Original article

Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling

Inge Riis KORSGAARD[a]*, Mogens Sandø LUND[a],
Daniel SORENSEN[a], Daniel GIANOLA[b],
Per MADSEN[a], Just JENSEN[a]

[a] Department of Animal Breeding and Genetics,
Danish Institute of Agricultural Sciences, PO Box 50, 8830 Tjele, Denmark
[b] Department of Meat and Animal Sciences, University of Wisconsin-Madison,
WI 53706-1284, USA

Abstract – A fully Bayesian analysis using Gibbs sampling and data augmentation in a multivariate model of Gaussian, right censored, and grouped Gaussian traits is described. The grouped Gaussian traits are either ordered categorical traits (with more than two categories) or binary traits, where the grouping is determined *via* thresholds on the underlying Gaussian scale, the liability scale. Allowances are made for unequal models, unknown covariance matrices and missing data. Having outlined the theory, strategies for implementation are reviewed. These include joint sampling of location parameters; efficient sampling from the fully conditional posterior distribution of augmented data, a multivariate truncated normal distribution; and sampling from the conditional inverse Wishart distribution, the fully conditional posterior distribution of the residual covariance matrix. Finally, a simulated dataset was analysed to illustrate the methodology. This paper concentrates on a model where residuals associated with liabilities of the binary traits are assumed to be independent. A Bayesian analysis using Gibbs sampling is outlined for the model where this assumption is relaxed.

categorical / Gaussian / multivariate Bayesian analysis / right censored Gaussian

1. INTRODUCTION

In a series of problems, it has been demonstrated that using the Gibbs sampler in conjunction with data augmentation makes it possible to obtain sampling-based estimates of analytically intractable features of posterior distributions.

---

* Correspondence and reprints
E-mail: IngeR.Korsgaard@agrsci.dk

Gibbs sampling [9, 10] is a Markov chain simulation method for generating samples from a multivariate distribution, and has its roots in the Metropolis-Hastings algorithm [11, 19]. The basic idea behind the Gibbs sampler, and other sampling based approaches, is to construct a Markov chain with the desired density as its invariant distribution [2]. The Gibbs sampler is implemented by sampling repeatedly from the fully conditional posterior distributions of parameters in the model. If the set of fully conditional posterior distributions do not have standard forms, it may be advantageous to use data augmentation [26], which as pointed out by Chib and Greenberg [3], is a strategy of enlarging the parameter space to include missing data and/or latent variables.

Bayesian inference in a Gaussian model using Gibbs sampling has been considered by *e.g.* [8] and with attention to applications in animal breeding, by [14, 23, 28, 30, 31]. Bayesian inference using Gibbs sampling in an ordered categorical threshold model was considered by [1, 24, 34]. In censored Gaussian and ordered categorical threshold models, Gibbs sampling in conjunction with data augmentation [25, 26] leads to fully conditional posterior distributions which are easy to sample from. This was demonstrated in Wei and Tanner [33] for the tobit model [27], and in right censored and interval censored regression models. A Gibbs sampler for Bayesian inference in a bivariate model with a binary threshold character and a Gaussian trait is given in [12]. This was extended to an ordered categorical threshold character by [32], and to several Gaussian, binary and ordered categorical threshold characters by [29]. In [29], the method for obtaining samples from the fully conditional posterior of the residual (co)variance matrix (associated with the normally distributed scale of the model) is described as being "ad hoc in nature".

The purpose of this paper was to present a fully Bayesian analysis of an arbitrary number of Gaussian, right censored Gaussian, ordered categorical (more than two categories) and binary traits. For example in dairy cattle, a four-variate analysis of a Gaussian, a right censored Gaussian, an ordered categorical and a binary trait might be relevant. The Gaussian trait could be milk yield. The right censored Gaussian trait could be log lifetime (if log lifetime is normally distributed). For cattle still alive, it is only known, that (log) lifetime will be higher than their current (log) age, *i.e.* these cattle have right censored records of (log) lifetime. The categorical trait could be calving ease score and the binary trait could be the outcome of a random variable indicating stillbirth or not. In general, allowances are made for unequal models and missing data. Throughout, we consider two models. In the first model, residuals associated with liabilities of the binary traits are assumed to be independent. This assumption may be relevant in applications where the different binary traits are measured on different groups of (related) animals. An example is infection trials, where some animals are infected with one pathogen and the remaining

animals with another pathogen. The two binary traits could be dead/alive three weeks after infection. (See *e.g.* [13] for a similar assumption in a bivariate analysis of two quantitative traits). In other applications and for a number of binary traits greater than one, however, the assumption of independence may be too restrictive. Therefore we also outline a Bayesian analysis using Gibbs sampling in the more general model where residuals associated with liabilities of the binary traits are correlated. (The two models are only different if the number of binary traits is greater than one).

The outline of the paper is the following: in Section 2, a fully Bayesian analysis of an arbitrary number of Gaussian, right censored Gaussian, ordered categorical and binary traits is presented for the particular case where all animals have observed values for all traits, *i.e.* no missing values. In Section 3, we extend the fully Bayesian analysis to allow for missing observations of the different traits. Strategies for implementation of the Gibbs sampler are given and/or reviewed in Section 4. These include univariate and joint sampling of location parameters, efficient sampling from a multivariate truncated normal distribution – necessary for sampling the augmented data, and sampling from an inverted Wishart distribution and from a conditional inverted Wishart distribution. Note that the conditional inverted Wishart distribution of the residual covariance matrix in the model assuming that residuals associated with liabilities of the binary traits are independent, is different from the conditional inverted Wishart distribution in the model where this assumption has been relaxed (if the number of binary traits is greater than one). The methods presented for obtaining samples from the fully conditional posterior of the residual covariance matrix are different from the method presented in [29]. To illustrate the developed methodology, simulated data are analysed in Section 5 which also outlines a way of choosing suitable starting values for the Gibbs sampler. The paper ends with a conclusion in Section 6.

2. THE MODEL WITHOUT MISSING DATA

2.1. The sampling model

Assume that m_1 Gaussian traits, m_2 right censored Gaussian traits, m_3 categorical traits with response in multiple ordered categories and m_4 binary traits are observed on each animal; $m_i \geq 0$, $i = 1, \ldots, 4$. The total number of traits is $m = m_1 + m_2 + m_3 + m_4$. In general, the data on animal i are $(\mathbf{y}_i, \boldsymbol{\delta}_i)$, $i = 1, \ldots, n$, where $\mathbf{y}_i = \big(y_{i1}, \ldots, y_{im_1}, y_{im_1+1}, \ldots, y_{im_1+m_2}, y_{im_1+m_2+1}, \cdots$ $\ldots, y_{im_1+m_2+m_3}, y_{im-m_4+1}, \ldots, y_{im}\big)$, and where $\boldsymbol{\delta}_i$ is a m_2 dimensional vector of censoring indicators of the right censored Gaussian traits. The number of animals with records is n and the data on all animals with records are $(\mathbf{y}, \boldsymbol{\delta})$. The observed vector of Gaussian traits of the animal i

is $(y_{i1}, \ldots, y_{im_1})$. For $j \in \{m_1 + 1, \ldots, m_1 + m_2\}$, y_{ij} is the observed value of $Y_{ij} = \min\{U_{ij}, C_{ij}\}$, where U_{ij} is normally distributed and C_{ij} is the point of censoring of the jth trait of animal i. The censoring indicator δ_{ij} is one iff U_{ij} is observed $(U_{ij} \le C_{ij})$ and zero otherwise. Δ_{oj} and Δ_{1j} will denote the sets of animals with δ_{ij} equal to zero and one, respectively, $j = m_1+1, \ldots, m_1+m_2$. The observed vector of categorical traits with response in three or more categories is $(y_{im_1+m_2+1}, \ldots, y_{im_1+m_2+m_3})$. The outcome y_{ij}, $j \in \{m_1 + m_2 + 1, \ldots, m_1 + m_2 + m_3\}$, is assumed to be determined by a grouping in an underlying Gaussian scale, the liability scale. The underlying Gaussian variable is U_{ij}, and the grouping is determined by threshold values. That is, $Y_{ij} = k$ iff $\tau_{jk-1} < U_{ij} \le \tau_{jk}$; $k = 1, \ldots, K_j$, where K_j $(K_j \ge 3)$ is the number of categories for trait j and $-\infty = \tau_{j0} \le \tau_{j1} \le \cdots \le \tau_{jK_j-1} \le \tau_{jK_j} = \infty$. The observed vector of binary traits is $(y_{im_1+m_2+m_3+1}, \ldots, y_{im})$. As for the ordered categorical traits, the observed value is assumed to be determined by a grouping in an underlying Gaussian scale. It is assumed that $Y_{ij} = 0$ iff $U_{ij} \le 0$ and $Y_{ij} = 1$ iff $U_{ij} > 0$.

Let $U_{ij} = Y_{ij}$ for $j = 1, \ldots, m_1$, that is for the Gaussian traits, and let $\mathbf{U}_i = (U_{i1}, \ldots, U_{im})'$ be the vector of Gaussian traits observed or associated with the right censored Gaussian traits, ordered categorical traits and binary traits of animal i. Define $\mathbf{U} = (\mathbf{U}_i)_{i=1,\ldots,n}$ as the nm-dimensional column vector containing the $\mathbf{U}_i's$. It is assumed that:

$$\mathbf{U}|\left(\mathbf{a}, \mathbf{b}, \mathbf{R} = \mathbf{r}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) \sim N_{nm}\left(\mathbf{Xb} + \mathbf{Za}, \mathbf{I}_n \otimes \begin{pmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} \\ \mathbf{r}_{21} & \mathbf{I}_{m_4} \end{pmatrix}\right) \quad (1)$$

where $\mathbf{b}$ is a p-dimensional vector of "fixed" effects. The vector $\mathbf{a}_i = (a_{i1}, \ldots, a_{im})'$ represents the additive genetic values of $\mathbf{U}_i$, $i = 1, \ldots, N$; $\mathbf{a} = (\mathbf{a}_i)_{i=1,\ldots,N}$, is the Nm dimensional column vector containing the $\mathbf{a}_i's$. N is the total number of animals in the pedigree; *i.e.* the dimension of the additive genetic relationship matrix, $\mathbf{A}$, is $N \times N$, and $\begin{pmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} \\ \mathbf{r}_{21} & \mathbf{I}_{m_4} \end{pmatrix}$ is the residual covariance matrix of $\mathbf{U}_i$ in the conditional distribution given $\left(\mathbf{a}, \mathbf{b}, \mathbf{R} = \mathbf{r}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)$. The usual condition that $R_{kk} = 1$ (*e.g.* [5]) has been imposed in the conditional probit model of Y_{ik} given $\mathbf{b}$ and $\mathbf{a}$, $k = m - m_4 + 1, \ldots, m$. Furthermore it is assumed that liabilities of the binary traits are conditionally independent, given $\mathbf{b}$ and $\mathbf{a}$. Note that we (in this section) carefully distinguish between the random (matrix) variable, $\mathbf{R}$, and an outcome, $\mathbf{r}$, of the random (matrix) variable, $\mathbf{R}$ (contrary to the way in which *e.g.* $\mathbf{b}$ and $\mathbf{a}$ are treated).

