

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

Inge Riis Korsgaard¹, Mogens Sandø Lund¹, Daniel Sorensen¹, Daniel Gianola², Per Madsen¹, Just Jensen¹. ¹Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, PO Box 50, DK-8830 Tjele
²Department 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 outlined. 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 the 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 between traits. The methodology has been implemented in generally available software.

1 Introduction

In a series of problems it has been demonstrated that using the Gibbs sampler in conjunction with data augmentation makes possible to obtain sampling-based estimates of analytically intractable densities. The basic idea behind the Gibbs sampler, and other sampling based approaches, is to construct a Markov chain with a tractable transition mechanism and having the desired density as its invariant distribution (Chan, 1993). 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 are intractable, it may be advantageous to use data augmentation, which as pointed out by Chib and Greenberg (1996) 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. Gelfand et al. (1990) and with attention to applications in animal breeding, by Jensen et al. (1994), Sorensen et al. (1994), Van Tassell et al. (1995) and Wang et al. (1993, 1994). Bayesian inference using Gibbs sampling in an ordered categorical threshold model was considered by Zeger

and Karim (1991), Albert and Chib (1993) and Sorensen et al. (1995). In censored Gaussian and ordered categorical threshold models, Gibbs sampling in conjunction with data augmentation (Tanner and Wong (1987) and Sorensen et al. (1998)) leads to fully conditional posterior distributions which are easy to sample from. This was demonstrated in Wei and Tanner (1990) for the tobit model (Tobin, 1958), 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 was given in Jensen (1994). This was extended to an ordered categorical threshold character by Wang et al. (1997), and to several Gaussian and ordered categorical threshold characters by Van Tassel et al. (1998).

The purpose of this paper is to present a fully Bayesian analysis of an arbitrary number of Gaussian, right censored Gaussian, ordered categorical (more than 2 categories) and binary traits. Allowances are made for unequal models and missing data. Furthermore, strategies for implementation are reviewed.

2 The model without missing data

2.1 The 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; $m_i \geq 0$, $i = 1, \dots, 4$. The total number of traits is $m = m_1 + m_2 + m_3 + m_4$. For example, in dairy cattle, if $m_i = 1$ for all i : Then 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 whether or not mastitis occurred in a given period. In general, data on animal i is $(\mathbf{y}_i, \delta_i)$, $i = 1, \dots, n$, where $\mathbf{y}_i = (y_{i1}, \dots, y_{im_1}, y_{im_1+1}, \dots, y_{im_1+m_2}, y_{im_1+m_2+1}, \dots, y_{im_1+m_2+m_3}, y_{im_1+m_2+m_3+1}, \dots, y_{im_1+m_2+m_3+m_4})$, and where δ_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 data on all animals with records is $(\mathbf{y}, \delta)$. The observed vector of Gaussian traits of animal i is $(y_{i1}, \dots, y_{im_1})$. For $j \in \{m_1 + 1, \dots, m_1 + m_2\}$, y_{ij} is the observed value of $Y_{ij} = \min\{U_{ij}, C_{ij}\}$, where U_{ij} is normally distributed and C_{ij} is time of censoring of the j 'th trait of animal i . The censoring indicator δ_{ij} is one iff U_{ij} is observed ($U_{ij} \leq C_{ij}$) and zero otherwise. Δ_{oj} and Δ_{1j} will denote the sets of animals with δ_{ij} equal to zero and one, respectively; $j = m_1 + 1, \dots, m_1 + m_2$. The observed vector of categorical traits with response in three or more categories is $(y_{im_1+m_2+1}, \dots, y_{im_1+m_2+m_3})$. The outcome y_{ij} , $j \in \{m_1 + m_2 + 1, \dots, m_1 + m_2 + m_3\}$, 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 $\tau_{jk-1} < U_{ij} \leq \tau_{jk}$; $k = 1, \dots, K_j$, where K_j ($K_j \geq 3$) is the number of categories and $-\infty = \tau_{j0} \leq \tau_{j1} \leq \dots \leq \tau_{jK_j-1} \leq \tau_{jK_j} = \infty$. The observed vector of binary traits is $(y_{im_1+m_2+m_3+1}, \dots, y_{im})$. 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 $U_{ij} \leq 0$ and $Y_{ij} = 1$ iff $U_{ij} > 0$.

