

Bayesian analysis of calving ease scores and birth weights

CS Wang*, RL Quaas, EJ Pollak

*Morrison Hall, Department of Animal Science, Cornell University,
Ithaca, NY 14853, USA*

(Received 3 January 1996; accepted 13 February 1997)

Summary – In a typical two stage procedure, breeding value prediction for calving ease in a threshold model is conditioned on estimated genetic and residual covariance matrices. These covariance matrices are traditionally estimated using analytical approximations. A Gibbs sampler for making full Bayesian inferences about fixed effects, breeding values, thresholds and genetic and residual covariance matrices to analyze jointly a discrete trait with multiple ordered categories (calving ease scores) and a continuously Gaussian distributed trait (birth weights) is described. The Gibbs sampler is implemented by drawing from a set of densities – (truncated) normal, uniform and inverted Wishart – making implementation of Gibbs sampling straightforward. The method should be useful for estimating genetic parameters based on features of their marginal posterior densities taking into full account uncertainties in estimating other parameters. For routine, large-scale estimation of location parameters (breeding values), Gibbs sampling is impractical. The joint posterior mode given the posterior mean estimates of thresholds and dispersion parameters is suggested. An analysis of simulated calving ease scores and birth weights is described.

dystocia / beef cattle / threshold model / Bayesian method / Gibbs sampling

Résumé – Analyse bayésienne des notes de difficultés de vêlage et des poids de naissance. Dans une procédure typique à deux étapes, l'évaluation génétique pour la difficulté de vêlage dans un modèle à seuil est conditionnée par les matrices de covariance génétiques et résiduelles. Ces matrices de covariance sont habituellement estimées au travers d'approximations analytiques. On décrit l'échantillonnage de Gibbs permettant d'effectuer des inférences bayésiennes complètes à propos des effets fixes, des valeurs génétiques, des seuils, et des matrices de covariance génétiques et résiduelles, pour analyser conjointement un caractère discret à catégories multiples ordonnées (note de difficulté de vêlage) et un caractère continu gaussien (poids de naissance). L'échantillonnage de Gibbs est assez simple à partir de densités de divers types : normale (tronquée), uniforme et Wishart inverse. La méthode est utile pour estimer les paramètres génétiques à partir de leurs distributions marginales a posteriori, après prise en compte des incertitudes

* Correspondence and reprints: Pfizer Central Research, T201, Eastern Point Rd, Groton, CT 06340, USA.

concernant les autres paramètres. L'échantillonnage de Gibbs n'est pas faisable en routine pour estimer les valeurs génétiques. On suggère le mode de la distribution conjointe a posteriori, pour des valeurs des seuils et des paramètres de dispersion correspondant à leurs moyennes a posteriori. On décrit une analyse de notes de difficulté de vêlage et de poids de naissance simulés.

dystocie / bovins à viande / modèle à seuil / méthode bayésienne / échantillonnage de Gibbs

INTRODUCTION

Calving ease is considered a calf trait and recorded subjectively as one of several exclusive ordered categories. For example, for American Simmental cattle, calving ease is scored as 1 (natural calving, no assistance), 2 (easy pull), 3 (hard pull) or 4 (mechanical force or Cesarean). Calf size (birth weight) affects ease of birth: the bigger the calf is, the more likely the birth will be difficult (Koger et al, 1967; Pollak, 1975).

In this paper, we consider joint modeling of calving ease scores and birth weights using the threshold model concept of Wright (1934). In a threshold model, an underlying continuous variable is postulated for calving ease. A set of thresholds divides this continuous variable into the discrete calving ease scores actually recorded. Gianola (1982) and Gianola and Foulley (1983) considered Bayesian analysis of single trait threshold models assuming known genetic variance. Harville and Mee (1984) and Foulley et al (1987) gave approximate methods for variance component estimation. Foulley et al (1983) developed a method to deal with a binary trait and two continuous traits without allowing for missing data, while Janss and Foulley (1993) extended the method to handle data with missing patterns. In 1990 at Cornell University, a system for routine sire evaluation of calving ease scores and birth weights jointly allowing for all possible missing data framework was implemented. This system assumed a sire-mgs (maternal grandsire) linear model for the underlying scale; it predicts the frequency of unassisted births for American Simmental cattle (Pollak et al, 1995 pers comm). This evaluation system also assumed that genetic and residual covariance matrices and thresholds were known. Variance components were estimated (Dong et al, 1991) by extension of Foulley et al (1987). Hoeschele et al (1995) described further extensions of Foulley et al (1983) and Janss and Foulley (1993) to a situation of one multiple ordered categorical trait and several continuous traits.

