

Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs

CS Wang*, JJ Rutledge, D Gianola

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

(Received 26 April 1993; accepted 17 December 1993)

Summary – The Gibbs sampling is a Monte-Carlo procedure for generating random samples from joint distributions through sampling from and updating conditional distributions. Inferences about unknown parameters are made by: 1) computing directly summary statistics from the samples; or 2) estimating the marginal density of an unknown, and then obtaining summary statistics from the density. All conditional distributions needed to implement the Gibbs sampling in a univariate Gaussian mixed linear model are presented in scalar algebra, so no matrix inversion is needed in the computations. For location parameters, all conditional distributions are univariate normal, whereas those for variance components are scaled inverted chi-squares. The procedure was applied to solve a Gaussian animal model for litter size in the Gamito strain of Iberian pigs. Data were 1 213 records from 426 dams. The model had farrowing season (72 levels) and parity (4) as fixed effects; breeding values (597), permanent environmental effects (426) and residuals were random. In CASE I, variances were assumed known, with REML (restricted maximum likelihood) estimates used as true parameter values. Here, means and variances of the posterior distributions of all effects were obtained, by inversion, from the mixed model equations. These exact solutions were used to check the Monte-Carlo estimates given by Gibbs, using 120 000 samples. Linear regression slopes of true posterior means on Gibbs means were almost exactly 1 for fixed, additive genetic and permanent environmental effects. Regression slopes of true posterior variances on Gibbs variances were 1.00, 1.01 and 0.96, respectively. In CASE II, variances were treated as unknown, with a flat prior assigned to these. Posterior densities of selected location parameters, variance components, heritability and repeatability were estimated. Marginal posterior distributions of dispersion parameters were skewed, save the residual variance; the means, modes and medians of these distributions differed from the REML estimates, as expected from theory. The conclusions are: 1) the Gibbs sampler converged to the true posterior distributions, as suggested by CASE I; 2) it provides a richer description of uncertainty about genetic

* Present address: Morrison Hall, Department of Animal Science, Cornell University, Ithaca, NY 14 853-4801, USA

parameters than REML; 3) it can be used successfully to study quantitative genetic variation taking into account uncertainty about all nuisance parameters, at least in moderately sized data sets. Hence, it should be useful in the analysis of experimental data.

Iberian pig / genetic parameters / linear model / Bayesian methods / Gibbs sampler

Résumé – Analyse bayésienne de modèles linéaires mixtes à l'aide de l'échantillonnage de Gibbs avec une application à la taille de portée de porcs ibériques. L'échantillonnage de Gibbs est une procédure de Monte-Carlo pour engendrer des échantillons aléatoires à partir de distributions conjointes, par échantillonnage dans des distributions conditionnelles réajustées itérativement. Les inférences relatives aux paramètres inconnus sont obtenues en calculant directement des statistiques récapitulatives à partir des échantillons générés, ou en estimant la densité marginale d'une inconnue, et en calculant des statistiques récapitulatives à partir de cette densité. Toutes les distributions conditionnelles nécessaires pour mettre en œuvre l'échantillonnage de Gibbs dans un modèle univarié linéaire mixte gaussien sont présentées en algèbre scalaire, si bien qu'aucune inversion matricielle n'est requise dans les calculs. Pour les paramètres de position, toutes les distributions conditionnelles sont normales univariées, alors que celles des composantes de variance sont des χ^2 inverses dimensionnés. La procédure a été appliquée à un modèle individuel gaussien de taille de portée dans la souche porcine ibérique Gamito. Les données représentaient 1 213 observations sur 426 mères. Le modèle incluait les effets fixes de la saison de mise bas (72 niveaux) et de la parité (4 niveaux); les valeurs génétiques individuelles (597), les effets de milieu permanent (426) et les résidus étaient aléatoires. Dans le CAS I, les variances étaient supposées connues, les estimées REML (maximum de vraisemblance restreinte) étant considérées comme les valeurs vraies des paramètres. Les moyennes et les variances des distributions a posteriori de tous les effets étaient alors obtenues par la résolution du système d'équations du modèle mixte. Ces solutions exactes étaient utilisées pour vérifier les estimées Monte-Carlo données par le Gibbs, en utilisant 120 000 échantillons. Les coefficients de régression linéaire des vraies moyennes a posteriori en fonction des moyennes de Gibbs étaient presque exactement de 1, pour les effets fixes, génétiques additifs et de milieu permanent. Les coefficients de régression des variances vraies a posteriori en fonction des variances de Gibbs étaient 1,00, 1,01, et 0,96 respectivement. Dans le CAS II, les variances étaient traitées comme des inconnues, avec une distribution a priori uniforme. Les densités a posteriori de paramètres de position choisis, des composantes de variance, de l'héritabilité et de la répétabilité ont été estimées. Les distributions a posteriori des paramètres de dispersion étaient dissymétriques, sauf la variance résiduelle; les moyennes, modes et médianes de ces distributions différaient des estimées REML, comme prévu d'après la théorie. On conclut que: i) l'échantillonneur de Gibbs converge vers les vraies distributions a posteriori, comme le suggère le CAS I, ii) il fournit une description de l'incertitude sur les paramètres génétiques plus riche que REML; iii) il peut être utilisé avec succès pour étudier la variation génétique quantitative avec prise en compte de l'incertitude sur tous les paramètres de nuisance, du moins avec un nombre de données modéré. Il devrait donc être utile dans l'analyse de données expérimentales.

porc ibérique / paramètres génétiques / modèle linéaire / analyse bayésienne / échantillonnage de Gibbs

INTRODUCTION

Prediction of merit or, equivalently, deriving a criterion for selection is an important theme in animal breeding. Cochran (1951), under certain assumptions, showed that the selection criterion that maximized the expected merit of the selected animals was the mean of the conditional distribution of merit given the data. The conditional mean is known as best predictor, or BP (Henderson, 1973), because it minimizes mean square error of prediction among all predictors. Computing BP requires knowing the joint distribution of predictands and data, which can seldom be met in practice. To simplify, attention may be restricted to linear predictors.

Henderson (1963, 1973) and Henderson *et al* (1959) developed the best linear unbiased prediction (BLUP), which removed the requirement of knowing the first moments of the distributions. BLUP is the linear function of the data that minimizes mean square error of prediction in the class of linear unbiased predictors.

Bulmer (1980), Gianola and Goffinet (1982), Goffinet (1983) and Fernando and Gianola (1986) showed that under multivariate normality, BLUP is the conditional mean of merit given a set of linearly independent error contrasts. This holds if the second moments of the joint distribution of the data and of the predictand are known. However, second moments are rarely known in practice, and must be estimated from the data at hand. If estimated dispersion parameters are used in lieu of the true values, the resulting predictors of merit are no longer BLUP.

In animal breeding, dispersion components are most often estimated using restricted maximum likelihood, or REML (Patterson and Thompson, 1971). Theoretical arguments (*eg*, Gianola *et al*, 1989; Im *et al*, 1989) and simulations (*eg*, Rothschild *et al*, 1979) suggest that likelihood-based methods have ability to account for some forms of nonrandom selection, which makes the procedure appealing in animal breeding. Thus, 2-stage predictors are constructed by, first, estimating variance and covariance components, and then obtaining BLUE and BLUP of fixed and random effects, respectively, with parameter values replaced by likelihood-based estimates. Under random selection, this 2-stage procedure should converge in probability to BLUE and BLUP as the information in the sample about variance components increases; however, its frequentist properties under nonrandom selection are unknown. One deficiency of this BLUE and BLUP procedure is that errors of estimation of dispersion components are not taken into account when predicting breeding values.

Gianola and Fernando (1986), Gianola *et al* (1986) and Gianola *et al* (1990a, b, 1992) advocate the use of Bayesian methods in animal breeding. The associated probability theory dictates that inferences should be based on marginal posterior distributions of parameters of interest, such that uncertainty about the remaining parameters is fully taken into account. The starting point is the joint posterior density of all unknowns. From the joint distribution, the marginal posterior distribution of a parameter, say the breeding value of an animal, is obtained by successively integrating out all nuisance parameters, these being the fixed effects, all the random effects other than the one of interest, and the variance and covariance components. This integration is difficult by analytical or numerical means, so approximations are usually sought (Gianola and Fernando, 1986; Gianola *et al*, 1986; Gianola *et al*, 1990a, b).

The posterior distributions are so complex that an analytical approach is often impossible, so attention has concentrated on numerical procedures (*eg*, Cantet *et al*, 1992). Recent breakthroughs are related to Monte-Carlo Markov chain procedures for multidimensional integrations and for sampling from joint distributions (Geman and Geman, 1984; Gelfand and Smith, 1990; Gelfand *et al*, 1990). One of these procedures, Gibbs sampling, has been studied extensively in statistics (Gelfand and Smith, 1990; Gelfand *et al*, 1990; Besag and Clifford, 1991; Gelfand and Carlin, 1991; Geyer and Thompson, 1992).

Wang *et al* (1993) described the Gibbs sampler for a univariate mixed linear model in an animal breeding context. They used simulated data to construct marginal densities of variance components, variance ratios and intraclass correlations, and noted that the marginal distributions of fixed and random effects could also be obtained.

However, their implementation was in matrix form. Clearly, some matrix computations are not feasible in many animal breeding data sets because inversion of large matrices is needed repeatedly.

