Original article

Bayesian estimation in animal breeding using the Dirichlet process prior for correlated random effects

Abraham Johannes VAN DER MERWE*,
Albertus Lodewikus PRETORIUS

Department of Mathematical Statistics, Faculty of Science,
University of the Free State, PO Box 339, Bloemfontein,
9300 Republic of South Africa

Abstract – In the case of the mixed linear model the random effects are usually assumed to be normally distributed in both the Bayesian and classical frameworks. In this paper, the Dirichlet process prior was used to provide nonparametric Bayesian estimates for correlated random effects. This goal was achieved by providing a Gibbs sampler algorithm that allows these correlated random effects to have a nonparametric prior distribution. A sampling based method is illustrated. This method which is employed by transforming the genetic covariance matrix to an identity matrix so that the random effects are uncorrelated, is an extension of the theory and the results of previous researchers. Also by using Gibbs sampling and data augmentation a simulation procedure was derived for estimating the precision parameter M associated with the Dirichlet process prior. All needed conditional posterior distributions are given. To illustrate the application, data from the Elsenburg Dormer sheep stud were analysed. A total of 3325 weaning weight records from the progeny of 101 sires were used.

Bayesian methods / mixed linear model / Dirichlet process prior / correlated random effects / Gibbs sampler

1. INTRODUCTION

In animal breeding applications, it is usually assumed that the data follows a mixed linear model. Mixed linear models are naturally modelled within the Bayesian framework. The main advantage of a Bayesian approach is that it allows explicit use of prior information, thereby giving new insights in problems where classical statistics fail.

In the case of the mixed linear model the random effects are usually assumed to be normally distributed in both the Bayesian and classical frameworks.

---

* Correspondence and reprints
E-mail: fay@wwg3.uovs.ac.za

According to Bush and MacEachern [3] the parametric form of the distribution of random effects can be a severe constraint. A larger class of models would allow for an arbitrary distribution of the random effects and would result in the effective estimation of fixed and random effects across a wide variety of distributions.

In this paper, the Dirichlet process prior was used to provide nonparametric Bayesian estimates for correlated random effects. The nonparametric Bayesian approach for the random effects is to specify a prior distribution on the space of all possible distribution functions. This prior is applied to the general prior distribution for the random effects. For the mixed linear model, this means that the usual normal prior on the random effects is replaced with a nonparametric prior. The foundation of this methodology is discussed in Ferguson [9], where the Dirichlet process and its usefulness as a prior distribution are discussed. The practical applications of such models, using the Gibbs sampler, has been pioneered by Doss [5], MacEachern [16], Escobar [7], Bush and MacEachern [3], Lui [15] and Müller, Erkani and West [18]. Other important work in this area was done by West *et al.* [24], Escobar and West [8] and MacEachern and Müller [17]. Kleinman and Ibrahim [14] and Ibrahim and Kleinman [13] considered a Dirichlet process prior for uncorrelated random effects.

Escobar [6] showed that for the random effects model a prior based on a finite mixture of the Dirichlet processes leads to an estimator of the random effects that has excellent behaviour. He compared his estimator to standard estimators under two distinct priors. When the prior of the random effects is normal, his estimator performs nearly as well as the standard Bayes estimator that requires the estimate of the prior to be normal. When the prior is a two point distribution, his estimator performs nearly as well as a nonparametric maximum likelihood estimator.

A mixture of the Dirichlet process priors can be of great importance in animal breeding experiments especially in the case of undeclared preferential treatment of animals. According to Strandén and Gianola [19,20] it is well known that in cattlebreeding the more valuable cows receive preferential treatment and to such an extent that the treatment cannot be accommodated in the model, this leads to bias in the prediction of breeding values. A "robust" mixed effects linear model based on the *t*-distribution for the "preferential treatment problem" has been suggested by them. The *t*-distribution, however, does not cover departures from symmetry while the Dirichlet process prior can accommodate an arbitrarily large range of model anomalies (multiple modes, heavy tails, skew distributions and so on). Despite the attractive features of the Dirichlet process, it was only recently investigated. Computational difficulties have precluded the widespread use of Dirichlet process mixtures of models until recently, when a series of papers (notably Escobar [6] and Escobar and West [8]) showed how Markov Chain Monte Carlo methods (and more

specifically Gibbs sampling) could be used to obtain the necessary posterior and predictive distributions.

In the next section a sampling based method is illustrated for correlated random effects. This method which is employed by transforming the numerator relationship matrix A to an identity matrix so that the random effects are uncorrelated, is an extension of the theory and results of Kleinman and Ibrahim [14] and Ibrahim and Kleinman [13] who considered uncorrelated random effects. Also by using Gibbs sampling and data augmentation a simulation procedure is derived for estimating the precision parameter M associated with the Dirichlet process prior.

2. MATERIALS AND METHODS

To illustrate the application, data from the Elsenburg Dormer sheep stud were analysed. A total of 3325 weaning records from the progeny of 101 sires were used.

2.1. Theory

A mixed linear model for this data structure is thus given by

$$y = X\beta + \tilde{Z}\gamma + \varepsilon \tag{1}$$

where y is a $n \times 1$ data vector, X is a known incidence matrix of order $n \times p$, β is a $p \times 1$ vector of fixed effects and uniquely defined so that X has a full column rank p, γ is a $q \times 1$ vector of unobservable random effects, (the breeding values of the sires). The distribution of γ is usually considered to be normal with a mean vector $\mathbf{0}$ and variance–covariance matrix $\sigma_\gamma^2 A$. $\tilde{Z}$ is a known, fixed matrix of order $n \times q$ and ε is a $n \times 1$ unobservable vector of random residuals such that the distribution of ε is n-dimensional normal with a mean vector $\mathbf{0}$ and variance-covariance matrix $\sigma_\varepsilon^2 I_n$. Also the vectors ε and γ are statistically independent and σ_γ^2 and σ_ε^2 are unknown variance components. In the case of a sire model, the $q \times q$ matrix A is the relationship (genetic covariance) matrix. Since A is known, equation (1) can be rewritten as

$$y = X\beta + Zu + \varepsilon \ldots$$

where $Z = \tilde{Z}B^{-1}$, $u = B\gamma$ and $BAB' = I$.

This transformation is quite common in animal breeding. A reference is Thompson [22]. The reason for making the transformation $u = B\gamma$ is to obtain independent random effects $u_i(i = 1, \ldots, q)$ and as will be shown

later the Dirichlet process prior for these random effects can then be easily implemented. The model for each sire can now be written as

$$y_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{u} + \boldsymbol{\varepsilon}_i \quad (i = 1, \ldots, q) \tag{2}$$

where y_i is $n_i \times 1$, the vector of weaning weights for the lambs (progeny) of the ith sire. X_i is a known incidence matrix of order $n_i \times p$, $Z_i = \mathbf{1}_{n_i}z_{(i)}$ is a matrix of order $n_i \times q$ where $\mathbf{1}_{n_i}$ is a $n_i \times 1$ vector of ones and $z_{(i)}$ is the ith row of B^{-1}. Also $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma_\varepsilon^2 I_{n_i})$ and $\sum_{i=1}^{q} n_i = n$.

The model defined in (2) is an extension of the model studied by Kleinman and Ibraham [14] and Ibrahim and Kleinman [13] where only one random effect, u_i and the fixed effects have an influence on the response y_i. This difference occurs because A was assumed an identity matrix by them.