With two or more binary traits included in the analysis, however, the assumption of independence between residuals associated with liabilities of the binary traits may be too restrictive. Therefore we also considered the model where it

is assumed that:

$$\mathbf{U} \mid \left(\mathbf{a}, \mathbf{b}, \mathbf{R} = \mathbf{r}, (R_{kk} = 1)_{k=m-m_4+1,\ldots,m}\right) \sim N_{nm}\left(\mathbf{Xb} + \mathbf{Za}, \mathbf{I}_n \otimes \begin{pmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} \\ \mathbf{r}_{21} & \tilde{\mathbf{r}}_{22} \end{pmatrix}\right)$$

(2)

with $(\tilde{\mathbf{r}}_{22})_{kl} = (\mathbf{r}_{22})_{kl}$ for $k, l = m - m_4 + 1, \ldots, m$ with $k \neq l$, and $(\tilde{\mathbf{r}}_{22})_{kk} = 1$ for $k, l = m - m_4 + 1, \ldots, m$.

In the following, first, the model associated with (1) is treated; second, the necessary modifications related to the model in (2) are outlined.

2.2. Prior distribution

Let the elements of $\mathbf{b}$ be ordered so that the first p_1 elements are regression effects and the remaining $p_2 = p - p_1$ elements are "fixed" classification effects. It is assumed, *a priori*, that $\mathbf{b} \mid (\sigma_1^2, \sigma_2^2) \sim N_p\left(\mathbf{0}, \begin{pmatrix} \mathbf{I}_{p_1}\sigma_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_2}\sigma_2^2 \end{pmatrix}\right)$, where σ_1^2 and σ_2^2 are known (alternatively, it can be assumed, that some elements of $\mathbf{b}$ follow a normal distribution and the remaining elements follow an improper uniform distribution). The *a priori* distribution of the additive genetic values is $\mathbf{a} \mid \mathbf{G} \sim N_{Nm}(\mathbf{0}, \mathbf{A} \otimes \mathbf{G})$, where $\mathbf{G}$ is the $m \times m$ additive genetic covariance matrix of $\mathbf{U}_i, i = 1, \ldots, N$. *A priori*, $\mathbf{G}$ is assumed to follow an m-dimensional inverted Wishart distribution: $\mathbf{G} \sim IW_m(\mathbf{\Sigma_G}, f_{\mathbf{G}})$. Assuming, for the model associated with (1), that $\mathbf{R}$ follows an inverted Wishart distribution: $\mathbf{R} \sim IW_m(\mathbf{\Sigma_R}, f_{\mathbf{R}})$, then the prior distribution of $\mathbf{R}$, in the conditional distribution given $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, is the conditional inverted Wishart distributed. All of $\mathbf{\Sigma_G}$, $f_{\mathbf{G}}$, $\mathbf{\Sigma_R}$ and $f_{\mathbf{R}}$ are assumed known. *A priori*, it is assumed that the elements of $\mathbf{\tau}_j = \left(\tau_{j2}, \ldots, \tau_{jK_j-2}\right)$ are distributed as order statistics from a uniform distribution in the interval $\left[\tau_{j1}; \tau_{jK_j-1}\right] = [0; 1]$, *i.e.*: $p\left(\tau_{j2}, \ldots, \tau_{jK_j-2}\right) = (K_j - 3)! \mathbf{1}\left(\mathbf{\tau}_j \in \mathfrak{I}_j\right)$, where

$$\mathfrak{I}_j = \left\{\left(s_2, \ldots, s_{K_j-2}\right) \mid 0 \leq s_2 \leq \cdots \leq s_{K_j-2} \leq 1\right\}$$

([20]).

Concerning prior independence, the following assumption was made:

(a) *A priori* $\mathbf{b}$, $(\mathbf{a}, \mathbf{G})$, $\mathbf{R}$ and $\mathbf{\tau}_j, j = m_1 + m_2 + 1, \ldots, m_1 + m_2 + m_3$ are mutually independent, and furthermore, the elements of $\mathbf{b}$ are mutually independent.

In the model associated with (2), the prior assumptions were similar except that, *a priori*, $\mathbf{R}$ conditional on $(R_{kk} = 1)_{k=m-m_4+1,\ldots,m}$ is assumed to follow a conditional inverse Wishart distribution (which for $m_4 > 1$ is different from the prior given in the model associated with (1)).

2.3. Joint posterior distribution

For each animal, the augmented variables are $U'_{ij}s$ of right censored $(\delta_{ij} = 0)$ Gaussian traits and liabilities of ordered categorical and binary traits. The following notation will be used: $\mathbf{U}_0^{RC} = \{U_{ij} : i \in \Delta_{0j};$ $j = m_1 + 1, \ldots, m_1 + m_2\}$, this is the set of $U'_{ij}s$ of the censored observations from the right censored Gaussian traits. $\mathbf{U}^{CAT}$ and $\mathbf{U}^{BIN}$ will denote the sets of liabilities of ordered categorical and binary traits, respectively. The following will be assumed concerning the censoring mechanism:

(b) Random censoring conditional on

$$\boldsymbol{\omega} = \left(\mathbf{b}, \mathbf{a}, \mathbf{G}, \mathbf{R}, \boldsymbol{\tau}_{m_1+m_2+1}, \ldots, \boldsymbol{\tau}_{m_1+m_2+m_3}\right),$$

i.e., $\mathbf{C} = (\mathbf{C}_i)_{i=1,\ldots,n}$, where $\mathbf{C}_i = \left(C_{im_1+1}, \ldots, C_{im_1+m_2}\right)$ is the m_2 dimensional random vector of censoring times of animal i, is stochastically independent of $\mathbf{U}$, given $\boldsymbol{\omega}$.

(c) Conditional on $\boldsymbol{\omega}$, censoring is noninformative on $\boldsymbol{\omega}$.

Having augmented with $\mathbf{U}_0^{RC}$, $\mathbf{U}^{CAT}$ and $\mathbf{U}^{BIN}$, it then follows that the joint posterior distribution of parameters and augmented data

$$\boldsymbol{\psi} = \left(\boldsymbol{\omega}, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}\right)$$

is given by

$$
\begin{aligned}
p\left(\boldsymbol{\psi} \mid \mathbf{y}, \boldsymbol{\delta}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) & \\
\propto\, & p\left(\mathbf{y}, \boldsymbol{\delta} \mid \boldsymbol{\psi}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) p\left(\boldsymbol{\psi} \mid \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) \\
=\, & \frac{p\left(\mathbf{y}, \boldsymbol{\delta}, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN} \mid \boldsymbol{\omega}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)}{p\left(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN} \mid \boldsymbol{\omega}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)} \\
& \times \left(p\left(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN} \mid \boldsymbol{\omega}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) p\left(\boldsymbol{\omega} \mid \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)\right) \\
=\, & p\left(\mathbf{y}, \boldsymbol{\delta}, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN} \mid \boldsymbol{\omega}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) \\
& \times p\left(\boldsymbol{\omega} \mid \mathbf{R}_{22} = \mathbf{I}_{m_4}\right).
\end{aligned}
$$

By assumption (a), it follows that the prior distribution of $\boldsymbol{\omega}$, conditional on $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, is given by

$$p\left(\boldsymbol{\omega} \mid \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) = p\left(\mathbf{b}\right) p\left(\mathbf{a} \mid \mathbf{G}\right) p\left(\mathbf{G}\right) p\left(\mathbf{R} \mid \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) \left[\prod_{j=m_1+m_2+1}^{m_1+m_2+m_3} p\left(\boldsymbol{\tau}_j\right)\right].$$

Let $\mathbf{x}_i$ $(m \times p)$ and $\mathbf{z}_i$ $(m \times Nm)$ be the submatrices of $\mathbf{X}$ and $\mathbf{Z}$ associated with animal i. Then, by assumptions (b) and (c), it follows that

$$p\left(\mathbf{y}, \boldsymbol{\delta}, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN} \mid \boldsymbol{\omega}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)$$

is given, up to proportionality, by:

$$\prod_{i=1}^{n}\left[\prod_{j=m_1+1}^{m_1+m_2}\left[\mathbf{1}\left(u_{ij}>y_{ij}\right)\right]^{1-\delta_{ij}}\right]$$

$$\times\prod_{i=1}^{n}\left[\prod_{j=m_1+m_2+1}^{m_1+m_2+m_3}\left\{\sum_{k=1}^{K_j}\left[\mathbf{1}\left(\tau_{jk-1}<u_{ij}\leq\tau_{jk}\right)\mathbf{1}\left(y_{ij}=k\right)\right]\right\}\right]$$

$$\times\prod_{i=1}^{n}\left[\prod_{j=m_1+m_2+m_3+1}^{m}\left[\mathbf{1}\left(u_{ij}\leq 0\right)\mathbf{1}\left(y_{ij}=0\right)+\mathbf{1}\left(0<u_{ij}\right)\mathbf{1}\left(y_{ij}=1\right)\right]\right]$$

$$\times\prod_{i=1}^{n}\left[(2\pi)^{-m/2}\left|\mathbf{R}\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathbf{u}_i-\mathbf{x}_i\mathbf{b}-\mathbf{z}_i\mathbf{a}\right)'\mathbf{R}^{-1}\left(\mathbf{u}_i-\mathbf{x}_i\mathbf{b}-\mathbf{z}_i\mathbf{a}\right)\right\}\right].$$

(Here the convention is adopted that, *e.g.*,

$$\left[\mathbf{1}\left(u_{ij}>y_{ij}\right)\right]^{0}=1\quad\text{and}\quad\left[\mathbf{1}\left(u_{ij}>y_{ij}\right)\right]^{1}=\left[\mathbf{1}\left(u_{ij}>y_{ij}\right)\right]).$$

In the model associated with (2) the joint posterior is derived similarly, with obvious modifications.