A difficulty in estimating parameters under threshold models is that the likelihood or marginal posterior distributions do not have closed forms and approximations are used. With the help of Monte Carlo methods, in particular Gibbs sampling (Geman and Geman, 1984; Gelfand et al, 1990), these approximations are no longer needed. Wang et al (1993, 1994a,b) described making Bayesian inferences in a univariate linear model in an animal breeding context using Gibbs sampling. Sorensen et al (1994) demonstrated how inference about response to selection in a linear model can be made. Berger et al (1995) applied the methods of Sorensen et al (1994) and Wang et al (1994b) to analyze a selection experiment of *Tribolium*. Jensen et al (1994) and Van Tassell (1994) extended the procedure to model maternal effects, while Van Tassell and Van Vleck (1996) further expanded

the scope to multitrait linear models. Bayesian analysis of univariate thresholds via Gibbs sampling in an animal breeding context was recently described by Sorensen et al (1995) by extending Albert and Chib (1993). For a binary trait, Hoeschele and Tier (1995) compared frequency properties of three variance component estimators: mode of approximate marginal likelihood (Foulley et al, 1987), marginal posterior mode and mean via Gibbs sampling. Jensen (1994) analyzed simulated data of one binary trait and one continuous trait via Gibbs sampling under a Bayesian framework. Wang et al (1995) gave a Bayesian method to analyze one multiple ordered categorical trait and one continuous trait with Gibbs sampling. Van Tassell et al (1996) presented Bayesian analysis of twinning and ovulation rate using Gibbs sampling.

The purpose of this paper is to extend the work of Sorensen et al (1995) and Wang et al (1995) to one multiple ordered categorical trait (calving ease) and one continuous trait (birth weight) with all possible missing patterns of data under a Bayesian setting via Gibbs sampling. A set of full conditional posterior densities will be derived in closed form facilitating straightforward implementation of Gibbs sampling. Simulated data are analyzed to illustrate the methodology.

MODEL

Let \mathbf{Y}_{1o} be a vector of birth weights (BW), with o denoting observed record, and \mathbf{Y}_{2o} be calving ease scores (CE, recorded as one of four scores (1 = no assistance, 2 = easy pull, 3 = hard pull and 4 = mechanical force or Cesarean), $\mathbf{U}_{1o} \equiv \mathbf{Y}_{1o}$ and \mathbf{U}_{2o} be the corresponding underlying variable for observed calving ease scores, which is also known as augmented 'data' (Tanner, 1993). A sire-mgs model used for the American Simmental calving ease sire evaluation (Quaas, 1994) is assumed on the underlying scale (an animal model can be assumed by appropriately defining terms):

$$\mathbf{U}_{1o} = \text{AgeDam} + \text{CG} + \mathbf{a}_{\text{sire}} + 1/2\mathbf{a}_{\text{mgs}} + \mathbf{e}_{1o}$$

$$\mathbf{U}_{2o} = \text{AgeDam} + \text{CG} + \mathbf{a}_{\text{sire}} + \mathbf{m}_{\text{mgs}} + \mathbf{e}_{2o}$$

where \mathbf{a}_i is sire effect (direct), and \mathbf{m}_{mgs} is maternal grandsire effect (1/4 direct BV plus 1/2 maternal BV) and \mathbf{e}_{1o} and \mathbf{e}_{2o} are residual effects. AgeDam is age of dam effect and CG is contemporary group effect. Note that maternal effects for BW are not modeled for the Simmental population because the maternal contribution to the total genetic variance was found to be negligible (Garrick et al, 1989). In matrix notation:

$$\mathbf{U}_{1o} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_1\mathbf{a}_1 + \mathbf{e}_{1o}$$

$$\mathbf{U}_{2o} = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{a}_2 + \mathbf{Z}_3\mathbf{m}_2 + \mathbf{e}_{2o} \quad [1]$$