In model (2) and for our data set, "flat" or uniform prior distributions are assigned to σ_ε^2 and $\boldsymbol{\beta}$ which means that all relevant prior information for these two parameters have been incorporated into the description of the model.

Therefore:

$$p(\boldsymbol{\beta}, \sigma_\varepsilon^2) = p(\boldsymbol{\beta})p(\sigma_\varepsilon^2) \propto \text{constant}$$

i.e. σ_ε^2 is a bounded flat prior $[0, \infty]$ and $\boldsymbol{\beta}$ is uniformly distributed on the interval $[-\infty, +\infty]$. Furthermore, the prior distribution for the uncorrelated random effects u_i $(i = 1, \ldots, q)$ is given by

$$u_i \sim G$$

where

$$G \sim DP(M{\cdot}G_0).$$

Such a model assumes that the prior distribution G itself is uncertain, but has been drawn from a Dirichlet process. The parameters of a Dirichlet process are G_0, the probability measure, and M, a positive scalar assigning mass to the real line. The parameter G_0, called the base measure or base prior, is a distribution that approximates the true nonparametric shape of G. It is the best guess of what G is believed to be and is the mean distribution of the Dirichlet process (see West *et al.* [24]). The parameter M on the contrary reflects our prior belief about how similar the nonparametric distribution G is to the base measure G_0. There are two special cases in which the mixture of the Dirichlet process (MDP) models leads to the fully parametric case. As $M \to \infty$, $G \to G_0$ so that the base prior is the prior distribution for u_i. Also if the true values of the random effects are identical, the same is true. The use of the Dirichlet process prior can be simplified by noting that when G is integrated over its prior distribution,

the sequence of u_i's follows a general Polya urn scheme, (Ferguson [9]), that is

$$u_1 \sim G_0$$

$$u_q | u_1, \ldots, u_{q-1} \begin{cases} = u_j \text{ with probability } \dfrac{1}{M+q-1}; \ j = 1, \ldots, q-1; \\ \sim G_0 \text{ with probability } \dfrac{M}{M+q-1}. \end{cases}$$

(3)

In other words, by analytically marginalising over this dimension of the model we avoid the infinite dimension of G. So marginally, the u_i's are distributed as the base measure along with the added property that $p(u_i = u_j \ i \neq j) > 0$. It is clear that the marginalisation implies that random effects $(u_i; \ i = 1, \ldots, q)$ are no longer conditionally independent. See Ferguson [9] for further details. Specifying a prior on M and the parameters of the base distribution G_0 completes the Bayesian model specification. In this note we will assume that

$$G_0 = N(0, \sigma_\gamma^2).$$

Marginal posterior distributions are needed to make inferences about the unknown parameters. This will be achieved by using the Gibbs sampler. The typical objective of the sampler is to collect a sufficiently large enough number of parameter realisations from conditional posterior densities in order to obtain accurate estimates of the marginal posterior densities, see Gelfand and Smith [10] and Gelfand *et al.* [11].

If "flat" or uniform priors are assigned to $\boldsymbol{\beta}$ and σ_ε^2, then the required conditionals for $\boldsymbol{\beta}$ and σ_ε^2 are

$$\boldsymbol{\beta} | \boldsymbol{u}, \sigma_\varepsilon^2, \boldsymbol{y} \sim N_p \left\{ \hat{\boldsymbol{\beta}}, \sigma_\varepsilon^2 (X'X)^{-1}, \right\}$$

(4)

where

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1} X'(\boldsymbol{y} - Z\boldsymbol{u})$$

and

$$p(\sigma_\varepsilon^2 | \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{y}) \propto \left\{ \prod_{i=1}^{q} \left(\frac{1}{\sigma_\varepsilon^2} \right)^{n_i/2} \right\} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} (\boldsymbol{y} - X\boldsymbol{\beta} - Z\boldsymbol{u})'(\boldsymbol{y} - X\boldsymbol{\beta} - Z\boldsymbol{u}) \right\} \cdot$$

(5)

The proof of the following theorem is contained in Appendix A.

Theorem 1

The conditional posterior distribution of the random effect u_ℓ is

$$p(u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y}, M)$$

$$\propto \sum_{j \neq \ell}^{q} \left\{ \prod_{i=1}^{q} \phi \left(\boldsymbol{y}_i | X_i\boldsymbol{\beta} + z_{i\ell}u_j\mathbf{1}_{n_i} + \sum_{m \neq \ell}^{q} z_{im}u_m\mathbf{1}_{n_i}; \sigma_\varepsilon^2 I_{n_i} \right) \right\} . \delta_{u_j}$$

$$+ J_\ell \left\{ \prod_{i=1}^{q} \phi \left(\boldsymbol{y}_i | X_i\boldsymbol{\beta} + z_{i\ell}u_\ell\mathbf{1}_{n_i} + \sum_{m \neq \ell}^{q} z_{im}u_m\mathbf{1}_{n_i}; \sigma_\varepsilon^2 I_{n_i} \right) \right\} \phi(u_\ell|0, \sigma_\gamma^2) \quad (6)$$

where $\phi(\bullet|\mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2, $\boldsymbol{u}_{(\ell)}$ denotes the vector of random effects for the subjects (sires) excluding subject ℓ, δ_s is a degenerate distribution with point mass at s and

$$J_\ell = M \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{q} \phi \left(\boldsymbol{y}_i | X_i\boldsymbol{\beta} + z_{i\ell}u_\ell\mathbf{1}_{n_i} + \sum_{m \neq \ell}^{q} z_{im}u_m\mathbf{1}_{n_i}; \sigma_\varepsilon I_{n_i} \right) \right\} \phi(u_\ell|0, \sigma_\gamma^2)du_\ell$$

$$= M(2\pi)^{-\frac{1}{2}n}(\sigma_\varepsilon^2)^{-\frac{1}{2}n}(\sigma_\gamma^2)^{-\frac{1}{2}} \left(\frac{\sum_{i=1}^{q} z_{i\ell}^2 n_i}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\gamma^2} \right)^{-\frac{1}{2}} \times \exp \left\{ -\frac{1}{2}\frac{1}{\sigma_\varepsilon^2} \right.$$

$$\times \left[\sum_{i=1}^{q} \left(\boldsymbol{y}_i - X_i\boldsymbol{\beta} - \sum_{m \neq \ell}^{q} z_{im}u_m\mathbf{1}_{n_i} \right)' \left(\boldsymbol{y}_i - X_i\boldsymbol{\beta} - \sum_{m \neq \ell}^{q} z_{im}u_m\mathbf{1}_{n_i} \right) \right.$$

$$\left. \left. - \left(\frac{1}{\sigma_\varepsilon^2} \right) \left(\frac{\sum_{i=1}^{q} z_{i\ell}^2 n_i}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\gamma^2} \right)^{-1} \left\{ \sum_{i=1}^{q} z_{i\ell}\mathbf{1}'_{n_i} \left(\boldsymbol{y}_i - X_i\boldsymbol{\beta} - \sum_{m \neq \ell}^{q} z_{im}u_m\mathbf{1}_{n_i} \right) \right\}^2 \right] \right\}.$$

$$(7)$$

Each summand in the conditional posterior distribution of u_ℓ given in (6) is therefore separated into two elements. The first element is a mixing probability, and the second is a distribution to be mixed. The conditional posterior

distribution of u_ℓ can be sampled according to the following rule:

