HAL open science

Gibbs sampling, adaptive rejection sampling and robustness to prior specification for a mixed linear model

Cm Theobald, Mz Firat, R Thompson

To cite this version:

Cm Theobald, Mz Firat, R Thompson. Gibbs sampling, adaptive rejection sampling and robustness to prior specification for a mixed linear model. Genetics Selection Evolution, 1997, 29 (1), pp.57-72. hal-00894156

HAL Id: hal-00894156

https://hal.science/hal-00894156v1

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genet Sel Evol (1997) 29, 57-72
© Elsevier/INRA

I.N.R.A. C.R.J.
Département de Génétique Animale
BIBLIOTHÈQUE
F-78352 JOUY EN JOSAS

Original article

Gibbs sampling, adaptive rejection sampling and robustness to prior specification for a mixed linear model

CM Theobald, MZ Firat *, R Thompson †

Department of Mathematics and Statistics, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, UK

(Received 4 October 1995; accepted 13 January 1997)

Summary – Markov chain Monte-Carlo methods are increasingly being applied to make inferences about the marginal posterior distributions of parameters in quantitative genetic models. This paper considers the application of one such method, Gibbs sampling, to Bayesian inferences about parameters in a normal mixed linear model when a restriction is imposed on the relative values of the variance components. Two prior distributions are proposed which incorporate this restriction. For one of them, the *adaptive rejection sampling* technique is used to sample from the conditional posterior distribution of a variance ratio. Simulated data from a balanced sire model are used to compare different implementations of the Gibbs sampler and also inferences based on the two prior distributions. Marginal posterior distributions of the parameters are illustrated. Numerical results suggest that it is not necessary to discard any iterates, and that similar inferences are made using the two prior specifications.

adaptive rejection sampling / Bayesian inference / Gibbs sampling / mixed linear model / sire model

Résumé – Échantillonnage de Gibbs, échantillonnage de rejet adapté et robustesse relative à l'a priori spécifié pour un modèle linéaire mixte. *On utilise de plus en plus des méthodes de Monte-Carlo basées sur des chaînes de Markov pour faire des inférences sur des distributions marginales a posteriori des paramètres de modèles génétiques quantitatifs. Cet article concerne l'application d'une telle méthode, l'échantillonnage de Gibbs, à des inférences bayésiennes sur des paramètres d'un modèle linéaire mixte normal, quand on impose une contrainte sur les valeurs relatives des composantes de variance. Deux distributions a priori incorporant cette contrainte sont proposées. Pour l'une d'elles, la technique de l'*échantillonnage de rejet adapté *est employée pour échantillonner dans une distribution conditionnelle a posteriori d'un rapport de variance. Des données simulées*

*Present address: Department of Animal Science, University of Çukurova, Faculty of Agriculture, 01330 Balcalı, Adana, Turkey.

†Present address: Rothamsted Experimental Station, Harpenden, Hertfordshire AL5 2JQ, UK.

dans un modèle père équilibré sont utilisées pour comparer différentes mises en œuvre de l'échantillonnage de Gibbs et aussi les inférences basées sur les deux distributions a priori. Les distributions marginales a posteriori des paramètres sont données comme illustrations. Les exemples numériques suggèrent qu'il n'est nécessaire d'éliminer aucun résultat d'itération et que des inférences similaires sont faites en utilisant l'un ou l'autre des deux a priori.

échantillonnage de rejet adapté / inférence bayésienne / échantillonnage de Gibbs / modèle linéaire mixte / modèle père

INTRODUCTION

Parameters such as heritability, and predictors of individual breeding values, depend on the unknown variance components in quantitative genetic models. The wide array of methods available for estimating variance components includes ANOVA, likelihood-based methods and Bayesian methods. A difficulty with ANOVA methods which has concerned many authors is that ANOVA estimates may lie outside the parameter space, for example estimates of heritabilities which are negative or greater than one. The use of such estimates for genetic parameters can lead to very inefficient selection decisions.

Owing to rapid advances in computing technology, methods based on likelihood have become common in animal breeding; in particular REML (Patterson and Thompson, 1971) has been widely used. REML estimators are by definition always within the parameter space, and they are consistent, asymptotically normal and efficient (Harville, 1977). However, interval estimates for variances based on the asymptotic distribution of REML estimators can include negative values (Gianola and Foulley, 1990).

In recent years, Bayesian methods have been applied to variance component estimation in animal breeding (Harville, 1977; Gianola and Fernando, 1986; Gianola et al, 1986; Foulley et al, 1987; Cantet et al, 1992; Wang et al, 1993, 1994). For normal models, the joint posterior distribution of the dispersion parameters can be derived, but numerical integration techniques are required to obtain the marginal distributions of functions of interest.

Gibbs sampling is a Markov chain Monte-Carlo method which operates by generating samples from a sequence of conditional distributions; under suitable conditions, its equilibrium distribution is the joint posterior distribution. Gilks et al (1996) provide a thorough introduction to the Gibbs sampler. The method is applied to variance component estimation in a general univariate mixed linear model by Wang et al (1993). Also, Wang et al (1994) obtain marginal inferences about fixed and random effects, variance components and functions of them through a scalar version of the Gibbs sampling algorithm.

The objective of this paper is to broaden the application of Gibbs sampling in an animal breeding context by considering prior distributions which incorporate a constraint on the relative values of variance components: such a constraint might arise from restricting heritability to values less than one. The next section defines the model to be considered and sets out two possible forms for the prior distribution: the second of these is in terms of one variance and a variance ratio. The following section describes how the techniques of Gibbs sampling and adaptive rejection

sampling can be applied to inferences about model parameters under the two specifications. Simulated data from a balanced univariate one-way sire model are used to demonstrate these methods, to compare three implementations of Gibbs sampling and to examine the robustness of the inferences to the choice of prior.

THE MODEL AND ALTERNATIVE PRIOR DISTRIBUTIONS

Model

We consider inferences about model parameters for a univariate mixed linear model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \qquad [1]$$

where $\mathbf{y}$ and $\mathbf{e}$ are n-vectors representing observations and residual errors, $\boldsymbol{\beta}$ is a p-vector of 'fixed effects', $\mathbf{u}$ is a q-vector of 'random effects', $\mathbf{X}$ is a known $n \times p$ matrix of full column rank and $\mathbf{Z}$ is a known $n \times q$ matrix. Here $\mathbf{e}$ is assumed to have the distribution $N_n(\mathbf{0}, \sigma_e^2\,\mathbf{I})$, independently of $\boldsymbol{\beta}$ and $\mathbf{u}$. Also $\mathbf{u}$ is taken to be $N_q(\mathbf{0}, \sigma_u^2\,\mathbf{G})$ with $\mathbf{G}$ a known positive definite matrix.