For reason of easy identification of conditional posterior distribution of the residual covariance matrix later, augmented data are further expanded to include residuals associated with missing data. Denote $\mathbf{U}'_1 = [\mathbf{U}'_{1o} \ \mathbf{e}'_{1m}]$, $\mathbf{U}'_2 = [\mathbf{U}'_{2o} \ \mathbf{e}'_{2m}]$, $\mathbf{e}'_1 = [\mathbf{e}_{1o} \ \mathbf{e}'_{1m}]$ and $\mathbf{e}'_2 = [\mathbf{e}_{2o} \ \mathbf{e}'_{2m}]$, where \mathbf{e}_{1m} and \mathbf{e}_{2m} are residuals associated with the

missing BW and CE, with m denoting missing records, respectively. Then [1] can be written as

$$\begin{bmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}$$

or

$$\mathbf{U} = \mathbf{W}\boldsymbol{\theta} + \mathbf{e}$$

where \mathbf{U} contains \mathbf{U}_1 and \mathbf{U}_2 , \mathbf{W} is composed of the design matrices – \mathbf{X}_i s, \mathbf{Z}_i s and rows of zeros associated with missing data, $\boldsymbol{\theta}$ contains location parameters $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{a}_1, \mathbf{a}_2$ and \mathbf{m}_2 , and \mathbf{e} , the residuals \mathbf{e}_1 and \mathbf{e}_2 , with all the elements properly ordered and matched. For \mathbf{U} , we assume

$$\mathbf{U}|\boldsymbol{\theta}, \mathbf{R} \sim N(\mathbf{W}\boldsymbol{\theta}, \mathbf{R}) \quad [2]$$

or

$$p(\mathbf{U}|\boldsymbol{\theta}, \mathbf{R}) \propto |\mathbf{R}|^{-1/2} \times \exp \left\{ -\frac{1}{2}(\mathbf{U} - \mathbf{W}\boldsymbol{\theta})'\mathbf{R}^{-1}(\mathbf{U} - \mathbf{W}\boldsymbol{\theta}) \right\}$$

where \mathbf{R} is a block diagonal matrix containing n submatrices of residual covariances (\mathbf{R}_0) for the record(s) of a particular animal with dimension 2, ie, $\mathbf{R} = \mathbf{R}_0 \otimes \mathbf{I}_n$, if the data are sorted by animal and trait, and n is the number of animals with at least one trait recorded.

A uniform prior distribution is assigned to $\boldsymbol{\beta}_i$ s, such that (eg, Gianola et al, 1990; Wang et al 1994a)

$$p(\boldsymbol{\beta}_i) \propto \text{constant} \quad [3]$$

which is similar to treating $\boldsymbol{\beta}$ as fixed in a traditional sense. We assume for the bull effects (genetic):

$$\mathbf{a}|\mathbf{G}_0 \sim N(\mathbf{0}, \mathbf{G}_0) \quad [4]$$

or

$$\begin{aligned} p(\mathbf{a}|\mathbf{G}_0) &\propto |\mathbf{G}|^{-1/2} \times \exp \left\{ -\frac{1}{2}\mathbf{a}'\mathbf{G}^{-1}\mathbf{a} \right\} \\ &\propto |\mathbf{G}_0|^{-q/2} \times \exp \left\{ -\frac{1}{2}\mathbf{a}'(\mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1})\mathbf{a} \right\} \end{aligned}$$

where \mathbf{a} contains $\mathbf{a}_1, \mathbf{a}_2$ and \mathbf{m} , q is the number of bulls (sires and mgs), $\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$ with $\mathbf{G}_0 = \{g_{ij}\}$, $i, j = 1, 2, 3$, the covariance matrix among three genetic effects for a particular animal and \mathbf{A} is the numerator relationship matrix among sires and mgs.

To describe prior uncertainty about \mathbf{G}_0 , an inverted Wishart distribution (Johnson and Kotz, 1972; Jensen et al, 1994; Van Tassell and Van Vleck, 1996) is assigned

$$\mathbf{G}_0|\mathbf{S}_g, v_g \sim IW_3((v_g - 3 - 1)\mathbf{S}_g, v_g) \quad [5]$$

with density

$$p(\mathbf{G}_0|\mathbf{S}_g, v_g) \propto |\mathbf{G}_0|^{-\frac{1}{2}(v_g+3+1)} \times \exp \left\{ \text{tr} \left(-\frac{v_g - 3 - 1}{2} \mathbf{S}_g \mathbf{G}_0^{-1} \right) \right\}$$

where \mathbf{S}_g is the location parameter matrix; v_g is the scalar shape parameter (degrees of belief); $\mathbf{S}_g = E(\mathbf{G}_0 | \mathbf{S}_g, v_g)$. A large value of v_g indicates relative certainty that \mathbf{G}_0 is similar to \mathbf{S}_g ; a small value, uncertainty, ie, a relatively flat distribution. (The subscript 3 of IW indicates the order of the covariance matrix.) Similarly for \mathbf{R}_0 ,

$$\mathbf{R}_0 | \mathbf{S}_e, v_e \sim IW_2((v_e - 2 - 1)\mathbf{S}_e, v_e) \quad [6]$$