$$p\{u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y}, M\}$$

$$
\begin{cases}
= u_j(j = 1, 2, \ldots, \ell - 1, \ell + 1, \ldots, q) \quad \text{with probability} \\[2mm]
\dfrac{\displaystyle\prod_{i=1}^q \phi\left(\boldsymbol{y}_i|X_i\boldsymbol{\beta} + z_{i\ell}u_j\mathbf{1}_{n_i} + \sum_{m\neq\ell}^q z_{im}u_m\mathbf{1}_{n_i}; \sigma_\varepsilon^2 I_{n_i}\right)}{J_\ell + \displaystyle\sum_{j\neq\ell}^q \prod_{i=1}^q \phi\left(\boldsymbol{y}_i|X_i\boldsymbol{\beta} + z_{i\ell}u_j\mathbf{1}_{n_i} + \sum_{m\neq\ell}^q z_{im}u_m\mathbf{1}_{n_i}; \sigma_\varepsilon^2 I_{n_i}\right)} \\[6mm]
\sim h(u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y}) \quad \text{with probability} \\[2mm]
\dfrac{J_\ell}{J_\ell + \displaystyle\sum_{j\neq\ell}^q \prod_{i=1}^q \phi\left(\boldsymbol{y}_i|X_i\boldsymbol{\beta} + z_{i\ell}u_j\mathbf{1}_{n_i} + \sum_{m\neq\ell}^q z_{im}u_m\mathbf{1}_{n_i}; \sigma_\varepsilon^2 I_{n_i}\right)}
\end{cases}
\tag{8}
$$

where

$$h(u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, y)$$

$$
= N\left\{\left[\frac{\left(\displaystyle\sum_{i=1}^q z_{i\ell}^2 n_i\right)}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\gamma^2}\right]^{-1} \frac{1}{\sigma_\varepsilon^2}\left\{\sum_{i=1}^q z_{i\ell}\mathbf{1}_{n_i}'\left(\boldsymbol{y}_i - X_i\boldsymbol{\beta}_i - \sum_{m\neq\ell}^q z_{im}u_m\mathbf{1}_{n_i}\right)\right\};\right.
$$

$$
\left.\left[\frac{\left(\displaystyle\sum_{i=1}^q z_{i\ell}^2 n_i\right)}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\gamma^2}\right]^{-1}\right\}.
\tag{9}
$$

Note that the function $h(u_\ell|\boldsymbol{\beta}, \sigma_\gamma^2, \sigma_\varepsilon^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y})$ is the conditional posterior density of u_ℓ if $G_0 = N(0, \sigma_\gamma^2)$ is the prior distribution of u_ℓ. For the procedure described in equation (8),the weights are proportional to

$$
\prod_{i=1}^q \phi\left(\boldsymbol{y}_i|X_i\boldsymbol{\beta} + z_{i\ell}u_j + \sum_{m\neq\ell}^q z_{im}u_m; \sigma_\varepsilon^2 I_{n_i}\right) \quad \text{and} \quad J_\ell.
$$

From the above sampling rule (equation (8)) it is clearer that the smaller the residual of the subject (sire) ℓ, the larger the probability that its new value will be selected from the conditional posterior density $h(u_\ell|\boldsymbol{\beta}, \sigma_\gamma^2, \sigma_\varepsilon^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y})$. On the contrary, if the residual of subject ℓ is relatively large, larger than the residual obtained using the random effect of subject j, then u_j is more likely to be chosen as the new random effect for subject ℓ.

The Gibbs sampler for $p(\boldsymbol{\beta}, \boldsymbol{u}, \sigma_\varepsilon^2|\sigma_\gamma^2, \boldsymbol{y}, M)$ can be summarised as follows:

(0)   Select starting values for $\boldsymbol{u}^{(0)}$ and $\sigma_\varepsilon^{2(0)}$. Set $\ell = 0$.
(1)   Sample $\boldsymbol{\beta}^{(\ell+1)}$ from $p(\boldsymbol{\beta}|\boldsymbol{u}^{(\ell)}, \sigma_\varepsilon^{2(\ell)}, \boldsymbol{y})$ according to equation (4).
(2)   Sample $\sigma_\varepsilon^{2(\ell+1)}$ from $p(\sigma_\varepsilon^2|\boldsymbol{\beta}^{(\ell+1)}, \boldsymbol{u}^{(\ell)}, \boldsymbol{y})$ according to equation (5).
(3.1) Sample $u_1^{(\ell+1)}$ from $p\{u_1|\boldsymbol{\beta}^{(\ell+1)}, \sigma_\varepsilon^{2(\ell+1)}, \sigma_\gamma^2, \boldsymbol{u}_{(1)}^{(\ell)}, \boldsymbol{y}, M\}$ according to equation (8).

$$\vdots$$

(3.q) Sample $u_q^{(\ell+1)}$ from $p\{u_q|\boldsymbol{\beta}^{(\ell+1)}, \sigma_\varepsilon^{2(\ell+1)}, \sigma_\gamma^2, \boldsymbol{u}_{(q)}^{(\ell+1)}, \boldsymbol{y}, M\}$ according to equation (8).
(4)   Set $\ell = \ell + 1$ and return to step (1).

The newly generated random effects for each subject (sire) will be grouped into clusters in which the subjects have equal u_ℓ's. That is, after selecting a new u_ℓ for each subject ℓ in the sample, there will be some number $k, 0 < k \le q$, of unique values among the u_ℓ's. Denote these unique values by $\delta_r, r = 1, \ldots, k$. Additionally let r represent the set of subjects with a common random effect δ_r. Note that knowing the random effects is equivalent to knowing k, all of the δ's and the cluster membership r. Bush and MacEachern [3], Kleinman and Ibrahim [14] and Ibrahim and Kleinman [13] recommended one additional piece of the model as an aid to convergence for the Gibbs sampler. To speed mixing over the entire parameter space, they suggest moving around the δ's after determining how the u_ℓ's are grouped. The conditional posterior distribution of the location of the cluster given the cluster structure is

$$\boldsymbol{\delta}|\boldsymbol{\beta}, \sigma_\gamma^2, \boldsymbol{y} \sim N\left\{\hat{\boldsymbol{\delta}}, \left(\tilde{Z}'\tilde{Z} + I_k\frac{\sigma_\varepsilon^2}{\sigma_\gamma^2}\right)^{-1}\sigma_\varepsilon^2\right\} \tag{10}$$

where

$$\boldsymbol{\delta} = [\delta_1, \delta_2 \ldots, \delta_k]', \quad I_k \text{ is a } k \times k \text{ identity matrix,}$$

$$\hat{\boldsymbol{\delta}} = \left(\tilde{Z}'\tilde{Z} + I_k\frac{\sigma_\varepsilon^2}{\sigma_\gamma^2}\right)^{-1}\tilde{Z}'(\boldsymbol{y} - X\boldsymbol{\beta})$$

and the matrix $\tilde{Z}(n \times k)$ is obtained by adding the row values of these columns of Z that correspond to the same cluster. After generating $\boldsymbol{\delta}^{(\ell+1)}$ these cluster locations are then assigned to the $\boldsymbol{u}^{(\ell+1)}$ according to the cluster structure.

When the algorithm is implemented without this step, we find that the locations of the clusters may not move from a small set of values for many iterations, resulting in very slow mixing over the posterior and leading to poor estimates of posterior quantities.