Restrictions on variance ratios

In animal breeding applications, interest may be in making inferences about ratios of variances or functions thereof, rather than about individual variance components themselves. For example, if $\mathbf{u}$ and $\mathbf{y}$ represent vectors of sire effects and of phenotypic values recorded on the sires' offspring in unrelated paternal half-sib families, and γ denotes the variance ratio σ_u^2/σ_e^2, then the heritability, h^2, of the trait is an increasing function of γ given by $h^2 = 4/(1+\gamma^{-1})$. The constraint that the heritability is between 0 and 1 restricts γ to lie in the interval $\left(0, \frac{1}{3}\right)$. A method of inference which ignores such restrictions may lead to ridiculous estimates of heritability. More generally, we might consider restricting γ to an interval $(0, g)$, corresponding to the constraint that $0 < \sigma_u^2 < g\sigma_e^2$, where g is a known constant.

Bayesian formulation with variance parameters σ_e^2 and σ_u^2

In the above model, the assumption made about the distribution of $\mathbf{u}$ must be supplemented by prior distributions for $\boldsymbol{\beta}$, σ_e^2 and σ_u^2. In specifying their form, we might aim to reflect the animal breeder's prior knowledge while permitting convenient calculations on the resulting posterior distributions.

The prior distribution of $\boldsymbol{\beta}$ is taken here to be uniform, corresponding to little being known about its value initially. Wang et al (1993, 1994) make the same prior assumption about $\boldsymbol{\beta}$ (in a model with further random effects), and consider taking σ_e^2 and σ_u^2 to be independent with scaled inverse-χ^2 distributions $\chi^{-2}(\nu_e, (\nu_e k_e)^{-1})$ and $\chi^{-2}(\nu_u, (\nu_u k_u)^{-1})$; k_e^{-1} and k_u^{-1} can then be interpreted as prior expectations of σ_e^{-2} and σ_u^{-2}, respectively, and ν_e and ν_u as precision parameters analogous to degrees of freedom.

This prior assumption may be modified to take account of an upper bound on the variance ratio, so that σ_e^2 and σ_u^2 are not independent, but have a joint prior density of the form

$$f(\sigma_e^2, \sigma_u^2 \mid \nu_e, k_e, \nu_u, k_u) \propto (\sigma_e^2)^{-\frac{1}{2}(\nu_e+2)} (\sigma_u^2)^{-\frac{1}{2}(\nu_u+2)} \exp\left[-\frac{1}{2}\left(\frac{\nu_e k_e}{\sigma_e^2} + \frac{\nu_u k_u}{\sigma_u^2}\right)\right]$$
$$(0 < \sigma_e^2 < \infty,\ 0 < \sigma_u^2 < g\sigma_e^2) \quad [2]$$

The conditional density of $\mathbf{y}$ given $\boldsymbol{\beta}$, $\mathbf{u}$ and σ_e^2 for the model given in [1] is proportional to $(\sigma_e^2)^{-\frac{1}{2}n} \exp[-\frac{1}{2}\sigma_e^{-2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})]$, so the joint posterior density is given by

$$f_1(\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \sigma_u^2 \mid \mathbf{y}) \propto (\sigma_e^2)^{-\frac{1}{2}(n+\nu_e+2)} (\sigma_u^2)^{-\frac{1}{2}(q+\nu_u+2)}$$
$$\times \exp\left\{-\frac{1}{2\sigma_e^2}\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \nu_e k_e\right] - \frac{1}{2\sigma_u^2}\left[\mathbf{u}'\mathbf{G}^{-1}\mathbf{u} + \nu_u k_u\right]\right\}$$
$$(\boldsymbol{\beta} \in R^p,\ \mathbf{u} \in R^q,\ 0 < \sigma_e^2 < \infty,\ 0 < \sigma_u^2 < g\sigma_e^2) \quad [3]$$

It is shown in the *Appendix* that if ν_u is positive and $n + \nu_e$ exceeds p then the marginal posterior distribution of γ is proper and hence this joint distribution is also.

To implement the Gibbs sampling algorithm, we require the conditional posterior distributions of each of β, $\mathbf{u}$, σ_e^2 and σ_u^2 given the remaining parameters, the so-called *full conditional distributions*, which are as follows (using the notation [.|.] for a conditional distribution)

$$[\boldsymbol{\beta} \mid \mathbf{u}, \sigma_e^2, \sigma_u^2, \mathbf{y}] = N_p\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Z}\mathbf{u}), \sigma_e^2(\mathbf{X}'\mathbf{X})^{-1}\right) \quad [4]$$

$$[\mathbf{u} \mid \boldsymbol{\beta}, \sigma_e^2, \sigma_u^2, \mathbf{y}] = N_q\left((\mathbf{Z}'\mathbf{Z} + \gamma^{-1}\mathbf{G}^{-1})^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \sigma_e^2(\mathbf{Z}'\mathbf{Z} + \gamma^{-1}\mathbf{G}^{-1})^{-1}\right) \quad [5]$$

$$[\sigma_e^2 \mid \boldsymbol{\beta}, \mathbf{u}, \sigma_u^2, \mathbf{y}] = \chi^{-2}\left(n + \nu_e, [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) + \nu_e k_e]^{-1}\right) \quad [6]$$

$$[\sigma_u^2 \mid \boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \mathbf{y}] = \chi^{-2}\left(q + \nu_u, [\mathbf{u}'\mathbf{G}^{-1}\mathbf{u} + \nu_u k_u]^{-1}\right) \quad [7]$$

(truncated at $g\sigma_e^2$). Apart from the use here of proper prior distributions for the variances, these are similar to the conditional distributions given in Wang et al (1993). Methods for sampling from these distributions are well known.