In this paper, we consider Bayesian marginal inferences about fixed and random effects, variance components and functions of variance components in a univariate Gaussian mixed linear model. Here, marginal inferences are obtained, in contrast to Wang *et al* (1993) through a scalar version of the Gibbs sampler, so inversion of matrices is not needed. Our implementation was applied to and validated with a data set on litter size of Iberian pigs.

THE GIBBS SAMPLER FOR THE GAUSSIAN MIXED LINEAR MODEL

Model

We consider a univariate mixed linear model with several independent random factors as in Henderson (1984), Macedo and Gianola (1987) and Gianola *et al* (1990a, b):

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

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} : $n \times 1$ vector of random residuals.

The conditional distribution that generates the data is:

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

where \mathbf{I} is an $n \times n$ identity matrix, and σ_e^2 is the variance of the random residuals.

Prior Distributions

An integral part of Bayesian analysis is the assignment of prior distributions to all unknowns in the model; here, these are $\boldsymbol{\beta}$, $\mathbf{u}_i (i = 1, 2, \dots, c)$ and the $c + 1$ variance components (one for each of the random vectors, plus the error). Usually, a flat or uniform 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, it is assumed that:

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

where \mathbf{G}_i is a known matrix and $\sigma_{\mathbf{u}_i}^2$ is the variance of the prior distribution of \mathbf{u}_i . In a genetic context, \mathbf{G}_i matrices can contain functions of known coefficients of coancestry. All \mathbf{u}_i 's are assumed to be mutually independent *a priori*, as well as independent of $\boldsymbol{\beta}$. Note that the priors for \mathbf{u}_i correspond to the assumptions made about these random vectors in the classical linear model.

Independent scaled inverted chi-square distributions are used as priors for variance components, so that:

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

and

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

Above, $\nu_e (\nu_{\mathbf{u}_i})$ is a 'degree of belief' parameter, and $s_e^2 (s_{\mathbf{u}_i}^2)$ can be thought of as a prior value of the appropriate variance.

Joint posterior density

Let

$$\boldsymbol{\theta}' = (\boldsymbol{\beta}', \mathbf{u}'_1, \dots, \mathbf{u}'_c) = (\theta_1, \theta_2, \dots, \theta_N)$$

where $N = p + \sum_{i=1}^c q_i = p + q$, and

$$\boldsymbol{\theta}'_{-i} = (\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_N)$$

be $\boldsymbol{\theta}'$ without θ_i . Further, let

$$\mathbf{v}' = (\sigma_{\mathbf{u}_1}^2, \sigma_{\mathbf{u}_2}^2, \dots, \sigma_{\mathbf{u}_c}^2)$$

be the vector of variance components other than the residual;

$$\mathbf{v}'_{-i} = \mathbf{v}' \text{ without component } i$$

and

$$\begin{aligned}\mathbf{s}' &= (s_{\mathbf{u}_1}^2, s_{\mathbf{u}_2}^2, \dots, s_{\mathbf{u}_c}^2, s_e^2) \\ \mathbf{v}' &= (\nu_{\mathbf{u}_1}, \nu_{\mathbf{u}_2}, \dots, \nu_{\mathbf{u}_c}, \nu_e)\end{aligned}$$

be the sets of all prior variances and degrees of belief, respectively. As shown, for example, by Macedo and Gianola (1987) and Gianola *et al* (1990a, b), the joint posterior density is in the normal-gamma form:

$$\begin{aligned}p(\boldsymbol{\theta}, \mathbf{v}, \sigma_e^2 | \mathbf{y}, \mathbf{s}, \mathbf{v}) \propto & (\sigma_e^2)^{-(n+\nu_e+2)/2} \exp \left\{ -\frac{1}{2\sigma_e^2} \left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i) + \nu_e s_e^2 \right] \right\} \\ & \times \prod_{i=1}^c \left[(\sigma_{\mathbf{u}_i}^2)^{-(q_i + \nu_{\mathbf{u}_i} + 2)/2} \exp \left\{ -\frac{1}{2\sigma_{\mathbf{u}_i}^2} [\mathbf{u}_i' \mathbf{G}_i^{-1} \mathbf{u}_i + \nu_{\mathbf{u}_i} s_{\mathbf{u}_i}^2] \right\} \right] \quad [7]\end{aligned}$$

Inferences about each of the unknowns $(\boldsymbol{\theta}, \mathbf{v}, \sigma_e^2)$ are based on their respective marginal densities. Conceptually, each of the marginal densities is obtained by successive integration of the joint density [7] with respect to parameters other than the one of interest. For example, the marginal density of σ_e^2 is

$$p(\sigma_e^2 | \mathbf{y}, \mathbf{s}, \mathbf{v}) = \int_{R_{\boldsymbol{\theta}, \mathbf{v}}} p(\boldsymbol{\theta}, \mathbf{v}, \sigma_e^2 | \mathbf{y}, \mathbf{s}, \mathbf{v}) d\boldsymbol{\theta} d\mathbf{v}.$$

It is difficult to carry out the needed integration analytically. Gibbs sampling is a Monte-Carlo procedure to overcome such difficulties.

Fully conditional posterior densities (Gibbs sampler)

The fully conditional posterior densities of all unknowns are needed for implementing the Gibbs sampling. Each of the full conditional densities can be obtained by regarding all other parameters in [7] as known. Let $\mathbf{W} = \{w_{ij}\}$, $i, j = 1, 2, \dots, N$, and $\mathbf{b} = \{b_i\}$, $i = 1, 2, \dots, N$ be the coefficient matrix and the right hand side of the mixed model equations, respectively. As proved in the *Appendix*, the conditional posterior distribution of each of the location parameters in $\boldsymbol{\theta}$ is normal, with mean and variance $\tilde{\theta}_i$ and \tilde{v}_i :

$$\theta_i | \mathbf{y}, \boldsymbol{\theta}_{-i}, \mathbf{v}, \sigma_e^2, \mathbf{s}, \mathbf{v} \sim N(\tilde{\theta}_i, \tilde{v}_i), i = 1, 2, \dots, N \quad [8]$$

where $\tilde{\theta}_i = (b_i - \sum_{j=1, j \neq i}^N w_{ij} \theta_j) / w_{ii}$ and $\tilde{v}_i = \sigma_e^2 / w_{ii}$. This is an important result

because all computations needed to implement Gibbs sampling are scalar, without any required inversion of matrices. This is in contrast with the matrix version of the conditional posterior distributions for the location parameters given by Wang *et al* (1993). It should be noted that distributions [8] do not depend on \mathbf{s} and \mathbf{v} , because \mathbf{v} is known in [8].

The conditional posterior density of σ_e^2 is in the scaled inverted chi-square form:

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

It can be readily seen that

$$\sigma_e^2 | \mathbf{y}, \boldsymbol{\theta}, \mathbf{v}, \mathbf{s}, \boldsymbol{\nu} \sim \tilde{\nu}_e \tilde{s}_e^2 \chi_{\tilde{\nu}_e}^{-2}$$

with parameters

$$\tilde{\nu}_e = n + \nu_e, \text{ and } \tilde{s}_e^2 = [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i) + \nu_e s_e^2] / \tilde{\nu}_e$$

Each condition posterior density of $\sigma_{\mathbf{u}_i}^2$ is also in the scaled inverted chi-square form:

$$p(\sigma_{\mathbf{u}_i}^2 | \mathbf{y}, \boldsymbol{\theta}, \mathbf{v}_{-i}, \sigma_e^2, \mathbf{s}, \boldsymbol{\nu}) \propto (\sigma_{\mathbf{u}_i}^2)^{-(q_i+\nu_{\mathbf{u}_i}+2)/2} \exp \left(-\frac{1}{2\sigma_{\mathbf{u}_i}^2} [\mathbf{u}_i' \mathbf{G}_i^{-1} \mathbf{u}_i + \nu_{\mathbf{u}_i} s_{\mathbf{u}_i}^2] \right) \quad [10]$$

or

$$\sigma_{\mathbf{u}_i}^2 | \mathbf{y}, \boldsymbol{\theta}, \mathbf{v}_{-i}, \sigma_e^2, \mathbf{s}, \boldsymbol{\nu} \sim \tilde{\nu}_{\mathbf{u}_i} \tilde{s}_{\mathbf{u}_i}^2 \chi_{\tilde{\nu}_{\mathbf{u}_i}}^{-2}, \quad i = 1, 2, \dots, c$$

with parameters $\tilde{\nu}_{\mathbf{u}_i} = q_i + \nu_{\mathbf{u}_i}$ and $\tilde{s}_{\mathbf{u}_i}^2 = (\mathbf{u}_i' \mathbf{G}_i^{-1} \mathbf{u}_i + \nu_{\mathbf{u}_i} s_{\mathbf{u}_i}^2) / \tilde{\nu}_{\mathbf{u}_i}$.

A set of the $N + c + 1$ conditional posterior distributions [8]–[10] is called the Gibbs sampler for our problem.

