

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1995 - 7th Annual Conference Proceedings

BAYESIAN INFERENCE ON VARIANCE COMPONENTS IN GENERALIZED LINEAR MIXED MODELS: AN EVALUATION OF DIFFERENT METHODS

Robert J. Tempelman

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Tempelman, Robert J. (1995). "BAYESIAN INFERENCE ON VARIANCE COMPONENTS IN GENERALIZED LINEAR MIXED MODELS: AN EVALUATION OF DIFFERENT METHODS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1334>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

BAYESIAN INFERENCE ON VARIANCE COMPONENTS IN GENERALIZED LINEAR MIXED MODELS: AN EVALUATION OF DIFFERENT METHODS

Robert J. Tempelman
Department of Experimental Statistics
Louisiana State University Agricultural Center
Baton Rouge, LA

ABSTRACT

Generalized linear mixed models are now popular in the animal breeding and biostatistics literature as these models allow inference on fixed and random effects for the exponential family of data distributions. In animal breeding, particular attention is directed towards variances of the random effects. We investigate three methods for marginal inference on variance components in binary data, including (1) the conventional expectation-maximization (EM) type algorithm, (2) Laplace's method, and (3) "exact" Gibbs sampling methods. A simulation study involving probit animal models was used to compare the modal estimates computed under the three methods. It was found that EM estimates were badly biased downwards in comparison to Laplacian estimates. An application of all methods within a repeated measures probit analysis of mastitis incidence in dairy cows suggests that Laplacian and Gibbs sampling posterior marginal modes are somewhat congruent in moderately sized data sets, although the tail of the posterior density was lighter for the Laplacian approximation.

1 Introduction

Mixed effects and repeated measures models are increasingly important for statistical inference in agricultural research. This is in part due to continued use of traditional experimental designs, such as split plot designs, along with recent consideration of designs involving potentially complex error structures (e.g. autocorrelated residuals, spatial analysis, etc.). Animal breeders, in particular, have recognized mixed models as the base technology for inference on correlated polygenic random effects for economically important traits in livestock (Henderson, 1984).

Although commercial software for linear mixed models is readily available, hypothesis testing on fixed and random effects require Gaussian distributional assumptions on the model residuals. This somewhat negates their use on categorical or discrete fertility and fitness traits, often observed in animal and plant research. Mixed effects parameterizations for the exponential family of sampling distributions have been addressed in the development of generalized linear mixed models or GLMM (Breslow and Clayton, 1993; Stroup and Kachman, 1994). GLMM defines a "link" function of the expected value of each observation to be a linear combination of the location parameters, specifically the fixed and random effects.

A GLMM allowing statistical inference on location parameters in ordinal categorical data for animal breeding applications was proposed by Gianola and Foulley (1983) and Harville and

Mee (1984). This GLMM was called the threshold mixed model in deference to Wright (1934), who postulated an underlying normal distribution for liabilities rendered discrete on an observed scale via fixed thresholds. This postulate can be shown to be equivalent to presuming a probit link for the expected probability of a particular response; hence the threshold model is often called a probit mixed model. Recent threshold mixed model applications in animal breeding are further discussed in Foulley *et al.* (1990).

A contentious issue in the implementation of GLMM for ordinal categorical data has been the estimation of variances of the random effects. Maximum likelihood estimation (MLE) of variance components requires integration of the random effects in order to derive the marginal likelihood of the data. Exact approaches based on numerical quadrature (Anderson and Aitken, 1985) are computationally unfeasible for most data situations in animal breeding. Approximate marginal inference on variance components has often been considered, in part due to the analogy drawn between Bayesian marginal inference and restricted maximum likelihood estimation (REML) in linear mixed models (Harville, 1974). Some animal breeders prefer the term maximum marginal likelihood (MML) to REML in order to distinguish between non-Gaussian and Gaussian data situations for marginal inference on variance components (Foulley *et al.*, 1990).

A common MML algorithm for variance component estimation in threshold mixed models is the "expectation-maximization" (EM) type algorithm proposed by Stiratelli *et al.* (1984). A multivariate normal assumption on the conditional distribution of the random effects is used to approximate the expectation step. This approximation has been blamed for poor frequency properties of the EM-type algorithm for variance component estimation in threshold sire models (Hoeschele *et al.*, 1987). Furthermore, the EM-algorithm is computationally onerous for large animal breeding data sets, thereby leading Hoeschele *et al.* (1987) to anticipate optimization methods analogous to derivative-free REML (Graser *et al.*, 1987) for variance component estimation in threshold mixed models. Tempelman and Gianola (1993) derived such an algorithm for a Poisson GLMM by applying Laplace's method to the estimation of variances of random effects. In an erroneous proof, they claimed that Laplace's method and the EM-type algorithm should lead to identical MML estimates of variance components for all GLMM. However, in a simulation study involving probit mixed models with uncorrelated random effects, Tempelman (1994) found that Laplace's approximation led to seemingly unbiased estimates of variance components, in direct contrast to badly downward biased estimates computed under the EM-type approximation. No research on assessing bias properties in probit mixed models with *a priori* correlated random effects has been pursued, using either the EM or Laplace's method. Inference on correlated random effects has become increasingly popularized due to the widespread use of the "animal" model in animal breeding (Henderson, 1984)

Gibbs sampling is currently touted as a computationally feasible method to obtain both likelihood (McCulloch, 1994) and Bayesian marginal inference (Albert and Chib, 1993) on variances of random effects in probit mixed models. This method which generates correlated samples of the parameters from the joint posterior density is simulation intensive. One requirement for Gibbs sampling is the ability to obtain samples from each parameter's full conditional distribution; i.e. the conditional distribution of each parameter given every other parameter in the model and the data. Inference based on Gibbs sampling is naturally subject to

Monte Carlo sampling error; however, this error can be controlled by increasing the length of the sampling chain. Unfortunately, computing costs using Gibbs sampling can be high relative to either the EM-type or Laplace's methods, particularly if the sampler mixes too slowly around the parameter space (Smith and Roberts, 1993). Slow mixing would be anticipated if parameters are highly correlated *a priori*, such as genetic effects in animal models (Wang *et al.*, 1994). Therefore, an assessment of those situations in which approximate methods may lead to reliable inference is desirable. Furthermore, approximate methods may provide good starting values that are crucial for the convergence of the Gibbs sampler.