2.4. Marginal posterior distributions, Gibbs sampling and fully conditional posterior distributions

From the joint posterior distribution of $\boldsymbol{\psi}$, the marginal posterior distribution of $\boldsymbol{\varphi}$, a single parameter or a subset of parameters of $\boldsymbol{\psi}$, can be obtained integrating out all the other parameters, $\boldsymbol{\psi}_{\backslash\varphi}$, including the augmented data. The notation $\boldsymbol{\psi}_{\backslash\varphi}$ denotes $\boldsymbol{\psi}$ excluding $\boldsymbol{\varphi}$. Here, we wish to obtain samples from the joint posterior distribution of $\boldsymbol{\omega}=\left(\mathbf{b},\mathbf{a},\mathbf{G},\mathbf{R},\tau_{m_1+m_2+1},\ldots,\tau_{m_1+m_2+m_3}\right)$ conditional on $\mathbf{R}_{22}=\mathbf{I}_{m_4}$. One possible implementation of the Gibbs sampler is as follows: Given an arbitrary starting value $\boldsymbol{\psi}^{(0)}$, then $(\mathbf{b},\mathbf{a})^{(1)}$ is generated from the fully conditional posterior distribution of $(\mathbf{b},\mathbf{a})$ given data, $(\mathbf{y},\boldsymbol{\delta})$, $\boldsymbol{\psi}_{\backslash(\mathbf{b},\mathbf{a})}$ and $\mathbf{R}_{22}=\mathbf{I}_{m_4}$. Superscript $^{(1)}$ (and later $^{(t)}$) refer to the sampling round of the implemented Gibbs sampler. Next, $\left(\mathbf{u}_0^{RC},\mathbf{u}^{CAT},\mathbf{u}^{BIN}\right)^{(1)}$ is generated from the fully conditional posterior distribution of $\left(\mathbf{U}_0^{RC},\mathbf{U}^{CAT},\mathbf{U}^{BIN}\right)$ given data, $\boldsymbol{\psi}_{\backslash(\mathbf{U}_0^{RC},\mathbf{U}^{CAT},\mathbf{U}^{BIN})}$ and $\mathbf{R}_{22}=\mathbf{I}_{m_4}$, and so on up to $\tau_{m_1+m_2+m_3,K_{m_1+m_2+m_3}-2}^{(1)}$, which is generated from the fully conditional posterior distribution of $\tau_{m_1+m_2+m_3,K_{m_1+m_2+m_3}-2}$ given data, $(\mathbf{y},\boldsymbol{\delta})$, $\boldsymbol{\psi}_{\backslash\left(\tau_{K_{m_1+m_2+m_3}-2}\right)}$ and $\mathbf{R}_{22}=\mathbf{I}_{m_4}$. This completes one cycle of the Gibbs sampler. After t cycles (t large) Geman and Geman [10] showed that $\boldsymbol{\psi}^{(t)}$, under mild conditions, can be viewed as a sample from the joint posterior distribution of $\boldsymbol{\psi}$ conditional on $\mathbf{R}_{22}=\mathbf{I}_{m_4}$.

The fully conditional posterior distributions that define one possible implementation of the Gibbs sampler are: Let $\boldsymbol{\theta} = \left(\mathbf{b}', \mathbf{a}'\right)'$, $\mathbf{W} = (\mathbf{X}, \mathbf{Z})$, and
$\mathbf{D}^{-1} = \begin{pmatrix} \mathbf{I}_{p_1}\left(\sigma_1^2\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_2}\left(\sigma_2^2\right)^{-1} \end{pmatrix}$; then

$$\boldsymbol{\theta}\mid\left((\mathbf{y}, \boldsymbol{\delta}), \boldsymbol{\psi}_{\backslash\boldsymbol{\theta}}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) \sim N_{p+Nm}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Lambda}_{\boldsymbol{\theta}}\right),$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = \boldsymbol{\Lambda}_{\boldsymbol{\theta}}\mathbf{W}'\left(\mathbf{I}_n \otimes \mathbf{R}\right)^{-1}\mathbf{u} \tag{3}$$

and

$$\begin{aligned}
\boldsymbol{\Lambda}_{\boldsymbol{\theta}}^{-1} &= \begin{pmatrix} \mathbf{X}'\left(\mathbf{I}_n \otimes \mathbf{R}\right)^{-1}\mathbf{X} + \mathbf{D}^{-1} & \mathbf{X}'\left(\mathbf{I}_n \otimes \mathbf{R}\right)^{-1}\mathbf{Z} \\ \mathbf{Z}'\left(\mathbf{I}_n \otimes \mathbf{R}\right)^{-1}\mathbf{X} & \mathbf{Z}'\left(\mathbf{I}_n \otimes \mathbf{R}\right)^{-1}\mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{pmatrix} \\
&= \mathbf{W}'\left(\mathbf{I}_n \otimes \mathbf{R}\right)^{-1}\mathbf{W} + \begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{pmatrix}.
\end{aligned} \tag{4}$$

Define $\mathbf{a}_M$ as the $N \times m$ matrix, where the jth row is $\mathbf{a}_j'$, $j = 1, \ldots, N$. Then,

$$\mathbf{G}\mid\left((\mathbf{y}, \boldsymbol{\delta}), \boldsymbol{\psi}_{\backslash\mathbf{G}}\right) \sim IW_m\left(\left[\boldsymbol{\Sigma}_{\mathbf{G}}^{-1} + \mathbf{a}_M'\mathbf{A}^{-1}\mathbf{a}_M\right]^{-1}, f_{\mathbf{G}} + N\right),$$

and the fully conditional posterior distribution of $\mathbf{R}$ conditional on data, $\boldsymbol{\psi}_{\backslash\mathbf{R}}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ is obtained from

$$\begin{aligned}
&\mathbf{R}\mid\left((\mathbf{y}, \boldsymbol{\delta}), \boldsymbol{\psi}_{\backslash\mathbf{R}}\right) \\
&\sim IW_m\left(\left[\boldsymbol{\Sigma}_{\mathbf{R}}^{-1} + \sum_{i=1}^n \left(\mathbf{u}_i - \mathbf{x}_i\mathbf{b} - \mathbf{z}_i\mathbf{a}\right)\left(\mathbf{u}_i - \mathbf{x}_i\mathbf{b} - \mathbf{z}_i\mathbf{a}\right)'\right]^{-1}, f_{\mathbf{R}} + n\right)
\end{aligned} \tag{5}$$

by conditioning on $\mathbf{R}_{22} = \mathbf{I}_{m_4}$.

The following notation will be used for augmented data of the animal i: $\mathbf{U}_i^{aug}$ is the vector of those $U_{ij}'s$ where j is the index of a censored observation $\left(\delta_{ij} = 0\right)$ from a right censored Gaussian trait, an ordered categorical or a binary trait. Therefore, $\mathbf{U}_i^{aug}$ may differ in dimension for different animals, depending on whether the observations for the right censored Gaussian traits are censored values. The dimension of $\mathbf{U}_i^{aug}$ is n_i^{aug}. The fully conditional posterior distribution of $\mathbf{U}_i^{aug}$ given data, $\boldsymbol{\psi}_{\backslash\mathbf{U}_i^{aug}}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ follows a

truncated n_i^{aug}-dimensional multivariate normal distribution on the interval

$$\prod_{j=m_1+1}^{m_1+m_2} \left[\mathbf{1}\left(u_{ij} > y_{ij}\right)\right]^{1-\delta_{ij}} \tag{6}$$

$$\times \prod_{j=m_1+m_2+1}^{m_1+m_2+m_3} \left\{ \sum_{k=1}^{K_j} \left[\mathbf{1}\left(\tau_{jk-1} < u_{ij} \leq \tau_{jk}\right) \mathbf{1}\left(y_{ij} = k\right)\right] \right\}$$

$$\times \prod_{j=m_1+m_2+m_3+1}^{m} \left[\mathbf{1}\left(u_{ij} \leq 0\right) \mathbf{1}\left(y_{ij} = 0\right) + \mathbf{1}\left(0 < u_{ij}\right) \mathbf{1}\left(y_{ij} = 1\right)\right].$$

The mean and variance of the corresponding normal distribution before truncation are given by

$$\left(\mathbf{x}_{i(aug)}\mathbf{b} + \mathbf{z}_{i(aug)}\mathbf{a}\right) + \mathbf{R}_{i(aug)(obs)}\mathbf{R}_{i(obs)}^{-1}\left(\mathbf{u}_{i(obs)} - \left(\mathbf{x}_{i(obs)}\mathbf{b} + \mathbf{z}_{i(obs)}\mathbf{a}\right)\right) \tag{7}$$

and

$$\mathbf{R}_{i(aug)} - \mathbf{R}_{i(aug)(obs)}\mathbf{R}_{i(obs)}^{-1}\mathbf{R}_{i(obs)(aug)}, \tag{8}$$

respectively. $\mathbf{x}_{i(obs)}$ and $\mathbf{x}_{i(aug)}$ are the $n_i^{obs} \times p$ and $n_i^{aug} \times p$ dimensional submatrices of $\mathbf{x}_i$ containing the rows associated with observed and uncensored continuous traits, and those associated with the augmented data of animal i, respectively. Similar definitions are given for $\mathbf{z}_{i(obs)}$ and $\mathbf{z}_{i(aug)}$. The dimension of observed and uncensored Gaussian traits, $\mathbf{u}_i^{obs}$, is $n_i^{obs} = m - n_i^{aug}$. $\mathbf{R}_{i(aug)}$ is $n_i^{aug} \times n_i^{aug}$ and is the part of $\mathbf{R}$ associated with augmented data of animal i. Similar definitions are given for $\mathbf{R}_{i(aug)(obs)}$, $\mathbf{R}_{i(obs)}$ and $\mathbf{R}_{i(obs)(aug)}$.

The fully conditional posterior distribution of τ_{jk} for $k = 2, \ldots, K_j - 2$ is uniform on the interval

$$\left[\max\left\{\max\left\{u_{ij} : y_{ij} = k\right\}, \tau_{jk-1}\right\}; \min\left\{\min\left\{u_{ij} : y_{ij} = k+1\right\}, \tau_{jk+1}\right\}\right],$$

for $j = m_1 + m_2 + 1, \ldots, m_1 + m_2 + m_3$.

Detailed derivations of the fully conditional posterior distributions can be found in, *e.g.*, [15].

In the model associated with (2) the fully conditional posterior distribution of the residual covariance matrix is also conditional inverse Wishart distributed, however the conditioning is on $(R_{kk} = 1)_{k=m-m_4+1,\ldots,m}$.

3. MODEL INCLUDING MISSING DATA

In this section allowance is made for missing data. First the notation is extended to deal with missing data. Let $\mathbf{J}(i) = (J_1(i), \ldots, J_m(i))'$ be the vector of response indicator random variables on animal i defined by $J_k(i) = 1$ if the kth trait is observed on animal i and $J_k(i) = 0$ otherwise, $k = 1, \ldots, m$. The observed data on animal i is $(\mathbf{y}_i, \boldsymbol{\delta}_i)_{\mathbf{J}(i)}$, where $(\mathbf{y}_i, \boldsymbol{\delta}_i)_{\mathbf{J}(i)}$ denotes the observed Gaussian, observed right censored Gaussian traits, with their censoring indicators, observed categorical and binary traits of animal i. An animal with a record is now defined as an animal with at least one of m traits observed of the Gaussian, right censored Gaussian, ordered categorical or binary traits. The vector of observed $y's$ of animal i is $\mathbf{y}_{i(obs)} = (\mathbf{y}_i)_{\mathbf{J}(i)}$, with $1 \leq \dim\left(\mathbf{y}_{i(obs)}\right) \leq m$. Data on all animals are $(\mathbf{y}, \boldsymbol{\delta})_{\mathbf{J}}$, where $\mathbf{J} = (\mathbf{J}(i))_{i=1,\ldots,n}$.