FULL CONDITIONAL POSTERIOR DENSITIES UNDER SPECIAL PRIORS

The Gibbs sampler with ‘naive’ priors for all variance components

The Gibbs sampler [8]–[10] given above is based on scaled inverted chi-squares used as priors for the variance components. These priors are proper and, therefore, informative, about the variance. A possible way of representing prior ignorance about variances would be to set the degree of belief parameters of the prior distributions for all the variance components to zero *ie*, $\nu_e = \nu_{\mathbf{u}_i} = 0$, for all i . These priors have been used, *inter alia*, by Gelfand *et al* (1990) and Gianola *et al* (1990a, b). In this case, the conditional posterior distributions of the location parameters are as in [8]:

$$\theta_i | \mathbf{y}, \boldsymbol{\theta}_{-i}, \mathbf{v}, \sigma_e^2 \sim N(\tilde{\theta}_i, \tilde{\nu}_i), i = 1, 2, \dots, N \quad [11]$$

because the distributions do not depend on \mathbf{s} and \mathbf{v} .

However, the conditional posterior distributions of the variance components no longer depend on hyper-parameters \mathbf{s} and \mathbf{v} . The conditional posterior distribution of the residual variance remains in the scaled inverted chi-square form:

$$\sigma_e^2 | \mathbf{y}, \boldsymbol{\theta}, \mathbf{v} \sim \tilde{\nu}_e \tilde{s}_e^2 \chi_{\tilde{\nu}_e}^{-2} \quad [12]$$

but now with parameters $\tilde{\nu}_e = n$, and $\tilde{s}_e^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i)'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \sum_{i=1}^c \mathbf{Z}_i \mathbf{u}_i) / \tilde{\nu}_e$. Each conditional posterior density of $\sigma_{\mathbf{u}_i}^2$ is again in the scaled inverted chi-square form:

$$\sigma_{\mathbf{u}_i}^2 | \mathbf{y}, \boldsymbol{\theta}, \mathbf{v}_{-i}, \sigma_e^2 \sim \tilde{\nu}_{\mathbf{u}_i} \tilde{s}_{\mathbf{u}_i}^2 \chi_{\nu_{\mathbf{u}_i}}^{-2} \quad i = 1, 2, \dots, c \quad [13]$$

with parameters $\tilde{\nu}_{\mathbf{u}_i} = q_i$ and $\tilde{s}_{\mathbf{u}_i}^2 = \mathbf{u}_i' \mathbf{G}_i^{-1} \mathbf{u}_i / \tilde{\nu}_{\mathbf{u}_i}$.

It has been noted recently (Besag *et al*, 1991; Raftery and Banfield, 1991) that under these priors, the joint posterior density [7] is improper because it does not integrate to 1. In the light of this, we do not recommend these 'naive' priors for variance component models.

The Gibbs sampler with flat priors for all variance components

Under flat priors for all variance components, *ie* $p(\mathbf{v}, \sigma_e^2) \propto \text{constant}$, the Gibbs sampler is as in [11]–[13], except that $\tilde{\nu}_e = n - 2$ and $\tilde{\nu}_{\mathbf{u}_i} = q_i - 2$ for $i = 1, 2, \dots, c$. This version of the sampler can also be obtained by setting $\nu_e = -2$ in [9] and $\mathbf{v} = (-2, -2, \dots, -2)'$ and $\mathbf{s} = (0, 0, \dots, 0)'$ in [10]. With flat priors, the joint posterior density [7] is proper.

The Gibbs sampler when all variance components are known

When variances are assumed known, the only conditional distributions needed are those for the location parameters, and these are as in [8] or [11].

INFERENCES ABOUT THE MARGINAL DISTRIBUTIONS THROUGH GIBBS SAMPLING

Gibbs sampling: an overview

The Gibbs sampler was used first in spatial statistics and presented formally by Geman and Geman (1984) in an image restoration context. Applications to Bayesian inference were described by Gelfand and Smith (1990) and Gelfand *et al* (1990). Since then, it has received extensive attention, as evidenced by recent discussion papers (Gelman and Rubin, 1992; Geyer, 1992; Besag and Green, 1993; Gilks *et al*, 1993; Smith and Roberts, 1993). Its power and usefulness as a general statistical tool to generate samples from complex distributions arising in some particular problems is unquestioned.

Our purpose is to generate random samples from the joint posterior distribution [7], through successively drawing samples from and updating the Gibbs sampler [8]–[10]. Formally, Gibbs sampling works as follows:

- (i) set arbitrary initial values for $\boldsymbol{\theta}, \mathbf{v}$ and σ_e^2 ;
- (ii) generate θ_i from [8] and update $\theta_i, i = 1, 2, \dots, N$;

- (iii) generate σ_e^2 from [9] and update σ_e^2 ;
- (iv) generate $\sigma_{\mathbf{u}_i}^2$ from [10] and update $\sigma_{\mathbf{u}_i}^2, i = 1, 2, \dots, c$;
- (v) repeat (ii)–(iv) k (length of the chain) times.

As $k \rightarrow \infty$, this creates a Markov chain with an equilibrium distribution that has [7] as its density. We shall call this procedure a single long-chain algorithm.

In practice, there are at least 2 ways of running the Gibbs sampler: a single long chain and multiple short chains. The multiple short-chain algorithm repeats steps (i)–(v) m times and saves only the k th iteration as sample (Gelfand and Smith, 1990; Gelfand *et al*, 1990). Based on theoretical arguments (Geyer, 1992) and on our experience, we used the single long-chain method in the present study.

Initial iterations are usually not stored as samples on grounds that the chain may not yet have reached the equilibrium distribution; this is called ‘warm-up’. After the warm-up, samples are stored every d iterations, where d is a small positive integer. Let the total number of samples saved be m , the sample size.

If the Gibbs sampler converges to the equilibrium distribution, the m samples are random drawings from the joint posterior distribution with density [7]. The i th sample

$$\{\boldsymbol{\theta}_i, \mathbf{v}_i \text{ and } (\sigma_e^2)_i\}, i = 1, 2, \dots, m \quad [14]$$

is then an $N + c + 1$ vector, and each of the elements of this vector is a drawing from the appropriate marginal distribution. Note that the m samples in [14] are identically but not independently distributed (*eg*, Geyer, 1992). We call m samples in [14] as Gibbs samples for reference.

Density estimation and Bayesian inference based on the Gibbs samples

Suppose $x_i, i = 1, 2, \dots, m$ is one of the components [14], *ie* a realization from running the Gibbs sampler of variable x . The m (dependent) samples can be used to compute features of the posterior distribution $P(x)$ by Monte-Carlo integration. An integral

$$u = \int g(x) dP(x) \quad [15]$$

can be approximated by

$$\hat{u} = \frac{1}{m} \sum_{i=1}^m g(x_i) \quad [16]$$

where $g(x)$ can be any feature of $P(x)$, such as its mean or variance. As $m \rightarrow \infty, \hat{u}$ converges almost surely to u (Geyer, 1992).

Another way to compute features of $P(x)$ is by first estimating the density $p(x)$, and then obtaining summary statistics from the estimated density using 1-dimensional numerical procedures. If $y_i (i = 1, 2, \dots, m)$ is another component of [14], an estimator of $p(x)$ is given by the average of the m conditional densities $p(x|y_i)$ (Gelfand and Smith, 1990):

$$\hat{p}(x) = \frac{1}{m} \sum_{i=1}^m p(x|y_i) \quad [17]$$

Note that this estimator does not use the samples $x_i, i = 1, 2, \dots, m$; instead, it uses the samples of variable y through the conditional density $p(x|y)$. This procedure, though developed primarily for identically and independently distributed (iid) data, can also be used for dependent samples, as noted by Liu *et al* (1991) and Diebolt and Robert (1993).

An alternative form of estimating $p(x)$ is to use samples $x_i (i = 1, 2, \dots, m)$ only. For example, a kernel density estimator is defined (Silverman, 1986) as:

$$\hat{p}(z) = \frac{1}{mh} \sum_{i=1}^m K[(z - x_i)/h] \quad [18]$$

where $\hat{p}(z)$ is the estimated density at point z , $K(\cdot)$ is a 'kernel function', and h is a fixed constant called window width; the latter determines the smoothness of the estimated curve. For example, if a normal density is chosen as kernel function, then [18] becomes:

$$\hat{p}(z) = \frac{1}{mh} \sum_{i=1}^m \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{z - x_i}{h} \right)^2 \right] \quad [19]$$

Again, though the kernel density estimator was developed for iid data, the work of Yu (1991) indicates that the method is valid for dependent data as well.

Once the density of $p(x)$ is estimated by either [17] or [19], summary statistics (*eg*, mean and variance) can be computed by a 1-dimensional numerical integration procedure, such as Simpson's rules. Probability statements about x can also be made, thus providing a full Bayesian solution to inferences about the distribution x .

Bayesian inference about functions of the original parameters

Suppose we want to make inference about the function:

$$z = x/y$$

The quantity

$$z_i = x_i/y_i, \quad i = 1, 2, \dots, m$$

is a random (dependent) sample of size m from a distribution with density $p(z)$. Formulae [16], [18] and [19] using such samples can also be used to make inferences about z .