For the Gibbs procedure described above, it is assumed that σ_γ^2 and M are known. Typically the variance σ_γ^2 in the base measure of the Dirichlet process is unknown and therefore a suitable prior distribution must be specified for it. Note that once this has been accomplished the base measure is no longer marginally normal.

For convenience, suppose $p(\sigma_\gamma^2) \propto$ constant to present lack of prior knowledge about σ_γ^2. The posterior distribution of σ_γ^2 is then an inverse gamma density

$$p(\sigma_\gamma^2 | \boldsymbol{\delta}, \boldsymbol{y}) \propto \left(\frac{1}{\sigma_\gamma^2}\right)^{k/2} \exp\left\{-\frac{1}{2\sigma_\gamma^2}\boldsymbol{\delta}'\boldsymbol{\delta}\right\} \sigma_\gamma^2 > 0 \tag{11}$$

The Gibbs sampling scheme is modified by sampling $\boldsymbol{\delta}$ from (10) and σ_γ^2 from (11).

The precision or total mass parameter M of the mixing Dirichlet process directly determines the prior distribution for k, the number of additional normal components in the mixture, and thus is a critical smoothing parameter for the model. The following theorem can now be stated.

Theorem 2

If the noninformative prior $p(M) \propto M^{-1}$ is used, then the posterior of M can be expressed as a mixture of two gamma posteriors, and the conditional distribution of the mixing parameter x given M and k is a simple beta. Therefore

$$p(M|x, k) \propto M^{k-1} \exp\left\{-M\big(\log(x)\big)\right\} + qM^{k-2} \exp\left\{-M\big(-\log(x)\big)\right\} \tag{12}$$

and

$$p(x|M, k) \propto x^M (1 - x)^{q-1} \quad 0 < x < 1. \tag{13}$$

The proof is given in the Appendix.

On completion of the simulation, we will have a series of sampled values of k, M, x and all the other parameters. Suppose that the Monte Carlo sample size is N, and denote the sampled values $k^{(\ell)}$, $x^{(\ell)}$, etc, for $\ell = 1, \ldots, N$. Only the sampled values $k^{(\ell)}$ and $x^{(\ell)}$ are needed in estimating the posterior $p(M|\boldsymbol{y})$ *via* the usual Monte Carlo average of conditional posteriors, *viz.*

$$p(M|\boldsymbol{y}) \simeq N^{-1} \sum_{\ell=1}^{N} p(M|x^{(\ell)}, k^{(\ell)})$$

where the summands are simply the conditional gamma mixtures in equation (12).

Finally the correlated random effects γ as defined in equation (1) can be obtained from the simulated u's by making the transformation $\gamma = B^{-1}u$.

Convergence was studied using the Gelman and Rubin [12] method. Multiple chains of the Gibbs sampler were run from different starting values and the scale reduction factor which evaluates between and within chain variation was calculated. Values of this statistic near one for all the model parameters was confirmation that the distribution of the Gibbs simulation was close to the true posterior distribution.

2.2. Illustration

Example: Elsenburg Dormer sheep stud

An animal breeding experiment was used to illustrate the nonparametric Bayesian procedure. The data are from the Dormer sheep stud started at the Elsenburg College of Agriculture near Stellenbosch, Western Cape, South Africa in 1940. The main object in developing the Dormer was the establishment of a mutton sheep breed which would be well adapted to the conditions prevailing in the Western Cape (winter rainfall) and which could produce the desired type of ram for crossbreeding purposes, Swart [21]. Single sire mating was practised with 25 to 30 ewes allocated to each ram. A spring breeding season (6 weeks duration) was used throughout the study. The season therefore had to be included as a fixed effect as a birth year-season concatenation. During lambing, the ewes were inspected daily and dam and sire numbers, date of birth, birth weight, age of dam, birth status (type of birth) and size of lamb were recorded. When the first lamb reached an age of 107 days, all the lambs 93 days of age and older were weaned and live weight was recorded. The same procedure was repeated every two weeks until all the lambs were weaned. All weaning weights were adjusted to a 100 day equivalent before analysis by using the following formula

$$\left(\frac{\text{Weaning weight} - \text{Birth weight}}{\text{Age at weaning}} \right) 100 + \text{Birth weight.}$$

As mentioned, a total of 3 325 weaning records from the progeny of 101 sires were used. In other words only a sample from the Elsenburg Dormer stud was used for calculation and illustration purposes. The model in this case is a sire model and the breeding values of the related sires are the random effects. Whenever appropriate, comparisons will be drawn in Section 3 between the Bayes estimates (using Gibbs sampling) obtained from a Matlab® programme, as well as the restricted maximum likelihood (REML) estimates. The classical (REML) estimates were obtained by using the MTDFREML programme developed by Boldman *et al.* [2].

For our example $\boldsymbol{\beta}(p \times 1) = (\beta_0, \boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_4)'$

where $\boldsymbol{\beta}'_1 = \beta_1$: sex of lamb effect;

       $\boldsymbol{\beta}'_2 = (\beta_2, \beta_3)$: birth status effects;

       $\boldsymbol{\beta}'_3 = (\beta_4, \ldots, \beta_8)$: age of dam effects;

and $\boldsymbol{\beta}'_4 = (\beta_9, \ldots, \beta_{27})$: year (season of birth) effects.

The sex of the lamb was male and female, birth status was individual, twins and triplets, age of dams were from 2 to 7 years and older and the years (season of birth) from 1980–1999.

The Gibbs sampler constructed to draw from the appropriate conditional posterior distributions is described in Section 2.1. Five different Gibbs sequences of length 404 000 were generated. The burn in period for each chain was 4000 and then every 250th draw was saved, thus giving a sample of 8000 uncorrelated draws. By examination of the scale reduction factor it was clear that convergence has been obtained. Draws 500 apart were also considered but no differences in the corresponding posterior distributions, parameter estimates or random effects could be detected.

3. RESULTS

3.1. Variance components

The estimates of the variance components obtained from REML, traditional and nonparametric Bayesian procedures are reported in Table I. Also given in Table I is $h^2 = \dfrac{4\sigma_\gamma^2}{\sigma_\gamma^2 + \sigma_\varepsilon^2}$, the heritability coefficient. In Table II the 95% credibility intervals are given and in Figures 1 and 2 the marginal posterior densities of σ_γ^2 and h^2 are illustrated.

From Tables I and II, it is clear that the point estimates and 95% credibility intervals of σ_ε^2 using REML or Bayesian methods are for all practical purposes the same. This comes as no surprise since the posterior density of the error variance is not directly influenced by the Dirichlet process prior.

Table I. REML and Bayesian estimates (posterior values) for the variance components and h^2.