Let $U_{ij} = Y_{ij}$ for $j = 1, \dots, m_1$, that is for the Gaussian traits, and let $\mathbf{U}_i = (U_{i1}, \dots, U_{im})'$ 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 $\mathbf{U} = (\mathbf{U}_i)_{i=1, \dots, n}$ as the nm -dimensional column vector containing the $\mathbf{U}_i$'s. It is assumed that:

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

where $\mathbf{b}$ is a p -dimensional vector of "fixed" effects. The vector $\mathbf{a}_i = (a_{i1}, \dots, a_{im})'$ represents the additive genetic values of $\mathbf{U}_i$, $i = 1, \dots, N$; $\mathbf{a} = (\mathbf{a}_i)_{i=1, \dots, N}$, 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 $\mathbf{R}$ is the residual covariance matrix of $\mathbf{U}_i$. The usual condition that $R_{kk} = 1$ (e.g. Cox and Snell (1989)) has been imposed in the conditional probit model of Y_{ik} given $\mathbf{b}$ and $\mathbf{a}$, $k = m - m_4 + 1, \dots, m$. Furthermore it is assumed that liabilities of the binary traits are conditionally independent, given $\mathbf{b}$ and $\mathbf{a}$.

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 $\mathbf{b}|\sigma_1^2, \sigma_2^2 \sim N_p \left(\mathbf{0}, \begin{pmatrix} \mathbf{I}_{p_1}\sigma_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_2}\sigma_2^2 \end{pmatrix} \right)$, 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 an improper uniform). The a priori distribution of the additive genetic values is $\mathbf{a}|\mathbf{G} \sim N_{Nm}(\mathbf{0}, \mathbf{A} \otimes \mathbf{G})$, where $\mathbf{G}$ is the $m \times m$ additive genetic covariance matrix of $\mathbf{U}_i$, $i = 1, \dots, N$. A priori, $\mathbf{G}$ is assumed to follow a m -dimensional inverted Wishart distribution: $\mathbf{G} \sim IW_m(\Sigma_{\mathbf{G}}, f_{\mathbf{G}})$. It is assumed that $\mathbf{R}$ conditional on $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ follows a conditional inverted Wishart distribution. The unconditional distribution of $\mathbf{R}$ is assumed to be inverted Wishart: $\mathbf{R} \sim IW_m(\Sigma_{\mathbf{R}}, f_{\mathbf{R}})$. All of $\Sigma_{\mathbf{G}}$, $f_{\mathbf{G}}$, $\Sigma_{\mathbf{R}}$ and $f_{\mathbf{R}}$ are assumed known. A priori, it is assumed, that the elements of $\tau_j = (\tau_{j2}, \dots, \tau_{jK_j-2})$ are distributed as order statistics from a uniform distribution in the interval $[\tau_{j1}; \tau_{jK_j-1}] = [0; 1]$, i.e.: $p(\tau_{j2}, \dots, \tau_{jK_j-2}) = (K_j - 3)! \mathbf{1}(\tau_j \in \mathfrak{S}_j)$,

where

$$\mathfrak{S}_j = \{(s_2, \dots, s_{K_j-2}) \mid 0 \leq s_2 \leq \dots \leq s_{K_j-2} \leq 1\}$$

(Mood et al. 1974).

Concerning prior independence, the following assumption is made:

- (a) A priori $\mathbf{b}$, $(\mathbf{a}, \mathbf{G})$, $\mathbf{R}$ and τ_j , $j = m_1 + m_2 + 1, \dots, m_1 + m_2 + m_3$ are mutually independent, and furthermore elements of $\mathbf{b}$ are mutually independent.

2.3 Joint posterior distribution

For each animal augmented variables are U'_{ij} s of censored ($\delta_{ij} = 0$) right censored Gaussian traits and with liabilities of ordered categorical and binary traits. The following notation will be used: $\mathbf{U}_0^{RC} = \{U_{ij} : i \in \Delta_{0j}; j = m_1 + 1, \dots, m_1 + m_2\}$, this is the set of U'_{ij} s of censored, 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 $\omega = (\mathbf{b}, \mathbf{a}, \mathbf{G}, \mathbf{R}, \tau_{m_1+m_2+1}, \dots, \tau_{m_1+m_2+m_3})$, i.e. $\mathbf{C} = (\mathbf{C}_i)_{i=1, \dots, n}$, where $\mathbf{C}_i = (C_{im_1+1}, \dots, C_{im_1+m_2})$ is the m_2 dimensional random vector of censoring times of animal i , is stochastically independent of $\mathbf{U}$ given ω .
- (c) Conditional on ω , censoring is noninformative on ω .