The first objective of this paper was to further extend the work of Tempelman (1994) and compare Laplace and EM-type methods for variance component estimation in probit animal models. Secondly, the degree of congruency between inference due to Laplacian and Gibbs sampling methods in probit mixed models is assessed. Finally, an application of all methods to a dataset involving binary records on mastitis incidence in dairy cows is presented.

2 Theory

In the threshold or probit mixed model, a linear combination of identified risk factors μ_j (i.e. levels of fixed and random effects) plus random Gaussian noise $e_j \sim N(0, \sigma_e^2)$ determine an underlying continuous random liability l_j for each animal j , such that

$$l_j = \mu_j + e_j; \quad j \in \{1, 2, \dots, n\}$$

The observed ordinal response $Y_j = k$ is defined by the ordinal category bordered by two adjacent threshold points t_{k-1} and t_k in which l_j is located. The response Y_j can take on any one of K possible outcomes delimited by the $K+1$ threshold points in $\mathbf{t} = [t_0 \ t_1 \ \dots \ t_{K-1} \ t_K]$ where $t_0 < t_1 < \dots < t_{K-1} < t_K$ with $t_0 = -\infty$ and $t_K = \infty$. To allow full rank estimability of the parameters, one additional level in \mathbf{t} is constrained to 0. The probability that a response on individual j falls into any category k can be written:

$$P_{jk} = \Pr(Y_j = k | \mu_j, \mathbf{t}) = \Pr(t_{k-1} < l_j < t_k | \mu_j, \mathbf{t}) = \Phi\left(\frac{t_k - \mu_j}{\sigma_e}\right) - \Phi\left(\frac{t_{k-1} - \mu_j}{\sigma_e}\right) \quad [2.1]$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. Here $\mu_j = \mathbf{w}_j' \boldsymbol{\eta} = \mathbf{x}_j' \boldsymbol{\beta} + \mathbf{z}_j' \mathbf{u}$ where $\boldsymbol{\beta}$ denotes the $p \times 1$ vector of fixed effects and \mathbf{u} denotes the $q \times 1$ vector of random effects and \mathbf{x}_j and \mathbf{z}_j are known incidence row vectors. For economy of notation, we let $\boldsymbol{\eta} = [\boldsymbol{\beta}' \ \mathbf{u}']'$ denote the vector of location parameters and $\mathbf{w}_j = [\mathbf{x}_j' \ \mathbf{z}_j']$. In Bayesian linear models, fixed effects are typically defined as those location parameters with flat uniform priors (i.e., $p(\beta_i) \propto \text{constant}; -\infty < \beta_i < \infty; i=1, 2, \dots, p$) whereas random effects are characterized by informative priors. Note that σ_e^2 is not identifiable in [2.1]; therefore we let $\sigma_e^2 = 1$ without any loss in generality.

The likelihood of the data is the product of one-trial multinomial distributions such that, except for a constant, the log-likelihood can be written:

$$\ln p(\mathbf{y} | \boldsymbol{\theta}) = \sum_j \sum_k 1(y_j = k) \ln(P_{jk}) \quad [2.2]$$

where $1(\cdot)$ denotes the indicator function and $\theta = [\mathbf{t}' \quad \eta']'$.

Let $p(\mathbf{u}|\mathbf{G})$ denote the prior density of the random effects whereby $\mathbf{u}|\mathbf{G} \sim N(\mathbf{0}, \mathbf{G})$. For additive genetic effects in an animal model, $\mathbf{G} = \mathbf{A}\sigma_u^2$, where \mathbf{A} is a known matrix of additive genetic relationships among animals (Henderson, 1976). The existence of non-zero prior correlations contrasts with specifications in many other hierarchical models such as split plot designs, conditionally independent hierarchical models (Albert, 1988), or simple sire models (Tempelman, 1994); in these cases, $\mathbf{G} = \mathbf{I}\sigma_u^2$ with \mathbf{I} as the identity matrix. We presume joint flat priors for \mathbf{t} , β and σ_u^2 ; i.e.

$$p(\beta, \mathbf{t}, \sigma_u^2) = p(\beta)p(\mathbf{t})p(\sigma_u^2) \propto c$$

where c denotes a constant. Letting $\vartheta = [\theta' \quad \sigma_u^2]'$ denote the vector of all parameters and $L(\vartheta|\mathbf{y})$ denote the log joint posterior density of all parameters, we find that

$$\begin{aligned} L(\vartheta|\mathbf{y}) &= \ln(p(\vartheta|\mathbf{y})) = \ln p(\mathbf{y}|\theta) + \ln p(\mathbf{u}|\mathbf{G}) \\ &= \sum_j \sum_k 1(y_j = k) \ln(P_{jk}) - \frac{\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}}{2\sigma_u^2} - \frac{q}{2} \ln(\sigma_u^2) + const. \end{aligned} \quad [2.3]$$

where *const.* denotes an arbitrary constant. Note that the marginal density of the variance component can be expressed as:

$$p(\sigma_u^2|\mathbf{y}) = \frac{p(\vartheta|\mathbf{y})}{p(\theta|\sigma_u^2, \mathbf{y})} \quad [2.4]$$

where $p(\vartheta|\mathbf{y})$ is the joint posterior density of the full parameter set and $p(\theta|\sigma_u^2, \mathbf{y})$ is the density of the location parameters conditional on the variances of the random effects. Define:

$$\tilde{\theta}_\sigma = \begin{bmatrix} \tilde{\mathbf{t}}_\sigma \\ \tilde{\beta}_\sigma \\ \tilde{\mathbf{u}}_\sigma \end{bmatrix} = \text{Arg}_\theta \max p(\theta|\sigma_u^2, \mathbf{y}) \quad [2.5]$$

to be the joint mode of $p(\theta|\sigma_u^2, \mathbf{y})$ and

$$\tilde{\mathbf{H}}_\sigma = \left[\frac{-\partial^2 \ln p(\vartheta|\mathbf{y})}{\partial\theta\partial\theta'} \right]_{\theta=\tilde{\theta}_\sigma} = \left[\frac{-\partial^2 \ln p(\theta|\sigma_u^2, \mathbf{y})}{\partial\theta\partial\theta'} \right]_{\theta=\tilde{\theta}_\sigma} \quad [2.6]$$

to be the negative Hessian or information matrix of the full joint posterior density with respect to θ and evaluated at $\theta = \tilde{\theta}_\sigma$.