For missing data, the idea of augmenting with residuals [32] is invoked. It is assumed that

$$
\begin{pmatrix} \mathbf{U}_{i(obs)} \\ \mathbf{U}_{i(aug)} \\ \mathbf{E}_{i(mis)} \end{pmatrix} \mid \left(\mathbf{b}, \mathbf{a}, \mathbf{R}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)
$$

$$
\sim N_m \left(\left(\begin{pmatrix} \left(\mathbf{x}_{i(obs)}\mathbf{b} + \mathbf{z}_{i(obs)}\mathbf{a}\right) \\ \left(\mathbf{x}_{i(aug)}\mathbf{b} + \mathbf{z}_{i(aug)}\mathbf{a}\right) \\ \mathbf{0} \end{pmatrix} \right), \begin{pmatrix} \mathbf{R}_{i(obs)} & \mathbf{R}_{i(obs)(aug)} & \mathbf{R}_{i(obs)(mis)} \\ \mathbf{R}_{i(aug)(obs)} & \mathbf{R}_{i(aug)} & \mathbf{R}_{i(aug)(mis)} \\ \mathbf{R}_{i(mis)(obs)} & \mathbf{R}_{i(mis)(aug)} & \mathbf{R}_{i(mis)} \end{pmatrix} \right).
$$

The dimensions of $\mathbf{U}_{i(obs)}$, $\mathbf{U}_{i(aug)}$ and $\mathbf{E}_{i(mis)}$ are n_i^{obs}, n_i^{aug} and n_i^{mis}, respectively, and $m = n_i^{obs} + n_i^{aug} + n_i^{mis}$. $\mathbf{U}_{i(obs)}$ is associated with observed and uncensored Gaussian traits, $\mathbf{U}_{i(aug)}$ is associated with augmented data of observed, censored right censored Gaussian and observed ordered categorical and binary traits. $\mathbf{E}_{i(mis)}$ is associated with residuals on the Gaussian scale of traits missing on animal i. The following will be assumed concerning the missing data pattern:

(d) Conditional on $\boldsymbol{\omega}$, data are missing at random, in the sense that $\mathbf{J}$ is stochastically independent of $(\mathbf{U}, \mathbf{C})$ conditional on $\boldsymbol{\omega}$.

(e) Conditional on $\boldsymbol{\omega}$, $\mathbf{J}$ is noninformative of $\boldsymbol{\omega}$.

Under the assumptions (a)–(e), and having augmented with $\mathbf{U}_{i(aug)}$ and $\mathbf{E}_{i(mis)}$ for all animals (*i.e.* with $\left(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}, \mathbf{E}^{MIS}\right)$), it then follows that the joint posterior distribution of parameters and augmented data

$\psi = \left(\omega, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}, \mathbf{E}^{MIS}\right)$ is given by:

$$p\left(\psi | (\mathbf{y}, \delta)_{\mathbf{J}}, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)$$

$$\propto p\left((\mathbf{y}, \delta)_{\mathbf{J}} | \psi, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) p\left(\psi | \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)$$

$$= p\left((\mathbf{y}, \delta)_{\mathbf{J}}, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}, \mathbf{E}^{MIS} | \omega, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) p\left(\omega | \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)$$

$$= p\left((\mathbf{y}, \delta)_{\mathbf{J}}, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}, \mathbf{E}^{MIS} | \omega, \mathbf{R}_{22} = \mathbf{I}_{m_4}\right) p\left(\omega | \mathbf{R}_{22} = \mathbf{I}_{m_4}\right)$$

$$\propto \prod_{i=1}^{n} \left[\prod_{j=m_1+1}^{m_1+m_2} \left(\left[\mathbf{1}\left(u_{ij} > y_{ij}\right) \right]^{1-\delta_{ij}} \right)^{J_j(i)} \right]$$

$$\times \prod_{i=1}^{n} \left[\prod_{j=m_1+m_2+1}^{m_1+m_2+m_3} \left\{ \sum_{k=1}^{K_j} \left[\mathbf{1}\left(\tau_{jk-1} < u_{ij} \leq \tau_{jk}\right) \mathbf{1}\left(y_{ij} = k\right) \right] \right\}^{J_j(i)} \right]$$

$$\times \prod_{i=1}^{n} \left[\prod_{j=m_1+m_2+m_3+1}^{m} \left[\mathbf{1}\left(u_{ij} \leq 0\right) \mathbf{1}\left(y_{ij} = 0\right) + \mathbf{1}\left(0 < u_{ij}\right) \mathbf{1}\left(y_{ij} = 1\right) \right]^{J_j(i)} \right]$$

$$\times \prod_{i=1}^{n} \left[(2\pi)^{-m/2} |\mathbf{R}|^{-1/2} \exp\left\{ -\frac{1}{2} \left(\mathbf{u}_i - \mathbf{x}_i \mathbf{b} - \mathbf{z}_i \mathbf{a}\right)' \mathbf{R}^{-1} \left(\mathbf{u}_i - \mathbf{x}_i \mathbf{b} - \mathbf{z}_i \mathbf{a}\right) \right\} \right],$$

where those rows of $\mathbf{x}_i$ and $\mathbf{z}_i$ associated with missing data are zero, and where u_{ij}, for j associated with missing data on animal i, is a residual, e_{ij}.

Deriving the fully conditional posterior distributions defining a Gibbs sampler proceeds as in the model with no missing data and with modifications according to the missing data pattern. (This is also true for the model associated with (2)).

Further details related to the derivation of the fully conditional posterior distributions can be found in, *e.g.*, [15].

4. STRATEGIES FOR IMPLEMENTATION OF THE GIBBS SAMPLER

Strategies for implementation are first outlined for the model associated with (1) for the case without missing data, and where, *a priori*, **b** conditional on σ_1^2 and σ_2^2 follows a multivariate normal distribution. The strategy is similar for the model associated with (2) except in obtaining samples from the fully conditional posterior of the residual covariance matrix.

4.1. Univariate sampling of location parameters

The fully conditional posterior distribution of θ given data, $\psi_{\backslash\theta}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ is $p + Nm$ dimensional multivariate normal distributed with mean

$\mu = \mu_\theta$ and covariance matrix $\Lambda = \Lambda_\theta$ given in (3) and (4) respectively. Let $\beta = (1, \dots, i-1, i+1, \dots, p+Nm)$, then using properties of the multivariate normal distribution and relationships between a matrix and its inverse, it follows, that the fully conditional posterior distribution of each element in θ is:

$$\theta_i | \left((\mathbf{y}, \delta) , \, \psi_{\backslash \theta_i}, \mathbf{R}_{22} = \mathbf{I}_{m_4} \right)$$
$$\sim N_1 \left(\mu_i + \Lambda_{i\beta} \Lambda_{\beta\beta}^{-1} \left(\theta_\beta - \mu_\beta \right) , \Lambda_{ii} - \Lambda_{i\beta} \Lambda_{\beta\beta}^{-1} \Lambda_{\beta i} \right)$$
$$= N_1 \left(C_{ii}^{-1} \left(r_i - \mathbf{C}_{i\beta} \theta_\beta \right) , C_{ii}^{-1} \right)$$

where r_i is the ith element of $\mathbf{r} = \mathbf{W}' \left(\mathbf{I} \otimes \mathbf{R}^{-1} \right) \mathbf{u}$ and $\mathbf{C} = \Lambda^{-1}$ is the coefficient matrix of the mixed model equations given by $\mathbf{C}\mu = \mathbf{r}$. The solution to these equations is $\mu = \Lambda \mathbf{r}$ and $\mathbf{C}_{i\beta} \theta_\beta = \mathbf{C}_i \theta - C_{ii} \theta_i$, where $\mathbf{C}_i$ is the ith row of the coefficient matrix and C_{ii} is the ith diagonal element.

4.2. Joint sampling of location parameters

Sampling univariately from the fully conditional posterior distribution of each location parameter in turn, may give poor mixing properties. García-Cortés and Sorensen [7] described a method to sample from the joint fully conditional posterior distribution of θ given data, $\psi_{\backslash \theta}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, that can avoid inverting the coefficient matrix $\mathbf{C} = \Lambda_\theta^{-1}$ of the mixed model equations. The idea behind this joint sampling scheme is that a linear combination of normally distributed random variables again is normally distributed and proceeds as follows: Let $\mathbf{b}_1^*$, $\mathbf{b}_2^*$, $\mathbf{a}^*$ and $\mathbf{e}^*$ be sampled independently from $N_{p_1} \left(\mathbf{0}, \mathbf{I}_{p_1} \sigma_1^2 \right)$, $N_{p_2} \left(\mathbf{0}, \mathbf{I}_{p_2} \sigma_2^2 \right)$, $N_{Nm} \left(\mathbf{0}, \mathbf{A} \otimes \mathbf{G} \right)$ and $N_{nm} \left(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{R} \right)$ distributions, respectively. Next let $\mathbf{b}^* = \left(\mathbf{b}_1^{*\prime}, \mathbf{b}_2^{*\prime} \right)'$ and $\theta^* = \left(\mathbf{b}^{*\prime}, \mathbf{a}^{*\prime} \right)'$ and define $\mathbf{u}^*$ as $\mathbf{W}\theta^* + \mathbf{e}^*$, then it follows that the linear combination of θ^* and $\mathbf{e}^*$ given by:

$$\theta^* + \Lambda_\theta \mathbf{W}' \left(\mathbf{I}_n \otimes \mathbf{R}^{-1} \right) \left(\mathbf{u} - \mathbf{u}^* \right)$$
$$= \Lambda_\theta \mathbf{W}' \left(\mathbf{I}_n \otimes \mathbf{R}^{-1} \right) \mathbf{u} + \left(\mathbf{I}_p - \Lambda_\theta \mathbf{W}' \left(\mathbf{I}_n \otimes \mathbf{R}^{-1} \right) \mathbf{W} \right) \theta^* - \Lambda_\theta \mathbf{W}' \left(\mathbf{I}_n \otimes \mathbf{R}^{-1} \right) \mathbf{e}^*$$

follows a $N_{p+Nm} \left(\mu_\theta, \Lambda_\theta \right)$-distribution. This is the fully conditional posterior distribution of location parameters, θ, given data and $\psi_{\backslash \theta}$. That is, having sampled θ^* and $\mathbf{e}^*$, then $\tilde{\theta} = \Lambda_\theta \mathbf{W}' \left(\mathbf{I}_n \otimes \mathbf{R}^{-1} \right) \left(\mathbf{u} - \mathbf{u}^* \right)$ can be found solving a set of mixed model equations given by: $\Lambda_\theta^{-1} \tilde{\theta} = \mathbf{W}' \left(\mathbf{I}_n \otimes \mathbf{R}^{-1} \right) \left(\mathbf{u} - \mathbf{u}^* \right)$. Finally θ^* is added to $\tilde{\theta}$ and the resulting value, $\theta^* + \tilde{\theta}$, is a sampled vector from the fully conditional posterior distribution of θ given data, $\psi_{\backslash \theta}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$.