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

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

is given by

$$\begin{aligned} p(\psi | \mathbf{y}, \delta, \mathbf{R}_{22} = \mathbf{I}_{m_4}) &\propto p(\mathbf{y}, \delta | \psi, \mathbf{R}_{22} = \mathbf{I}_{m_4}) p(\psi | \mathbf{R}_{22} = \mathbf{I}_{m_4}) \\ &= p(\mathbf{y}, \delta, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN} | \omega, \mathbf{R}_{22} = \mathbf{I}_{m_4}) \\ &\quad \times p(\omega | \mathbf{R}_{22} = \mathbf{I}_{m_4}) \end{aligned}$$

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

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

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

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

up to proportionality, is given by:

$$\begin{aligned} & \prod_{i=1}^n \left[\prod_{j=m_1+1}^{m_1+m_2} [\mathbf{1}(u_{ij} > y_{ij})]^{1-\delta_{ij}} \right] \\ & \times \prod_{i=1}^n \left[\prod_{j=m_1+m_2+1}^{m_1+m_2+m_3} \left\{ \sum_{k=1}^{K_j} [\mathbf{1}(\tau_{jk-1} < u_{ij} \leq \tau_{jk}) \mathbf{1}(y_{ij} = k)] \right\} \right] \\ & \times \prod_{i=1}^n \left[\prod_{j=m_1+m_2+m_3+1}^m [\mathbf{1}(u_{ij} \leq 0) \mathbf{1}(y_{ij} = 0) + \mathbf{1}(0 < u_{ij}) \mathbf{1}(y_{ij} = 1)] \right] \\ & \times \prod_{i=1}^n \left[(2\pi)^{-m/2} |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{u}_i - \mathbf{x}_i \mathbf{b} - \mathbf{z}_i \mathbf{a})' \mathbf{R}^{-1} (\mathbf{u}_i - \mathbf{x}_i \mathbf{b} - \mathbf{z}_i \mathbf{a}) \right\} \right] \end{aligned}$$

(Here the convention is adopted that e.g. $[\mathbf{1}(u_{ij} > y_{ij})]^0 = 1$ and $[\mathbf{1}(u_{ij} > y_{ij})]^1 = [\mathbf{1}(u_{ij} > y_{ij})]$)

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

From the joint posterior distribution of ψ , marginal posterior distributions of φ , a single parameter of a subset of parameters of ψ , can be obtained integrating out all the other parameters, $\psi_{\setminus \varphi}$, including augmented data. Here $\psi_{\setminus \varphi}$ denotes ψ excluding φ . This integration is implicitly performed using Gibbs sampling, which is an iterative method for generating samples from a multivariate distribution, and has its roots in the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). Here we wish to obtain samples from the joint posterior distribution of $\omega = (\mathbf{b}, \mathbf{a}, \mathbf{G}, \mathbf{R}, \tau_{m_1+m_2+1}, \dots, \tau_{m_1+m_2+m_3})$. One possible implementation of the Gibbs sampler is: Given an arbitrary starting value $\psi^{(0)}$, $(\mathbf{b}, \mathbf{a})^{(1)}$ is generated from the fully conditional posterior distribution of $(\mathbf{b}, \mathbf{a})$ given data, $(\mathbf{y}, \delta)$, $\psi_{\setminus (\mathbf{b}, \mathbf{a})}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, next $(\mathbf{u}_0^{RC}, \mathbf{u}^{CAT}, \mathbf{u}^{BIN})^{(1)}$ is generated from the fully conditional posterior distribution of $(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN})$ given data, $\psi_{\setminus (\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN})}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, and so on up to $\tau_{K_{m_1+m_2+m_3}-2}^{(1)}$, which is generated from the fully conditional posterior distribution of $\tau_{K_{m_1+m_2+m_3}-2}$ given data, $(\mathbf{y}, \delta)$, $\psi_{\setminus (\tau_{K_{m_1+m_2+m_3}-2})}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, to complete one iteration of the Gibbs sampler. After t iterations (t large) Geman and Geman (1984)

showed that $\psi^{(t)}$ under mild conditions, is a sample from the joint posterior distribution of ψ .

The fully conditional posterior distributions that define one possible implementation of the Gibbs sampler are: Let $\theta = (\mathbf{b}', \mathbf{a}')'$, $\mathbf{W} = (\mathbf{X}, \mathbf{Z})$, and $\mathbf{D}^{-1} = \begin{pmatrix} \mathbf{I}_{p_1} (\sigma_1^2)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_2} (\sigma_2^2)^{-1} \end{pmatrix}$, then $\theta | (\mathbf{y}, \delta), \psi_{\setminus \theta}, \mathbf{R}_{22} = \mathbf{I}_{m_4} \sim N_{p+Nm}(\mu_\theta, \Lambda_\theta)$, where

$$\mu_\theta = \Lambda_\theta \mathbf{W}' (\mathbf{I}_n \otimes \mathbf{R})^{-1} \mathbf{y} \quad (1)$$