Laplace's method involves approximating both the numerator and the denominator in [2.4] by a second order Taylor series expansion about $\tilde{\theta}_\sigma$. It can be seen (Tempelman and Gianola, 1993) that the Laplacian approximation of the marginal density of the variance component is:

$$\hat{p}(\sigma_u^2|\mathbf{y}) \propto p(\tilde{\theta}_\sigma, \sigma_u^2|\mathbf{y}) |\tilde{\mathbf{H}}_\sigma|^{-1/2} \quad [2.7]$$

In other words, using [2.6], the log marginal density of the variance components in threshold mixed models can be written, except for a constant, by

$$L(\sigma_u^2 | \mathbf{y}) = \ln \hat{p}(\sigma_u^2 | \mathbf{y}) = \sum_j \sum_k 1(y_j = k) \ln(\tilde{P}_{jk})_\sigma - \frac{\tilde{\mathbf{u}}_\sigma' \mathbf{A}^{-1} \tilde{\mathbf{u}}_\sigma}{2\sigma_u^2} - \frac{q}{2} \ln(\sigma_u^2) \frac{1}{2} \ln |\tilde{\mathbf{H}}_\sigma| \quad [2.8]$$

where $(\tilde{P}_{jk})_\sigma = \Phi(\tilde{t}_k - \mathbf{x}'_j \tilde{\boldsymbol{\beta}}_\sigma - \mathbf{z}'_j \tilde{\mathbf{u}}_\sigma) - \Phi(\tilde{t}_{k-1} - \mathbf{x}'_j \tilde{\boldsymbol{\beta}}_\sigma - \mathbf{z}'_j \tilde{\mathbf{u}}_\sigma)$. Any optimization routine (e.g. see Meyer, 1989) could be used to compute the mode of $L(\sigma_u^2 | \mathbf{y})$.

A set of iterative Newton Raphson equations is used to compute $\tilde{\boldsymbol{\theta}}_\sigma$, i.e.

$$[\mathbf{H}_\sigma]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[t]}} [\boldsymbol{\theta}^{[t+1]} - \boldsymbol{\theta}^{[t]}] = \left[\frac{\partial \log p(\boldsymbol{\theta} | \sigma_u^2, \mathbf{y})}{\partial \boldsymbol{\theta}} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{[t]}} \quad [2.9]$$

where t denotes the Newton Raphson iteration conditional on σ_u^2 . The set of equations in [2.9] can be seen to resemble Henderson's mixed model equations with \mathbf{H}_σ as the mixed model coefficient matrix. For ease of computation, Gianola and Foulley (1983) advocated using the expected information or $E(\mathbf{H}_\sigma)$, rather than the observed information \mathbf{H}_σ , as the coefficient matrix for Fisher scoring computation of $\tilde{\boldsymbol{\theta}}_\sigma$. While $\tilde{\boldsymbol{\theta}}_\sigma$ should be generally identical whether the $E(\mathbf{H}_\sigma)$ or \mathbf{H}_σ is used, these two matrices are nevertheless different for probit link GLMM and may lead to slightly different inferences on σ_u^2 based on [2.7]. Asymptotic inference based on $E(\mathbf{H}_\sigma)$ should be inferior to that based on \mathbf{H}_σ , in agreement with conventional wisdom (Efron and Hinckley, 1978).

The expectation step of the EM-type algorithm for computing MML of σ_u^2 is based on approximating the conditional density of the random effects by a multivariate Gaussian distribution:

$$\mathbf{u} | \sigma_u^2, \mathbf{y} \sim N(\tilde{\mathbf{u}}_\sigma, \mathbf{C}_{\mathbf{uu}})$$

(Foulley *et al.*, 1990) where $\mathbf{C}_{\mathbf{uu}}$ is the block diagonal component of $[\tilde{\mathbf{H}}_\sigma]^{-1}$ corresponding to the random effects; if Fisher scoring is used then $\mathbf{C}_{\mathbf{uu}}$ is the respective block diagonal component of $[E(\tilde{\mathbf{H}}_\sigma)]^{-1}$. The EM algorithm can be rewritten in terms of the following iterative algorithm for σ_u^2 :

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

where m denotes iteration number. Iteration continues until the difference between $\sigma_u^{2[m+1]}$ and $\sigma_u^{2[m]}$ is arbitrarily small.

Gibbs sampling has been seen to be particularly amenable for use in probit mixed models (Albert and Chib, 1993). This is because the full conditional distributions of all parameters are easily characterized by augmenting the parameter set by the latent variables or liabilities $\mathbf{l} = [l_1 \quad l_2 \quad \dots \quad l_n]'$ described previously. Details on generating Gibbs samples for all

parameters in probit mixed models for animal breeding applications are provided by Sorensen *et al* (1995).

The extension of all methods described above to handle multiple sources of random effects, i.e. more than one variance component, is relatively straightforward.

3 Simulation study

3.1 Laplace versus EM estimation of variance components

An animal model was used to generate binary data; e.g. incidence of clinical mastitis in dairy cows. Data for 1280 dams and their female offspring were generated from Bernoulli distributions with parameters $p_{ijk} = \Phi(\eta_{ijk})$ where