The final parameters are the thresholds: $\mathbf{t} = (t_1, t_2, t_3)$, with $t_0 = -\infty$ and $t_4 = \infty$. These are assumed to be distributed as order statistics from a uniform distribution in the interval $[t_{\min}, t_{\max}]$ (Sorensen et al, 1995):

$$p(\mathbf{t}) = (c - 1)!(t_{\max} - t_{\min})^{-(c-1)} I(\mathbf{t} \in T) \quad [7]$$

where $I(\cdot)$ is an indicator function and

$$T = \{(t_1, t_2, \dots, t_{c-1}) | t_{\min} \leq t_1 \leq t_2 \leq \dots \leq t_{c-1} \leq t_{\max}\}$$

$c = 4$ in our case, the number of categories.

Applying Bayesian theorem, the joint posterior density of all the parameters including the augmented data $(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2)$ given the observed data $(\mathbf{Y}' = [\mathbf{Y}'_{1o}, \mathbf{Y}'_{2o}])$ and prior parameters, assuming prior independence of \mathbf{t} , \mathbf{G}_0 and \mathbf{R}_0 is:

$$\begin{aligned} & p(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2 | \mathbf{Y}_{1o}, \mathbf{Y}_{2o}, \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ & \propto p(\mathbf{Y}_{1o}, \mathbf{Y}_{2o} | \boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2, \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ & \times p(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2 | \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ & = p(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{Y}_{1o}, \mathbf{e}_{1m}, \mathbf{U}_2, \mathbf{Y}_{2o} | \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ & = p(\mathbf{Y}_{1o}, \mathbf{e}_{1m}, \mathbf{U}_2, \mathbf{Y}_{2o} | \boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0) \times p(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0 | \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ & = p(\mathbf{U} | \boldsymbol{\theta}, \mathbf{R}_0) \times p(\mathbf{Y}_{2o} | \mathbf{U}_2, \mathbf{t}) \times p(\boldsymbol{\theta} | \mathbf{G}_0) \times p(\mathbf{t}) \times p(\mathbf{G}_0 | \mathbf{S}_g, v_g) \times p(\mathbf{R}_0 | \mathbf{S}_e, v_e) \\ & \propto p(\mathbf{Y}_{2o} | \mathbf{U}_{2o}, \mathbf{t}) \times p(\mathbf{t}) \\ & \times |\mathbf{R}_0|^{-n/2} \times \exp \left\{ -\frac{1}{2} (\mathbf{U} - \mathbf{W}\boldsymbol{\theta})' (\mathbf{R}_0^{-1} \otimes \mathbf{I}_n) (\mathbf{U} - \mathbf{W}\boldsymbol{\theta}) \right\} \\ & \times |\mathbf{G}_0|^{-q/2} \times \exp \left\{ -\frac{1}{2} \mathbf{a}' (\mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1}) \mathbf{a} \right\} \\ & \times |\mathbf{G}_0|^{-\frac{1}{2}(v_g+3+1)} \times \exp \left\{ \text{tr} \left(-\frac{v_g - 3 - 1}{2} \mathbf{S}_g \mathbf{G}_0^{-1} \right) \right\} \\ & \times |\mathbf{R}_0|^{-\frac{1}{2}(v_e+2+1)} \times \exp \left\{ \text{tr} \left(-\frac{v_e - 2 - 1}{2} \mathbf{S}_e \mathbf{R}_0^{-1} \right) \right\} \end{aligned} \quad [8]$$

Combining terms in [8]:

$$\begin{aligned} & p(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2 | \mathbf{Y}_{1o}, \mathbf{Y}_{2o}, \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ & \propto p(\mathbf{Y}_{2o} | \mathbf{U}_{2o}, \mathbf{t}) \times p(\mathbf{t}) \end{aligned}$$

$$\begin{aligned} &\times |\mathbf{G}_0|^{-\frac{1}{2}(q+v_g+3+1)} \times \exp \left\{ -\frac{1}{2} \text{tr} [((v_g - 3 - 1)\mathbf{S}_g + \mathbf{S}\mathbf{S}_g)\mathbf{G}_0^{-1}] \right\} \\ &\times |\mathbf{R}_0|^{-\frac{1}{2}(n+v_e+2+1)} \times \exp \left\{ -\frac{1}{2} \text{tr} [((v_e - 2 - 1)\mathbf{S}_e + \mathbf{S}\mathbf{S}_e)\mathbf{R}_0^{-1}] \right\} \end{aligned} \quad [9]$$