and

$$\begin{aligned} \Lambda_\theta^{-1} &= \begin{pmatrix} \mathbf{X}' (\mathbf{I}_n \otimes \mathbf{R})^{-1} \mathbf{X} + \mathbf{D}^{-1} & \mathbf{X}' (\mathbf{I}_n \otimes \mathbf{R})^{-1} \mathbf{Z} \\ \mathbf{Z}' (\mathbf{I}_n \otimes \mathbf{R})^{-1} \mathbf{X} & \mathbf{Z}' (\mathbf{I}_n \otimes \mathbf{R})^{-1} \mathbf{Z} + \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{pmatrix} \\ &= \mathbf{W}' (\mathbf{I}_n \otimes \mathbf{R})^{-1} \mathbf{W} + \begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \otimes \mathbf{G}^{-1} \end{pmatrix} \end{aligned} \quad (2)$$

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

$$\mathbf{G} | (\mathbf{y}, \delta), \psi_{\setminus \mathbf{G}} \sim IW_m \left([\Sigma_{\mathbf{G}}^{-1} + \mathbf{a}'_M \mathbf{A}^{-1} \mathbf{a}_M]^{-1}, f_{\mathbf{G}} + N \right)$$

and the fully conditional posterior distribution of $\mathbf{R}$ conditional on data, $\psi_{\setminus \mathbf{R}}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ follows a conditional inverted Wishart distribution; where, unconditionally on $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, we have:

$$\begin{aligned} &\mathbf{R} | (\mathbf{y}, \delta), \psi_{\setminus \mathbf{R}} \\ &\sim IW_m \left(\left[\Sigma_{\mathbf{R}}^{-1} + \sum_{i=1}^n (\mathbf{u}_i - \mathbf{x}_i \mathbf{b} - \mathbf{z}_i \mathbf{a}) (\mathbf{u}_i - \mathbf{x}_i \mathbf{b} - \mathbf{z}_i \mathbf{a})' \right]^{-1}, f_{\mathbf{R}} + n \right) \end{aligned}$$

The following notation will be used for augmented data of animal i : $\mathbf{U}_i^{aug}$ is the vector of those U_{ij}' s where j indices a censored ($\delta_{ij} = 0$) 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 the number of censored, right censored Gaussian traits. The dimension of $\mathbf{U}_i^{aug}$ is n_i^{aug} . The fully conditional posterior distribution of $\mathbf{U}_i^{aug}$ given data, $\psi_{\setminus \mathbf{U}_i^{aug}}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ follows a truncated n_i^{aug} -dimensional multivariate normal distribution on the interval:

$$\begin{aligned}
& \prod_{j=m_1+1}^{m_1+m_2} [\mathbf{1}(u_{ij} > y_{ij})]^{1-\delta_{ij}} \\
& \times \prod_{j=m_1+m_2+1}^{m_1+m_2+m_3} \left\{ \sum_{k=1}^{K_j} [\mathbf{1}(\tau_{jk-1} < u_{ij} \leq \tau_{jk}) \mathbf{1}(y_{ij} = k)] \right\} \\
& \times \prod_{j=m_1+m_2+m_3+1}^m [\mathbf{1}(u_{ij} \leq 0) \mathbf{1}(y_{ij} = 0) + \mathbf{1}(0 < u_{ij}) \mathbf{1}(y_{ij} = 1)]
\end{aligned} \tag{3}$$

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

$$(\mathbf{x}_{i(aug)} \mathbf{b} + \mathbf{z}_{i(aug)} \mathbf{a}) + \mathbf{R}_{i(aug)(obs)} \mathbf{R}_{i(obs)}^{-1} (\mathbf{u}_{i(obs)} - (\mathbf{x}_{i(obs)} \mathbf{b} - \mathbf{z}_{i(obs)} \mathbf{a})) \tag{4}$$

and

$$\mathbf{R}_{i(aug)} - \mathbf{R}_{i(aug)(obs)} \mathbf{R}_{i(obs)}^{-1} \mathbf{R}_{i(obs)(aug)} \tag{5}$$

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 augmented data of animal i , respectively. Similar definitions are given to $\mathbf{z}_{i(obs)}$ 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 to $\mathbf{R}_{i(aug)(obs)}$, $\mathbf{R}_{i(obs)}$ and $\mathbf{R}_{i(obs)(aug)}$.

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

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

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

Detailed derivations of the fully conditional posterior distributions can be found in e.g. Korsgaard (1997).