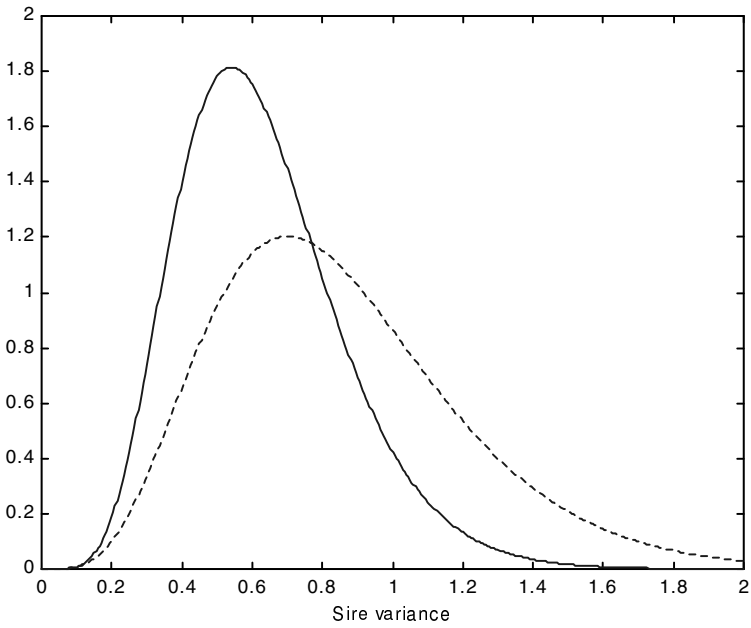| Parameter	REML	Traditional Bayes		Nonparametric Bayes – Dirichlet process prior	
		Mean	Mode	Mean	Mode
σ_ε^2	18.81	18.83	18.80	18.73	18.71
σ_γ^2	0.54	0.63	0.54	0.85	0.70
h^2	0.11	0.13	0.12	0.17	0.15

Figure 1. Estimated marginal posterior densities of σ_γ^2 (sire variance) for the traditional Bayes (solid) and non parametric method (- -).
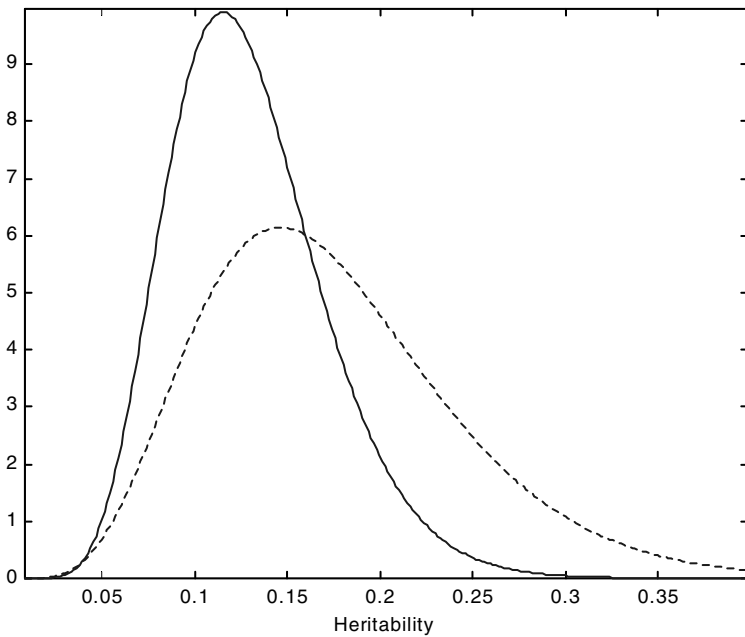

Figure 2. Estimated marginal posterior densities of h^2 (heritability) for the traditional Bayes (solid) and non parametric method (- -).

Table II. 95% credibility intervals for the variance components and h^2.

Parameter	Traditional Bayes	Nonparametric Bayes – Dirichlet process prior
σ_ε^2	17.87 ; 19.70	17.77 ; 19.71
σ_γ^2	0.26 ; 1.16	0.29 ; 1.69
h^2	0.06 ; 0.22	0.07 ; 0.33

The situation for the sire variance was however, somewhat different. The posterior distributions of σ_γ^2 are displayed in Figure 1. The differences in the densities are evident from the figure. These differences were expected, since the sire variance component is directly affected by the relaxation of the normal assumption.

The posterior densities of h^2 are given in Figure 2. The differences in the spread of these densities are similar to those of Figure 1, *i.e.* the posterior density in the case of the nonparametric Bayes method is more spread out than the corresponding density for the traditional Bayes procedure.

3.2. Fixed effects

The emphasis in breeding experiments is on the variance components and on the prediction of particular random effects, but estimation of the fixed effects is also important. Once estimates of the variance components have been obtained, they can be used to derive estimates of the fixed effects. These estimates are given in the REML/BLUE column of Table III where the estimates of the first ten fixed effects are given. The "Traditional Bayes" and "Nonparametric Bayes" estimates as well as the 95% credibility intervals were calculated using the Gibbs sampler. As expected the estimates of the fixed effects for the different methods are for all practical purposes the same. The Dirichlet process prior did not directly influence the posterior distributions of the fixed effects to the same extent as that for the random effects.

The fixed effect β_1 for example measures the expected difference in average weaning weight between male and female lambs, β_2 the expected difference between "single births" and "triplets" and β_3 between "twins" and triplets. β_6, however, measures the expected difference in average weaning weight of lambs for dams 4 and 7 years of age and β_9 measures the expected difference in average weaning weight between the lambs born in 1980 and 1999.

3.3. Random effects

Rather than reporting the results for all 101 sires, we will focus our discussion on the worst five and best five sires. The rankings of these ten sires were the same for the REML and traditional Bayes procedures but differed somewhat

Table III. REML and Bayesian estimates (modes of the posterior densities) as well as the 95% credibility intervals.

Parameter	REML/ BLUE	Traditional Bayes		Nonparametric Bayes – Dirichlet process prior	
		Mode	95% Credibility interval	Mode	95% Credibility interval
β_0	18.83	18.80	17.30 ; 20.21	18.83	17.28 ; 20.26
β_1	2.09	2.08	1.78 ; 2.39	2.09	1.74 ; 2.42
β_2	7.29	7.29	6.61 ; 7.99	7.21	6.59 ; 8.01
β_3	2.08	2.09	1.40 ; 2.72	1.99	1.39 ; 2.76
β_4	0.01	0.02	−0.69 ; 0.68	0.03	−0.68 ; 0.71
β_5	1.73	1.72	1.08 ; 2.44	1.72	1.01 ; 2.45
β_6	2.29	2.29	1.62 ; 2.94	2.28	1.60 ; 2.95
β_7	2.06	2.05	1.38 ; 2.74	2.03	1.29 ; 2.79
β_8	0.94	0.90	0.21 ; 1.67	0.91	0.20 ; 1.69
β_9	7.16	7.18	5.85 ; 8.65	7.23	5.81 ; 8.70

from the rankings of the nonparametric Bayesian method. For example sire No. 63 had the fifth best ranking according to the traditional Bayes method but was only ranked eighth using the Dirichlet process prior method. On the contrary, sire No. 59 was ranked the fifth best using the nonparametric procedure but was only the seventh best according to REML and traditional Bayes. The breeding values and 95% credibility intervals are listed in Table IV while the estimated posterior densities of the breeding values for sires No. 35 ("best sire") and 36 ("worst sire") are illustrated in Figures 3 and 4.

Unlike the densities of the fixed effects, the Dirichlet process prior had a large effect on the posterior densities of the different breeding values. Also from the posterior distributions of the selected breeding values, different facts were revealed. Not only were the credibility intervals wider in the case of the nonparametric Bayesian procedure, but the central values (modes and means) were different as well. A key difference between REML/BLUP prediction and Bayesian inference is the treatment of the variance components. To obtain the BLUP estimates, the variance components are fixed at a single value, ignoring uncertainty associated with estimating their values. The Bayesian analysis incorporates this uncertainty by averaging over the plausible values of the variance components.

