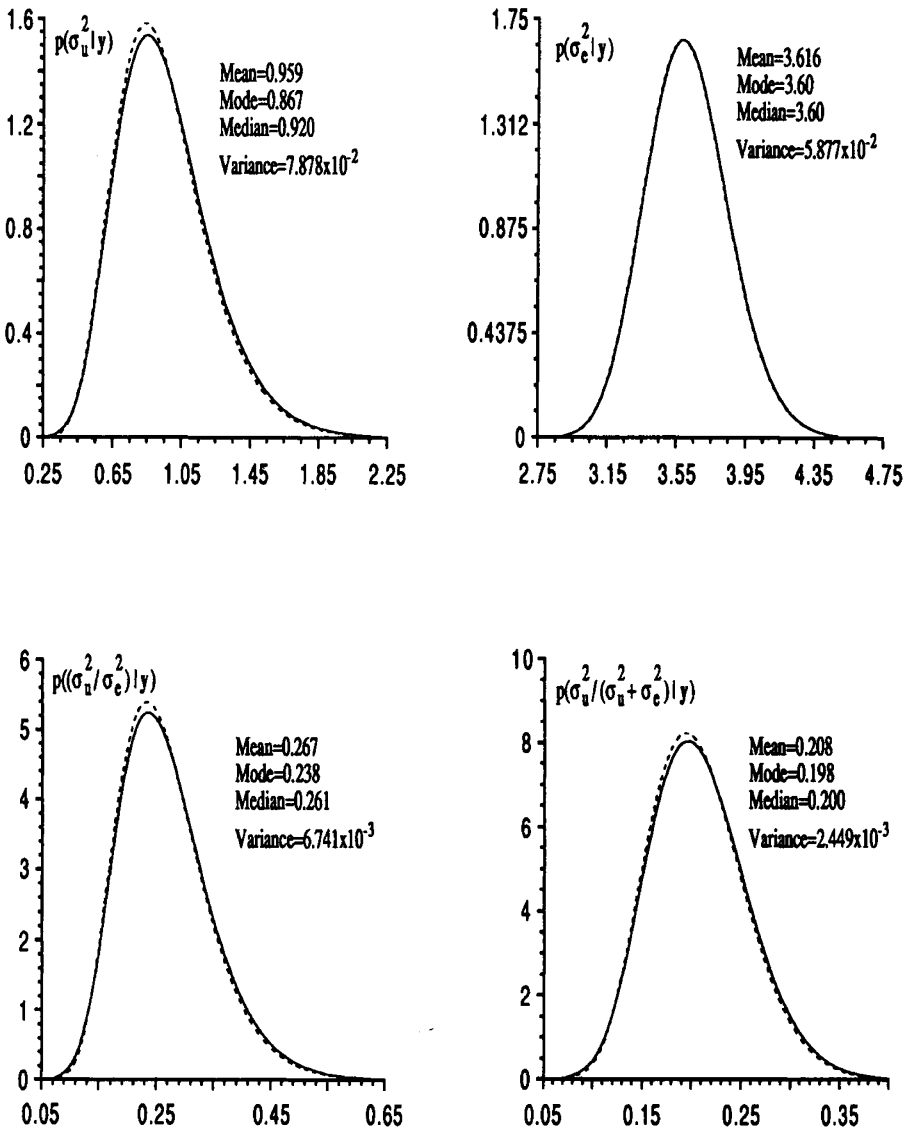


Fig 2. Marginal densities of variance components and of functions of variance components from data set II. The dotted curve was generated with $k = 20$ and $m = 300$, while the solid curve was with $k = 20$ and $m = 1000$. See footnotes to table I for definitions of symbols.