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 $\mathbf{J}(i) = (J_1(i), \dots, J_m(i))'$ be the vector of response indicator random variables on animal i defined by $J_k(i) = 1$ if the k 'th trait is observed on animal i and $J_k(i) = 0$ otherwise, $k = 1, \dots, m$. The observed data on animal i is $(\mathbf{y}_i, \delta_i)_{\mathbf{J}(i)}$, where $(\mathbf{y}_i, \delta_i)_{\mathbf{J}(i)}$ denotes the observed

Gaussian, observed right censored Gaussian traits, with their censoring indicators, observed categorical and binary traits of animal i . An animal with 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 $1 \leq \dim(\mathbf{y}_{i(obs)}) \leq m$. Data on all animals are $(\mathbf{y}, \delta)_{\mathbf{J}}$, where $\mathbf{J} = (\mathbf{J}(i))_{i=1, \dots, n}$.

For missing data, the idea of augmenting with residuals (Wang et al., 1997) is invoked. It is assumed that

$$\begin{pmatrix} \mathbf{U}_{i(obs)} \\ \mathbf{U}_{i(aug)} \\ \mathbf{E}_{i(mis)} \end{pmatrix} | \mathbf{b}, \mathbf{a}, \mathbf{R}, \mathbf{R}_{22} = \mathbf{I}_{m_4} \\ \sim N_m \left(\begin{pmatrix} (\mathbf{x}_{i(obs)} \mathbf{b} + \mathbf{z}_{i(obs)} \mathbf{a}) \\ (\mathbf{x}_{i(aug)} \mathbf{b} + \mathbf{z}_{i(aug)} \mathbf{a}) \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{R}_{i(obs)} & \mathbf{R}_{i(obs)(aug)} & \mathbf{R}_{i(obs)(mis)} \\ \mathbf{R}_{i(aug)(obs)} & \mathbf{R}_{i(aug)} & \mathbf{R}_{i(aug)(mis)} \\ \mathbf{R}_{i(mis)(obs)} & \mathbf{R}_{i(mis)(aug)} & \mathbf{R}_{i(mis)} \end{pmatrix} \right)$$

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 ω , data are observed/missing at random in the sense, that $\mathbf{J}$ is stochastically independent of $(\mathbf{U}, \mathbf{C})$ conditional on ω .
- (e) Conditional on ω , $\mathbf{J}$ is noninformative of ω .

Under the assumptions (a)-(e), and having augmented with $\mathbf{U}_{i(aug)}$ and $\mathbf{E}_{i(mis)}$ for all animals (i.e. with $(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}, \mathbf{E}^{MIS})$), it then follows, that the joint posterior distribution of parameters and augmented data $\psi = (\omega, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}, \mathbf{E}^{MIS})$ is given by:

$$\begin{aligned}
& p(\psi | (\mathbf{y}, \delta)_{\mathbf{J}}, \mathbf{R}_{22} = \mathbf{I}_{m_4}) \\
& \propto p((\mathbf{y}, \delta)_{\mathbf{J}} | \psi, \mathbf{R}_{22} = \mathbf{I}_{m_4}) p(\psi | \mathbf{R}_{22} = \mathbf{I}_{m_4}) \\
& = p((\mathbf{y}, \delta)_{\mathbf{J}}, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}, \mathbf{E}^{MIS} | \omega, \mathbf{R}_{22} = \mathbf{I}_{m_4}) p(\omega | \mathbf{R}_{22} = \mathbf{I}_{m_4}) \\
& = p((\mathbf{y}, \delta)_{\mathbf{J}}, \mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN}, \mathbf{E}^{MIS} | \omega, \mathbf{R}_{22} = \mathbf{I}_{m_4}) p(\omega | \mathbf{R}_{22} = \mathbf{I}_{m_4}) \\
& \propto \prod_{i=1}^n \left[\prod_{j=m_1+1}^{m_1+m_2} \left([\mathbf{1}(u_{ij} > y_{ij})]^{1-\delta_{ij}} \right)^{J_j(i)} \right] \\
& \times \prod_{i=1}^n \left[\prod_{j=m_1+m_2+1}^{m_1+m_2+m_3} \left\{ \sum_{k=1}^{K_j} [\mathbf{1}(\tau_{jk-1} < u_{ij} \leq \tau_{jk}) \mathbf{1}(y_{ij} = k)] \right\}^{J_j(i)} \right] \\
& \times \prod_{i=1}^n \left[\prod_{j=m_1+m_2+m_3+1}^m [\mathbf{1}(u_{ij} \leq 0) \mathbf{1}(y_{ij} = 0) + \mathbf{1}(0 < u_{ij}) \mathbf{1}(y_{ij} = 1)]^{J_j(i)} \right] \\
& \times \prod_{i=1}^n \left[(2\pi)^{-m/2} |\mathbf{R}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{u}_i - \mathbf{x}_i \mathbf{b} - \mathbf{z}_i \mathbf{a})' \mathbf{R}^{-1} (\mathbf{u}_i - \mathbf{x}_i \mathbf{b} - \mathbf{z}_i \mathbf{a}) \right\} \right]
\end{aligned}$$

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 associated with 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.