where $\mathbf{S}\mathbf{S}_g = \begin{bmatrix} \mathbf{a}'_1 \mathbf{A}^{-1} \mathbf{a}_1 & \mathbf{a}'_1 \mathbf{A}^{-1} \mathbf{a}_2 & \mathbf{a}'_1 \mathbf{A}^{-1} \mathbf{m}_2 \\ \mathbf{a}'_2 \mathbf{A}^{-1} \mathbf{a}_1 & \mathbf{a}'_2 \mathbf{A}^{-1} \mathbf{a}_2 & \mathbf{a}'_2 \mathbf{A}^{-1} \mathbf{m}_2 \\ \mathbf{m}'_2 \mathbf{A}^{-1} \mathbf{a}_1 & \mathbf{m}'_2 \mathbf{A}^{-1} \mathbf{a}_2 & \mathbf{m}'_2 \mathbf{A}^{-1} \mathbf{m}_2 \end{bmatrix}$ and $\mathbf{S}\mathbf{S}_e = \sum_{i=1}^n \mathbf{e}_i \mathbf{e}'_i$ with

$$\mathbf{e}_i = \begin{cases} \begin{bmatrix} U_{1o,i} - \mathbf{w}'_{1i} \boldsymbol{\theta}_1 \\ U_{2o,i} - \mathbf{w}'_{2i} \boldsymbol{\theta}_2 \end{bmatrix} & \text{if both BW and CE observed} \\ \begin{bmatrix} U_{1o,i} - \mathbf{w}'_{1i} \boldsymbol{\theta}_1 \\ e_{2m,i} \end{bmatrix} & \text{if only BW observed} \\ \begin{bmatrix} e_{1m,i} \\ U_{2o,i} - \mathbf{w}'_{2i} \boldsymbol{\theta}_2 \end{bmatrix} & \text{if only CE observed} \end{cases}$$

with \mathbf{w}_{1i} and/or \mathbf{w}_{2i} the incidence vectors associated with $\boldsymbol{\theta}_1$ and/or $\boldsymbol{\theta}_2$ for the i th animal's record(s), respectively.

The first term in [8], $p(\mathbf{Y}_{2o} | \mathbf{U}_{2o}, \mathbf{t})$, has a degenerate density (Albert and Chib, 1993; Sorensen et al, 1995):

$$p(\mathbf{Y}_{2o} | \mathbf{U}_{2o}, \mathbf{t}) = \prod_{i=1}^n \left\{ \sum_{j=1}^4 I(t_{j-1} < U_{2o,i} \leq t_j) I(Y_{2o,i} = j) \right\} \quad [10]$$

where $I(\cdot)$ is an indicator function. For example, for a particular CE score ($= k$), we have

$$p(Y_{2o,i} = k | \mathbf{U}_{2o}, \mathbf{t}) = p(Y_{2o,i} = k | U_{2o,i}, t_{k-1}, t_k) = \begin{cases} 1 & \text{if } t_{k-1} < U_{2o,i} \leq t_k \\ 0 & \text{otherwise} \end{cases} \quad [11]$$

One way to ensure identifiability of parameters is to set constraints. Assuming full column rank for the incidence matrix \mathbf{W} , two constraints need be specified. Usually, one threshold and the residual variance for CE on the underlying scale are set to 0 and 1, respectively (Harville and Mee, 1984). An equivalent parameterization (Sorensen et al, 1995) is to fix two thresholds and to estimate the CE residual variance. We followed the latter because it allows easy specification of the conditional density of \mathbf{R}_0 . The two parameterizations, though equivalent, may not yield the same joint posterior density owing to different sets of priors specified.