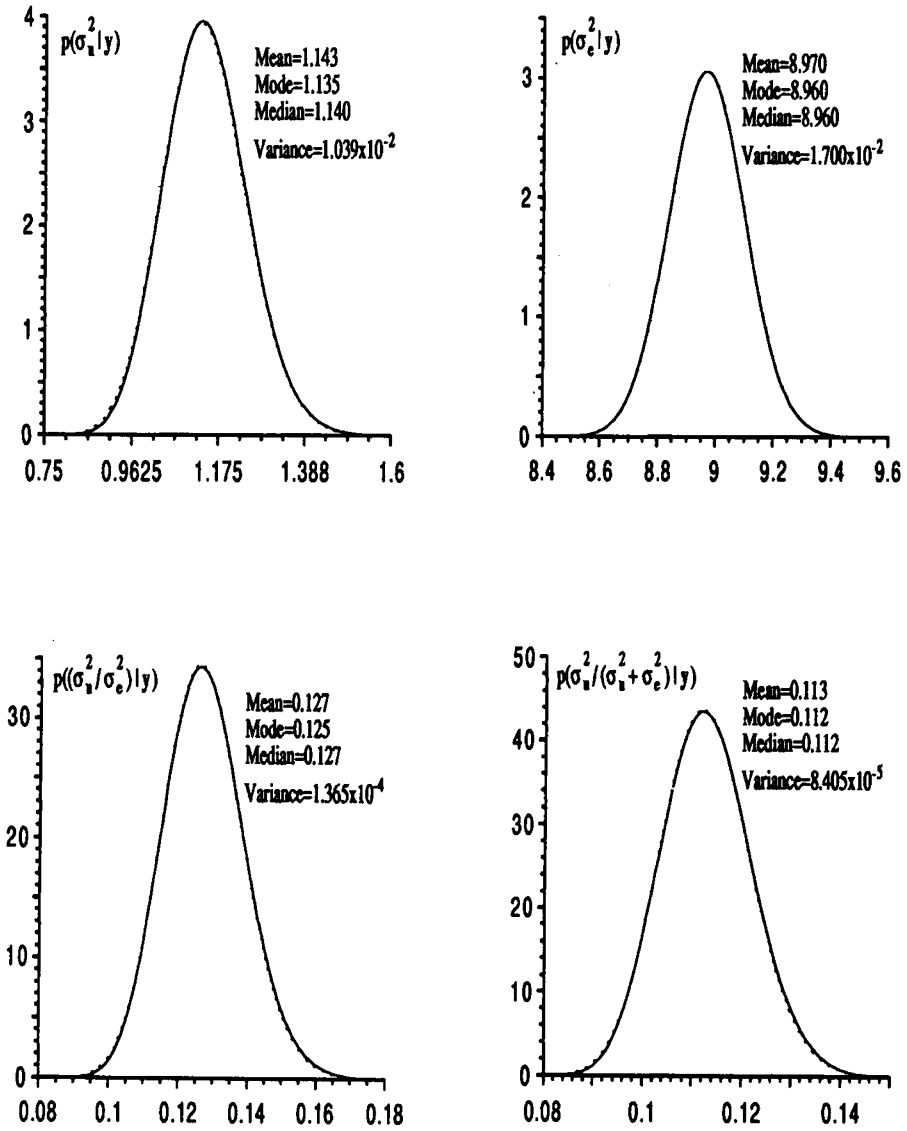


Fig 3. Marginal densities of variance components and of functions of variance components from data set III. The dotted curve was generated with $k = 10$ and $m = 1000$, while the solid curve was with $k = 50$ and $m = 500$. See footnotes to table I for definitions of symbols.