Further details related to the derivation of the fully conditional posterior distributions can be found in e.g. Korsgaard (1997).

4 Strategies for implementation of the Gibbs sampler

Strategies for implementation are outlined for the case without missing data, and where a priori $\mathbf{b}$ conditional on σ_1^2 and σ_2^2 follows a multivariate normal distribution.

4.1 Univariate sampling of location parameters

The fully conditional posterior distribution of θ given data, $\psi_{\setminus \theta}$ 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 (1) and (2) 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:

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

where r_i is the i 'th element of $\mathbf{r} = \mathbf{W}' (\mathbf{I} \otimes \mathbf{R}^{-1}) \mathbf{u}$ 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 i 'th row of the coefficient matrix and C_{ii} is the i 'th 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. Garcia-Cortés and Sorensen (1996) described a method to sample from the joint fully conditional posterior distribution of θ given data, $\psi_{\setminus \theta}$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, that can avoid inverting the coefficient matrix $\mathbf{C} = \Lambda_\theta^{-1}$ of 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 $N_{p_1}(\mathbf{0}, \mathbf{I}_{p_1} \sigma_1^2)$, $N_{p_2}(\mathbf{0}, \mathbf{I}_{p_2} \sigma_2^2)$, $N_{Nm}(\mathbf{0}, \mathbf{A} \otimes \mathbf{G})$ and $N_{nm}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{R})$ distributions, respectively. Next let $\mathbf{b}^* = (\mathbf{b}_1^{*'}, \mathbf{b}_2^{*'})'$ and $\theta^* = (\mathbf{b}^{*'}, \mathbf{a}^{*'})'$ and define $\mathbf{u}^*$ as $\mathbf{W}\theta^* + \mathbf{e}^*$, then it follows that the linear combination of θ^* and $\mathbf{e}^*$ given by:

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

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

4.3 Sampling of augmented data

The fully conditional posterior distribution of augmented Gaussian traits,

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

given data, $\psi_{\setminus}(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN})$ and $\mathbf{R}_{22} = \mathbf{I}_{m_4}$ will be sampled jointly. The dimension of $(\mathbf{U}_0^{RC}, \mathbf{U}^{CAT}, \mathbf{U}^{BIN})$ is $\sum_{i=1}^n n_i^{aug}$. Realising that $\mathbf{U}_i^{aug}$'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 $(\mathbf{y}_i, \delta_i)$ and ω conditional on $\mathbf{R}_{22} = \mathbf{I}_{m_4}$. This is a n_i^{aug} -dimensional multivariate truncated Gaussian distribution on the interval given in (3). Before truncation, mean and variance are given by (4) and (5), respectively.

Let ξ and Σ 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 to 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 $N_1(\xi_2 + \Sigma_{21}\Sigma_{11}^{-1}(u_{i1}^{aug} - \xi_1), \Sigma_{22.1})$ -distribution. Finally, proceeding this way, $u_{in_i}^{aug}$ is sampled from a truncated univariate normal distribution with mean and variance before truncation given by

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

and

$$\Sigma_{n_i^{aug} n_i^{aug}} - \Sigma_{n_i^{aug}(1:n_i^{aug}-1)} \Sigma_{(1:n_i^{aug}-1)}^{-1} \Sigma_{(1:n_i^{aug}-1) n_i^{aug}}$$

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 sampling 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. Devroye, 1986) that avoids rejections is implemented. First x is sampled from a $R(0, 1)$ -distributed random variable, X . Next, let z be given by

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

where F_Y is the distribution function of Y , then z is a realised value from the truncated $N_1(\mu, \sigma^2)$ -distribution on I . The proof follows from (6) given below, where Z is the random variable from which z is generated; z is a value between s_1 and s_2 :

$$\begin{aligned}
P(Z \leq z) &= P(F_Y^{-1}[F_Y(s_1) + X(F_Y(s_2) - F_Y(s_1))] \leq z) \\
&= P(F_Y(s_1) + X(F_Y(s_2) - F_Y(s_1)) \leq F_Y(z)) \\
&= P\left(X \leq \frac{F_Y(z) - F_Y(s_1)}{F_Y(s_2) - F_Y(s_1)}\right) \\
&= \frac{F_Y(z) - F_Y(s_1)}{F_Y(s_2) - F_Y(s_1)}
\end{aligned} \tag{6}$$