Inference about location and dispersion parameters will be based on the joint posterior density [9], or on their respective marginal posterior densities. For example, if interest of inference is on the location parameters, we need to integrate out all other parameters in [9] other than $\boldsymbol{\theta}$ to obtain its marginal posterior density:

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \int_{\mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2} p(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2 | \mathbf{Y}, \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) d\mathbf{t} d\mathbf{G}_0 d\mathbf{R}_0 d\mathbf{e}_{1m} d\mathbf{U}_2 \quad [12]$$

Similarly, inference about \mathbf{G}_0 is based on:

$$p(\mathbf{G}_0|\mathbf{Y}) = \int_{\boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2} p(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2 | \mathbf{Y}, \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) d\boldsymbol{\theta} d\mathbf{t} d\mathbf{R}_0 d\mathbf{e}_{1m} d\mathbf{U}_2 \quad [13]$$

These densities cannot be derived analytically. Monte Carlo methods, such as Gibbs sampling, draw samples from [9]. Such samples, if considered jointly, are from the joint posterior distribution or, viewed marginally, from an appropriate marginal posterior distribution. Inferences can be based on these drawn samples. Inferences about functions of parameters, such as heritabilities and genetic correlations, can be made based on transformed samples.

Fully conditional posterior densities (Gibbs sampler)

The Gibbs sampler consists of a set of fully conditional posterior densities of unknown parameters in the model, ie, the conditional density of a parameter given all other parameters and the data. These can be derived from the joint posterior density [8] or [9].

For location parameters ($\boldsymbol{\theta}$), we keep terms in [8] that are functions of $\boldsymbol{\theta}$ such that:

$$\begin{aligned} & p(\boldsymbol{\theta} | \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2, \mathbf{Y}_{1o}, \mathbf{Y}_{2o}, \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ & \propto p(\mathbf{U} | \boldsymbol{\theta}, \mathbf{R}_0) \times p(\boldsymbol{\theta} | \mathbf{G}_0) \\ & \propto p(\boldsymbol{\theta} | \mathbf{G}_0, \mathbf{R}_0, \mathbf{U}) \\ & \propto \exp \left\{ -\frac{1}{2} (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' (\mathbf{W}'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n) \mathbf{W} + \boldsymbol{\Omega}) (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \right\} \end{aligned}$$

where $\boldsymbol{\Omega} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1} \end{bmatrix}$, with blocks of $\mathbf{0}$ s corresponding to $\boldsymbol{\beta}$ (Gianola et al, 1990). This is a normal density, so

$$\boldsymbol{\theta} | \mathbf{U}, \mathbf{G}_0, \mathbf{R}_0 \sim N(\tilde{\boldsymbol{\theta}}, (\mathbf{W}'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n) \mathbf{W} + \boldsymbol{\Omega})^{-1}) \quad [14]$$

where $\tilde{\boldsymbol{\theta}}$ satisfies Henderson's mixed model equations (MME) (Henderson, 1973, 1984):

$$(\mathbf{W}'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n) \mathbf{W} + \boldsymbol{\Omega}) \tilde{\boldsymbol{\theta}} = \mathbf{W}'(\mathbf{R}_0^{-1} \otimes \mathbf{I}_n) \mathbf{U}$$

To sample a subvector or a scalar of $\boldsymbol{\theta}$, rewrite the MME as

$$\mathbf{C} \tilde{\boldsymbol{\theta}} = \mathbf{b}$$

where $\mathbf{C} = \{\mathbf{C}_{ij}\}$, $i, j = 1, 2, \dots, N$, is the coefficient matrix of the MME, with \mathbf{C}_{ij} s as blocks of \mathbf{C} , and $\mathbf{b} = \{\mathbf{b}_i\}$, $i = 1, 2, \dots, N$ is the corresponding right-hand. The conditional posterior distribution for the location parameters is:

$$\boldsymbol{\theta}_i | \mathbf{U}, \boldsymbol{\theta}_{-i}, \mathbf{G}_0, \mathbf{R}_0 \sim N(\tilde{\boldsymbol{\theta}}_i, \tilde{\mathbf{v}}_i), i = 1, 2, \dots, N \quad [15]$$

where

$$\tilde{\boldsymbol{\theta}}_i = \mathbf{C}_{ii}^{-1}(\mathbf{b}_i - \sum_{j \neq i} \mathbf{C}_{ij} \boldsymbol{\theta}_j) \quad [15.1]$$

$\tilde{\mathbf{v}}_i = \mathbf{C}_{ii}^{-1}$, $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$ are subvectors, possibly scalars, of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{-i}$ is $\boldsymbol{\theta}$ with $\boldsymbol{\theta}_i$ deleted. If $\boldsymbol{\theta}_i$ is a scalar, [15] is the scalar version of the sampler for the location parameters (Wang et al, 1994a). Note the similarity of [15.1] to an update in (block) Gauss-Seidel iteration. It may be advantageous to sample a subvector jointly to speed up convergence of Gibbs chain. For example, sampling all genetic effects for an animal may reduce serial correlations among Gibbs samples (Van Tassell, 1994; Garcia-Cortes and Sorensen, 1996).