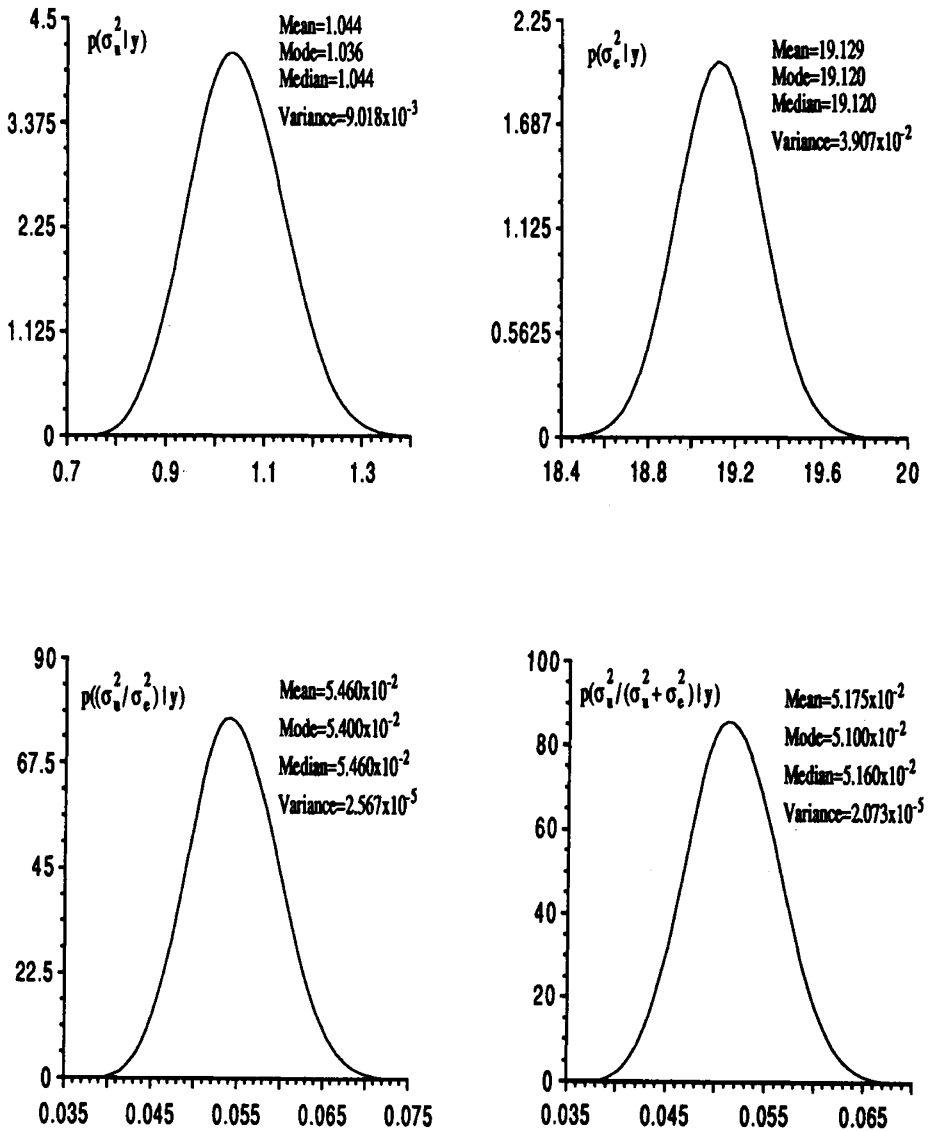


Fig 4. Marginal densities of variance components and of functions of variance components from data set IV. The curve was generated with $k = 30$ and $m = 300$. See footnotes to table I for definitions of symbols.