4.3. Sampling of augmented data

The fully conditional posterior distribution of augmented Gaussian traits,

$$\left(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN} \right)$$

given data, $\boldsymbol{\psi}_{\backslash(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN})}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ will be sampled jointly. The dimension of $\left(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}\right)$ is $\sum_{i=1}^{n} n_i^{aug}$. Realising that $\mathbf{U}_i^{aug\prime}s$ of different animals are independent conditional on "fixed" and random effects, it follows that joint sampling of augmented Gaussian traits can be decomposed into n steps. One step is to sample from the fully conditional posterior distribution of $\mathbf{U}_i^{aug}$ given $(\mathbf{y}_i, \boldsymbol{\delta}_i)$, $\boldsymbol{\omega}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$. This is a n_i^{aug}-dimensional multivariate truncated Gaussian distribution on the interval given in (6). Before truncation, mean and variance are given by (7) and (8), respectively.

Let $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}$ be shorthand notation for the mean and variance of the fully conditional posterior distribution of $\mathbf{U}_i^{aug}$ before truncation. Then first u_{i1}^{aug} is sampled from a $N_1(\xi_1, \Sigma_{11})$-distribution, truncated at the relevant interval. Next u_{i2}^{aug} is sampled from the fully conditional posterior distribution of U_{i2}^{aug} given $U_{i1}^{aug} = u_{i1}^{aug}$; this is from a truncated $N_1\left(\xi_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(u_{i1}^{aug} - \xi_1\right), \Sigma_{22\cdot1}\right)$ -distribution. Finally, proceeding in this way, $u_{in_i^{aug}}^{aug}$ is sampled from a truncated univariate normal distribution with mean and variance before truncation given by

$$\xi_{n_i^{aug}} + \boldsymbol{\Sigma}_{n_i^{aug}\left(1:n_i^{aug}-1\right)}\boldsymbol{\Sigma}_{\left(1:n_i^{aug}-1\right)}^{-1}\left[\begin{pmatrix} u_{i1}^{aug} \\ \vdots \\ u_{i\left(n_i^{aug}-1\right)}^{aug} \end{pmatrix} - \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_{\left(n_i^{aug}-1\right)} \end{pmatrix}\right]$$

and

$$\Sigma_{n_i^{aug}n_i^{aug}} - \boldsymbol{\Sigma}_{n_i^{aug}\left(1:n_i^{aug}-1\right)}\boldsymbol{\Sigma}_{\left(1:n_i^{aug}-1\right)}^{-1}\boldsymbol{\Sigma}_{\left(1:n_i^{aug}-1\right)n_i^{aug}}$$

respectively.

Different ways can be chosen to sample from a univariate truncated $N_1(\mu, \sigma^2)$-distribution on the interval $I = \,]s_1; s_2]$. One possibility is to sample independently from the untruncated $N_1(\mu, \sigma^2)$-distribution and then only accept sampled values that belong to the interval I. Let $Y \sim N_1(\mu, \sigma^2)$, if $P(Y \in I)$ is very small this procedure is inefficient. The following procedure (e.g. [6]) that avoids rejections is implemented. First x is sampled from a $R(0, 1)$-distributed random variable, X. Let F_Y denote the distribution function of Y, then z given by:

$$z = F_Y^{-1}\Big(F_Y(s_1) + x\big(F_Y(s_2) - F_Y(s_1)\big)\Big)$$

is a realised value from the truncated $N_1(\mu, \sigma^2)$-distribution on I. The proof follows from (9) given below, where Z is the random variable from which z is

generated; z is a value between s_1 and s_2:

$$P\left(Z \leq z\right) = P\left(F_Y^{-1}\left[F_Y\left(s_1\right) + X\left(F_Y\left(s_2\right) - F_Y\left(s_1\right)\right)\right] \leq z\right) \tag{9}$$
$$= P\left(F_Y\left(s_1\right) + X\left(F_Y\left(s_2\right) - F_Y\left(s_1\right)\right) \leq F_Y\left(z\right)\right)$$
$$= P\left(X \leq \frac{F_Y\left(z\right) - F_Y\left(s_1\right)}{F_Y\left(s_2\right) - F_Y\left(s_1\right)}\right)$$
$$= \frac{F_Y\left(z\right) - F_Y\left(s_1\right)}{F_Y\left(s_2\right) - F_Y\left(s_1\right)} .$$

4.4. Sampling of covariance matrices

The strategy, for obtaining samples from the fully conditional posterior of the residual covariance matrix in the model associated with (1), is presented in Section 4.4.1. For the model associated with (2), the strategy is slightly different and is presented in Section 4.4.2.

4.4.1. Model associated with (1)

The fully conditional posterior distribution of the residual covariance matrix, $\mathbf{R}$, of $\mathbf{U}_i$, is conditional inverse Wishart distributed. The conditioning is on a block diagonal submatrix, $\mathbf{R}_{22}$, equal to the identity matrix of the inverse Wishart distributed matrix, $\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{pmatrix}$. Note that if the number of binary traits is equal to zero, the fully conditional posterior of $\mathbf{R}$ is inverse Wishart distributed. In order to obtain samples from the conditional inverse Wishart distribution, the approach described in [16] is implemented. The method relies on well-known relationships between a partitioned matrix and its inverse, and properties of Wishart distributions. The method is as follows: Let $\mathbf{R} \sim IW_m\left(\mathbf{\Sigma}, f\right)$ and let $\mathbf{V} = \mathbf{R}^{-1}$, where $\mathbf{V}$ by definition is Wishart distributed, $\mathbf{V} \sim W_m\left(\mathbf{\Sigma}, f\right)$. Next $\mathbf{R}$ is expressed in terms of $\mathbf{V}$: $\mathbf{R} = \begin{pmatrix} \mathbf{V}_{11}^{-1} + \left(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\right)\mathbf{V}_{22\cdot1}^{-1}\left(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\right)' & -\left(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\right)\mathbf{V}_{22\cdot1} \\ -\mathbf{V}_{22\cdot1}\left(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\right)' & \mathbf{V}_{22\cdot1}^{-1} \end{pmatrix}$, where $\mathbf{V}_{22\cdot1} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} = \mathbf{R}_{22}^{-1}$. From properties of the Wishart distribution, it is known that $\mathbf{V}_{11} \sim W_{m-m_4}\left(\mathbf{\Sigma}_{11}, f\right)$, $\left(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\right)|\mathbf{V}_{11} = \mathbf{v}_{11} \sim N_{(m-m_4)\times m_4}\left(\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}, \mathbf{v}_{11}^{-1} \otimes \mathbf{\Sigma}_{22\cdot1}\right)$, where $\mathbf{\Sigma}_{22\cdot1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ and that $\mathbf{V}_{22\cdot1} \sim W_{m_4}\left(\mathbf{\Sigma}_{22\cdot1}, f - (m - m_4)\right)$. Furthermore $\left(\mathbf{V}_{11}, \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\right)$ is stochastically independent of $\mathbf{V}_{22\cdot1}$. Realising that $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, is equivalent to $\mathbf{V}_{22\cdot1} = \mathbf{I}_{m_4}$, it follows that a matrix sampled from the conditional inverse Wishart distribution of $\mathbf{R}$ given $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ can be obtained in the following way: First $\mathbf{v}_{11}$ is sampled from the marginal distribution of $\mathbf{V}_{11}$. Next $\mathbf{t}_2$ is sampled from the conditional distribution of $\left(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\right)$ given $\mathbf{V}_{11} = \mathbf{v}_{11}$.

The matrix $\mathbf{r} = \begin{pmatrix} \mathbf{v}_{11}^{-1} + \mathbf{t}_2\mathbf{t}_2' & -\mathbf{t}_2 \\ -\mathbf{t}_2' & \mathbf{I}_{m_4} \end{pmatrix}$ is then a realised matrix from the conditional inverse Wishart distribution of $\mathbf{R}$ given $\mathbf{R}_{22} = \mathbf{I}_{m_4}$.

In order to obtain samples from a Wishart distribution, the algorithm of Odell and Feiveson [21] is implemented. The basic idea in their algorithm can be summarised as follows: Let $\mathbf{V} \sim W_m(\boldsymbol{\Sigma}, f)$ and let $\mathbf{LL}'$ be a Cholesky factorisation of $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\Sigma} = \mathbf{LL}'$. A realised matrix, $\mathbf{v}$, can be generated from the distribution of $\mathbf{V}$, by sampling $\mathbf{w}$ from a $W_m(\mathbf{I}_m, f)$-distribution, then $\mathbf{v}$ given by $\mathbf{LwL}'$ is a realised matrix from the desired Wishart distribution.

Using successively the properties already given of the Wishart distribution, a realised matrix, $\mathbf{w}$, from $\mathbf{W} \sim W_m(\mathbf{I}_m, f)$ can be generated as follows:

w_{11} is sampled from $W_{11} \sim W_1(1, f) = \chi^2(f)$;

t_2 is sampled from $W_{11}^{-1}W_{12}|W_{11} = w_{11} \sim N_1\left(0, w_{11}^{-1}\right)$;

$w_{22\cdot 1}$ is sampled from a $W_1(1, f-1)$-distribution;

$\mathbf{w}_{22}$ given by $\begin{pmatrix} w_{11} & w_{11}t_2 \\ (w_{11}t_2)' & w_{22\cdot 1} + t_2'w_{11}t_2 \end{pmatrix}$ is then a realised matrix from the distribution of $\mathbf{W}_{22} \sim W_2(\mathbf{I}_2, f)$.

For $i = 3$ and up to m, the dimension of $\mathbf{W}$, we proceed as follows:

$\mathbf{t}_i$ is sampled from

$$\mathbf{T}_i = \mathbf{W}_{(i-1)(i-1)}^{-1}\mathbf{W}_{(1:i-1)i}|\mathbf{W}_{(i-1)(i-1)} = \mathbf{w}_{(i-1)(i-1)} \sim N_{i-1}\left(\mathbf{0}, \mathbf{w}_{(i-1)(i-1)}^{-1}\right).$$

$\mathbf{W}_{(1:i-1)i}$ is used as the notation for the $((i-1) \times 1)$-dimensional vector of elements $\left(W_{ji}\right)_{j=1,i-1}$ of $\mathbf{W}$ and $\mathbf{W}_{(i-1)(i-1)}$ is the $(i-1)$-dimensional square matrix of $\mathbf{W}$, with elements $\left(W_{jk}\right)_{j,k=1,i-1}$;

$w_{ii\cdot(i-1)}$ is sampled from a $W_1(1, f-(i-1)) = \chi^2(f-(i-1))$-distribution;

$\mathbf{w}_{ii}$ given by $\begin{pmatrix} \mathbf{W}_{(i-1)(i-1)} & \mathbf{W}_{(i-1)(i-1)}\mathbf{t}_i \\ \left(\mathbf{W}_{(i-1)(i-1)}\mathbf{t}_i\right)' & w_{ii\cdot(i-1)} + \mathbf{t}_i'\mathbf{W}_{(i-1)(i-1)}\mathbf{t}_i \end{pmatrix}$ is then a realised matrix from the distribution of $\mathbf{W}_{ii} \sim W_i(\mathbf{I}_i, f)$.