An alternative Bayesian formulation with prior independence of σ_e^2 and γ

As Wang et al (1993) show for a model with a slightly simpler prior for σ_e^2 and σ_u^2, the prior specification in terms of these two parameters is a convenient one for applying the method of Gibbs sampling. It is also easily generalized when several traits are recorded. It may not, however, be the most useful specification for eliciting the prior opinions of animal breeders about parameter values. The animal breeder

may prefer to think in terms of the heritability of a trait, equivalent for a paternal half-sib model to the variance ratio $\gamma = \sigma_u^2/\sigma_e^2$ defined above. A second function of the two variances must also be considered in order to specify their joint prior distribution, and σ_e^2 may be an appropriate choice because inferences about σ_e^2 from previous data are likely to be much more precise than those about σ_u^2. To specify a joint prior distribution for the sire and residual variance components, it may therefore be convenient to assign priors to σ_e^2 and the variance ratio γ. We thus consider the alternative parameter vector $(\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \gamma)$, but make the same prior assumptions for $\boldsymbol{\beta}$ and $\mathbf{u}$ as before, noting that σ_u^2 becomes $\gamma\sigma_e^2$. We can incorporate an upper bound g on the range of γ more naturally than with the previous formulation by using the family of Beta distributions on the interval $(0, g)$: this is a fairly flexible family with probability density of the form

$$f(\gamma \mid a, b) \propto \gamma^{a-1}\,(g-\gamma)^{b-1} \quad (0 < \gamma < g) \qquad [8]$$

If γ is taken to be independent of $\boldsymbol{\beta}$ and σ_e a priori, and σ_e^2 assumed to be $\chi^{-2}(\nu_e, (\nu_e k_e)^{-1})$ then the joint posterior density is given by

$$f_2(\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \gamma \mid \mathbf{y}) \propto \left(\sigma_e^2\right)^{-\frac{1}{2}(n+q+\nu_e+2)} \gamma^{a-\frac{1}{2}q-1}\,(g-\gamma)^{b-1}$$
$$\times \exp\left\{-\frac{1}{2\sigma_e^2}\left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) + \nu_e k_e + \gamma^{-1}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}\right]\right\}$$
$$(\boldsymbol{\beta} \in R^p,\ \mathbf{u} \in R^q,\ 0 < \sigma_e^2 < \infty,\ 0 < \gamma < g) \qquad [9]$$

The full conditional posterior distributions for $\boldsymbol{\beta}$ and $\mathbf{u}$ are the same as [4] and [5], while that for σ_e^2 is

$$\chi^{-2}\left(n+q+\nu_e, [(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) + \nu_e k_e + \gamma^{-1}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}]^{-1}\right) \qquad [10]$$

The full conditional density for γ is proportional to

$$\gamma^{a-\frac{1}{2}q-1}\,(g-\gamma)^{b-1} \exp\left(-\frac{\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}}{2\gamma\sigma_e^2}\right) \quad (0 < \gamma < g) \qquad [11]$$

This family is not a well-known one, so it is not immediately clear how to sample from it. However, the *adaptive rejection sampling* technique of Gilks and Wild (1992) provides a reasonably efficient method for sampling from the distribution of $\ln\gamma$.

Given these two methods of including the restriction on σ_u^2/σ_e^2, the question arises of how posterior inferences, especially about γ, are affected when the second prior specification is used instead of the first. To answer this, we consider alternative specifications in which the priors for the variance parameters are similar: to match these distributions we choose a and b in [8] to give γ the same upper and lower quartiles under both. Details are given in the *Appendix*.

CALCULATIONS ON THE POSTERIOR DISTRIBUTIONS

Gibbs sampling

Markov chain Monte-Carlo methods, particularly Gibbs sampling, are now widely used for making inferences from posterior distributions of parameters. Gelfand and Smith (1990) review the Gibbs sampler and other Monte-Carlo methods, and Gelfand et al (1990) illustrate the use of the Gibbs sampler on a range of statistical models, including variance components models. Wang et al (1993, 1994) give two versions of the Gibbs sampler for a univariate normal mixed linear model assuming prior independence of the variance components.

The Gibbs sampler algorithm generates samples iteratively from the full conditional distributions of disjoint subsets of the parameters given the current value of the complementary subset.

The following version of the Gibbs sampler can be applied to sample from the posterior distribution defined in [3] using the full conditional distributions given in [4]–[7]. It incorporates the restriction that σ_u^2/σ_e^2 cannot exceed an upper bound g by discarding pairs of variances which do not satisfy this condition.

(i) Choose an arbitrary starting value $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)}, \mathbf{u}^{(0)}, \sigma_e^{2(0)}, \sigma_u^{2(0)})$ for the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{u}, \sigma_e^2, \sigma_u^2)$;

(ii) draw a value $\boldsymbol{\beta}^{(1)}$ from the distribution $[\boldsymbol{\beta} \mid \mathbf{u}^{(0)}, \sigma_e^{2(0)}, \sigma_u^{2(0)}, \mathbf{y}]$;

(iii) draw $\mathbf{u}^{(1)}$ from $[\mathbf{u} \mid \boldsymbol{\beta}^{(1)}, \sigma_e^{2(0)}, \sigma_u^{2(0)}, \mathbf{y}]$;

(iv) draw $\sigma_u^{2(1)}$ from $[\sigma_u^2 \mid \boldsymbol{\beta}^{(1)}, \mathbf{u}^{(1)}, \sigma_e^{2(0)}, \mathbf{y}]$ ignoring the restriction on σ_u^2/σ_e^2;

(v) draw $\sigma_e^{2(1)}$ from $[\sigma_e^2 \mid \boldsymbol{\beta}^{(1)}, \mathbf{u}^{(1)}, \sigma_u^{2(1)}, \mathbf{y}]$ ignoring the restriction on σ_u^2/σ_e^2;

(vi) if $\sigma_u^{2(1)}/\sigma_e^{2(1)} \geqslant g$ then repeat (iv) and (v) until $\sigma_u^{2(1)}/\sigma_e^{2(1)} < g$.

Thus, subsets of the parameters are sampled in an arbitrary order and this cycle defines the transition from $\boldsymbol{\theta}^{(0)}$ to $\boldsymbol{\theta}^{(1)}$. Subsequent cycles replace $\boldsymbol{\theta}^{(1)}$ to give further vectors $\boldsymbol{\theta}^{(2)}$, $\boldsymbol{\theta}^{(3)}$, and so on.

Opinions differ on how the sequence of vectors used in subsequent analyses should be defined. The following three methods of implementation have been proposed, and are compared on simulated data in the next section.

(a) *A single long chain*: one generates a single run of the chain as practised by Geman and Geman (1984), ie,
(vii) repeat (ii)–(vi) m times using updated values and store all the values.
A common modification is to discard vectors from the start of the sequence to allow it to approach equilibrium. The remaining vectors $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(m)}$ approximate simulated values from the posterior distribution of $\boldsymbol{\theta}$, but successive vectors may be correlated and the sequence may be strongly influenced by the starting values.