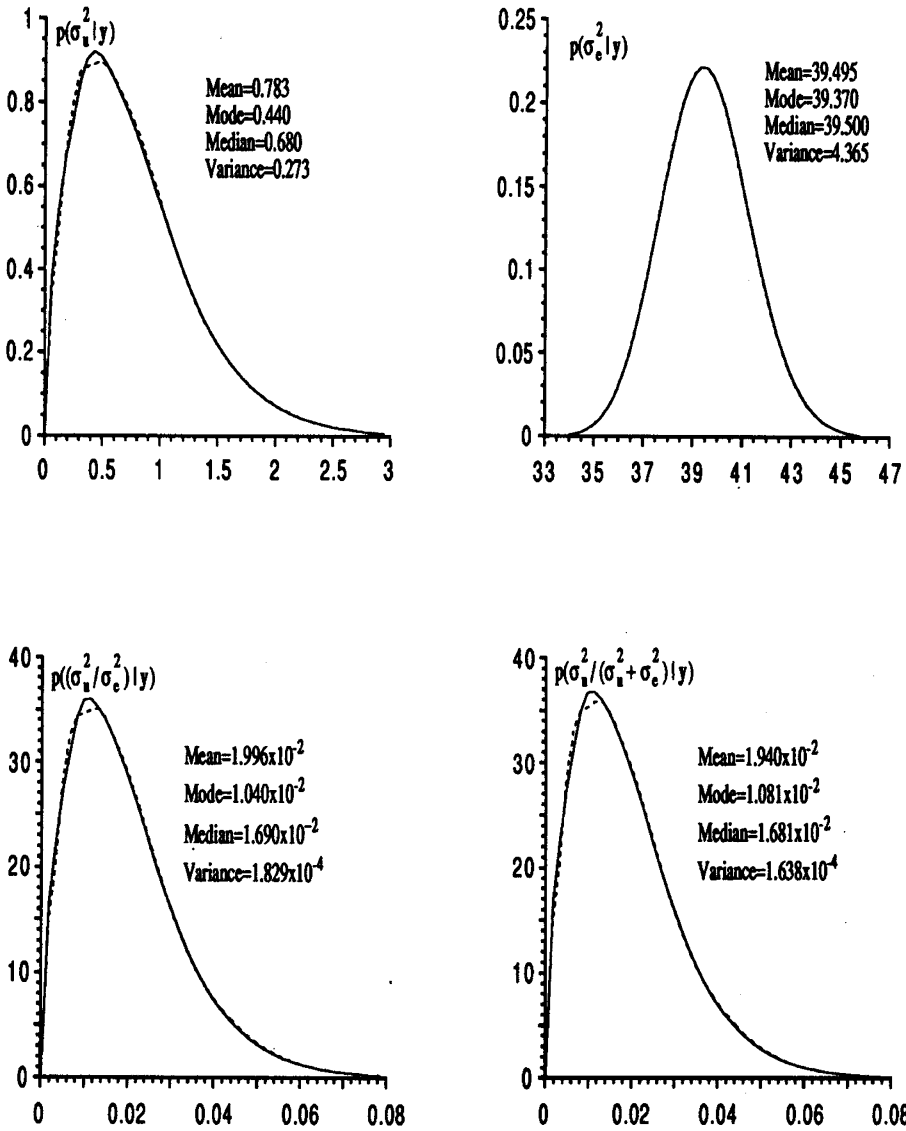


Fig 5. Marginal densities of variance components and of functions of variance components from data set V. The dotted curve was generated with $k = 20$ and $k = 2000$, while the solid curve was with $k = 20$ and $m = 3000$. See footnotes to table I for definitions of symbols.