Finally, $\mathbf{w} = \mathbf{w}_{mm}$ is a realised matrix from the distribution of $\mathbf{W} \sim W_m(\mathbf{I}_m, f)$.

4.4.2. Model associated with (2)

In the following we outline a method for sampling from the fully conditional posterior distribution of $\mathbf{R}$ in the model associated with (2) for $m_4 \geq 1$. (Note, if the number of binary traits is equal to zero or one, $m_4 = 0$ or $m_4 = 1$, then the model associated with (2) is identical to the model described by (1). Thus for $m_4 = 1$ we end up with two different methods for obtaining samples from the fully conditional posterior distribution of $\mathbf{R}$). Now consider the partitioning of

R described earlier in this section. Because

$$p\big(\mathbf{r}|\,(R_{kk}=1)_{k=m-m_4+1,...,m}\big)$$
$$= p\big(\mathbf{r}|\mathbf{R}_{22}=\mathbf{r}_{22},\,(R_{kk}=1)_{k=m-m_4+1,...,m}\big)p\big(\mathbf{r}_{22}|\,(R_{kk}=1)_{k=m-m_4+1,...,m}\big)$$

it follows that a matrix $\tilde{\mathbf{r}}$, from the conditional distribution of **R** given $(R_{kk}=1)_{k=m-m_4+1,...,m}$, can be sampled as follows. First $\tilde{\mathbf{r}}_{22}$ is sampled from the conditional distribution of $\mathbf{R}_{22}$ given $(R_{kk}=1)_{k=m-m_4+1,...,m}$, *i.e.* from a conditional inverse Wishart distribution conditional on all diagonal elements equal to one (*i.e.* $\tilde{\mathbf{r}}_{22}$ is a correlation matrix). Second, $\mathbf{v}_{11}$ is sampled from the marginal distribution of $\mathbf{V}_{11}$ and $\mathbf{t}_2$ from the conditional distribution of $\big(\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\big)$ given $\mathbf{V}_{11}=\mathbf{v}_{11}$. Finally, the matrix $\tilde{\mathbf{r}} = \begin{pmatrix} \mathbf{v}_{11}^{-1}+\mathbf{t}_2\tilde{\mathbf{r}}_{22}\mathbf{t}_2' & -\mathbf{t}_2\tilde{\mathbf{r}}_{22}^{-1} \\ -\tilde{\mathbf{r}}_{22}^{-1}\mathbf{t}_2' & \tilde{\mathbf{r}}_{22} \end{pmatrix}$ is a realised matrix from the conditional inverse Wishart distribution of **R** given $(R_{kk}=1)_{k=m-m_4+1,...,m}$.

Obtaining samples from the fully conditional posterior of $\mathbf{R}_{22}$ given $(R_{kk}=1)_{k=m-m_4+1,...,m}$ is not trivial. Therefore, inspired by Chib and Greenberg [4], we suggest the following Metropolis-Hastings algorithm for obtaining samples from the fully conditional posterior distribution of **R**. Let $q_1\big(\mathbf{r}_{22}|\tilde{\mathbf{r}}_{22}\big)$ denote a density that generates candidate values, $\tilde{\tilde{\mathbf{r}}}_{22}$, *i.e.* candidate correlation matrices given the current value (correlation matrix), $\tilde{\mathbf{r}}_{22}$ (and $(\mathbf{y},\delta)$, $\boldsymbol{\psi}_{\backslash \mathbf{R}}$) (see *e.g.* [18] for generating random correlation matrices). As proposal density, $q\big(\mathbf{r}|\tilde{\mathbf{r}}\big)$, for generating candidate values, $\tilde{\tilde{\mathbf{r}}}$ (*i.e.* candidate covariance matrices given the current value/covariance matrix, $\tilde{\mathbf{r}}$) we suggest taking

$$q\big(\mathbf{r}|\tilde{\mathbf{r}}\big) = p(\mathbf{r}|\,(\mathbf{y},\delta)\,,\boldsymbol{\psi}_{\backslash \mathbf{R}},\mathbf{R}_{22}=\mathbf{r}_{22},\,(R_{kk}=1)_{k=m-m_4+1,...,m})q_1\big(\mathbf{r}_{22}|\tilde{\mathbf{r}}_{22}\big).$$

This results in the following algorithm:

1. Sample a proposal value, $\tilde{\tilde{\mathbf{r}}}_{22}$, from the density $q_1\big(\mathbf{r}_{22}|\tilde{\mathbf{r}}_{22}\big)$. Next sample $\mathbf{v}_{11}$ and $\mathbf{t}_2$ as described above (and with parameters given in (5)). Then the matrix $\tilde{\tilde{\mathbf{r}}} = \begin{pmatrix} \mathbf{v}_{11}^{-1}+\mathbf{t}_2\tilde{\tilde{\mathbf{r}}}_{22}\mathbf{t}_2' & -\mathbf{t}_2\tilde{\tilde{\mathbf{r}}}_{22}^{-1} \\ -\tilde{\tilde{\mathbf{r}}}_{22}^{-1}\mathbf{t}_2' & \tilde{\tilde{\mathbf{r}}}_{22} \end{pmatrix}$ is a realised matrix from $q\big(\mathbf{r}|\tilde{\mathbf{r}}\big)$.

2. Move to $\tilde{\tilde{\mathbf{r}}}$ with probability $\alpha\big(\tilde{\mathbf{r}},\tilde{\tilde{\mathbf{r}}}\big)$ given by

$$\alpha\big(\tilde{\mathbf{r}},\tilde{\tilde{\mathbf{r}}}\big) = \min\left\{ \frac{p\big(\tilde{\tilde{\mathbf{r}}}|\,(\mathbf{y},\delta)\,,\boldsymbol{\psi}_{\backslash \mathbf{R}},\,(R_{kk}=1)_{k=m-m_4+1,...,m}\big)\,q\big(\tilde{\mathbf{r}}|\tilde{\tilde{\mathbf{r}}}\big)}{p\big(\tilde{\mathbf{r}}|\,(\mathbf{y},\delta)\,,\boldsymbol{\psi}_{\backslash \mathbf{R}},\,(R_{kk}=1)_{k=m-m_4+1,...,m}\big)\,q\big(\tilde{\tilde{\mathbf{r}}}|\tilde{\mathbf{r}}\big)},\,1 \right\}$$

and stay at $\tilde{\mathbf{r}}$ with probability $1-\alpha\big(\tilde{\mathbf{r}},\tilde{\tilde{\mathbf{r}}}\big)$.

Note that with the suggested proposal density, then

$$\alpha\left(\tilde{\mathbf{r}}, \tilde{\tilde{\mathbf{r}}}\right)$$

$$= \min\left\{\frac{p\left(\tilde{\tilde{\mathbf{r}}}_{22} \mid \left((\mathbf{y}, \boldsymbol{\delta}), \boldsymbol{\psi}_{\backslash\mathbf{R}}, (R_{kk} = 1)_{k=m-m_4+1,\ldots,m}\right)\right) q_1\left(\tilde{\mathbf{r}}_{22} \mid \tilde{\tilde{\mathbf{r}}}_{22}\right)}{p\left(\tilde{\mathbf{r}}_{22} \mid \left((\mathbf{y}, \boldsymbol{\delta}), \boldsymbol{\psi}_{\backslash\mathbf{R}}, (R_{kk} = 1)_{k=m-m_4+1,\ldots,m}\right)\right) q_1\left(\tilde{\tilde{\mathbf{r}}}_{22} \mid \tilde{\mathbf{r}}_{22}\right)}, 1\right\}.$$

5. EXAMPLE

In order to illustrate the methodology, a simulated dataset was analysed. The simulated data and results are presented below.

5.1. Simulated data

The simulated data consist of records on five-thousand animals. First the complete data consisting of a Gaussian, a right censored Gaussian, an ordered categorical, and a binary trait are generated for each animal (described in detail below). Next the missing data pattern is generated independently of the random vector associated with the complete data.

The complete data are simulated as follows: First records at the normally distributed level of the model are generated. The animals are assumed to be located in one herd and to be offspring of fifty unrelated sires and five-thousand unrelated dams (all dams and sires of the animals with records are assumed to be mutually unrelated). The fifty 4-dimensional sire effects, $\mathbf{s}_l$, $l = 1, \ldots, 50$ are generated independently from a $N_4\left(\mathbf{0}, \mathbf{G}_S\right)$-distribution. The number of offspring per sire was 100 on average. Residuals were generated independently ($\mathbf{e}_i \sim N_4\left(\mathbf{0}, \mathbf{R}_S\right)$, $i = 1, \ldots, 5000$), so that the ith 4-dimensional normally distributed "record", u_i, is equal to:

$$\mathbf{u}_i = \boldsymbol{\mu}_H + \mathbf{s}_{l(i)} + \mathbf{e}_i$$

$i = 1, \ldots, 5000$ (these will be called the normally distributed data); where $\boldsymbol{\mu}_H' = (8000, 900, 0.5, -0.2562)$. The covariance matrices $\mathbf{G}_S$ and $\mathbf{R}_S$ are:

$$\begin{pmatrix} 108000 & 12728 & 1.6822 & -18.418 \\ 12728 & 9375 & -0.99124 & -6.9767 \\ 1.6822 & -0.99124 & 0.010481 & 0.001639 \\ -18.418 & -6.9767 & 0.001639 & 0.025639 \end{pmatrix}$$

and

$$\begin{pmatrix} 1332000 & 47272 & 22.888 & 54.875 \\ 47272 & 240620 & -19.484 & -43.658 \\ 22.888 & -19.484 & 0.15721 & -0.0016392 \\ 54.875 & -43.658 & -0.0016392 & 1 \end{pmatrix},$$

respectively. The complete data are generated by the following procedure:

$$Y_{i1} = U_{i1},$$

$$(Y_{i2}, \delta_i) = \begin{cases} (U_{i2}, 1) & \text{if } U_{i2} \leq 1300 \\ (1300, 0) & \text{if } U_{i2} > 1300 \end{cases},$$

$$Y_{i3} = \begin{cases} 1 & \text{if } U_{i3} \leq 0 \\ 2 & \text{if } 0 < U_{i3} \leq 0.3231 \\ 3 & \text{if } 0.3231 < U_{i3} \leq 0.6769 \\ 4 & \text{if } 0.6769 < U_{i3} \leq 1 \\ 5 & \text{if } U_{i3} > 1 \end{cases}$$

and

$$Y_{i4} = \begin{cases} 0 & \text{if } U_{i4} \leq 0 \\ 1 & \text{if } U_{i4} > 0 \end{cases}.$$

Finally, the pattern for the missing data is generated. Let $\mathbf{J}(i)$ denote the 4-dimensional vector of response indicator random variables of animal i. Then the missing data pattern is generated so that $\mathbf{J}(i)$, $i = 1, \ldots, 5000$ are independently and identically distributed with

$$P\left(\mathbf{J}(i) = \mathbf{j}\right) = \begin{cases} 3/4 & \text{if } \mathbf{j} = (1, 1, 1, 1) \\ 1/56 & \text{if } \mathbf{j} = (0, 0, 0, 1), (0, 0, 1, 0), \\ & \quad (0, 0, 1, 1), (0, 1, 0, 0), (0, 1, 0, 1), \\ & \quad (0, 1, 1, 0), (0, 1, 1, 1), (1, 0, 0, 0), \\ & \quad (1, 0, 0, 1), (1, 0, 1, 0), (1, 0, 1, 1), \\ & \quad (1, 1, 0, 0), (1, 1, 0, 1) \text{ or } (1, 1, 1, 0) \\ 0 & \text{otherwise} \end{cases}$$

It follows that three quarters of the animals are expected to have observed (or censored) values on all four traits.