$$\eta_{ijk} = \mu + f_i + h_j + a_k$$

Here μ is the overall mean, $f_i = -1 + 0.5(i-1)$ for $i=1,2,\dots,5$ denotes a fixed effect with five levels, $\mathbf{h} = \{h_j\} \sim N(\mathbf{0}, \mathbf{I}\sigma_h^2)$ is a 64×1 vector of herd effects, and $\mathbf{a} = \{a_k\} \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2)$ is a 1344×1 vector of random additive genetic effects. The Bernoulli parameter p_{ijk} denotes the expected probability of non-infection for the individual ijk with location parameter η_{ijk} . The base population consisted of 64 unrelated sires without records and 256 dams with records. Dams were nested within herd, each herd containing 4 dams. Sires were randomly mated to dams to generate 1024 progeny, where each progeny was located in the same herd as her dam. That is, the only across herd genetic connections were that due to sire; this roughly characterizes the current mating situation in North American dairy cattle populations. Levels of the fixed effect were randomly assigned to both dams and progeny in generating the records. Three populations pertaining to different values of σ_a^2 were considered: Population I. $\sigma_a^2=0.20$; Population II. $\sigma_a^2=0.60$; and Population III. $\sigma_a^2=1.00$. For all populations, $\sigma_h^2=0.40$. Each of these populations were further considered at four different environments or values of the overall mean μ : $\mu = \Phi^{-1}(0.50)$, $\mu = \Phi^{-1}(0.65)$, $\mu = \Phi^{-1}(0.80)$ and $\mu = \Phi^{-1}(0.95)$. These different values of μ represent different levels of overall liability or expected rates of infection (i.e. corresponding to ~50%, ~35%, ~20% and ~5% rates of infection, respectively, ignoring the influence of random effects on the overall incidence). MML estimates of the variance components were computed using three different approximate methods: (1) the EM-method based on the inverse expected information matrix $\left[\mathbf{E}(\tilde{\mathbf{H}}_\sigma) \right]^{-1}$, (2) the EM-method based on the inverse observed information matrix $\left[\tilde{\mathbf{H}}_\sigma \right]^{-1}$ and (3) Laplace's method. Fifty replicates of each population for each of the four different values for μ were used to assess the empirical bias and empirical relative error (root mean square error) of the estimates.

3.2 Laplace versus Gibbs sampling inference on variance components

In linear animal models, it has been found that Gibbs sampling is particularly computationally intensive due to the high correlations between successive draws of Gibbs samples (Wang *et al.*, 1994). High correlations necessitates a very large number of draws in order to generate an effectively large number of independent samples. This problem is

particularly intensified with Gibbs sampling in probit mixed models (Sorensen *et al.*, 1995). Therefore to limit computing costs, only two datasets based on the simulation procedure described in the previous section were investigated. The overall mean considered in both datasets was $\mu = \Phi^{-1}(0.80)$. The first dataset (Dataset I) was generated with a genetic variance $\sigma_a^2 = 0.20$ while the second dataset (Dataset II) was generated with $\sigma_a^2 = 1.00$. For both datasets, $\sigma_h^2 = 0.40$ and was presumed known in order to allow a direct comparison of marginal inference on σ_a^2 . A proper prior yet highly dispersed inverted gamma density on σ_a^2 was used for both data sets. In the Dataset I, the prior mean and standard deviation on σ_a^2 was 0.075 and 2.23, respectively. In Dataset II, the prior mean and standard deviation on σ_a^2 was 0.75 and 23.69, respectively. Proper prior densities on variance components in a non-Gaussian GLMM may be required to allow a proper full joint posterior density. A total of 5,000 samples from each conditional distribution in a Gibbs cycle was generated before samples were collected. This ensures the necessary "burn-in" required for convergence to the equilibrium joint distribution in the Gibbs sampler (Smith and Roberts, 1993). For Dataset I, 200,000 samples were generated with every second sample saved for a total of 100,000 Gibbs samples. This was necessary to generate an effectively large enough number of samples due to the stickiness of the chain, particularly when variance components are small. For Dataset II, 100,000 samples were generated with every second sample saved for a total of 50,000 Gibbs samples. Both sampling strategies were used to economically compromise computing time with hard disk space. Posterior densities were generated using the "Rao-Blackwellization" method (Smith and Roberts, 1993).

4 Results

4.1 Laplace versus EM estimation of variance components

In all cases, the EM-method based on expected information led to variance component estimates with slightly more empirical relative bias and relative error than the EM-method based on observed information. As both EM methods performed poorly compared to Laplace's method, results for the EM-method based on expected information are not presented. Furthermore, as the relative performance of all methods did not vary greatly over all three populations, only graphical results for Population II were presented.

Empirical relative biases and empirical relative errors of the two approximate methods for variance component estimates are given in Figures 1 and 2, respectively, for Population II. In general, the relative biases for Laplace estimates of σ_h^2 were negative but small, except for Population III where the relative biases were as large as (-20%). In contrast, estimates of σ_h^2 under the EM method showed large negative bias ((-15%)-(-45%)) that generally increased with σ_a^2 (i.e. from Population I to Population III). There was no trend in relative biases for σ_h^2 estimates over increasing $\Phi(\mu)$ (i.e. decreasing incidence rates) for either EM or Laplace's method within each population, except perhaps for an unexpected decrease in bias for Laplacian estimates at the most extreme incidence level ($\Phi(\mu)=.95$) within the first two populations. The differences between relative errors of σ_h^2 estimates for the two methods remained somewhat constant over increasing $\Phi(\mu)$ within each population; however, they widened as the true values of σ_a^2 increased.

Estimates of genetic variance were always characterized by larger relative bias and relative error than estimates of herd variance. Laplace estimates of σ_a^2 were biased by (-20%)-(-50%) with absolute bias increasing with σ_a^2 . Relative biases on EM based estimates of σ_a^2 were substantially larger than Laplacian estimates, ranging from (-50%)-(-80%), and also increased in absolute measure with σ_a^2 . Relative biases were somewhat proportional to $\Phi(\mu)$ or increasing extremity in incidence rates. Larger relative errors for estimates of σ_a^2 were also associated with larger $\Phi(\mu)$ for both methods although relative errors showed no trend with σ_a^2 . Nevertheless, Laplace's method generally yielded estimates with 20-30% less relative error than EM.

4.2 Laplacian versus Gibbs sampling inference on variance components

Figure 3 shows the posterior densities of the genetic variance component for Dataset I ($\sigma_a^2 = 0.20$) under both Laplacian and Gibbs sampling inference. In addition, the prior density used in deriving both posterior inferences is shown for comparison. The modes (~ 0.03) of both posterior densities are virtually identical and may have been strongly influenced by the prior density on σ_a^2 , which has a mode at 0.025. However, the right tail of the posterior density under Gibbs sampling was much heavier than that derived under Laplacian integration. This heavier tail resulted in a large posterior mean (relative to the posterior mode) of 0.119 and a posterior standard deviation of 0.102 under Gibbs sampling. A flat uniform prior on σ_a^2 was used to further assess the impact of the informative and proper inverted gamma prior on the posterior density computed under Laplace's method. This density is also shown in Figure 3. This density is much more dispersed than the other posterior densities and shows a mode near 0.18. Since a seemingly highly dispersed prior influences the posterior density relative to a flat prior specification, this data is perhaps not too informative for sharp inference on σ_a^2 .