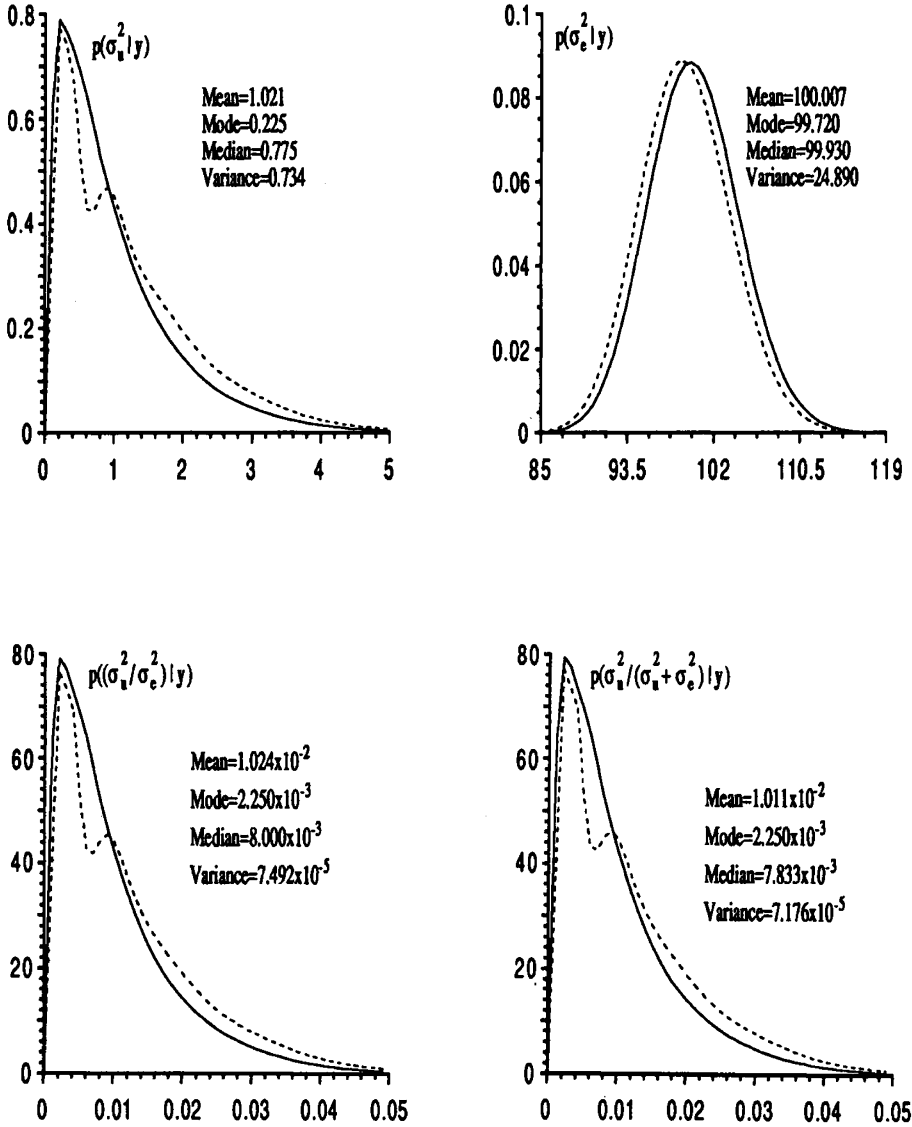


Fig 6. Marginal densities of variance components and of functions of variance components from data set VI. The dotted curve was generated with $k = 20$ and $m = 300$, while the solid curve was with $k = 20$ and $m = 3000$. See footnotes to table I for definitions of symbols.