3.4. Precision parameter

Let us now turn to the important parameter of the Dirichlet process, M. Recall that the parameter M, a type of dispersion parameter for the Dirichlet
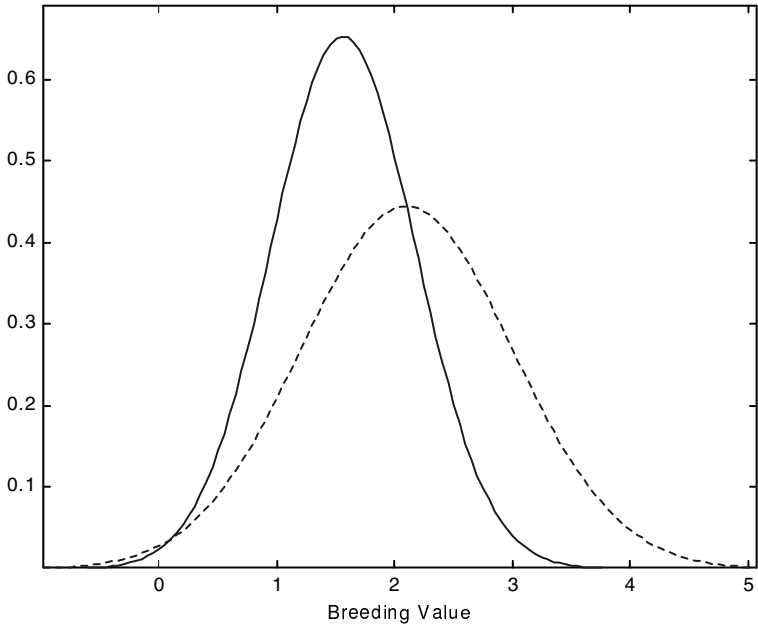
Figure 3. Estimated marginal posterior densities of γ_{35} – breeding values of "best sire" for the traditional Bayes (solid) and non parametric method (- -).
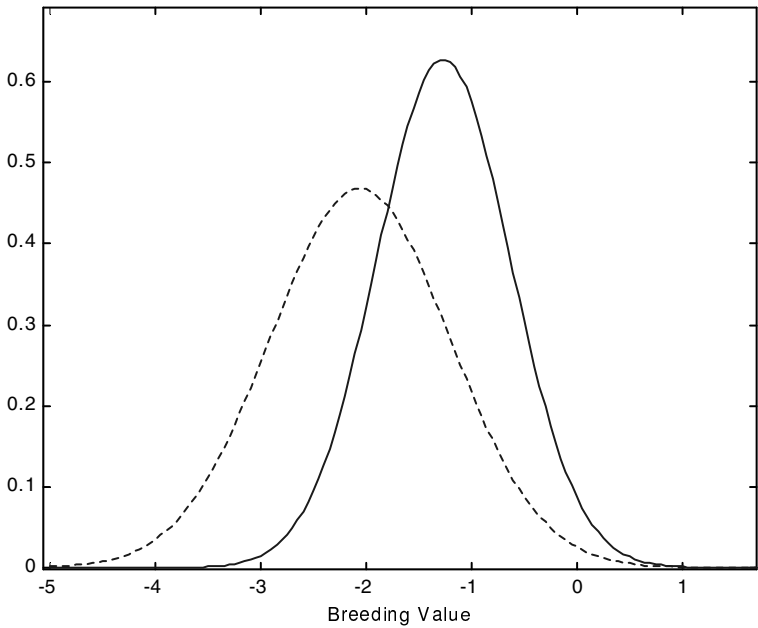

Figure 4. Estimated marginal posterior densities of γ_{36} – breeding value of "worst sire" for the traditional Bayes (solid) and non parametric method (- -).

Table IV. Breeding values (modes of the posterior densities) for the worst and best five sires between 1980 and 1999.

REML/ Trad Bayes	REML/ BLUP	Traditional Bayes		Nonparametric Bayes Dirichlet process prior		Non-parametric Bayes
Sire No.	Mode	Mode	95% credibility interval	Mode	95% credibility interval	Sire No.
36	−1.22	−1.26	−2.66 ; −0.12	−2.06	−3.70 ; −0.48	36
34	−1.04	−1.07	−2.32 ; −0.03	−1.69	−3.33 ; −0.01	34
23	−0.99	−1.01	−2.08 ; −0.08	−1.51	−2.84 ; −0.21	23
89	−0.92	−0.98	−2.25 ; 0.12	−1.35	−3.09 ; 0.41	9
9	−0.86	−0.96	−2.40 ; 0.22	−1.33	−3.00 ; 0.25	89
63	0.78	0.82	−0.24 ; 2.01	1.28	0.03 ; 2.66	59
24	0.81	0.84	−0.23 ; 1.88	1.32	−0.34 ; 3.06	92
92	0.81	0.87	−0.33 ; 2.17	1.33	−0.34 ; 3.06	43
43	1.14	1.17	0.19 ; 2.24	1.83	0.11 ; 3.93	55
35	1.50	1.56	0.40 ; 2.81	2.10	0.55 ; 3.64	35

process prior, is a measure of the strength in the belief that G is G_0. Although it may be hard to quantify, M is a positive scalar that is related to how "clumpy" the data are (often called a precision parameter). Clumpy data occur when the different subjects (sires) are concentrated into a few clusters. Ibrahim and Kleinman [13] chose a range of values for the parameter M to reflect small, moderate and large departures from normality for the distribution of the random effects. In practice it is, however, difficult to select appropriate values for this parameter. In this example, we extended the results of Ibrahim and Kleinman [13] by placing a vague prior on M and simulating its posterior distribution. Recall also that M determines the prior distribution of k, the number of additional normal components in the mixture and is a critical smoothing parameter for the mixed linear model.