Figure 4 displays posterior marginal inferences for Data Set II ($\sigma_a^2 = 1.00$). The Laplacian modal estimates under both informative and flat priors are highly biased downwards; this is consistent with the increasing relative bias of modal estimates found with larger values of σ_a^2 in the simulation study results of Section 4.1. It further appears that there is a relatively minor effect of the prior on posterior inference. This is again shown in the two Laplacian posterior densities where the difference in the two modes was relatively small. The posterior density generated under Gibbs sampling had a mode near 0.90, a mean of $0.973 \pm$ a posterior standard deviation of 0.363.

5 An Application

Binary data on mastitis teat infection by *Streptococcus agalactiae* in dairy cows from the Hill Farm Research Station near Homer, LA., a unit of the Louisiana Agricultural Experiment Station, Louisiana Agricultural Center, was collected to determine the efficacy of a teat germicide in preventing clinical mastitis. Two diagonally opposite teats on each of 130 cows were dipped full length with the germicidal teat dip with the remaining two teats serving as undipped controls. The generalized linear mixed model corresponding to the expected incidence rate p_{ijk} of mastitis for the ijk^{th} experimental unit was:

$$\Phi^{-1}(p_{ijk}) = \mu_{ijk}$$

where

$$\mu_{ijk} = \mu + \tau_i + l_j + \tau l_{ij} + c_k$$

Here,

μ denotes the overall mean on the underlying continuous liability scale,

τ_i denotes the fixed effect of i th treatment (dipped versus not dipped), $i \in \{1,2\}$,

l_j denotes the fixed effect of j th teat location on udder (front versus back), $j \in \{1,2\}$,

τl_{ij} denotes the interaction of the i th treatment with the j th location effect, and

c_k denotes the random effect of cow k , $k \in \{1,2,\dots,130\}$.

As pedigree information was not available, it was presumed that the cow effects were identically and independently normally distributed with mean 0 and variance σ_c^2 . Since observations for all teats were taken, there were four repeated measurements for each cow. The posterior marginal density of σ_c^2 was computed using Laplace's method and Gibbs sampling. Two inverted gamma priors on σ_c^2 were considered. The prior mean and standard deviation under the first prior (Prior I) was 0.25 and 249, respectively. The second prior (Prior II) was based on a mean and standard deviation of 0.75 and 749. Any substantial difference in the shape of the posterior densities using these two different priors is suggestive of little information provided by the data likelihood. An initial 5000 Gibbs samples were drawn and discarded. Then 100,000 Gibbs samples were further generated with every 5th sample saved for a total of 20,000 samples.

The posterior density of cow variance under Prior I is given in Figure 5. The modal estimates of the two posterior densities are quite similar (0.10-0.20) and appear to be substantially influenced by the prior mode. The Gibbs density, nevertheless, has a heavier right tail; leading to a posterior mean of $0.343 \pm$ a standard deviation of 0.277. Figure 6 shows the influence of Prior II on the posterior inference. Both posterior modes (0.20-0.30) increased appreciably compared to inference under Prior I, even though both priors were highly dispersed. Again the Gibbs sampling density was characterized by a heavier tail, reflecting in a large posterior mean of $0.529 \pm$ a standard deviation of 0.357. It appears that relatively little information on cow variance is provided by the data likelihood, as indicated by the influence of vaguely informative priors.

6 Discussion

Many algorithms have been proposed for approximate marginal inference on variance components in generalized linear mixed models. An exact method based on Gibbs sampling is attractive; however, it is computationally intensive and can be easily misused if sampling is not adequate and mixing is slow. Approximate methods have generally required a Gaussian assumption on the conditional distribution of the random effects. This assumption appears to be particularly tenuous for small samples sizes (Breslow and Clayton, 1993; Hoeschele *et al.*, 1987) and for large values of the dispersion parameter (Drum and McCullagh, 1993). This latter relationship was also seen in this study as differences in point estimates and posterior densities between all methods widened with increasing values of the variance components. However,

marginal densities also tended toward normality with larger true values of the variance components as the data likelihood then seemingly become more informative.

Wolfinger (1993) recently claimed that Laplacian and EM-based methods lead to identical estimates of the joint marginal modes of variance components in generalized linear mixed models under quasi-likelihood settings. It is clear, however, from the results of this study and that in Tempelman (1994) that this is not true in fully parameterized GLMM's. Laplace's method yielded point estimates with smaller relative bias and mean square error than the EM-based method. Let n denote the sample size. In EM, one Gaussian approximation is invoked resulting in an error of approximation term that is proportional to $O(n^{-1/2})$. Conversely, two approximations are invoked in the Laplacian procedure, one for each of the numerator and the denominator of [2.4]. This allows cancellation of some terms thereby resulting in an error of approximation that is proportional to $O(n^{-1})$ (Tierney and Kadane, 1986).

In small samples, posterior marginal densities of the variance components are not symmetric but generally skewed positively (see also Wang *et al.*, 1994). In other words, the posterior mean will generally be greater than the posterior mode. The question then arises as to which measure one should use for a point estimate of variance. In Bayesian inference, this decision depends on the appropriate loss functions which cannot be easily determined. Therefore, the visualization of posterior marginal densities and the determination of quantiles afforded by Gibbs sampling provide a particularly invaluable supplement to point estimation. In this study, "exact" marginal modes due to Gibbs sampling and Laplacian marginal modes were similar. However, as noted by Leonard *et al.* (1994) and by the results in this study, Laplace's method may not adequately describe the tail behavior in small data sets.

Binary data was chosen in this study in order to consider the potentially most nonlinear of all cases in GLMM. Sorensen *et al.* (1995) recently compared the EM algorithm to Gibbs sampling and found little difference between these two procedures for point estimates on variance components in a threshold mixed model analysis of hip dysplasia in dogs. However, they analyzed a data set consisting of seven possible ordinal categories of response and a sire model. Generally, asymptotic inference on variances of uncorrelated random effects could be very reliable in threshold mixed models when a large number of ordinal categories exist. Conversely, all methods may lead to highly biased modal estimates or highly skewed posterior densities as a result of increased nonlinearity in binary data.