(b) *Equally spaced samples*: to reduce serial correlations, one can choose suitable integers k and m, perform a single run of length km, and then form a sample from every kth vector, ie, use $\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(2k)}, \ldots, \boldsymbol{\theta}^{(km)}$.

(c) *Multiple short chains*: by contrast, Gelfand and Smith (1990) and Gelfand *et al* (1990) perform m runs each of length l, with different starting values, forming a sample by using the last iterate from each run.

In support of implementation (c), Gelman and Rubin (1992) argue that, even if using one long run may be more efficient, it may still be important to use several different starting points.

It has been common practice to discard a substantial number of initial iterations, and to use implementations (b) and (c) (especially (b), storing only every kth, usually 10th or 20th, iterate). The results of Raftery and Lewis (1992) suggest that in many cases this is rather profligate.

For any individual parameter, the collection of m values can be viewed as a simulated sample from the appropriate marginal distribution. This sample can be used to calculate summaries, such as the marginal posterior mean, or to estimate marginal posterior distributions of parameters or of functions of them using kernel density estimation. We use kernel density estimation below with normal kernels and window width suggested by Silverman (1986).

Adaptive rejection sampling

The *adaptive rejection sampling* technique introduced by Gilks and Wild (1992) is a method of rejection sampling intended for situations in which evaluation of a univariate density is computationally expensive. It uses a squeezing function (Devroye, 1986), and is adaptive in the sense that the rejection envelope and squeezing function (which are piecewise exponential) are revised as sampling proceeds, and approach the density function. In Gibbs sampling, the full conditional distribution for each parameter changes as the iteration proceeds, so that usually only one value is required from a particular density. Adaptive rejection sampling has the advantage that it does not require an optimization step to locate the mode of the density before a sample can be generated.

The method requires that the density is log-concave, that is that the derivative of the logarithm of the density is monotone decreasing. It is shown in the *Appendix* that the density of γ in expression [11] is guaranteed to be log-concave only if a is greater than $\frac{1}{2}q + 1$, but that if we transform to $\delta = \ln\gamma$ then the full conditional density of δ is log-concave if b is at least one. The condition on a is not a reasonable one, since a prior with a maximum at zero must have a no greater than one, for example. However, a value of b less than one implies that the prior density in [8] has a maximum at its upper bound, g (which might correspond to a heritability of one), and it is *not* unreasonable to exclude such values. Hence we use adaptive rejection sampling with $\ln\gamma$.

APPLICATION TO A ONE-WAY CLASSIFICATION

The methods described above were applied to a one-way sire model of the form

$$y_{ij} = \beta + u_i + e_{ij} \quad (i = 1, \ldots, s;\ j = 1, \ldots, d) \qquad [12]$$

This is intended to represent observations of a trait on the d half-sib offspring of each of s sires: β and the u_i are the overall expectation and the ith sire effect, respectively. It is a particular case of the model given in [1]

in which p and q equal 1 and s, and $\mathbf{X}, \mathbf{Z}$ and $\mathbf{G}$ become $\mathbf{1}_{sd}$ (a vector of 1s with length sd), $\mathrm{diag}(\mathbf{1}_d, \ldots, \mathbf{1}_d)$ and $\mathbf{I}_s$. The full conditional distributions for β, $\mathbf{u}$, σ_e^2 and σ_u^2 defined in expressions [4]–[7] become $N\left(\bar{y}_{..} - \bar{u}_{.}, \frac{\sigma_e^2}{sd}\right)$, $N_s\left(\frac{d\gamma}{d\gamma+1}(\bar{\mathbf{y}}_{.} - \beta\mathbf{1}_s), \frac{\gamma\sigma_e^2}{d\gamma+1}\mathbf{I}_s\right)$, $\chi^{-2}\left(sd + \nu_e, \left[\sum_i \sum_j (y_{ij} - \beta - u_i)^2 + \nu_e k_e\right]^{-1}\right)$ and $\chi^{-2}\left(s + \nu_u, \left[\sum_i u_i^2 + \nu_u k_u\right]^{-1}\right)$, where $\bar{\mathbf{y}}_{.}$, $\bar{y}_{..}$ and $\bar{u}_{.}$ denote the vector of family means and the overall means of the y_{ij} and the u_i.

Four sets of data were generated using this model with parameter values $\beta = 0$, $\sigma_e^2 = 0.975$, $\sigma_u^2 = 0.025$ and hence $h^2 = 0.1$ with 25 families and 20 offspring per sire. ANOVA estimates are summarized in table I: the data sets are ordered by the estimates of heritability, which range from −0.073 to 0.306. The first ANOVA estimate of σ_u^2 is negative, making conventional inference about this parameter difficult.

Table I. ANOVA estimates for the four data sets.

Data sets	β	σ_u^2	σ_e^2	γ	h^2
1	0.072	−0.018	1.002	−0.018	−0.073
2	−0.032	0.013	0.966	0.013	0.052
3	0.040	0.029	1.013	0.028	0.110
4	−0.025	0.087	1.056	0.083	0.306

Analysis using the prior for σ_e^2 and σ_u^2

Bayesian analyses of the four data sets are given using the prior distribution in [2] with $\nu_e = \nu_u = 1$, $k_e = 0.975$, $k_u = 0.025$ and $g = \frac{1}{3}$. The prior distribution for σ_e^2 and σ_u^2 is proper and leads to a proper joint posterior distribution (as shown in the *Appendix*), but the small values chosen for ν_e and ν_u correspond to a very dispersed distribution; this is intended to reveal any problems with convergence of the Gibbs sampler algorithm. The sample data provide little information about σ_u^2 as there are only 25 sires. The algorithm is carried out for the four sets of data with the following implementations, which each provide 1 000 Gibbs samples:

(a) a single run of 1 000 iterations;
(b) taking every (1) 10th, (2) 20th and (3) 30th value in single runs of length 10 000, 20 000 and 30 000, respectively, using only one starting value for the parameter vector;
(c) short runs of length (1) 10, (2) 20 and (3) 30, storing the last iterate and replicating this process 1 000 times in each case.

The parameter values used to generate the data were also employed as starting values for all three implementations. No iterates were discarded since the true parameter values were known.

The resulting marginal posterior means and standard deviations for β, σ_u^2, σ_e^2, γ and h^2 for the four data sets are shown in table II. This table reveals little difference between the implementations in the marginal posterior means, except for a slight tendency for standard deviations to be lower using (a) for β, σ_u^2, γ and h^2. Figures 1–4 show the estimated marginal posterior densities for β, σ_u^2, σ_e^2 and h^2 using density estimators with normal kernels under the prior specification of this subsection (with implementation (a)) and that considered in the next subsection.