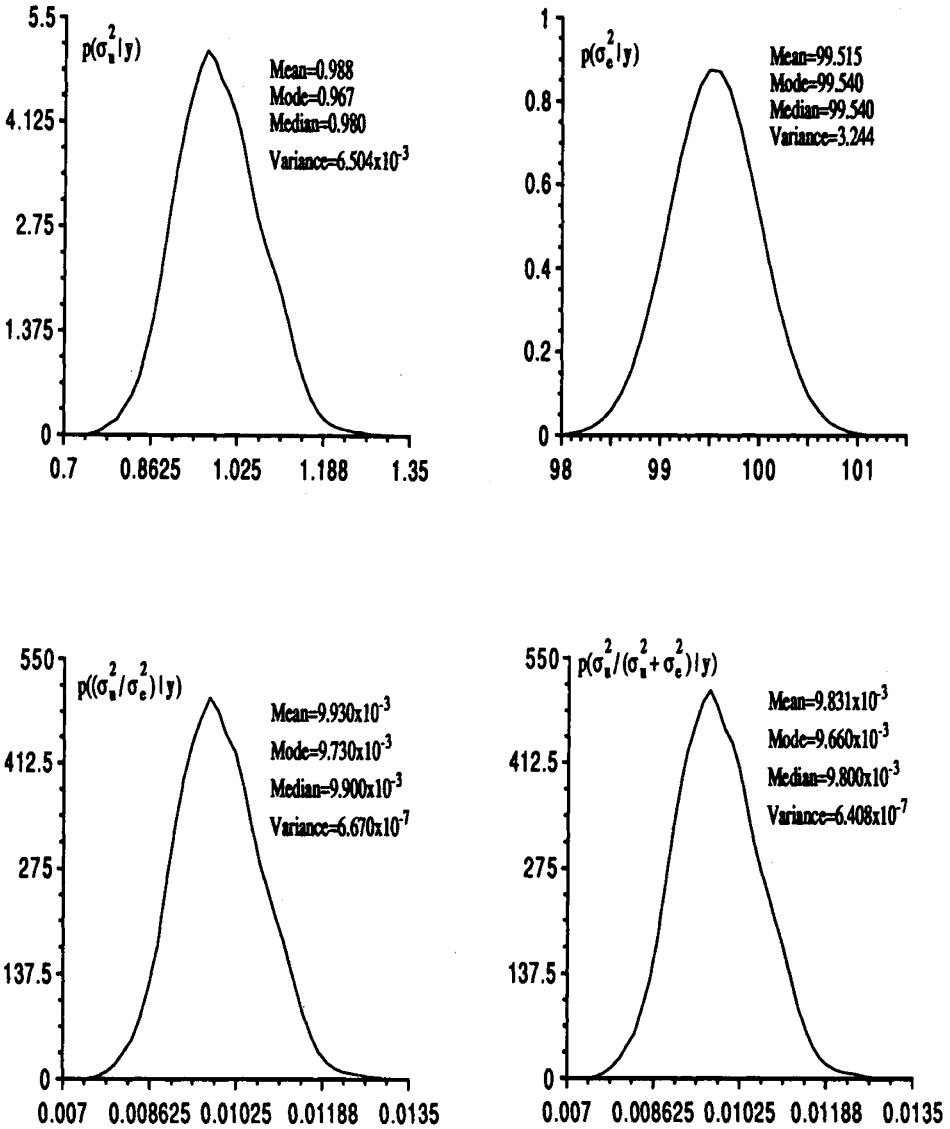


Fig 7. Marginal densities of variance components and of functions of variance components from data set VII. The curve was generated with $k = 20$ and $m = 2000$. See footnotes to table I for definitions of symbols.

DISCUSSION

Gianola and Foulley (1990) gave approximations to the marginal distribution of variance components, which should be adequate provided that the joint posterior density of the ratios between variance components is sharp or symmetric about its mode. As in any approximation, its goodness depends on the amount of information contained in the data, and on the number of parameters in the model. An important practical question is to what extent the approximations hold. From our experiments, the information in data sets I, II, V and VI is not sufficient to justify use of the approximation, even in a model with 2 variance components; this is so because the marginal distribution of the variance ratio was neither sharp nor symmetric. However, the approximation would hold in the remaining data sets. A more detailed comparative study between the exact and the approximate method is needed to ascertain when the latter can be used confidently. In the meantime, the Gibbs sampling provides a way to estimate the marginal distributions without using complicated numerical integration procedures.

The present work is an extension of that of Gelfand *et al* (1990), in which they illustrated the Gibbs sampler with several normal data models, including variance component estimation in a balanced one-way random effects layout. In our paper, formulae were presented to implement Gibbs sampling in a more general univariate mixed linear model. With these extensions, a full Bayesian solution to the problem of inference about variance components, and functions thereof in such a mixed linear model is possible. Gibbs sampling turns an analytically intractable multidimensional integration problem into a feasible numerical one.

Gibbs sampling is relatively easy to implement. Given the likelihood function and the priors, one can always obtain the joint posterior density of the unknowns under consideration. From this density, at least in the normal linear model, one can directly get the full conditional distribution of a particular variable by fixing the rest of the variables in the joint posterior distribution. The set of all full conditional densities gives the expressions needed for implementing the Gibbs sampler. The full conditional densities in this case are in families, such as normal and inverted χ^2 , where generating random numbers is not exceedingly complicated. The limiting factor is the efficiency with which random numbers can be generated, because the method requires generating a large quantity of random numbers; $m \times k \times r$, where r is the number of parameters in the model; $r = q + 3$, in our case. It is difficult to specify the values of m and k , *a priori*. Gelfand *et al* (1990) described a method for assessing convergence under different values of m and k , and suggested using $k = 10 - 20$ and $m = 100$ for a variance component problem in a balanced one-way model. However, they increased m to 1 000 when the variance ratio was under consideration. Our numerical results for the same model support their suggestion for the value of k , but indicate that Gibbs sample sizes of 2 000 to 3 000 may be needed for badly behaved marginal distributions (figs 6, 7). The most difficult situation encountered in our experiments was when intraclass correlation and sample size were both small. In general the appropriate values of m and k depend on the number of variables in the model, the shapes of the marginal distributions and the accuracy required to estimate densities.

An appealing aspect of Gibbs sampling is its flexibility. For example, densities of functions of the original variables included in the posterior distribution can be estimated using standard theory of random variable transformation, with minimal additional calculations.