From [9], the full conditional posterior density of the genetic covariance matrix, as in Gaussian linear models (Jensen et al, 1994; Van Tassell and Van Vleck, 1996), is

$$\begin{aligned} & p(\mathbf{G}_0 | \boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0, \mathbf{e}_{1m}, \mathbf{U}_2, \mathbf{Y}_{1o}, \mathbf{Y}_{2o}, \mathbf{S}_g, \mathbf{S}_c, v_g, v_e) \\ &= p(\mathbf{G}_0 | \boldsymbol{\theta}, \mathbf{S}_g, v_g) \\ &= p(\boldsymbol{\theta} | \mathbf{G}_0) \times p(\mathbf{G}_0 | \mathbf{S}_g, v_g) \\ &\propto |\mathbf{G}_0|^{-\frac{1}{2}(q+v_g+3+1)} \times \exp \left\{ -\frac{1}{2} \text{tr}[(v_g - 3 - 1)\mathbf{S}_g + \mathbf{S}\mathbf{S}_g]\mathbf{G}_0^{-1} \right\}, \quad [16] \end{aligned}$$

or

$$\mathbf{G}_0 | \boldsymbol{\theta}, \mathbf{S}_g, v_g \sim IW_3((v_g - 3 - 1)\mathbf{S}_g + \mathbf{S}\mathbf{S}_g), q + v_g)$$

Similarly, the fully conditional posterior density of the residual covariance matrix is

$$\begin{aligned} & p(\mathbf{R}_0 | \boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{e}_{1m}, \mathbf{U}_2, \mathbf{Y}_{1o}, \mathbf{Y}_{2o}, \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ &= p(\mathbf{R}_0 | \mathbf{U}, \boldsymbol{\theta}, \mathbf{S}_e, v_e) \\ &\propto p(\mathbf{U} | \boldsymbol{\theta}, \mathbf{R}_0) \times p(\mathbf{R}_0 | \mathbf{S}_e, v_e) \\ &\propto |\mathbf{R}_0|^{-\frac{1}{2}(n+v_e+2+1)} \times \exp \left\{ -\frac{1}{2} \text{tr}[(v_e - 2 - 1)\mathbf{S}_e + \mathbf{S}\mathbf{S}_e]\mathbf{R}_0^{-1} \right\} \quad [17] \end{aligned}$$

or

$$\mathbf{R}_0 | \mathbf{U}, \boldsymbol{\theta}, \mathbf{S}_e, v_e \sim IW_2((v_e - 2 - 1)\mathbf{S}_e + \mathbf{S}\mathbf{S}_e), n + v_e)$$

Now we proceed to derive the full conditional posterior densities of the underlying variable for CE, \mathbf{U}_{2o} , used in [15] and of the missing residuals, \mathbf{e}_{1m} and \mathbf{e}_{2m} , needed

for \mathbf{SS}_e in [17]. From [8], in general,

$$\begin{aligned} & p(\mathbf{e}_{1m}, \mathbf{U}_2 | \boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0, \mathbf{Y}_{1o}, \mathbf{Y}_{2o}, \mathbf{S}_g, \mathbf{S}_e, v_g, v_e) \\ &= p(\mathbf{e}_{1m}, \mathbf{U}_2 | \boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0, \mathbf{Y}_{1o}, \mathbf{Y}_{2o}) \\ &= p(\mathbf{U} | \boldsymbol{\theta}, \mathbf{R}_0) \times p(\mathbf{Y}_{2o} | \mathbf{U}_{2o}, \mathbf{t}) \end{aligned}$$

These distributions depend on which combination of records is observed for a calf: BW only, CE only or both BW and CE. For a particular calf, if a BW is observed ($U_{1o,i} = Y_{1o,i}$) but CE is not, we need only to sample $e_{2m,i}$ for CE, the distribution is only involved with $p(\mathbf{U} | \boldsymbol{\theta}, \mathbf{R}_0)$, which follows a univariate normal distribution with density:

$$\begin{aligned} & p(e_{2m,i} | \mathbf{U}_{-1}, \boldsymbol{\theta}, \mathbf{R}_0) \\ &= p(e_{2m,i} | Y_{1o,i}, \boldsymbol{\theta}, \mathbf{R}_0) \\ &= p(e_{2m,i} | e_{1o,i}, \mathbf{R}_0) \\ &= \phi(\mu, \sigma^2) \end{aligned} \tag{18}$$

where $\phi(\cdot)$ is a normal density function; $\mu = be_{1o,i} = b(u_{1o,i} - \mathbf{w}'_{1i}\boldsymbol{\theta}_1)$; \mathbf{w}_{1i} is the incidence vector associated with $\boldsymbol{\theta}_1$; $\sigma^2 = r_{22} - r_{12}^2/r_{11}$; $b = r_{12}/r_{11}$; $\{r_{ij}\} = \mathbf{R}_0$.