5.2. Gibbs sampling implementation and starting values

The Gibbs sampler was run as a single chain with joint updating of location parameters. After discarding the first 40 000 rounds of the Gibbs sampler (burn-in), 10 000 samples of selected model parameters were saved with a sampling interval of 100. The Gibbs sampler was implemented with improper uniform prior distributions on elements of $\boldsymbol{\mu}_H$, and on (co)variance matrices $\mathbf{G}_S$ and $\mathbf{R}_S$. It was assumed that the vector of sire effects, conditional on $\mathbf{G}_S$, followed a $N_{4N}\left(\mathbf{0}, \mathbf{I}_N \otimes \mathbf{G}_S\right)$-distribution, with $N = 50$. Finally the two thresholds τ_2 and

τ_3 were *a priori* assumed to be distributed as order statistics from a uniform distribution in the interval [0, 1], as described in Section 2.2.

Starting values for the location parameters were found as the solution to the mixed model equations given by

$$
\begin{pmatrix} \mathbf{X}'\,(\mathbf{I}_n \otimes \mathbf{R}_S)^{-1}\,\mathbf{X} & \mathbf{X}'\,(\mathbf{I}_n \otimes \mathbf{R}_S)^{-1}\,\mathbf{Z} \\ \mathbf{Z}'\,(\mathbf{I}_n \otimes \mathbf{R}_S)^{-1}\,\mathbf{X} & \mathbf{Z}'\,(\mathbf{I}_n \otimes \mathbf{R})^{-1}\,\mathbf{Z} + \mathbf{I}_n^{-1} \otimes \mathbf{G}_S^{-1} \end{pmatrix} \boldsymbol{\mu}_\theta = \mathbf{W}'\,(\mathbf{I}_n \otimes \mathbf{R}_S)^{-1}\,\mathbf{u}^{(0)}
$$

with initial values for (co)variance matrices inserted and $\mathbf{u}^{(0)}$, being a vector of observed Gaussian traits and starting values for augmented data, given as follows: Let $\mathbf{u}^{(0)} = \left(\mathbf{u}_1^{(0)\prime}, \ldots, \mathbf{u}_n^{(0)\prime} \right)'$ with $\mathbf{u}_i^{(0)\prime} = \left(u_{i1}^{(0)}, u_{i2}^{(0)}, u_{i3}^{(0)}, u_{i4}^{(0)} \right)$, then

$$
u_{i1}^{(0)} = \begin{cases} y_{i1} & \text{if } J_1(i) = 1 \\ 0 & \text{if } J_1(i) = 0 \end{cases},
$$

$$
u_{i2}^{(0)} = \begin{cases} y_{i2} & \text{if } J_2(i) = 1 \ \text{and} \ \delta_i = 1 \\ y_{i2} + 0.2\sigma_{uo} & \text{if } J_2(i) = 1 \ \text{and} \ \delta_i = 0 \\ 0 & \text{if } J_2(i) = 0 \end{cases},
$$

where σ_{uo} is the standard deviation of uncensored ($J_2(i) = 1$ and $\delta_i = 1$) observations of trait 2,

$$
u_{i3}^{(0)} = \begin{cases} -0.05 & \text{if } J_3(i) = 1 \ \text{and} \ y_{i2} = 1 \\ \dfrac{\left(\tau_k^{(0)} - \tau_{k-1}^{(0)} \right)}{2} & \text{if } J_3(i) = 1 \ \text{and} \ y_{i2} = k, \quad k = 2, 3, 4 \\ 1.05 & \text{if } J_3(i) = 1 \ \text{and} \ y_{i2} = 5 \\ 0 & \text{if } J_3(i) = 0 \end{cases}
$$

and

$$
u_{i4}^{(0)} = \begin{cases} -0.5 & \text{if } J_4(i) = 1 \ \text{and} \ y_{i3} = 0 \\ 0.5 & \text{if } J_4(i) = 1 \ \text{and} \ y_{i3} = 1 \\ 0 & \text{if } J_4(i) = 0 \end{cases}
$$

where $\tau_2^{(0)}$ and $\tau_3^{(0)}$ are the initial values for non-fixed thresholds. Let n_k denote the number of observations in category k, $k = 1, \ldots, 5$, of the ordered categorical trait, then the initial values, $\tau_2^{(0)}$ and $\tau_3^{(0)}$, of τ_2 and τ_3 are obtained from the following equations

$$
\Phi\left(\frac{\tau_k^{(0)} - \mu}{\sigma} \right) - \Phi\left(\frac{\tau_{k-1}^{(0)} - \mu}{\sigma} \right) = \frac{n_k}{\sum_{j=1}^{5} n_j}, \quad k = 1, \ldots, 5,
$$

with $\tau_k^{(0)} = -\infty$, $\tau_1^{(0)} = 0$, $\tau_4^{(0)} = 1$ and $\tau_5^{(0)} = \infty$ (*i.e.* by equating observed and expected frequencies in a (generally simpler) model, where liabilities of the ordered categorical trait are assumed to be independent and identically distributed with mean μ and variance σ^2).

5.3. Post Gibbs analysis and results

For each selected parameter, ψ, let $\psi^{(1)}, \dots, \psi^{(n)}$ denote the saved sampled values from the Gibbs sampler. The marginal posterior mean, $\hat{\psi}_{PM}$, and variance, $\hat{\sigma}_{PSTD}^2$, were estimated by $\frac{1}{n}\sum_{i=1}^n \psi^{(i)}$ and $\frac{1}{(n-1)}\sum_{i=1}^n \left(\psi^{(i)} - \hat{\psi}_{PM}\right)^2$, respectively. The method of batching (*e.g.* [11]) was chosen for estimating Monte Carlo variance, *MCV*, and effective sample size, N_e. The saved sampled values were divided into B batches (here $B = 20$) of equal size, n_b (here $n_b = 500$). For each batch b, $b = 1, \dots, B$, the batch mean is given by $\hat{\psi}_{BM}^{(b)} = \frac{1}{n_b}\sum_{i=(k-1)n_b+1}^{kn_b} \psi^{(i)}$. Assuming that the estimators associated with the batch means are approximately independently and identically distributed, then because $\hat{\psi}_{PM} = \frac{1}{B}\sum_{b=1}^B \hat{\psi}_{BM}^{(b)}$, it follows that the $MCV = Var\left(\hat{\psi}_{PM}\right) = \frac{1}{B}Var\left(\hat{\psi}_{BM}^{(1)}\right)$. The variance of $\hat{\psi}_{BM}^{(1)}$, $Var\left(\hat{\psi}_{BM}^{(1)}\right)$, is estimated by $\frac{1}{(B-1)}\sum_{b=1}^B \left(\hat{\psi}_{BM}^{(b)} - \hat{\psi}_{PM}\right)^2$. The effective sample size, N_e, is calculated by $\hat{\sigma}_{PSTD}^2/MCV$ (follows by equating $\frac{1}{N_e}\hat{\sigma}_{PSTD}^2$ and *MCV*).

Convergence of the Gibbs sampler was assessed *via* trace plots. The rate of mixing of the Gibbs sampler was investigated estimating lag correlations (between saved sampled values) in a standard time series analysis. The rate of mixing was good for all parameters except for the two thresholds of the categorical trait (with 5 categories). Lag 30 correlations between saved sampled values of elements of $\mathbf{\mu}_H$, of $\mathbf{G}_S$ and $\mathbf{R}_S$, for genetic, residual and intraclass correlations were all numerically close to zero. The intraclass correlations, ρ_1, ρ_2, ρ_3 and ρ_4 are calculated by $\rho_i = \mathbf{G}_{Sii}/(\mathbf{G}_{Sii} + \mathbf{R}_{Sii})$, $i = 1, 2, 3, 4$. Lag 30 correlations between saved sampled values of τ_2 and τ_3 were 0.31 and 0.25, respectively. Summary statistics of selected parameters are shown in Table I.

Let c_p^ψ denote the pth quantile of the (empirical) marginal posterior distribution of ψ. For all of the parameters in Table I the value of ψ used for simulating data is included in the interval $\left[c_{0.025}^\psi; c_{0.975}^\psi\right]$ – except one of the thresholds.

Inferences concerning a subset of the parameters from the present Bayesian analysis were compared with those obtained using restricted maximum likelihood (REML). This comparison is restricted to the covariance matrices associated with "the normally distributed data". The normally distributed data were analysed using the Gibbs sampler and REML [22]. Burn-in, sampling interval and the number of saved sampled values for the Gibbs sampler implemented for

Table I. Mean, standard deviation, 0.025 and 0.975 quantiles ($c_{0.025}$ and $c_{0.975}$) of the marginal posterior distributions of selected parameters. N_e is the effective sample size. "True" refers to the values used in the sampling model for generating data.

Parameter	True	Mean	Std.	$c_{0.025}$	$c_{0.975}$	N_e
$\mathbf{G}_{S11}$	108000	104850	26355	63943	146150	8979
$\mathbf{G}_{S22}$	9375	12059	3308	6881	17135	5568
$\mathbf{G}_{S33}$	0.010481	0.0128	0.0033	0.0078	0.0179	6758
$\mathbf{G}_{S44}$	0.025639	0.0279	0.0100	0.0127	0.0429	12793
$\rho_{\mathbf{G}_{S21}}$	0.40	0.2613	0.1659	-0.0791	0.4937	27388
$\rho_{\mathbf{G}_{S31}}$	0.05	0.1181	0.1654	-0.2176	0.3520	12458
$\rho_{\mathbf{G}_{S32}}$	-0.10	-0.0578	0.1752	-0.3935	0.1991	4200
$\rho_{\mathbf{G}_{S41}}$	-0.35	-0.2933	0.1926	-0.6359	-0.0034	10211
$\rho_{\mathbf{G}_{S42}}$	-0.45	-0.6062	0.1618	-0.8671	-0.3561	8734
$\rho_{\mathbf{G}_{S43}}$	0.10	0.3706	0.1815	-0.0071	0.6196	10166
ρ_1	0.075	0.0743	0.0172	0.0466	0.1013	9099
ρ_2	0.0375	0.0479	0.0125	0.0278	0.0669	5329
ρ_3	0.0625	0.0795	0.0184	0.0500	0.1079	5801
ρ_4	0.025	0.0270	0.0093	0.0125	0.0411	12697
τ_2	0.3231	0.3403	0.0079	0.3255	0.3516	178
τ_3	0.6769	0.6878	0.0072	0.6737	0.6982	243

analysing the normally distributed data were 4000, 10, and 10 000 respectively. Again improper uniform prior distributions were assumed for elements of $\boldsymbol{\mu}_H$, and for (co)variance matrices $\mathbf{G}_S$ and $\mathbf{R}_S$. The results from this part of the analysis are shown in Table II.