As in Tempelman and Gianola (1993) and Tempelman (1994), the programming strategy for Laplacian inference on variance components in GLMM is virtually identical to derivative-free REML procedures introduced by Graser *et al.* (1987) and programmed in *DFREML* for linear mixed models by Meyer (1989). In addition to the computing strategies presented by Meyer (1989), one must consider the following for GLMM inference on variance components: (1) the choice of a link function (2) the sampling density of the data and (3) iterative solution of the fixed and random effects. For moderate number of variance components (<5), the computational advantages of Laplacian over EM inference could be substantial with large $\dim(\theta)$.

The estimation of fixed and random effects has not been addressed in this paper. Generally, mixed model solutions for the fixed and random effects are computed conditionally on variance component estimates; from a Bayesian context, this is also known as empirical Bayes estimation (Foulley *et al.*, 1990). The question arises as to the implications of

approximate methods of variance component estimation on empirical Bayes estimation versus fully Bayesian inference via Gibbs sampling on fixed and random effects in a threshold mixed model. This will be addressed in a future study.

7 Summary

It seems useful to have approximate methods for marginal inference on variance components in GLMM which are not as computationally intensive as Gibbs sampling. Laplacian integration leads to modal estimates that are very similar to modal estimates computed under Gibbs sampling but which are nevertheless biased downwards when information is sparse as in animal models. Furthermore, Laplacian posterior densities are somewhat underdispersed relative to the "true" posterior density. The EM approximation used by most animal breeders should not be considered for point estimation of variance components in threshold animal models, particularly in highly nonlinear situations (i.e. binary data with extreme incidence rates, large variance components, and animal model parameterizations).

8 Acknowledgements

Dr. Steve Nickerson is gratefully acknowledged for providing the data.

9 List of References

- Albert, J.H. (1988) Computational methods using a Bayesian hierarchical generalized linear model. *Journal of the American Statistical Association* **83**, 1037-1044.
- Albert, J.H., and Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669-679.
- Anderson, D.A. and Aitken, M. (1985) Variance components models with binary response: interviewer variability. *Journal of the Royal Statistical Society Ser. B* **47**, 203-210.
- Breslow, N.E. and Clayton, D.G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9-25.
- Drum, M. and McCullagh, P. (1993) REML estimation with exact covariance in the logistic mixed model. *Biometrics* **49**, 677-689.
- Efron, B.F. and Hinckley, D.V. (1978) Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika* **65**, 457-487.
- Foulley, J.L., Gianola, D., and Im S. (1990) Genetic evaluation for discrete polygenic traits in animal breeding. In: *Advances in statistical methods for genetic improvement of livestock* (Gianola D., Hammond, K., eds.) Springer Verlag, Heidelberg, pp.361-409.
- Gianola D., and J.L. Foulley (1983) Sire evaluation for ordered categorical data with threshold model. *Genetics, Selection, Evolution* **15**, 201-224.
- Graser H.U., Smith, S.P. and Tier, B. (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *Journal of Animal Science* **64**, 1362-1370
- Harville, D.A. (1974) Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383-385.

- Harville, D.A. and R.W. Mee (1984) A mixed model procedure for analyzing ordered categorical data. *Biometrics* **40**, 393-398.
- Henderson, C.R. (1976) A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**, 69-83
- Henderson, C.R. (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph Press, Guelph, CANADA
- Hoeschele I, Gianola D., and Foulley, J.L. (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *Journal of Animal Breeding and Genetics* **104**, 334-339.
- Leonard, T., Hsu, J.S.J., and Ritter, C. (1994) The Laplacian t -approximation in Bayesian inference. *Statistica Sinica* **4**, 127-142.
- McCulloch, C.E. (1994) Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association* **89**, 330-335.
- Meyer, K. (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genetics, Selection, Evolution* **21**, 317-330.
- Sorensen, D.A., Andersen, S., Gianola, D. and Korsgaard, I. (1995) Bayesian inference in threshold mixed models using Gibbs sampling. *Genetics, Selection, Evolution* **27**, 229-249
- Smith, A.F.M. and Roberts, G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov Carlo methods. *Journal of the Royal Statistical Society Ser. B* **55**, 3-23
- Stiratelli, R., Laird, N.M. and Ware, J.H. (1984) Random-effects models for serial observations with binary response. *Biometrics* **40**, 961-971.
- Stroup, W.W. and Kachman, S. (1994) Generalized linear mixed models: an overview, *Proceedings of the Sixth Annual Kansas State University Conference on Applied Statistics in Agriculture*.
- Tempelman, R.J. (1994) Evaluation of derivative-free marginal maximum likelihood estimation of variance components in threshold mixed models. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production Vol* **22**, 15-18.
- Tempelman, R.J. and Gianola D. (1993) Marginal maximum likelihood estimation of variance components in Poisson mixed models using Laplacian integration. *Genetics, Selection, Evolution* **25**, 305-319.
- Tierney, L., and Kadane, J.B. (1986) Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* **81**, 82-86.
- Wang, C.S., Rutledge, J.J., and Gianola, D. (1993) Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genetics, Selection, Evolution* **26**, 91-115.
- Wolfinger, R. (1993) Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791-795.
- Wright, S (1934) An analysis of variability in number of digits in an inbred strain of guinea pig. *Genetics* **19**, 506-536.