Analysis assuming prior independence of σ_e^2 and γ

We now compare the above analyses with those based on the second prior specification, which assumes prior independence of σ_e^2 and γ and prior density [8] for γ; we use adaptive rejection sampling on the full conditional posterior distribution of $\delta = \ln \gamma$, as defined in the *Appendix*. The *Appendix* also describes how the values of a and b in the Beta prior for γ are determined from those of ν_e, ν_u, k_e and k_u.

At each Gibbs sampling iteration, adaptive rejection sampling requires at least two points for the construction of the rejection envelope and squeezing function (Gilks and Wild, 1992). The parameter space for δ is unbounded on the left, so Gilks and Wild (1992) advise that the smallest of these initial points be chosen where the derivative of the log posterior density is positive. The 15th and 85th percentiles of the sampling density from the previous iteration were tried as initial points, additional points being supplied if the derivatives at these values had the same sign. Three evaluations of the log posterior density of δ were required on average at each iteration of adaptive rejection sampling.

Convergence of the Gibbs sampler, assessed by monitoring summary statistics for each parameter based on every ten iterations, had clearly occurred before 1 000 iterations. The marginal posterior summaries shown in table II and the density estimates for β, σ_u^2, σ_e^2 and h^2 in figures 1–4 are based on this number of iterations. The results obtained using the two prior specifications agree very closely, especially those for β and σ_e^2.

DISCUSSION

Our results from using Gibbs sampling on a balanced one-way sire model with the prior density of expression [2] for σ_e^2 and σ_u^2 suggest that it is unnecessary to discard all but the 10th, 20th or 30th iterate; doing so is probably quite wasteful. A much larger proportion of the iterates should be used unless storage is an issue. Using 1 000 successive iterates seems adequate to make inferences about the parameters of interest, although the slightly lower standard deviations obtained with implementation (a) may reflect the need to discard 'burn-in' iterations. We encountered no difficulties with the convergence of the Gibbs sampling algorithms used. With real data, we would use REML estimates as starting values for the Gibbs iterations and monitor convergence diagnostics to decide how many 'burn-in' iterations to discard (Gilks et al, 1996).

Table II. Marginal posterior means and standard deviations (SD) of individual parameters for four data sets using 1 000 Gibbs samples. Implementations (a), (b1)–(b3) and (c1)–(c3) are based on the prior distribution for σ_e^2 and σ_u^2, but use different implementations of the Gibbs sampler; (d) assumes prior independence of σ_e^2 and γ and also uses adaptive rejection sampling.

	β		σ_u^2		σ_e^2		γ		h^2	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
					Data set 1					
a	0.0712	0.0489	0.0120	0.0079	0.9868	0.0636	0.0121	0.0079	0.0477	0.0306
b1	0.0724	0.0494	0.0117	0.0077	0.9879	0.0629	0.0119	0.0079	0.0467	0.0306
b2	0.0722	0.0494	0.0117	0.0078	0.9869	0.0614	0.0119	0.0079	0.0467	0.0306
b3	0.0719	0.0502	0.0117	0.0079	0.9837	0.0628	0.0120	0.0083	0.0472	0.0317
c1	0.0708	0.0501	0.0115	0.0079	0.9861	0.0631	0.0117	0.0081	0.0461	0.0311
c2	0.0702	0.0501	0.0118	0.0082	0.9855	0.0621	0.0119	0.0083	0.0471	0.0318
c3	0.0694	0.0482	0.0117	0.0079	0.9880	0.0651	0.0119	0.0082	0.0468	0.0316
d	0.0716	0.0489	0.0119	0.0099	0.9836	0.0625	0.0121	0.0100	0.0472	0.0383
					Data set 2					
a	−0.0324	0.0508	0.0182	0.0127	0.9690	0.0626	0.0189	0.0132	0.0736	0.0498
b1	−0.0313	0.0540	0.0198	0.0146	0.9686	0.0620	0.0206	0.0154	0.0798	0.0575
b2	−0.0318	0.0525	0.0201	0.0145	0.9678	0.0606	0.0209	0.0152	0.0809	0.0571
b3	−0.0325	0.0538	0.0202	0.0145	0.9640	0.0617	0.0212	0.0155	0.0820	0.0578
c1	−0.0344	0.0522	0.0207	0.0144	0.9664	0.0627	0.0216	0.0153	0.0838	0.0572
c2	−0.0321	0.0516	0.0203	0.0150	0.9674	0.0630	0.0210	0.0156	0.0815	0.0582
c3	−0.0346	0.0515	0.0205	0.0146	0.9661	0.0626	0.0214	0.0156	0.0830	0.0583
d	−0.0311	0.0500	0.0182	0.0146	0.9677	0.0620	0.0188	0.0151	0.0730	0.0568
					Data set 3					
a	0.0390	0.0553	0.0288	0.0191	1.0200	0.0670	0.0284	0.0188	0.1093	0.0695
b1	0.0406	0.0589	0.0296	0.0208	1.0205	0.0664	0.0293	0.0209	0.1121	0.0763
b2	0.0398	0.0569	0.0294	0.0199	1.0195	0.0647	0.0290	0.0198	0.0115	0.0729
b3	0.0389	0.0576	0.0295	0.0198	1.0155	0.0661	0.0294	0.0202	0.1127	0.0742
c1	0.0399	0.0573	0.0303	0.0215	1.0209	0.0641	0.0299	0.0217	0.1148	0.0792
c2	0.0381	0.0565	0.0281	0.0194	1.0319	0.0684	0.0277	0.0193	0.1065	0.0708
c3	0.0420	0.0576	0.0289	0.0202	1.0214	0.0669	0.0286	0.0204	0.1097	0.0744
d	0.0407	0.0558	0.0288	0.0219	1.0204	0.0660	0.0284	0.0217	0.1087	0.0797
					Data set 4					
a	−0.0272	0.0715	0.0861	0.0407	1.0625	0.0704	0.0815	0.0389	0.2969	0.1288
b1	−0.0217	0.0777	0.0843	0.0426	1.0608	0.0694	0.0799	0.0406	0.2908	0.1349
b2	−0.0257	0.0768	0.0847	0.0430	1.0603	0.0666	0.0803	0.0411	0.2921	0.1357
b3	−0.0237	0.0781	0.0828	0.0406	1.0651	0.0692	0.0783	0.0390	0.2857	0.1298
c1	−0.0217	0.0713	0.0826	0.0413	1.9661	0.0725	0.0782	0.0404	0.2851	0.1335
c2	−0.0223	0.0734	0.0822	0.0410	1.0650	0.0706	0.0778	0.0397	0.2839	0.1321
c3	−0.0257	0.0763	0.0850	0.0431	1.0634	0.0689	0.0806	0.0417	0.2929	0.1382
d	−0.0261	0.0731	0.0887	0.0388	1.0632	0.0686	0.0837	0.0362	0.3047	0.1209