An alternative is to use standard techniques to transform from either the conditional densities $p(x|y)$ or $p(y|x)$, to $p(z|y)$ or $p(z|x)$. Let the transformation be from $x|y$ to $z|y$; the Jacobian of the transformation is $|y|$, so the conditional density of $z|y$ is:

$$p(z|y) = |y|p(zy|y) \quad [20]$$

An estimator of $p(z)$, obtained by averaging m conditional densities of $p(z|y)$, is

$$\hat{p}(z) = \frac{1}{m} \sum_{i=1}^m |y_i|p(zy_i|y_i) \quad [21]$$

APPLICATION OF GIBBS SAMPLING TO LITTER SIZE IN PIGS

Data

Records were from the Gamito strain of Iberian pigs, Spain. The trait considered was number of pigs born alive per litter. Details about this strain and the data are in Dobao *et al* (1983) and Toro *et al* (1988); Perez-Enciso and Gianola (1992) gave REML estimates of genetic parameters. Briefly, the data were 1 213 records from 426 dams (including 68 crossfostered females). There were 72 farrowing seasons and 4 parity classes as defined by Perez-Enciso and Gianola (1992).

Model

A mixed linear model similar to that of Perez-Enciso and Gianola (1992) was:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{c} + \mathbf{e} \quad [22]$$

where \mathbf{y} is a vector of observations (number of pigs born alive per litter); \mathbf{X} , \mathbf{Z}_1 and \mathbf{Z}_2 are known incidence matrices relating $\boldsymbol{\beta}$, \mathbf{u} and \mathbf{c} , respectively, to \mathbf{y} ; $\boldsymbol{\beta}$ is a vector of fixed effects, including a mean, farrowing season (72 levels) and parity (4 levels); \mathbf{u} is a random vector of additive genetic effects (597 levels); \mathbf{c} is a random vector of permanent environmental effects (426 levels); and \mathbf{e} is a random vector of residuals. Distributional assumptions were:

$$\mathbf{u} \sim N(0, \mathbf{A}\sigma_u^2), \mathbf{c} \sim N(0, \mathbf{I}\sigma_c^2) \quad \text{and} \quad \mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2), \quad \text{respectively,} \quad [23]$$

where σ_u^2 , σ_c^2 and σ_e^2 are variance components and \mathbf{A} is the numerator of Wright's relationship matrix; the vectors \mathbf{u} , \mathbf{c} and \mathbf{e} were assumed to be pairwise independent. After reparameterization, the rank (p) of \mathbf{X} was $1 + 71 + 3 = 75$; the rank of the mixed model equations was then: $N = 75 + 597 + 426 = 1\,098$.

Gibbs sampling

We ran 2 separate Gibbs samplers with this data set, and we refer to these analyses as CASES I and II. In CASE I, the 3 variance components were assumed known, with REML estimates (Meyer, 1988) used as true parameter values. In CASE II, the variance components were unknown, and flat priors were assigned to them. For each of the 2 cases, a single chain of size 1 205 000 was run. After discarding the first 5 000 iterations, samples were saved every 10 iterations ($d = 10$), so the total number of samples (m) saved was 120 000. This specification (mainly the length of a chain) of running the Gibbs sampler was based on our own experience with this data and with others. It may be different for other problems.

Due to computer storage limitation, not all Gibbs samples and conditional means and variances could be saved for all location parameters. Instead, for further analysis and illustration, we selected 4 location parameters arbitrarily, one from each of the 4 factors (farrowing season, parity, additive genetic effect and permanent environmental effect). For each of these 4 location parameters, the following quantities were stored:

$$x_1, x_2, \dots, x_m \quad [24]$$

$$\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_m \quad [25]$$

$$\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_m \quad [26]$$

where x_i is a Gibbs sample from an appropriate marginal distribution, and $\tilde{\theta}_i$ and \tilde{v}_i are the mean and variance of the conditional distribution, [8] or [11], used for generating x_i at each of the Gibbs steps.

In CASE II, we also saved the m Gibbs samples for each of the variance components, and

$$s_1, s_2, \dots, s_m \quad [27]$$

where s_i is the scale parameter appearing in the conditional density [9] or [10] at each of the Gibbs iterations.

A FORTRAN program was written to generate the samples, with IMSL subroutines used for drawing random numbers (IMSL, INC, 1989).

Density estimation and inferences

For each of the 4 selected location parameters (CASES I and II) and the 3 variance components, we estimated the marginal posterior with estimators [17] and [19], *ie* by averaging m conditional densities and by the normal kernel density estimation method, respectively. Estimator [17] of the density of a location parameter was explicitly:

$$\hat{p}(z|\mathbf{y}) = \frac{1}{m} \sum_{j=1}^m \frac{1}{\sqrt{2\pi\tilde{v}_j}} \exp \left[-\frac{1}{2} \frac{(z - \tilde{\theta}_j)^2}{\tilde{v}_j} \right] \quad [28]$$

where $\tilde{\theta}_j$ and \tilde{v}_j are the conditional mean and variance of the conditional posterior density of z . For each of the variance components, the estimator was:

$$\hat{p}(z|\mathbf{y}) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} (z)^{-\nu/2-1} \frac{1}{m} \sum_{j=1}^m (s_j)^{\nu/2} \exp(-\frac{1}{2}\nu s_j/z) \quad [29]$$

where ν is the degree of belief, s_j is the scale parameter of the conditional posterior distribution of the variance of interest, and $\Gamma(\cdot)$ is the gamma function. The normal kernel estimator [19] was applied directly to the samples for location and dispersion parameters.

To estimate the densities, we divided the ‘effective domain’ of each parameter into 100 equally spaced intervals; the effective domain contained at least 99.5% of the density mass. Running through the effective domain, a sequence of pairs $(\hat{p}(z_i), z_i), i = 1, 2, \dots, 101$, was generated. Densities were displayed graphically by plotting $(\hat{p}(z_i), z_i)$ pairs. For the normal kernel density estimator [19], window width was specified as: $h = (\text{range of effective domain})/75$.

For the 4 selected location parameters, the mean, mode, median and variance of each of the marginal distributions were computed as summary features. The mean,

median and variance were obtained with Simpson's integration rules by further dividing the effective domain into 1 000 equally spaced intervals, and using a cubic spline technique (IMSL, INC, 1989); the mode was located through a grid search. For location parameters other than the 4 selected ones, the posterior means and variances were computed directly from the Gibbs samples.

Density estimation for functions of variance components

As mentioned previously, the posterior density of any function of the original parameters can be made through transformation of random variables without rerunning the Gibbs sampler, provided appropriate samples are saved. In this section, we summarize methods for density estimation of functions of variance components; see also Wang *et al* (1993).

Let the Gibbs samples for σ_u^2 , σ_c^2 and σ_e^2 , be respectively:

$$\begin{aligned} x_{u1}, x_{u2}, \dots, x_{um} \\ x_{c1}, x_{c2}, \dots, x_{cm} \end{aligned} \quad [30]$$

and

$$x_{e1}, x_{e2}, \dots, x_{em}$$

Also, let the scale parameters of the corresponding densities be:

$$\begin{aligned} s_{u1}, s_{u2}, \dots, s_{um} \\ s_{c1}, s_{c2}, \dots, s_{cm} \end{aligned} \quad [31]$$

and

$$s_{e1}, s_{e2}, \dots, s_{em}$$

Consider first estimating the marginal posterior density of the total variance: $\sigma_p^2 = \sigma_u^2 + \sigma_c^2 + \sigma_e^2$. Let the transformation be from the conditional posterior density, $(\sigma_e^2 | \mathbf{y}, \boldsymbol{\theta}, \mathbf{s}, \boldsymbol{\nu})$ to $(\sigma_p^2 | \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{s}, \boldsymbol{\nu})$. The Jacobian of the transformation is 1. Thus, using [17], the estimator of the marginal posterior density of σ_p^2 is:

$$\begin{aligned} \hat{p}(\sigma_p^2 | \mathbf{y}) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{1}{m} \sum_{j=1}^m (\sigma_p^2 - x_{uj} - x_{cj})^{-\nu/2-1} \\ (s_{ej})^{\nu/2} \exp \left[-\frac{1}{2} \nu s_{ej} / (\sigma_p^2 - x_{uj} - x_{cj}) \right] \end{aligned} \quad [32]$$

where $\nu = \tilde{\nu}_e$. Further, $x_{pi} = x_{ui} + x_{ci} + x_{ei}$, ($i = 1, 2, \dots, m$), is a sample from the marginal distribution of σ_p^2 .

Similarly, the estimator of the marginal posterior density of $h^2 = \sigma_u^2 / \sigma_p^2$ is obtained by using the transformation $\sigma_u^2 \rightarrow h^2$ in the conditional posterior density of σ_u^2 . One obtains

$$\begin{aligned} \hat{p}(h^2 | \mathbf{y}) = c (h^2)^{-\nu/2-1} (1 - h^2)^{\nu/2-1} \\ \sum_{j=1}^m (s_{uj} / (x_{cj} + x_{ej}))^{\nu/2} \exp \left[-\frac{1}{2} \nu s_{uj} / [(x_{cj} + x_{ej}) h^2 / (1 - h^2)] \right] \end{aligned} \quad [33]$$

where $\nu = \tilde{\nu}_u$ and

$$c = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \frac{1}{m}$$

For repeatability, $r = (\sigma_u^2 + \sigma_c^2)/\sigma_p^2$, we used the transformation $\sigma_e^2 \rightarrow r$ in the conditional posterior density of σ_e^2 to obtain

$$\hat{p}(r|\mathbf{y}) = cr^{-2} \left(\frac{1-r}{r} \right)^{-\nu/2-1} \sum_{j=1}^m (s_{ej}/(x_{uj} + x_{cj}))^{\nu/2} \exp \left[-\frac{1}{2} \nu s_{ej} / [(x_{uj} + x_{cj})(1-r)/r] \right] \quad [34]$$

where c is as in [33] but with $\nu = \tilde{\nu}_e$.