Figure 1: Empirical relative bias (based on 50 replicates) of herd and genetic variance component estimates using either Laplace's method or EM in an animal model simulation study: $\sigma_a^2=0.60$, $\sigma_h^2 = 0.40$

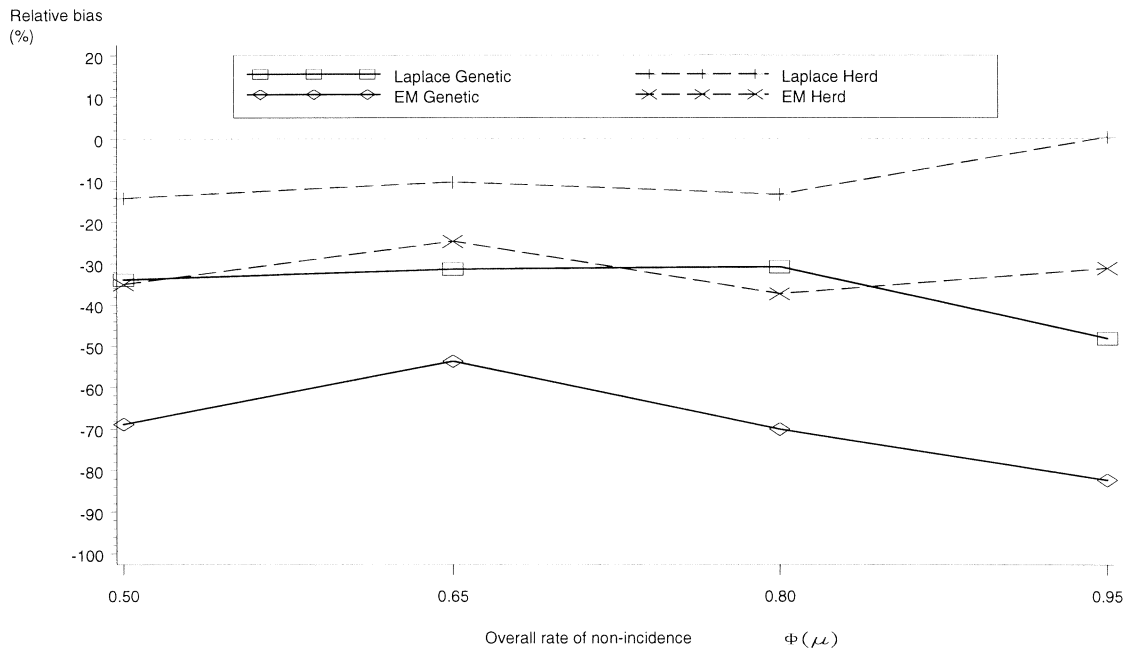


Figure 2: Empirical relative error (based on 50 replicates) of herd and genetic variance component estimates using either Laplace's method or EM in an animal model simulation study: $\sigma_a^2=0.60$, $\sigma_h^2 = 0.40$

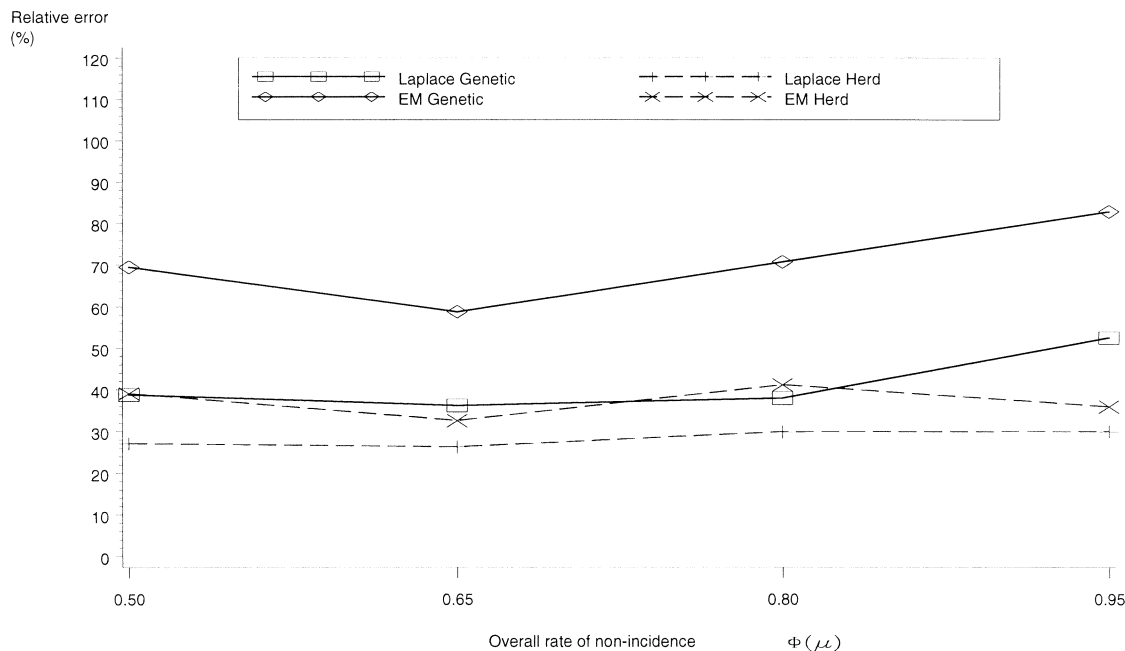


Figure 3: Prior and posterior marginal densities of σ_a^2 under Laplace's method and Gibbs sampling for Simulated DataSet I ($\sigma_h^2 = 0.40$ presumed known; $\sigma_a^2 = 0.20$)