4.4 Sampling of covariance matrices

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 distribution of $\mathbf{R}$ is inverse Wishart distributed. In order to obtain samples from the conditional inverse Wishart distribution, the method described in Korsgaard et al. (1999) 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 $\mathbf{R} \sim IW_m(\Sigma, f)$ and let $\mathbf{V} = \mathbf{R}^{-1}$, where $\mathbf{V}$ by definition is Wishart distributed, $\mathbf{V} \sim W_m(\Sigma, f)$. Next $\mathbf{R}$ is expressed in terms of $\mathbf{V}$: $\mathbf{R} = \begin{pmatrix} \mathbf{V}_{11}^{-1} + (\mathbf{V}_{11}^{-1}\mathbf{V}_{12})\mathbf{V}_{22}^{-1}(\mathbf{V}_{11}^{-1}\mathbf{V}_{12})' & -(\mathbf{V}_{11}^{-1}\mathbf{V}_{12})\mathbf{V}_{22}^{-1} \\ -\mathbf{V}_{22}^{-1}(\mathbf{V}_{11}^{-1}\mathbf{V}_{12})' & \mathbf{V}_{22}^{-1} \end{pmatrix}$, where $\mathbf{V}_{22}^{-1} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12} = \mathbf{R}_{22}^{-1}$. From properties of the Wishart distribution, it is known that $\mathbf{V}_{11} \sim W_{m-m_4}(\Sigma_{11}, f)$, $(\mathbf{V}_{11}^{-1}\mathbf{V}_{12})|\mathbf{V}_{11} = \mathbf{v}_{11} \sim N_{(m-m_4) \times m_4}(\Sigma_{11}^{-1}\Sigma_{12}, \mathbf{v}_{11}^{-1} \otimes \Sigma_{22.1})$, where $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and that $\mathbf{V}_{22.1} \sim W_{m_4}(\Sigma_{22.1}, f - (m - m_4))$. Furthermore $(\mathbf{V}_{11}, \mathbf{V}_{11}^{-1}\mathbf{V}_{12})$ is stochastically independent of $\mathbf{V}_{22.1}$. Realising that $\mathbf{R}_{22} = \mathbf{I}_{m_4}$, is equivalent to $\mathbf{V}_{22.1} = \mathbf{I}_{m_4}$, 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 $(\mathbf{V}_{11}^{-1}\mathbf{V}_{12})$ 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 (1966) is implemented. The basic idea in their algorithm can be summarised as follows: Let $\mathbf{V} \sim W_m(\Sigma, f)$ and let $\mathbf{LL}'$ be a Cholesky factorisation of Σ , i.e. $\Sigma = \mathbf{LL}'$. 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 $\mathbf{L}\mathbf{w}\mathbf{L}'$ 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 $W_{11}^{-1}W_{12}|W_{11} = w_{11} \sim N_1(0, w_{11}^{-1})$

$w_{22 \cdot 1}$ 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 $\mathbf{T}_i = \mathbf{W}_{(i-1)(i-1)}^{-1} \mathbf{W}_{(1:i-1)i} | \mathbf{W}_{(i-1)(i-1)} = \mathbf{w}_{(i-1)(i-1)} \sim N_{i-1}(\mathbf{0}, \mathbf{w}_{(i-1)(i-1)}^{-1})$. $\mathbf{W}_{(1:i-1)i}$ is used as notation for the $((i-1) \times 1)$ -dimensional vector of elements $(W_{ji})_{j=1, i-1}$ of $\mathbf{W}$ and $\mathbf{W}_{(i-1)(i-1)}$ is the $(i-1)$ -dimensional square matrix of $\mathbf{W}$, with elements $(W_{jk})_{j,k=1, i-1}$

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

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

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

5 Conclusion

A fully Bayesian analysis of Gaussian, right censored Gaussian, categorical and binary traits using the Gibbs sampler and data augmentation has been outlined. The method has been implemented as a module to DMU (Jensen and Madsen, 1994) following the strategies for implementation outlined. In the program 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. The Gibbs sampler has only been outlined for models with additive genetic effects. It is easy to generalise to more independent random effects and this has been implemented. The models are also allowed to include maternal effects correlated with animal effects. For models including binary traits, the residuals of liabilities of binary traits are assumed to be independent. In cases with two or more binary traits included in the analysis, this is a restriction. It is not a restriction with only one binary trait included in the analysis and in this case, the method implemented avoids the ad hoc scaling procedure described in Van Tassell (1998).