If one wishes to make inferences about the variance ratio, $\gamma = \sigma_u^2/\sigma_e^2$, the transformation $\sigma_u^2 \rightarrow \gamma$ yields the estimator of the marginal posterior density of γ .

$$\hat{p}(\gamma|\mathbf{y}) = \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} \gamma^{-\nu/2-1} \frac{1}{m} \sum_{j=1}^m (s_{uj}/x_{ej})^{\nu/2} \exp(-\frac{1}{2\gamma} \nu s_{uj}/x_{ej}) \quad [35]$$

where $\nu = \tilde{\nu}_u$. The variance ratio, $\delta = \sigma_c^2/\sigma_e^2$, is estimated in the same manner as for γ in [35] with the samples s_{cj} substituted in place of s_{uj} and $\nu = \tilde{\nu}_c$, using the transformation $\sigma_c^2 \rightarrow \delta$.

RESULTS

When variance components are known (CASE I), the marginal posterior distributions of all location parameters are normal (Gianola and Fernando, 1986). The mean (mode or median) of the marginal distribution of a location parameter is given by the corresponding component of the solution vector of the mixed model equations, and the variance of the distribution is equal to the corresponding diagonal element of the inverted mixed model coefficient matrix, multiplied by the residual variance. These are mathematical facts, and do not relate in any way to the Gibbs sampler. We used this knowledge to assess the convergence of the Gibbs sampler, which gives Monte-Carlo estimates of the posterior means and variances. In CASE I, for the data at hand, the posterior distributions can be arrived at more efficiently by direct inversion or iterative methods than *via* the Gibbs sampler.

Table I contains results of a comparison between the posterior means and variances estimated by Gibbs sampling (GIBBS) with the exact values found by solving directly the mixed model equations (TRUE). Several criteria were used to compare the 2 sets of results: absolute difference (bias) between TRUE and GIBBS; absolute relative bias (RB) defined as bias divided by TRUE; the slopes of the linear regression of TRUE on GIBBS, and *vice versa*; and the correlation between TRUE and GIBBS. Of course, GIBBS and TRUE were not exactly the same because GIBBS is subject to Monte-Carlo sampling errors; as $m(k)$ goes to infinity, GIBBS is expected to converge to TRUE. We found excellent agreement

between TRUE and GIBBS for all these criteria. The average absolute RB did not exceed 1%, except for the posterior means of additive genetic and permanent environmental effects. For these effects, the RB criterion can be misleading, because the true posterior means were very small in value, so that even small biases made the RB very large. The regressions and correlation coefficients between GIBBS and TRUE were all close to 1.0 for both means and variances. All these results indicate that the Gibbs sampler converged in this application.

Table I. A comparison between the true posterior means and variances with Monte-Carlo estimates.

	<i>Fixed effects</i>		<i>Additive effects</i>		<i>PE effects</i>	
	<i>Mean</i>	<i>Variance</i>	<i>Mean</i>	<i>Variance</i>	<i>Mean</i>	<i>Variance</i>
Ave Bias	0.0121	0.0100	0.00233	0.000686	0.000545	0.000182
Ave RB (%)	0.712	0.716	6.680	0.372	9.162	0.324
Regression slope						
GIBBS on TRUE	1.002	1.000	1.000	0.993	0.999	1.014
TRUE on GIBBS	0.998	0.999	1.000	1.005	1.001	0.964
Corr(GIBBS, TRUE)	1.000	1.000	1.000	0.999	1.000	0.989

TRUE: exact solutions by direct inversion; GIBBS: Monte-Carlo estimates *via* Gibbs sampling; PE effects: permanent environmental effects; Ave |Bias|: average absolute difference between TRUE and GIBBS; Ave |RB|: average absolute bias relative to TRUE; Corr(GIBBS, TRUE): correlation between TRUE and GIBBS.

The true and estimated posterior distributions of the 4 selected location parameters are depicted in figure 1. The true densities are simply normal density plots with means and variances from the TRUE analysis. The estimated densities were obtained with [17] and [19]. The 3 curves overlapped perfectly, indicating convergence of the Gibbs sampler to the true posterior distributions.

Figure 2 depicts the marginal posterior densities of the same 4 selected location parameters for CASE II, *ie* with unknown variances. For the 2 fixed effects, the distributions were essentially symmetric, and similar to those found for CASE I (fig 1). This indicated, for this application, that replacing the unknown variances by REML estimates, and then completing a Bayesian analysis of fixed effects as if variances were known (Gianola *et al*, 1986), would give a good approximation to inferences about fixed effects in the absence of knowledge about variances. Note that the variances of the posterior distributions of the fixed effects shown in figures 1 and 2 are similar. In theory, one would expect the posterior variances to be somewhat larger in CASE II. However, it should be borne in mind that the Gibbs variances are Monte-Carlo estimates, therefore subject to sampling error.

A noteworthy feature of figure 2 was that densities of the additive genetic and permanent environmental effects were skewed, in contrast to the normal distributions found in CASE I. Further, the posterior densities were sharply peaked in the neighborhood of 0, the prior mean; this is consistent with the fact (as discussed later) that in this data there was considerable density near 0 for the additive genetic and permanent environmental components of variance. For these 2 parameters, an analysis using REML estimates would tend to give misleading

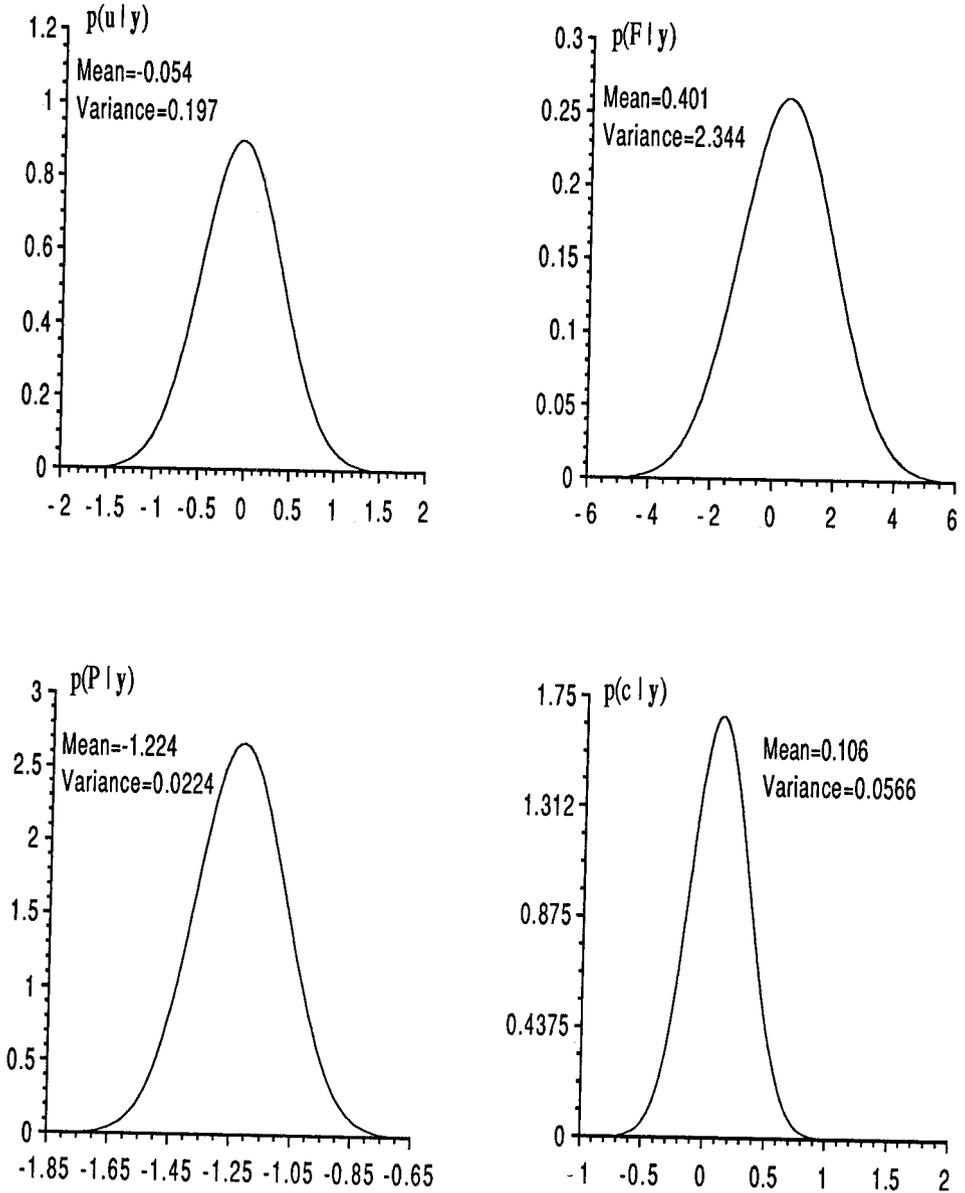


Fig 1. Estimated marginal densities, $p(\cdot|y)$, of 4 selected location parameters with variance components known (CASE I); u = additive genetic effect, F = farrowing season, P = parity and c = permanent environment.