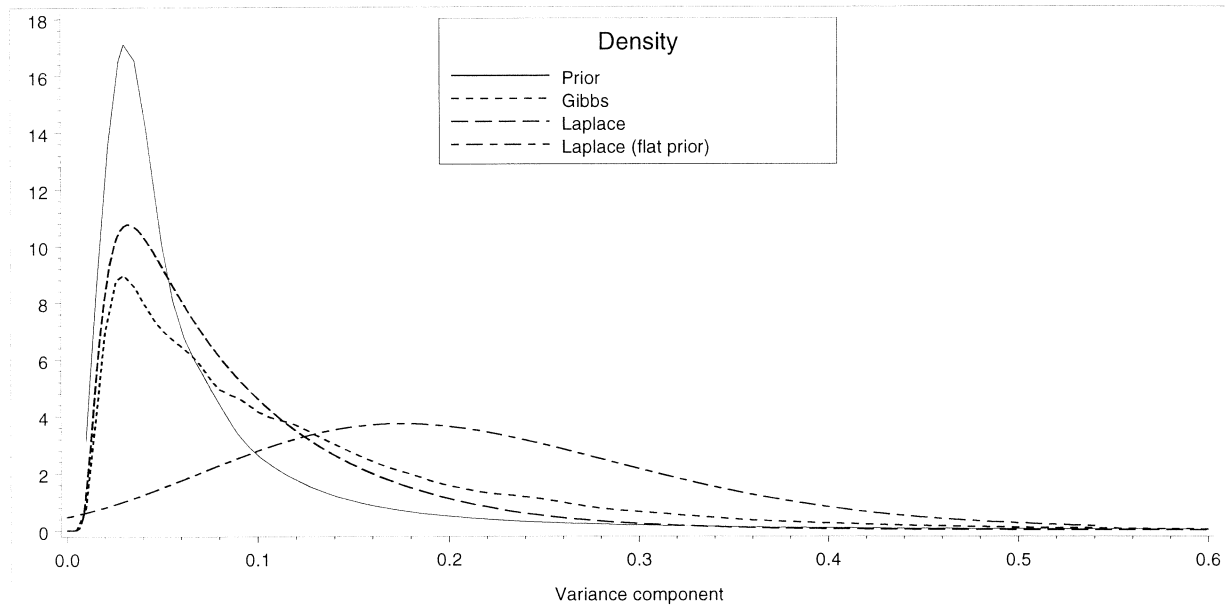


Figure 4: Prior and posterior marginal densities of σ_a^2 under Laplace's method and Gibbs sampling for Simulated Dataset II ($\sigma_h^2 = 0.40$ presumed known; $\sigma_a^2 = 1.00$)

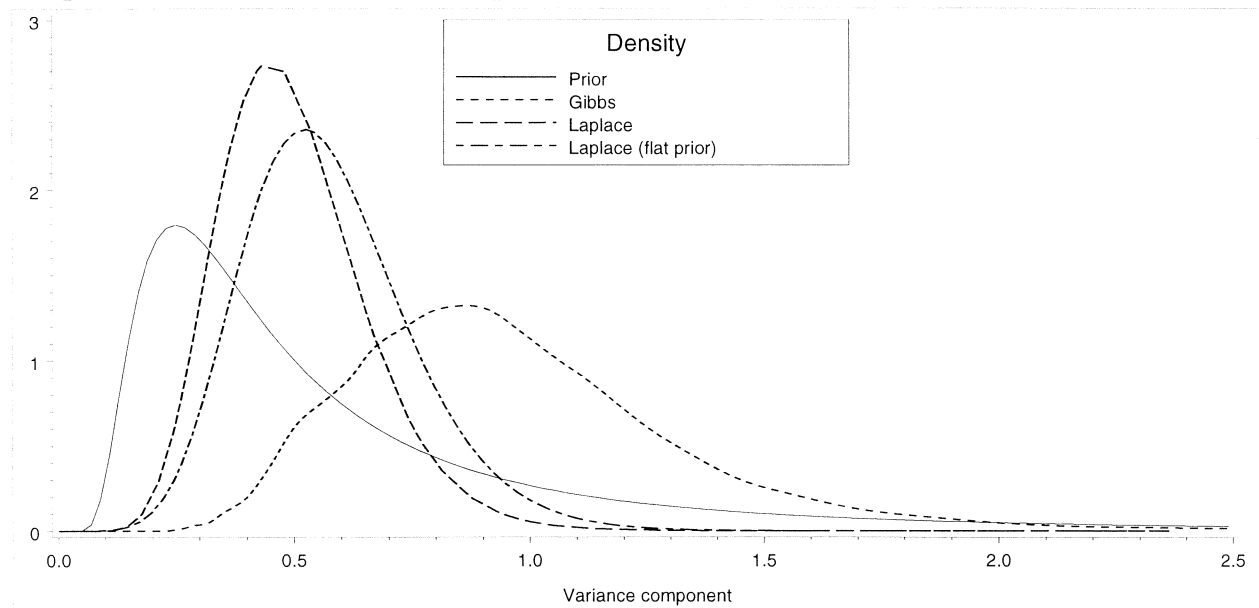


Figure 5: Posterior densities of cow variance using Gibbs sampling and Laplace's method for mastitis data under Prior I (prior mean = 0.25; prior standard deviation = 249)

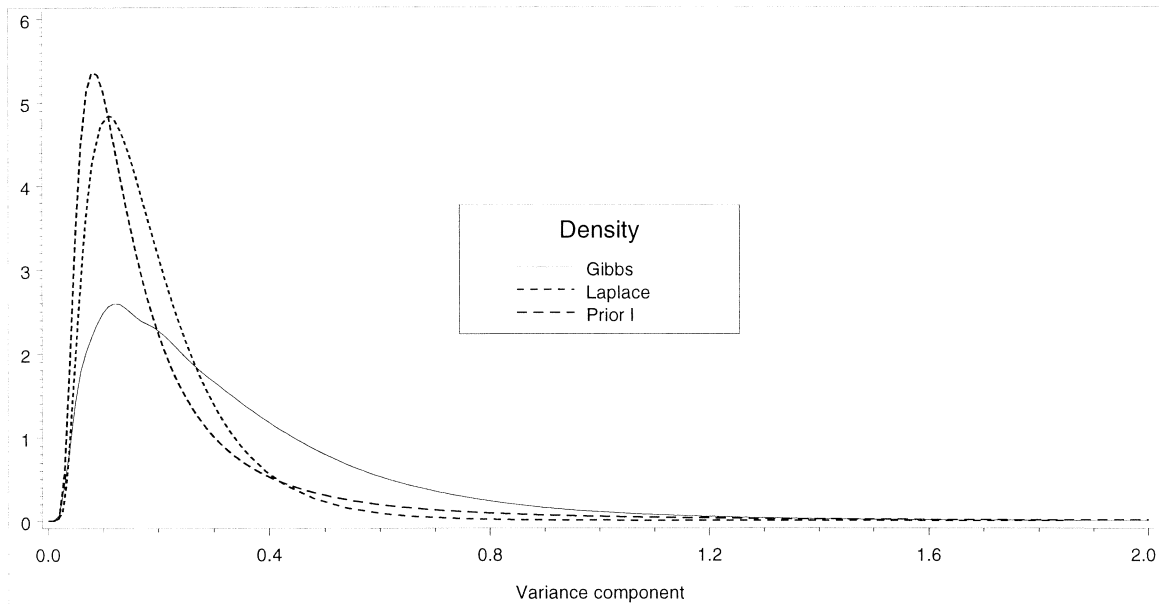


Figure 6: Posterior densities of cow variance using Gibbs sampling and Laplacian integration for mastitis data under Prior II (prior mean = 0.75; prior standard deviation = 749)