In Figure 5, the posterior distribution of k, the number of "sire clusters" for the Nonparametric Bayesian method is illustrated. The mean value is $\bar{k} = 70$ and the mode of the posterior distribution of M is $M_0 = 110.5$. As mentioned, large values of M favour the base measure G_0 as our prior, *i.e.* the traditional Bayes procedure. For this example M_0 is relatively small which is an indication that the nonparametric Bayes procedure will give better estimates and predictions than the traditional method. The posterior distribution of M is given in Figure 6.
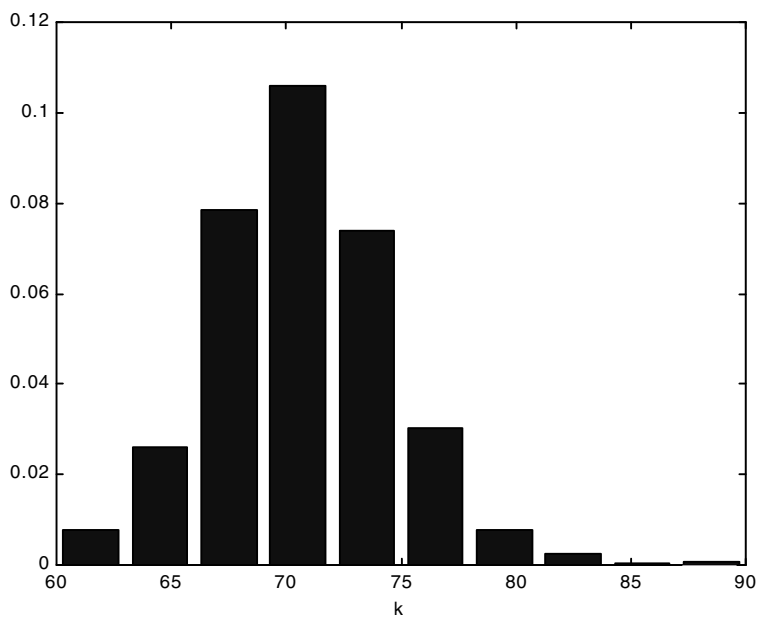
Figure 5. Observed histogram for k, number of groups. $\bar{k} = 70$.
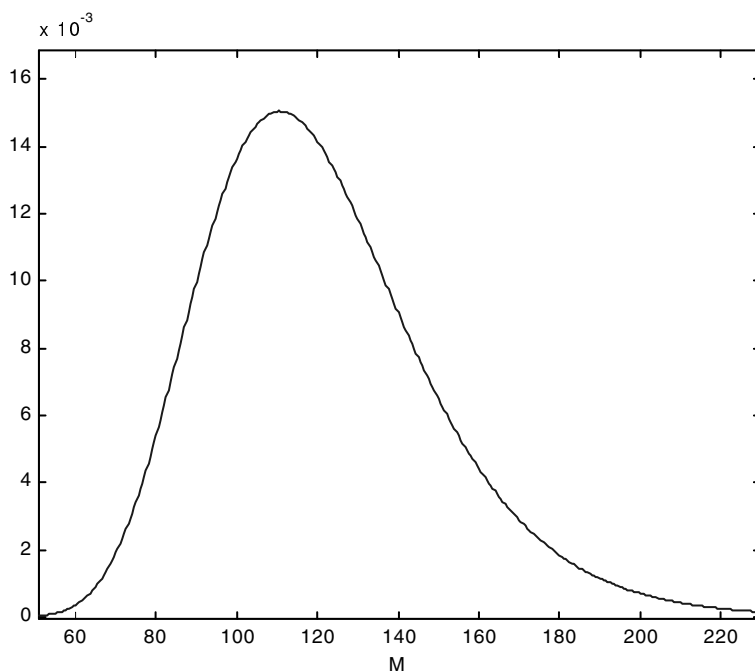

Figure 6. Estimated marginal posterior density of M (precision parameter). $M_0 = 110.5$.

4. DISCUSSION

There is a general feeling among many statisticians that it is desirable in many contexts to make fewer assumptions about the underlying populations from which the data are obtained than is required for a parametric analysis. It is of practical interest, therefore, to study models that are less sensitive than Gaussian ones to departures from assumptions. For example, in dairy cattle breeding, it is known that more valuable cows receive preferential treatment (better housing, hormonal treatment, better feeding, etc.) and to such an extent that the treatment cannot be accommodated in the model, this leads to bias in the prediction of breeding values. Not much work so far has been undertaken on how to cope with a preferential treatment in practice. Strandén and Gianola [19,20] applied a Bayesian approach to fit mixed linear models with t-distributed random and residual terms both in the univariate and multivariate forms. The problem with the t-distribution, however, is that it does not cover departures from symmetry. It might also be that no single parametric model could fully capture the variability inherent in the response variable, suggesting that we average over a collection of densities (see Carlin and Louis [4]). Unfortunately inference using mixture distributions is sometimes difficult since the parameters of the models are often not fully identified by the data. An appealing alternative to discrete finite mixtures would be continuous mixtures but this creates the problem of choosing an appropriate mixing distribution.

In this note and for our example, the mixing distribution was specified nonparametrically by using the Dirichlet process prior, in fact we allowed the data to suggest the appropriate mixing distribution. The Dirichlet process prior for the random effects is a more flexible prior than the t-distribution because it can accommodate an arbitrarily large range of model anomalies (multiple modes, heavy tails, skew distributions and so on).

In our opinion the best solution (for the preferential treatment of the animals problem) will be to combine the t-distribution and the Dirichlet process prior, *i.e.* the student t-distribution for the errors and the Dirichlet process prior for the random effects. This will result in a model that is also robust to outlying observations. At the moment, it is not yet possible to apply these robust methods to large data sets in animal breeding. Hence, if it is established that such a nonparametric model improves the estimation of variance components and fixed effects and the prediction of random effects, simpler and faster computer programmes should be developed.

We must, however, state that the discussion of the "preferential treatment of animal problems" has no linkage to our particular data set problem. The data from the Elsenburg Dormer sheep stud was mainly included for illustration and comparison purposes: *i.e.* to illustrate the application of the Dirichlet process prior for correlated random effects (which as far as we know has never been

done before) and to compare the corresponding results of the REML, traditional Bayes and nonparametric Bayesian procedures. We are, however, quite sure that the Dirichlet process prior will in the future play an important role in cases where undeclared preferential treatment of animals occurs.

5. CONCLUSION

In this note we have applied a Bayesian nonparametrics method to the mixed linear model. A Gibbs sampling procedure to calculate posterior distributions for the genetic parameters of this model has been presented. The important contribution of this paper revolves around the nonparametric prior distribution, the Dirichlet process prior, for the random effects and to correctly model and interpret the estimated effects and variance components.

It is clear from the example that the error variance and fixed effects are not directly influenced by the Dirichlet process prior. The situation for the sire variance and breeding values is, however, quite different. The posterior densities for the nonparametric Bayes method are more spread out. These differences are to be expected, since the sire variance component and random effects are directly affected by the relaxation of the normal assumption.

It is also well known that the value of the precision parameter M largely influences the posterior distribution of the random effects, the sire variance and the heritability coefficient. The value of M will further determine whether the estimates from the Dirichlet process behave like the standard Bayes estimate or like a nonparametric likelihood estimator. Further research and simulation experiments should be done to determine if the Dirichlet process prior can be used for solving the "undeclared preferential treatment of animals" problem.

REFERENCES

[1] Antoniak C.E., Mixtures of Dirichlet processes with applications to nonparametric problems, Ann. Statist. 2 (1974) 1152–1174.
[2] Boldman K.G., Kriese L.A., Van Vleck L.D., Van Tassell C.P., Kachman S.D., A Manual for use of MTDFREML. A set of Programs to obtain Estimates of Variances and Covariances, U.S. Dept. of Agric., Agricultural Research Service, 1995.
[3] Bush C.A., MacEachern S.N., A semiparametric Bayesian model for randomized block designs, Biometrika 83 (1996) 275–285.

[4]   Carlin B.P., Louis T.A., Bayes and Empirical Bayes Methods for Data Analysis, Chapman & Hall/CRC, New York, 2002.

[5]   Doss H., Bayesian nonparametric estimation for incomplete data via successive substitution sampling, Ann. Stat. 22 (1994) 1763–1786.

[6]   Escobar M.D., Estimating normal means with a Dirichlet process prior, Department of Statistics, Carnegie Mellon University, Technical Report No. 512, 1991.

[7]   Escobar M.D., Estimating normal means with a Dirichlet process prior, J. Am. Stat. Assoc. 89 (1994) 268–275.

[8]   Escobar M.D., West M., Bayesian density estimation and inference using mixtures, J. Am. Stat. Assoc. 90 (1995) 578–588.

[9]   Ferguson T.S., A Bayesian analysis of some nonparametric problems, Ann. Stat. 1 (1973) 209–230.

[10] Gelfand A.E., Smith A.F.M., Gibbs sampling for marginal posterior expectations, Comm. Statist. Theory Methods 20 (1991) 1747–1766.

[11] Gelfand A.F., Smith A.F.M, Lee T.M., Bayesian analysis of constrained parameters and truncated data problems using Gibbs sampling, J. Am. Stat. Assoc. 87 (1992) 523–532.