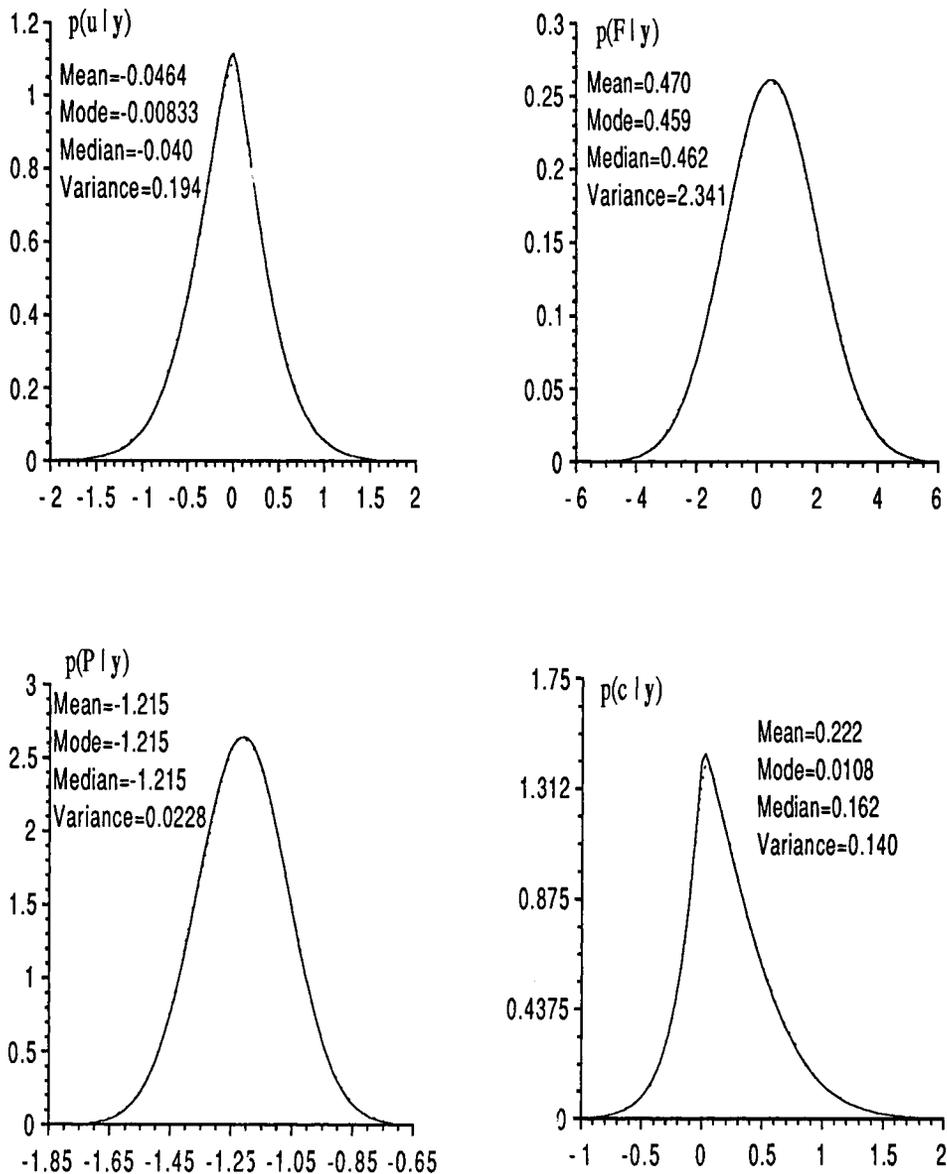


Fig 2. Estimated marginal densities, $p(\cdot|y)$, of the same 4 selected location parameters as in fig 1 with variance components unknown (CASE II); u = additive genetic effect, F = farrowing season, P = parity and c = permanent environment. Solid curve = density estimated with [17]; dotted curve = density estimated with [19].

probability statements and posterior confidence intervals. In particular, in CASE I, the posterior mean (variance) of the selected permanent environmental effect was 0.106 (0.0566); in CASE II, the corresponding figure was 0.22 (0.140). The data contained little information about the permanent environmental effect of this animal, and this is proportional to the number of litters produced by the sow in question. It is precisely in these instances that 'errors' in variance component estimates are crucial. The posterior variance of the permanent environmental effect in CASE II was almost twice as large as in CASE I, illustrating the impact of errors in estimating variance components on inferences. The point is that variances, interval estimates and probability statements about location parameters based on normal approximations, with variance components assumed known in the mixed model equations, can be misleading when information about location parameters and variances is scant in the data. The more information one has about a location parameter, the less influential are the assumed values for the variance components. The problem could be serious if the number of location parameters is large relative to the number of observations, *eg*, in an animal model, and if the data do not contain sufficient information about the variance parameters. An exact Bayesian analysis such as the one conducted here for CASE II would correct all these problems.

Estimated densities for variance components and their functions are presented in figure 3. REML estimates (Meyer, 1988) are also included for comparison purposes. The striking feature was that all distributions, with the exception of those of the residual and phenotypic variances, were skewed. Hence, the mean, mode and median tended to differ from each other. Consider, for example, the posterior distribution of the additive genetic variance; the REML estimate was 0.206, identical to the estimated posterior mean, but quite different from the marginal mode (0.165). For σ_c^2 , the REML estimate was closer to the marginal mode than to the mean or median. A naive 95% 'confidence interval' for σ_c^2 based on the mean and variance of the posterior distribution and asymptotic theory would be $(-0.052, 0.309)$; inference of this type is typical in likelihood based analyses. In the light of the Bayesian analysis depicted in figure 3, the hypothesis that the permanent environmental variance is zero could not be rejected, although there is considerable posterior probability that $\sigma_c^2 > 0.04$. Further, whereas any reasonable Bayesian confidence interval would be in the permissible parameter space, an REML interval would not in this case. For this data set, 95% asymptotic confidence intervals for σ_u^2 and σ_c^2 based on the REML analysis were $(-0.971, 1.356)$ and $(-0.267, 0.384)$, respectively.

The estimated densities using [17] were less smooth than those based on [19]. This was due to Monte-Carlo sampling errors; the curves can be smoothed by increasing the length of the chain.

DISCUSSION

A Gibbs sampling scheme was developed for a univariate Gaussian mixed linear model with correlated observations, such as those arising in quantitative genetics. With this implementation, a full Bayesian analysis of the location and dispersion parameters, or of their functions, in a real-life mixed linear model was possible. The Gibbs sampler made feasible integration of all nuisance parameters, and gave

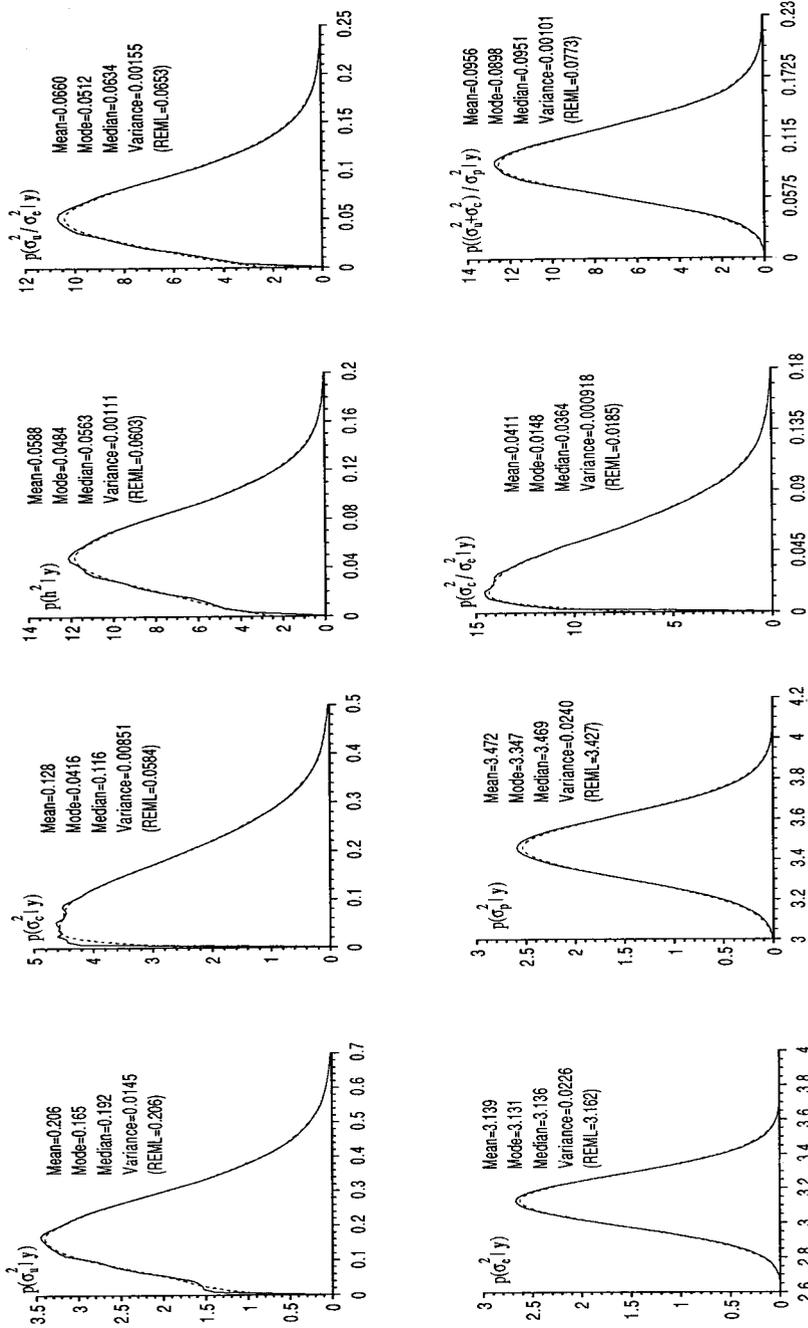


Fig 3. Estimated marginal densities, $p(\cdot | y)$, of variances components and functions thereof (CASE II) $\sigma_p^2 = \sigma_u^2 + \sigma_c^2 + \sigma_e^2 + h^2 = \sigma_u^2 / \sigma_p^2$. Solid curve = density estimated with [17]; dotted curve = density estimated with [19]