Gibbs sampling is iterative. In this respect, there are 2 issues of concern: convergence and uniqueness. However, Geman and Geman (1984) showed that under mild regularity conditions, the Gibbs sampler converges uniquely to the appropriate marginal distributions. Casella and George (1990) discuss numerical means to accelerate convergence. Another way to speed up convergence is to integrate out analytically some nuisance parameters from the joint posterior distribution before running the Gibbs sampler. In our case, inference is sought on variance components and on their functions. Here, \mathbf{u} and $\boldsymbol{\beta}$ would be integrated out before running the Gibbs sampler. The necessary conditional distributions are given by Gianola and Foulley (1990).

The finite mixture density estimators of σ_e^2 and σ_u^2 in [39] and [40] can be thought of as average of a finite number of inverted χ^2 . This ensures that "point estimates" of variance components, *eg*, mean, mode and median will always be within the allowable parameter space. Likewise, interval estimates are also within the parameter space, in contrast to the asymptotic confidence intervals obtained from full or restricted maximum likelihood, which may include negative values.

Once the marginal densities are obtained, it is easy to calculate summary statistics from the posterior distributions. From a decision theoretical viewpoint, the optimum Bayes estimator under quadratic loss is the posterior mean; the posterior median is optimal if the loss functions is proportional to the absolute value of the error of estimation and the posterior mode is optimal if a step loss function is adopted.

Some caution is needed in variance component problems in genetics. For example, some genetic models dictate bounds for a particular variable. If one employs a "sire" model, the intraclass correlation (ρ) must lie inside the $[0, 1/4]$ interval, because heritability is between 0 and 1. This implies that the variance ratio σ_u^2/σ_e^2 is between 0 and 1/3, and that $0 \leq \sigma_u^2 \leq \sigma_e^2/3$. Hence, use of truncated inverted χ^2 densities in the Gibbs sampler would be more sensible in such a model. On the other hand, if one considers an "animal" model, the bounds for ρ and σ_u^2/σ_e^2 are from 0 to 1, and from 0 to ∞ respectively, and the variance components are unbounded. Since any "sire" model is expressible as an "animal" model, this would solve the problem mentioned above, though at some computational expense.

Gibbs sampling is computer intensive, but in some simple models, such as the sire model employed here, large data sets can be handled (*eg*, data set VII and fig 7). The feasibility of Gibbs sampling in large "animal" model is a subject for further research.

ACKNOWLEDGMENTS

We thank the college of Agriculture and Life Sciences, University of Wisconsin-Madison, for supporting this work. C Ritter of the Department of Statistics is thanked for useful suggestions. Computational resources were generously provided by the San Diego

Supercomputer Center, San Diego, CA. We thank two anonymous reviewers for comments on the paper.

REFERENCES

- Casella G, George EI (1990) *Gibbs for Kids (An Introduction to Gibbs Sampling)*. Paper BU-1098-M, Biometrics Unit, Cornell Univ
- Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85, 398-409
- Gelfand AE, Hills SE, Racine-Poon A, Smith AFM (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *J Am Stat Assoc* 85, 972-985
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans Pattern Anal Machine Intelligence* 6, 721-741
- Gianola D, Foulley JL (1990) Variance estimation from integrated likelihood (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 Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, New York, 15-40
- Gianola D, Im S, Macedo FWM (1990b) A framework for prediction of breeding value. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, New York, 210-238
- Harville DA (1974) Bayesian inference for variance components using only error contrasts. *Biometrika* 61, 383-385
- Henderson CR (1953) Estimation of variance and covariance components. *Biometrics* 9, 226-252
- IMSL Inc (1989) *IMSL[®] Stat/Library: Fortran Subroutines for Statistical Analysis*. Houston, TX
- Macedo FWM, Gianola D (1987) Bayesian analysis of univariate mixed models with informative priors. In: *Eur Assoc Anim Prod, 38th Annu Meet*. Lisbon, Portugal, 35 p
- Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545-554
- Thompson WA (1962) The problem of negative estimates of variance components. *Ann Math Stat* 33, 273-289
- Tierney L (1991) Exploring posterior distributions using Markov chains. In: *Computing Science and Statistics: Proc 23rd Sym Interface* (Keramidas E, ed) Am Stat Assoc, Alexandria, VA, 8 p