[12] Gelman A., Rubin D.R., A single series from the Gibbs sampler provides a false sense of security, in: Bernardo J.M., Berger J., Dawid A.P., Smith A.F.M. (Eds.), Bayesian Statistics 4, Oxford University Press, Oxford, 1992, pp. 627–635.

[13] Ibrahim J.G., Kleinman K.P., Semiparametric Bayesian methods for random effects models, in: Dey D., Müller P., Sinha D. (Eds.), Practical Nonparametric and Semiparametric Bayesian Statistics, Lectures Notes in Statistics, 133, 2000, pp. 89–113.

[14] Kleinman K.P., Ibrahim J.G., A Semiparametric Bayesian approach to the random effects model, Biometrics 54 (1998) 921–938.

[15] Lui J., Nonparametric hierarchical Bayes via sequential imputation, Ann. Stat. 24 (1996) 911–930.

[16] MacEachern S.N., Estimation normal means with a conjugate style Dirichlet process prior, Commun. Stat. – Simula 23 (1994) 727–741.

[17] MacEachern S.N., Müller P., Estimating mixture of Dirichlet process models, J. Comput. Graph. Stat. 7 (1998) 223–238.

[18] Müller P., Erkani A., West M., Bayesian curve fitting using multivariate normal mixtures, Biometrika 83 (1996) 67–79.

[19] Strandén I., Gianola D., Attenuating effects of preferential treatment with student-t mixed linear models: a simulation study, Genet. Sel. Evol. 30 (1998) 565–583.

[20] Strandén I., Gianola D., Mixed effects linear models with t-distributions for quantitative genetic analysis: a Bayesian approach, Genet. Sel. Evol. 31 (1999) 25–42.

[21] Swart J.C., The origin and development of the Dormer sheep breed, Meat Ind. 31 May–June (1967).

[22] Thompson R., Sire evalution, Biometrics 35 (1979) 339–353.

[23] West M., Hyperparameter estimation in Dirichlet process mixture models, Duke University, ISDS, Technical Report 92 – A03, 1992.

[24] West M., Müller P., Escobar M.D., Hierarchical priors and mixture models, with applications in regression and density estimation, in: Smith A.F.M, Freeman P.R. (Eds.), Aspects of Uncertainty: a tribute to Lindley D.V., Wiley, London, 1994, pp. 363–386.

APPENDIX

A.1 Proof of Theorem 1

From equation (3) it follows (see also Ferguson [9], p. 217), that the prior of the random effect u_q given $u_1, \ldots, u_{q-1}$ is

$$p(u_q|u_1, \ldots, u_{q-1}) = \frac{1}{M + q - 1} \sum_{j=1}^{q-1} \delta_{u_j} + \frac{M}{M + q - 1} G_0.$$

By exchangeability it follows that

$$p(u_\ell|\boldsymbol{u}_{(\ell)}) = \frac{1}{M + q - 1} \sum_{j \neq \ell}^{q} \delta_{u_j} + \frac{M}{M + q - 1} G_0.$$

The conditional likelihood of u_ℓ, is given by

$$L(u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y}, M) = \prod_{i=1}^{q} \left(\boldsymbol{y}_i | X_i \boldsymbol{\beta} + z_{i\ell} u_\ell \mathbf{1}_{n_i} + \sum_{m \neq \ell}^{q} z_{im} u_m \mathbf{1}_{n_i}; \sigma_\varepsilon^2 I_{n_i} \right).$$

Now the Bayes'theorem gives

$$p(u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y}, M) = \frac{p(u_\ell|\boldsymbol{u}_{(\ell)}) L(u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y}, M)}{\displaystyle\int_{-\infty}^{\infty} p(u_\ell|\boldsymbol{u}_{(\ell)}) L(u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y}, M) du_\ell}.$$

$$\text{(A.1)}$$

Equation (A.1) is proportional to (6) and proves the theorem.

Furthermore,

$$\int_{-\infty}^{\infty} p(u_\ell|\boldsymbol{u}_{(\ell)}) L(u_\ell|\boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{u}_{(\ell)}, \boldsymbol{y}, M) du$$

$$= \frac{1}{M + q - 1} \left[\sum_{j \neq \ell} \left\{ \prod_{i=1}^{q} \left(\boldsymbol{y}_i | X_i \boldsymbol{\beta} + z_{i\ell} u_j \mathbf{1}_{n_i} + \sum_{m \neq \ell}^{q} z_{im} u_m \mathbf{1}_{n_i}; \sigma_\varepsilon^2 I_{n_i} \right) \right\} + J_\ell \right]$$

where J_ℓ is defined in equation (7).

A.2 Proof of Theorem 2

In Antoniak [1] it is shown that the prior distribution of k may be written as

$$p(k|M, q) = c_q(k)q!M^k \frac{\Gamma(M)}{\Gamma(M + q)} \quad (k = 1, 2, \ldots, q). \tag{A.2}$$

West [23] mentioned that if required, the factors $c_q(k)$ can be easily computed using recurrence formulae for Stirling numbers. He also showed that the conditional posterior distribution of M,

$$p(M|k, \boldsymbol{\beta}, \boldsymbol{u}, \sigma_\varepsilon^2, \sigma_\gamma^2, \boldsymbol{y}) = p(M|k) \propto p(M)p(k|M) \tag{A.3}$$

where $p(M)$ is the prior for M and $p(k|M)$ is defined in (A.2). West [23] assumed $M \sim G(a, b)$, a Gamma prior with shape $a > 0$ and scale $b > 0$ (which we may extend to include a reference prior (Uniform for $\log(M)$) by letting $a \to 0$ and $b \to 0$). In this note we will use the latter which means that

$$p(M) \propto M^{-1} \quad M > 0.$$

For $M > 0$, the gamma functions in (A.2) can be written as

$$\frac{\Gamma(M)}{\Gamma(m + q)} = \frac{(M + q)B(M + 1, q)}{M\Gamma(q)}$$

where $B(\cdot, \cdot)$ is the beta function. Then in (A.3) and for any $k = 1, 2, \ldots q$, the posterior

$$p(M|k) \propto p(M)M^{k-1}(M + q)B(M + 1, q)$$

$$\propto M^{k-2}(M + q) \int_0^1 x^M(1 - x)^{q-1}dx, \tag{A.4}$$

using the definition of the beta function. From (A.3), it also follows that the joint posterior density of M and x is

$$p(M, x|k) \propto M^{k-2}(M + q)x^M(1 - x)^{q-1} \quad (0 < M, 0 < x < 1).$$

The conditional posteriors $p(M|x, k)$ and $p(x|M, k)$ can be determined as follows. Firstly

$$p(M|x, k) \propto M^{k-2}(M + q)e^{-M(-\log(x))}.$$

For $M > 0$, the latter reduces easily to a mixture of two gamma densities, *viz.*

$$M|x, k \sim \pi_x G\big(k, -\log(x)\big) + (1 - \pi_x)G\big(k - 1, -\log(x)\big) \tag{A.5}$$

with weights π_x defined by $\dfrac{\pi_x}{1 - \pi_x} = \dfrac{k - 1}{q(-\log(x))}$.

Secondly

$$p(x|M, k) \propto x^M(1 - x)^{q-1} \quad 0 < x < 1 \tag{A.6}$$

so that $x|M, k \sim B(M+1, q)$, a beta distribution with mean $(M+1)/(M+q+1)$.