a Monte-Carlo estimate of the marginal posterior distribution of the parameter of interest. In the classical sense, this is equivalent to taking into account errors in estimation of all other parameters in the model when inferences about a parameter of interest are made. This is precisely why REML was advanced over maximum likelihood: errors in estimating fixed effects are taken into account in REML. In this sense, a marginal Bayesian analysis with flat priors can be thought of as an analysis of a marginal likelihood, with additional richness brought by the probability calculus on which Bayesian inference is based.

Bayesian analysis *via* Gibbs sampling provides the complete marginal posterior distribution of an unknown. Any features of this distribution can be computed, including probability statements. Because Bayes theorem operates within the space in which parameters are defined, all statistics fall in the permissible parameter space. This is a serious problem of frequentist procedures such as REML. Although the REML estimates are defined within the permissible parameter space, interval estimates based on asymptotic theory can include values outside its boundaries, as illustrated previously.

Unfortunately, a richer analysis requires more intensive computations. For the problem studied in this paper, it took about 14.5 and 23 hours of CPU for CASES I and II, respectively, on a HP9000/827 running HPUNIX 8.02, with a Gibbs chain length of 1 205 000. This certainly limits the applicability of the procedure to large problems, at least at present. Our experience suggests that the procedure is feasible with as many as 10 000 location parameters; hence, analysis of data from designed experiments would be feasible. With the fast advances in computing technology, it is likely that much larger models could be handled efficiently in the near future. We do not advocate at this time Gibbs sampling as a computing method for routine genetic evaluation. However, it is appealing for scientific purposes, *eg*, when many simplifying assumptions must be relaxed, or when model flexibility is needed.

The chain length of 1 205 000 was deliberately long. Summary statistics of a marginal distribution can be computed from a much shorter chain in practice, with relatively high precision. This, of course, would reduce computing cost.

Theoretical results guarantee that an irreducible Markov chain converges to its equilibrium distribution (Kipnis and Varadhan, 1986; Tierney, 1991). However, this does not translate easily into practical guidelines for convergence checking. Our heuristic convergence checking procedure was to run chains under different specifications (starting values, chain length and number of samples saved) of the sampler. If they produced similar results, convergence was assumed. Our experience suggested that if one can obtain a smooth density curve by averaging conditionals, as in [17], the sampler has converged. In CASE II, a smooth curve using [17] was much harder to obtain for dispersion than for location parameters; with [17], very long chains were needed to obtain smooth estimated densities. In CASE I, convergence was sure, because the estimated densities were almost identical to those derived by analytical means. At any rate, checking for convergence is a difficult problem in most areas of numerical analysis, and Gibbs sampling is no exception. Here, there are additional complications stemming from Monte-Carlo errors and from convergence in probability to the true distributions. In the process of monitoring convergence, we observed that this was slower for σ_u^2 and σ_c^2 than for other parameters.

We used 2 density estimators: averaging conditionals [17] and the normal density estimation [19]. Theoretically, [17] is expected to be more efficient than its counterpart because of the use of conditional information (Gelfand and Smith, 1990). However, we did not observe sizable differences in our analysis, as can be ascertained from figures 1–3. In fact, we found that similar density estimates could be obtained with [19], but using much fewer samples than 120 000. This would favor [19] over [17], in this situation. The naive fixed window length used in [19] throughout performed well in all cases.

The procedure can be divided into 2 stages: Gibbs sampling and post-Gibbs analysis. In our case, because sample size was large ($m = 120\,000$), the post Gibbs analysis was onerous. In general, large sample sizes are needed because of high serial correlations between consecutive samples. The effective sample size, perhaps measured as $m(1 - \rho)/(1 + \rho)$ (Tierney, 1991), where ρ is the lag-one serial correlation, could be much smaller than m . For example, if $\rho = 0.9$, the effective sample size would be 6 316, or 5.26% of 120 000. In our study, we monitored serial correlations between consecutive drawings for the variance components. The estimated lag-300 correlations for σ_u^2 and σ_c^2 were 0.6 and 0.3, respectively, while the lag-one correlation for σ_e^2 was almost 0. This is why the chains were so long, and so many samples were saved.

One possible way to reduce dependence between samples would be to embed a Hasting or Metropolis updating step in the basic Gibbs sampling scheme, as used, for example, in pedigree analysis (Lin and Thompson, 1993). Further research is needed in this area for the type of models applied in animal breeding.

We have demonstrated in this paper that the Gibbs sampling scheme can be used successfully to carry out an exact Bayesian analysis of all parameters in a general univariate mixed linear model. The method, however, could also be used in classical situations for problems where analytical integration is intractable. Examples are Besag and Clifford (1989, 1991) on Monte-Carlo tests, and Geyer and Thompson (1992) or Gelfand and Carlin (1991) on Monte-Carlo maximum likelihood. Extensions to multivariate problems, *eg*, genetic maternal effects, are in progress (Jensen *et al*, 1994). An application of Gibbs sampling to the analysis of selection experiment is given by Wang *et al* (1994).

ACKNOWLEDGMENTS

We thank the College of Agricultural and Life Sciences (CALS), University of Wisconsin-Madison, and the National Pork Producer Council, USA, for supporting this work. Computations were made on the HP9000/827 of CALS, with additional computing resources generously provided by the San Diego Supercomputer Center. M Perez-Enciso is thanked for kindly providing the data, and M Newton is thanked for useful suggestions. We thank 2 anonymous reviewers for comments on the paper.

REFERENCES

- Besag J, Clifford P (1989) Generalized Monte-Carlo significance tests. *Biometrika* 76, 633-642
Besag J, Clifford P (1991) Sequential Monte-Carlo p -values. *Biometrika* 78, 301-304

- Besag J, Green PJ (1993) Spatial statistics and Bayesian computation (with discussion). *J R Statist Soc B* 55, 25-37
- Besag J, York J, Mollie A (1991) Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann Inst Statist Math* 43, 1-59
- Bulmer MG (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford
- 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
- Cochran WG (1951) Improvement by means of selection. *Proc 2nd Berkeley Symposium* (J Neyman ed) Math Stat and Prob, University of California Press, Los Angeles, CA, 449-470
- Diebolt J, Robert CP (1993) *J R Statist Soc B* 55, 71-72
- Dobao MT, Rodriganez J, Silio L (1983) Seasonal influence on fecundity and litter performance characteristics in Iberian pigs. *Livest Prod Sci* 10, 601-610
- Fernando RL, Gianola D (1986) Effect of assortative mating on genetic change due to selection. *Theor Appl Genet* 72, 395-404
- Gelfand AE, Carlin BP (1991) Maximum likelihood estimation for constrained or missing data models. Research report 91-002, Division of Biostatistics, University of Minnesota
- 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
- Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85, 398-409
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences (with discussion). *Stat Sci* 7, 467-511
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence* 6, 721-741
- Geyer CJ (1992) Practical Markov chain Monte Carlo. *Stat Sci* 7, 467-511
- Geyer CJ, Thompson EA (1992) Constrained Monte Carlo maximum likelihood for dependent data. *J R Stat Soc B* 54, 657-699
- Gianola D, Goffinet B (1982) Sire evaluation with best linear unbiased predictors. *Biometrics* 38, 1085-1088
- Gianola D, Fernando RF (1986) Bayesian methods in animal breeding theory. *J Anim Sci* 63, 217-244
- Gianola D, Foulley JL, Fernando RL (1986) Prediction of breeding values when variances are not known. *Genet Sel Evol* 18, 485-498
- Gianola D, Fernando RL, Im S, Foulley JL (1989) Likelihood estimation of quantitative genetic parameters when selection occurs: models and problems. *Genome* 31, 768-777
- Gianola D, Im S, Macedo FWM (1990a) A framework for prediction of breeding value. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds) Springer-Verlag, New York, 210-239
- Gianola D, Im S, Fernando RL, Foulley JL (1990b) Mixed model methodology and the Box-Cox theory of transformations: a Bayesian approach. In: *Advances in*