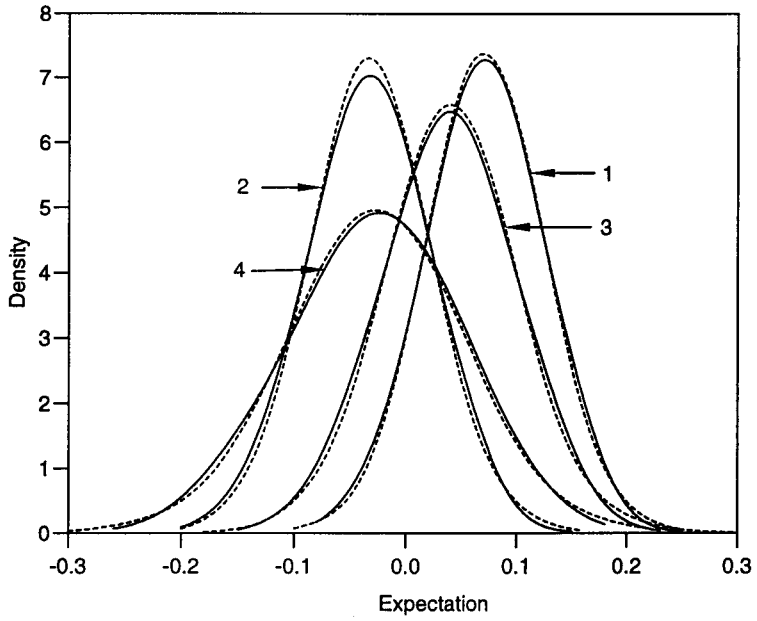

Fig 1. Estimated marginal posterior densities of β for four sets of data; —, prior specification for σ_e^2 and σ_u^2; - - - -, prior specification for σ_e^2 and γ.

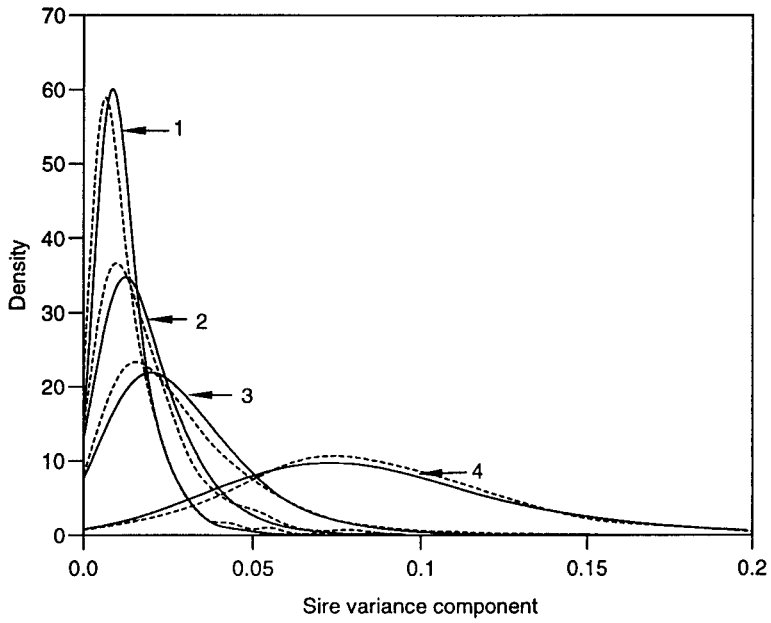

Fig 2. Estimated marginal posterior densities of σ_u^2 for four sets of data; —, prior specification for σ_e^2 and σ_u^2; - - - -, prior specification for σ_e^2 and γ.

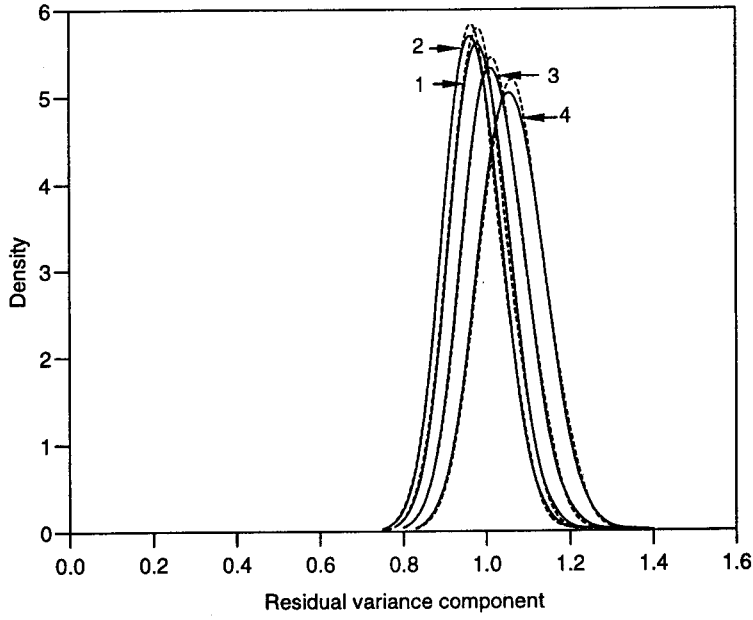

Fig 3. Estimated marginal posterior densities of σ_e^2 for four sets of data; —, prior specification for σ_e^2 and σ_u^2; - - -, prior specification for σ_e^2 and γ.