6 Appendix

The convention used for Wishart and inverted Wishart distributions follows Mardia et al. (1979). Let $\mathbf{M} \sim W_p(\Sigma, f)$, the density of $\mathbf{M}$ is up to proportionality given by (for $\Sigma > 0$ and $f \geq p$):

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

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

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:

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

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

References

- [1] Albert JH and Chib S (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 88, 669-679.
- [2] Chib S and Greenberg E (1996). Markov Chain Monte Carlo simulation methods in econometrics. *Econometric Theory* 12, 409-431.
- [3] Chan KS (1993). Asymptotic behavior of the Gibbs sampler. *Journal of the American Statistical Association* 88, 320-326
- [4] Cox DR and Snell EJ (1989). *Analysis of binary data*. Chapman and Hall, London.
- [5] Devroye L (1986). *Non-uniform random variate generation*. Springer-Verlag, New York
- [6] García-Cortés LA and Sorensen D (1996). On a multivariate implementation of the Gibbs sampler. *Genet. Sel. Evol.* 28, 121-126.
- [7] Gelfand AE, Hills SE, Racine-Poon A and Smith AFM (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* 85, 972-985.
- [8] Gelfand AE and Smith AFM (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398-409.

- [9] Geman S and Geman D (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721-741.
- [10] Hastings WK (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97-109.
- [11] Jensen J (1994). Bayesian analysis of bivariate mixed models with one continuous and one binary trait using the Gibbs sampler. *Proceedings of the 5th world congress on genetics applied to livestock production* 18, 333-336.
- [12] Jensen J and Madsen P (1994). DMU: A package for the analysis of multivariate mixed models. In: *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production*. Vol. 22, 45-46
- [13] Jensen J, Wang CS, Sorensen DA and Gianola D (1994). Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. *Acta Agric. Scand.* 44, 193-201.
- [14] Korsgaard IR (1997). Genetic analysis of survival data. Ph.D. Thesis, Department of Theoretical Statistics, University of Aarhus.
- [15] Korsgaard IR, Andersen AH and Sorensen D (1999). A useful reparameterisation to obtain samples from conditional inverse Wishart distributions. *Genet. Sel. Evol.* 31. (To appear)
- [16] Mardia KV, Kent JT and Bibby JM (1979). *Multivariate analysis*. Academic Press, London.
- [17] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH and Teller E (1953) Equations of state calculations by fast computing machines. *Journal Of Chemical Physics* 21, 1087-1091.
- [18] Mood AM, Graybill FA and Boes DC (1974). *Introduction to the Theory of Statistics*. McGraw-Hill, NY, USA.
- [19] Odell PL and Feiveson AH (1966). A numerical procedure to generate a sample covariance matrix. *Am. Stat. Ass. J.*, 199-203
- [20] Sorensen DA, Wang CS, Jensen J and Gianola D (1994). Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet. Sel. Evol.* 26, 333-360.
- [21] Sorensen DA, Andersen S, Gianola D and Korsgaard IR (1995). Bayesian inference in threshold models using Gibbs sampling. *Genet. Sel. Evol.* 27, 229-249.
- [22] Sorensen DA, Gianola D and Korsgaard IR (1998). Bayesian mixed-effects model analysis of a censored normal distribution with animal breeding applications. *Acta Agric. Scand., Sect. A, Animal Sci.* 48, 222-229

- [23] Tanner MA and Wong WH (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82, 528-540.
- [24] Tobin J (1958). Estimation of relationships for limited dependent variables. *Econometrica* 26, 24-36.
- [25] Van Tassell CP, Casella G and Pollak EJ (1995). Effects of selection on estimates of variance components using Gibbs sampling and restricted maximum likelihood. *J. Dairy Sci.* 78, 678-692.
- [26] Van Tassell CP, Van Vleck LD and Gregory KE (1998). Bayesian analysis of twinning and ovulation rates using a multiple-trait threshold model and Gibbs sampling. *J. Anim Sci.* 76, 2048-2061.
- [27] Wang CS, Rutledge JJ and Gianola D (1993). Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet. Sel. Evol.* 25, 41-62.
- [28] Wang CS, Rutledge JJ and Gianola D (1994). Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet. Sel. Evol.* 26, 91-115.
- [29] Wang CS, Quass RL and Pollak EJ (1997). Bayesian analysis of calving ease scores and birth weights. *Genet. Sel. Evol.* 29, 117-143.
- [30] Wei GCG and Tanner MA (1990). Posterior computations for censored regression data. *Journal of the American Statistical Association* 85, 829-839.
- [31] Zeger SL and Karim MR (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association* 86, 79-86.