Statistical Methods for Genetic Improvement of Livestock (D Gianola, K Hammond, eds), Springer-Verlag, New York, 15-40

Gianola D, Foulley JL, Fernando RL, Henderson CR, Weigel KA (1992) Estimation of Heterogeneous variances using empirical Bayes methods: theoretical considerations. *J Dairy Sci* 75, 2805-2823

Gilks WR, Clayton DG, Spiegelhalter DJ, Best NG (1993) Modeling complexity: applications of Gibbs sampling in medicine (with discussion). *J R Statist Soc B* 55, 39-52

Goffinet B (1983) Selection on selected records. *Genet Sel Evol* 15, 91-98

Henderson CR (1963) Selection index and expected genetic advance. In: *Statistical Genetics and Plant Breeding* (WD Hanson, HF Robinson, eds) NAS-NRC 982, Washington, DC

Henderson CR (1973) Sire evaluation and genetic trends. In: *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr Jay L Lush*. ASAS and ADSA, Champaign, IL

Henderson CR (1984) Applications of linear models in animal breeding. University of Guelph, Guelph, Canada

Henderson CR, Kempthorne O, Searle SR, von Krosigk CM (1959) The estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15, 192-218

Im S, Fernando RL, Gianola D (1989) Likelihood inferences in animal breeding under selection: a missing data theory view point *Genet Sel Evol* 21, 399-414

IMSL, INC (1989) IMSL₂, STAT/LIBRARY: FORTRAN Subroutines for Statistical Analysis. Houston, Texas

Jensen J, Wang CS, Sorensen DA, Gianola D (1994) Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects using Gibbs sampler. *Acta Agric Scandinavica* (in press)

Kipnis C, Varadhan SRS (1986) Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun Math Phys* 104, 1-19

Lin S, Thompson EA (1993) *J R Statist Soc B* 55, 81-82

Liu J, Wong W, Kong A (1991) Correlation structure and convergence rate of the Gibbs sampler. Technical report 304. Dept of statistics, University of Chicago

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

Meyer K (1988) DFREML - A set of programs to estimate variance components under an individual animal model. *J Dairy Sci* 71 (Supple 2), 33-34

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

Perez-Enciso M, Gianola D (1992) Estimates of genetic parameters for litter size in six strains of Iberian pigs. *Livest Prod Sci* 32, 283-293

Raftery AE, Banfield JD (1991) Stopping the Gibbs sampler, the use of morphology, and other issues in spatial statistics. *Ann Inst Statist Math* 43, 32-43

Rothschild MF, Henderson CR, Quaas RL (1979) Effects of selection on variances and covariances of simulated first and second lactations. *J Dairy Sci* 62, 996-1002

Silverman BW (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London and New York

Smith AFM, Roberts GO (1993) Bayesian computation *via* Gibbs sampler and related Markov chain Monte-Carlo methods (with discussion). *J R Statist Soc B* 55, 3-23

Tierney L (1991) Markov chains for exploring posterior distributions. Technical report No 560. School of Statistics, University of Minnesota

Toro MA, Silio L, Rodriguez J, Dobao MT (1988) Inbreeding and family index selection for prolificacy in pigs. *Anim Prod* 46, 79-85

Wang CS, Rutledge JJ, Gianola D (1993) Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet Sel Evol* 25, 41-62

Wang CS, Gianola D, Sorensen DA, Jensen J, Christensen A, Rutledge JJ (1994) Response to selection for litter size in Danish Landrace pigs: a Bayesian analysis. *Theor Appl Genet* (in press)

Yu B (1991) Density estimation in the L^∞ norm for dependent data with applications to the Gibbs sampler. Technical report 882. Dept of Statistics, University of Wisconsin

APPENDIX

Derivation of conditional posterior distributions of location parameters

Consider model [1], and write Henderson's mixed model equations as:

$$\mathbf{W}\hat{\boldsymbol{\theta}} = \mathbf{b} \tag{A1}$$

It is well known (*eg*, Gianola and Fernando, 1986) that the posterior distribution of $\boldsymbol{\theta}$ given the variance components \mathbf{v} is multivariate normal with mean $\hat{\boldsymbol{\theta}}$ and variance-covariance matrix $\hat{\mathbf{V}}$, *ie*

$$\boldsymbol{\theta}|\mathbf{y}, \mathbf{v}, \sigma_e^2 \sim N(\hat{\boldsymbol{\theta}}, \hat{\mathbf{V}}),$$

where $\hat{\boldsymbol{\theta}} = \mathbf{W}^{-1}\mathbf{b}$ and $\hat{\mathbf{V}} = \mathbf{W}^{-1}\sigma_e^2$.

Now, partition the location parameter vector into 2 parts: $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}'_2)'$, where θ_1 is a scalar, and express the mixed model equations above as:

$$\begin{bmatrix} w_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & \mathbf{W}_{22} \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ \mathbf{b}_2 \end{bmatrix} \tag{A2}$$

Note that

$$\hat{\theta}_1 = (b_1 - \mathbf{w}_{12}\hat{\boldsymbol{\theta}}_2)/w_{11} \tag{A3}$$

also

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{bmatrix} = \mathbf{W}^{-1}\mathbf{b} = \begin{bmatrix} w^{11} & \mathbf{w}^{12} \\ \mathbf{w}^{21} & \mathbf{W}^{22} \end{bmatrix} \begin{bmatrix} b_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} w^{11}b_1 + \mathbf{w}^{12}\mathbf{b}_2 \\ \mathbf{w}^{21}b_1 + \mathbf{W}^{22}\mathbf{b}_2 \end{bmatrix} \tag{A4}$$

and

$$\hat{\mathbf{V}} = \mathbf{W}^{-1}\sigma_e^2 = \begin{bmatrix} w^{11} & \mathbf{w}^{12} \\ \mathbf{w}^{21} & \mathbf{W}^{22} \end{bmatrix} \sigma_e^2 \tag{A5}$$

Using standard theory, the conditional posterior distribution of θ_1 given $\boldsymbol{\theta}_2$ is also normal with parameters $\tilde{\theta}_1$ and \tilde{v}_1 :

$$\theta_1 | \mathbf{y}, \boldsymbol{\theta}_2, \mathbf{v}, \sigma_e^2 \sim N(\tilde{\theta}_1, \tilde{v}_1). \quad [\text{A6}]$$

Now,

$$\begin{aligned} \tilde{\theta}_1 &= E(\theta_1 | \mathbf{y}, \boldsymbol{\theta}_2, \mathbf{v}, \sigma_e^2) \\ &= \hat{\theta}_1 + \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} (\boldsymbol{\theta}_2 - \hat{\boldsymbol{\theta}}_2) \end{aligned} \quad [\text{A7}]$$

Expressing $\hat{\theta}_1$ and $\hat{\boldsymbol{\theta}}_2$ as in [A4], and using Schur complements, we have

$$\begin{aligned} \tilde{\theta}_1 &= \hat{\theta}_1 - \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \hat{\boldsymbol{\theta}}_2 + \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \boldsymbol{\theta}_2 \\ &= w^{11} b_1 + \mathbf{w}^{12} \mathbf{b}_2 - \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \hat{\boldsymbol{\theta}}_2 + \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \boldsymbol{\theta}_2 \\ &= w^{11} b_1 + \mathbf{w}^{12} \mathbf{b}_2 - \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} (\mathbf{w}^{21} b_1 + \mathbf{W}^{22} \mathbf{b}_2) + \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \boldsymbol{\theta}_2 \\ &= w^{11} b_1 - \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \mathbf{w}^{21} b_1 + \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \boldsymbol{\theta}_2 \\ &= \left(w^{11} - \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \mathbf{w}^{21} \right) b_1 + \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \boldsymbol{\theta}_2 \\ &= (w_{11})^{-1} b_1 + \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \boldsymbol{\theta}_2 \\ &= (w_{11})^{-1} [b_1 + w_{11} \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \boldsymbol{\theta}_2] \\ &= (w_{11})^{-1} (b_1 - \mathbf{w}_{12} \boldsymbol{\theta}_2) \\ &= [b_1 - \sum_{j=2}^N w_{1j} \theta_j] / w_{11} \end{aligned} \quad [\text{A8}]$$

Likewise, from [A5]

$$\begin{aligned} \tilde{v}_1 &= \left(w^{11} - \mathbf{w}^{12} (\mathbf{W}^{22})^{-1} \mathbf{w}^{21} \right) \sigma_e^2 \\ &= \mathbf{W}^{-1} \sigma_e^2 \\ &= \sigma_e^2 / w_{11} \end{aligned} \quad [\text{A9}]$$

Since the matrix partition in [A2] is arbitrary, we have

$$\theta_i | \mathbf{y}, \boldsymbol{\theta}_{-i}, \mathbf{v}, \sigma_e^2 \sim N(\tilde{\theta}_i, \tilde{v}_i) \quad i = 1, 2, \dots, N \quad [\text{A10}]$$

where $\tilde{\theta}_i = (b_i - \sum_{j=1, j \neq i}^N w_{ij} \theta_j) / w_{ii}$ and $\tilde{v}_i = \sigma_e^2 / w_{ii}$. We also note that the conditional distributions in [A10] are independent of the priors used for the variance components.