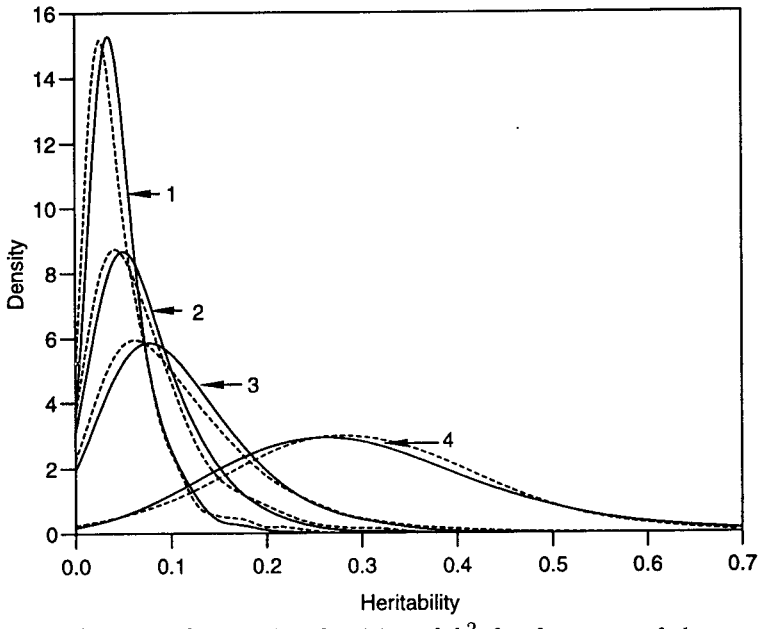

Fig 4. Estimated marginal posterior densities of h^2 for four sets of data; —, prior specification for σ_e^2 and σ_u^2; - - -, prior specification for σ_e^2 and γ.

The prior distribution which assumes σ_e^2 and γ to be independent provides a potentially useful alternative to the usual prior in which σ_e^2 and σ_u^2 are taken to be independent. The adaptive rejection sampling algorithm of Gilks and Wild (1992) is found to work efficiently with the resulting full conditional distribution for $\ln\gamma$.

The sensitivity of a Bayesian analysis to the choice of prior hyperparameters and to different functional forms of prior can be investigated. For our simulated data, the use of an alternative form of prior distribution led to very similar marginal posterior inferences on each parameter.

The use of the alternative prior distribution for σ_e^2 and γ can be extended to models with further variance components, for example if [1] is generalized to

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{T}\mathbf{v} + \mathbf{e} \qquad [13]$$

where $\mathbf{v}$ is an r-vector of random effects, $\mathbf{T}$ is a known $n \times r$ matrix and $\mathbf{v}$ is taken to be $N_r(\mathbf{0},\, \sigma_v^2\,\mathbf{H})$ independently of $\boldsymbol{\beta}, \mathbf{u}$ and $\mathbf{e}$ with $\mathbf{H}$ a known matrix. If $\boldsymbol{\beta}$, σ_e^2 and σ_v^2 are taken to be mutually independent a priori, with uniform, $\chi^{-2}(\nu_e, (\nu_e k_e)^{-1})$ and $\chi^{-2}(\nu_v, (\nu_v k_v)^{-1})$ distributions, respectively, and γ is independent of them with the Beta distribution defined in [8], then the full conditional distributions for $\boldsymbol{\beta}, \mathbf{u}, \mathbf{v}, \sigma_e^2$ and σ_v^2 are, respectively,

$$N_p\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Z}\mathbf{u} - \mathbf{T}\mathbf{v}), \sigma_e^2(\mathbf{X}'\mathbf{X})^{-1}\right) \qquad [14]$$

$$N_q\left((\mathbf{Z}'\mathbf{Z} + \gamma^{-1}\mathbf{G}^{-1})^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{T}\mathbf{v}), \sigma_e^2(\mathbf{Z}'\mathbf{Z} + \gamma^{-1}\mathbf{G}^{-1})^{-1}\right) \qquad [15]$$

$$N_r\left((\mathbf{T}'\mathbf{T} + \varepsilon^{-1}\mathbf{H}^{-1})^{-1}\mathbf{T}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}), \sigma_e^2(\mathbf{T}'\mathbf{T} + \varepsilon^{-1}\mathbf{H}^{-1})^{-1}\right) \qquad [16]$$

$$\chi^{-2}\left(n + q + \nu_e, [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u} - \mathbf{T}\mathbf{v})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u} - \mathbf{T}\mathbf{v}) + \nu_e k_e + \gamma^{-1}\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}]^{-1}\right) \qquad [17]$$

and

$$= \chi^{-2}\left(r + \nu_v, [\mathbf{v}'\mathbf{H}^{-1}\mathbf{v} + \nu_v k_v]^{-1}\right) \qquad [18]$$

where ε denotes σ_v^2/σ_e^2, while that for γ is the same as [11].

ACKNOWLEDGEMENTS

The authors are grateful to W Gilks for providing a FORTRAN program to perform adaptive rejection sampling. MZ Firat was supported by the Turkish Government.

REFERENCES

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

Devroye L (1986) *Non-Uniform Random Variate Generation.* Springer-Verlag, New York

Foulley JL, Im S, Gianola D, Hoschele I (1987) Empirical Bayes estimation of parameters for n polygenic binary traits. *Genet Sel Evol* 19, 197–224

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

Gelman, A, Rubin DB (1992) A single series from the Gibbs sampler provides a false sense of security. In: *Bayesian Statistics 4* (JM Bernardo, JO Berger, AP David, AFM Smith, eds), Clarendon Press, Oxford, 625–631

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

Gianola D, Fernando RL (1986) Bayesian methods in animal breeding theory. *J Anim Sci* 63, 217–244

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

Gianola D, Foulley JL, Fernando RL (1986) Prediction of breeding values when variances are not known. *Genet Sel Evol* 18, 485–498

Gilks WR, Richardson S, Spiegelhalter DJ (1996) *Markov Chain Monte Carlo in Practice.* Chapman and Hall, London

Gilks WR, Wild P (1992) Adaptive rejection sampling for Gibbs sampling. *Appl Statist* 41, 337–348

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

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

Raftery AE, Lewis SM (1992) How many iterations in the Gibbs sampler? In: *Bayesian Statistics 4* (JM Bernardo, JO Berger, AP David, AFM Smith, eds), Clarendon Press, Oxford, 763–773

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

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

APPENDIX

Marginal posterior probability density function of γ using the prior for σ_e^2 and σ_u^2

Integrating the posterior density for $\boldsymbol{\beta}$, $\mathbf{u}$, σ_e^2 and σ_u^2 given in [3] with respect to $\boldsymbol{\beta}$ and $\mathbf{u}$, the posterior density of σ_e^2 and σ_u^2 is found to be proportional to

$$(\sigma_e^2)^{-\frac{1}{2}(n+\nu_e-p-q+2)}\,(\sigma_u^2)^{-\frac{1}{2}(q+\nu_u+2)}\,\left|\mathbf{Z}'\mathbf{Q}\mathbf{Z}+\gamma^{-1}\mathbf{G}^{-1}\right|^{-\frac{1}{2}}\,\exp\left[-\sigma_e^{-2}D(\gamma)\right]$$

where $\mathbf{Q}$ denotes $\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and

$$D(\gamma) = \frac{1}{2}\left\{\nu_e k_e + \gamma^{-1}\nu_u k_u + \mathbf{y}'\left[\mathbf{Q} - \mathbf{Q}\mathbf{Z}(\mathbf{Z}'\mathbf{Q}\mathbf{Z}+\gamma^{-1}\mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{Q}\right]\mathbf{y}\right\}$$