REML estimates are joint mode estimates of the (joint) marginal posterior distribution of (co)variance matrices. If the (joint) marginal posterior distribution of (co)variance matrices is symmetric, then joint posterior mode estimates and marginal posterior mean estimates would be equal – except for numerical and/or Monte Carlo error. Based on "the normally distributed data", marginal posterior means and REML estimates of genetic correlations are remarkably close to each other. Marginal posterior mean estimates of intraclass correlations are all slightly higher, compared to the REML estimates. This is because the marginal posterior distributions of intraclass correlations are all skewed to the right; *i.e.* posterior mode estimates are expected to be lower compared to posterior mean estimates.

In conclusion, the Gibbs sampler implementation of the Bayesian analysis of the rather complicated data (model) shows satisfactory behaviour.

Table II. Mean, 0.025 and 0.975 quantiles ($c_{0.025}$ and $c_{0.975}$) of the marginal posterior distributions of selected parameters; and REML estimates for the same parameters. "True" refers to the values used in the sampling model for generating data.

Parameter	True	Mean	$c_{0.025}$	$c_{0.975}$	REML
$\rho_{\mathbf{G}_{S21}}$	0.40	0.2741	−0.0542	0.4962	0.2730
$\rho_{\mathbf{G}_{S31}}$	0.05	0.1215	−0.1946	0.3498	0.1244
$\rho_{\mathbf{G}_{S32}}$	−0.10	−0.0216	−0.3490	0.2247	−0.0225
$\rho_{\mathbf{G}_{S41}}$	−0.35	−0.2508	−0.5714	0.0108	−0.2504
$\rho_{\mathbf{G}_{S42}}$	−0.45	−0.4795	−0.7561	−0.2331	−0.4652
$\rho_{\mathbf{G}_{S43}}$	0.10	0.2578	−0.1056	0.5037	0.2569
ρ_1	0.0750	0.0749	0.0469	0.1007	0.068
ρ_2	0.0375	0.0488	0.0294	0.0673	0.045
ρ_3	0.0625	0.0762	0.0491	0.1026	0.069
ρ_4	0.025	0.0245	0.01297	0.0356	0.023

6. CONCLUSION

During the last decade, a major change of emphasis in animal breeding research has taken place. Rather than focusing singly on productivity, there is now an interest in understanding the complex biological and statistical interrelationships among traits related to product quality, disease resistance, behaviour and production. Addressing these problems requires the development of probability models which properly describe the underlying structures in the data perceived by the experimenter. These models are highly complex and often cannot be implemented *via* traditional methods. However an increase in computer power and the introduction of modern computer-based inference methods are making this implementation possible. In this paper we have developed and implemented a fully Bayesian analysis of Gaussian, right censored Gaussian, categorical and binary traits using the Gibbs sampler and data augmentation. The methodology was applied to analyse a simulated dataset and the results show that posterior distributions cover well the values of the parameters used in the simulations.

The computer programme (available upon request), which has been developed for models associated with (1), allows analyses based on models with several random effects, including maternal genetic effects. In the programme, it is possible to choose between univariate or joint sampling of all location parameters. Augmented data are sampled jointly, using the method of composition, from their truncated multivariate normal distribution. Covariance matrices are sampled from inverted or conditional inverted Wishart distributions depending on the absence or presence of binary traits, respectively. In most

applications of models including at least two binary traits, it is not reasonable to assume that the residuals of liabilities of the binary traits are independent, *i.e.* the model associated with (2) is to be preferred. The Gibbs sampler outlined for the model associated with (2) is almost identical to the one associated with (1); the only real difference is the Metropolis-Hastings step invoked for sampling the residual covariance matrix associated with the residuals of liabilities (this step has not yet been implemented in the programme).

REFERENCES

[1] Albert J.H., Chib S., Bayesian analysis of binary and polychotomous response data, J. Am. Stat. Assoc. 88 (1993) 669–679.

[2] Chan K.S., Asymptotic behavior of the Gibbs sampler, J. Am. Stat. Assoc. 88 (1993) 320–326.

[3] Chib S., Greenberg E., Markov Chain Monte Carlo simulation methods in econometrics, Econom. Theory 12 (1996) 409–431.

[4] Chib S., Greenberg E., Analysis of multivariate probit models, Biometrika 85 (1998) 347–361.

[5] Cox D.R., Snell E.J., Analysis of binary data, Chapman and Hall, London, 1989.

[6] Devroye L., Non-uniform random variate generation, Springer-Verlag, New York, 1986.

[7] García-Cortés L.A., Sorensen D., On a multivariate implementation of the Gibbs sampler, Genet. Sel. Evol. 28 (1996) 121–126.

[8] Gelfand A.E., Hills S.E., Racine-Poon A., Smith A.F.M., Illustration of Bayesian inference in normal data models using Gibbs sampling, J. Am. Stat. Assoc. 85 (1990) 972–985.

[9] Gelfand A.E., Smith A.F.M., Sampling-based approaches to calculating marginal densities, J. Am. Stat. Assoc. 85 (1990) 398–409.

[10] Geman S., Geman D., Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (1984) 721–741.

[11] Hastings W.K., Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.

[12] Jensen J., Bayesian analysis of bivariate mixed models with one continuous and one binary trait using the Gibbs sampler, in: Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Guelph, 7–12 August 1994, University of Guelph, Vol XVIII, pp. 333–336.

[13] Jensen J., Hohenboken W.D., Madsen P., Andersen B.B., Sire × nutrition interactions and genetic parameters for energy intake, production and efficiency of nutrient utilization in young bulls, heifers and lactating cows, Acta Agric. Scand. 45 (1995) 81–91.

[14] Jensen J., Wang C.S., Sorensen D.A., Gianola D., Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler, Acta Agric. Scand. 44 (1994) 193–201.

[15] Korsgaard I.R., Genetic analysis of survival data, Ph.D. Thesis, Department of Theoretical Statistics, University of Aarhus, 1997.

[16] Korsgaard I.R., Andersen A.H., Sorensen D., A useful reparameterisation to obtain samples from conditional inverse Wishart distributions, Genet. Sel. Evol. 31 (1999) 177–181.

[17] Mardia K.V., Kent J.T., Bibby J.M., Multivariate analysis, Academic Press, London, 1979.

[18] Marsaglia G., Olkin I., Generating correlation matrices, Siam J. Sci. Stat. Comput. 5 (1984) 470–475.

[19] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., Equations of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1091.

[20] Mood A.M., Graybill F.A., Boes D.C., Introduction to the Theory of Statistics, McGraw-Hill, NY, USA, 1974.

[21] Odell P.L., Feiveson A.H., A numerical procedure to generate a sample covariance matrix, Am. Stat. Ass. J. (1966) 199–203.

[22] Patterson H.D., Thompson R., Recovery of inter-block information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[23] Sorensen D.A., Wang C.S., Jensen J., Gianola D., Bayesian analysis of genetic change due to selection using Gibbs sampling, Genet. Sel. Evol. 26 (1994) 333–360.

[24] Sorensen D.A., Andersen S., Gianola D., Korsgaard I.R., Bayesian inference in threshold models using Gibbs sampling, Genet. Sel. Evol. 27 (1995) 229–249.

[25] Sorensen D.A., Gianola D., Korsgaard I.R., Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications, Acta Agric. Scand., Sect. A, Animal Sci. 48 (1998) 222–229.

[26] Tanner M.A., Wong W.H., The calculation of posterior distributions by data augmentation, J. Am. Stat. Assoc. 82 (1987) 528–540.

[27] Tobin J., Estimation of relationships for limited dependent variables, Econometrica 26 (1958) 24–36.

[28] Van Tassell C.P., Casella G., Pollak E.J., Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood, J. Dairy Sci. 78 (1995) 678–692.

[29] Van Tassell C.P., Van Vleck L.D., Gregory K.E., Bayesian analysis of twinning and ovulation rates using a multiple-trait threshold model and Gibbs sampling, J. Anim. Sci. 76 (1998) 2048–2061.

[30] Wang C.S., Rutledge J.J., Gianola D., Marginal inferences about variance components in a mixed linear model using Gibbs sampling, Genet. Sel. Evol. 25 (1993) 41–62.

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

[32] Wang C.S., Quaas R.L., Pollak E.J., Bayesian analysis of calving ease scores and birth weights, Genet. Sel. Evol. 29 (1997) 117–143.

[33] Wei G.C.G., Tanner M.A., Posterior computations for censored regression data, J. Am. Stat. Assoc. 85 (1990) 829–839.

[34] Zeger S.L., Karim M.R., Generalized linear models with random effects; a Gibbs sampling approach, J. Am. Stat. Assoc. 86 (1991) 79–86.

APPENDIX

The convention used for Wishart and inverted Wishart distributions follows Mardia *et al.* [17]. Let $\mathbf{M} \sim W_p\left(\mathbf{\Sigma}, f\right)$, the density of $\mathbf{M}$ is up to proportionality given by (for $\mathbf{\Sigma} > 0$ and $f \geq p$):

$$p\left(\mathbf{M}\right) \propto |\mathbf{\Sigma}|^{-\frac{f}{2}}\, |\mathbf{M}|^{\frac{(f-p-1)}{2}}\, \exp\left\{-\frac{1}{2}tr\left(\mathbf{\Sigma}^{-1}\mathbf{M}\right)\right\} .$$

The mean and variance of $\mathbf{M}$ are given by: $E\left(\mathbf{M}\right) = f\mathbf{\Sigma}$ and $Var\left(\mathbf{M}\right) = 2f\mathbf{\Sigma} \otimes \mathbf{\Sigma}$.

Let $\mathbf{U} = \mathbf{M}^{-1}$, then $\mathbf{U}$ is said to have an inverted Wishart distribution. The density of $\mathbf{U}$ is, up to proportionality, given by:

$$p\left(\mathbf{U}\right) \propto |\mathbf{\Sigma}|^{-\frac{f}{2}}\, |\mathbf{U}|^{-\frac{(f+p+1)}{2}}\, \exp\left\{-\frac{1}{2}tr\left(\mathbf{\Sigma}^{-1}\mathbf{U}^{-1}\right)\right\} .$$

The mean of $\mathbf{U}$ is given by $E\left(\mathbf{U}\right) = \mathbf{\Sigma}^{-1}/\left(f - p - 1\right)$ if $f \geq p + 2$.