The joint density of σ_e^2 and γ is therefore proportional to

$$(\sigma_e^2)^{-\frac{1}{2}(n+\nu_e+\nu_u-p+2)}\,\gamma^{-\frac{1}{2}(\nu_u+2)}\,\left|\mathbf{G}^{-1}+\gamma\,\mathbf{Z}'\mathbf{Q}\mathbf{Z}\right|^{-\frac{1}{2}}\exp\left[-\sigma_e^{-2}D(\gamma)\right]$$

and the marginal posterior density of γ is proportional to

$$\gamma^{-\frac{1}{2}(\nu_u+2)}\,\left|\mathbf{G}^{-1}+\gamma\,\mathbf{Z}'\mathbf{Q}\mathbf{Z}\right|^{-\frac{1}{2}}\,[D(\gamma)]^{-\frac{1}{2}(n+\nu_e+\nu_u-p)}\qquad(0<\gamma<g)$$

provided that $n+\nu_e+\nu_u-p$ is positive. Whether or not this is a proper density function depends on its behaviour near zero. For small γ, $\left|\mathbf{G}^{-1}+\gamma\,\mathbf{Z}'\mathbf{Q}\mathbf{Z}\right|$ is approximately constant, while the $\gamma^{-1}\nu_u k_u$ term is dominant in $D(\gamma)$ so long as ν_u is positive. So the density of γ behaves like $\gamma^{\frac{1}{2}(n+\nu_e-p-2)}$ near zero, and has a finite integral if the exponent exceeds -1, that is if $n+\nu_e$ exceeds p.

Matching quartiles of γ in the two prior distributions

To compare inferences under the two prior distributions considered in this article, we determine values of the hyperparameters a and b defining the Beta distribution for γ in the second prior so that the two specifications are matched in some sense. The first specification for the variance parameters (in terms of σ_e^2 and σ_u^2) implies that γ (which equals σ_u^2/σ_e^2) has probability density function proportional to

$$\gamma^{\frac{1}{2}(\nu_e-2)}(1+\frac{\nu_e k_e}{\nu_u k_u}\gamma)^{-\frac{1}{2}(\nu_u+\nu_e)}\qquad(0<\gamma<g)$$

Hence $\phi=\gamma k_e/k_u$ has an $F(\nu_e,\nu_u)$ distribution truncated at gk_e/k_u. We choose a and b to match the Beta distribution used for γ in the second prior to this by equating their quartiles. If G denotes the distribution function of $F(\nu_e,\nu_u)$ then ϕ has lower and upper quartiles ϕ_{L}, ϕ_{U} which are solutions of

$$G(\phi_{\mathrm{L}})=0.25\,G(gk_e/k_u)\,,\quad G(\phi_{\mathrm{U}})=0.75\,G(gk_e/k_u)$$

and γ has quartiles $\phi_{\mathrm{L}}k_e/k_u$ and $\phi_{\mathrm{U}}k_e/k_u$. Thus if x_{L} and x_{U} denote the lower and upper quartiles of the Beta distribution $B(a,b)$, we equate x_{L} and x_{U} to $\phi_{\mathrm{L}}k_u/(gk_e)$ and $\phi_{\mathrm{U}}k_u/(gk_e)$, respectively, by solving for a and b the two equations

$$\frac{1}{B(a,b)}\int_0^{x_{\mathrm{L}}}x^{a-1}(1-x)^{b-1}\,\mathrm{d}x=0.25\,,\quad\frac{1}{B(a,b)}\int_0^{x_{\mathrm{U}}}x^{a-1}(1-x)^{b-1}\,\mathrm{d}x=0.75$$

Example: We use the prior specification for σ_e^2 and σ_u^2 with $\nu_e=\nu_u=1$, $k_e=0.975$, $k_u=0.025$ and $g=\frac{1}{3}$, so that $\gamma k_e/k_u$ has an $F(1,1)$ distribution truncated at $\frac{1}{3}k_e/k_u=13$; $G(13)=0.8278$, so

$$\phi_{\mathrm{L}}=G^{-1}(0.25\times0.8278)=0.1136\,,\quad\phi_{\mathrm{U}}=G^{-1}(0.75\times0.8278)=2.1777$$

$$x_{\mathrm{L}}=0.1136/13=0.0087\,,\quad x_{\mathrm{U}}=2.1777/13=0.1675$$

The corresponding values of a and b are 0.4038 and 3.0678.

Log-concavity of the full conditional distributions of γ ***and*** $\ln\gamma$

A density is defined to be log-concave if the derivative of its logarithm is monotone decreasing. For the density of γ in [11], the required derivative is

$$\frac{1}{2}\frac{2a-2-q}{\gamma} - \frac{b-1}{g-\gamma} + \frac{1}{2}\frac{\mathbf{u}'\mathbf{G}^{-1}\mathbf{u}}{\gamma^2\sigma_e^2} \quad (0<\gamma<g)$$

The second and third terms are decreasing in γ if b exceeds 1, but the first is not unless a is greater than $\frac{1}{2}q+1$; log-concavity cannot be guaranteed for all σ_e^2 without this condition.

If we transform to $\delta = \ln\gamma$ (and include the Jacobian of the transformation), the logarithm of the density is, apart from an additive constant,

$$(a-\frac{1}{2}q)\,\delta + (b-1)\ln(g-e^{\delta}) - \frac{1}{2}\sigma_e^{-2}\,e^{-\delta}\,\mathbf{u}'\mathbf{G}^{-1}\mathbf{u} \quad (-\infty<\delta<\ln g)$$

and this has derivative

$$(a-\frac{1}{2}q) - \frac{b-1}{g-e^{\delta}} + \frac{1}{2}\sigma_e^{-2}\,e^{-\delta}\,\mathbf{u}'\mathbf{G}^{-1}\mathbf{u} \quad (-\infty<\delta<\ln g)$$

The first term is constant and the second and third are decreasing with respect to δ (so that the density is log-concave) as long as b is at least 1.