If both BW ($U_{1i} = Y_{1i}$) and CE ($Y_{2i} = k$) are observed, then only $U_{2o,i}$ needs to be sampled,

$$\begin{aligned} & p(U_{2o,i} | \mathbf{U}_{-i}, \boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0) \\ &= p(U_{2o,i} | Y_{1o,i}, Y_{2o,i}, \boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0) \\ &= p(U_{2o,i} | Y_{1o,i}, \boldsymbol{\theta}, \mathbf{R}_0) p(Y_{2o,i} | U_{2o,i}, \mathbf{t}) \\ &= p(U_{2o,i} | Y_{1o,i}, \boldsymbol{\theta}, \mathbf{R}_0) \left\{ \sum_{j=1}^4 I(t_{j-1} < U_{2o,i} \leq t_j) I(Y_{2o,i} = j) \right\} \end{aligned}$$

This is in a form of univariate truncated normal (TN) distribution such that

$$U_{2o,i} | Y_{1o,i}, Y_{2o,i} = k, \boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0 \sim TN(\mu, \sigma^2) \tag{19}$$

with $t_{k-1} < U_{2o,i} \leq t_k$, where $\mu = \mathbf{w}'_{2i}\boldsymbol{\theta}_2 + b(U_{1o,i} - \mathbf{w}'_{1i}\boldsymbol{\theta}_1)$, σ^2 as in [18] and \mathbf{w}_{2i} is the incidence vector associated with $\boldsymbol{\theta}_2$.

If only a CE ($Y_{2i} = k$) is observed, then both $e_{1m,i}$ and $U_{2o,i}$ need to be sampled from a truncated bivariate normal, ie,

$$\begin{aligned} & p(e_{1m,i}, U_{2o,i} | \mathbf{U}_{-i}, \boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0) \\ &= p(e_{1m,i}, U_{2o,i} | Y_{2o,i}, \boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0) \\ &= p(e_{1m,i}, U_{2o,i} | Y_{2o,i}, \boldsymbol{\theta}, \mathbf{R}_0) p(Y_{2o,i} | U_{2o,i}, \mathbf{t}) \\ &= p(e_{1m,i}, U_{2o,i} | Y_{2o,i}, \boldsymbol{\theta}, \mathbf{R}_0) \left\{ \sum_{j=1}^4 I(t_{j-1} < U_{2o,i} \leq t_j) I(Y_{2o,i} = j) \right\} \end{aligned}$$

or

$$\begin{bmatrix} e_{1m,i} \\ U_{2o,i} \end{bmatrix} \Big| Y_{2o,i} = k, \boldsymbol{\theta}, \mathbf{t}, \mathbf{R}_0 \sim TN \left(\begin{bmatrix} 0 \\ \mathbf{w}'_{2i} \boldsymbol{\theta}_2 \end{bmatrix}, \mathbf{R}_0 \right) \quad [20]$$

with $t_{k-1} < U_{2o,i} \leq t_k$.

Finally, the conditional posterior distribution of a threshold is uniform (Albert and Chib, 1993; Sorensen et al, 1995), if CE is not missing:

$$t_k | \mathbf{U}, \mathbf{t}_{-k} \sim U(\min(U_{2o,i} | Y_{2o,i} = k + 1), \max(U_{2o,i} | Y_{2o,i} = k)) \quad [21]$$

or

$$t_k | \mathbf{U}, \mathbf{t}_{-k} \sim U(\min[\min(U_{2o,i} | Y_{2o,i} = k + 1), t_{k+1}], \max[\max(U_{2o,i} | Y_{2o,i} = k), t_{k-1}]) \quad [21.1]$$

if CE is missing.

As mentioned previously, for $\mathbf{t} = (t_1, t_2, t_3)$, there is only one estimable threshold, which to estimate is arbitrary. We took:

$$\mathbf{t} = (t_1 = 0, t_2 = \text{constant} > 0, t_3) \quad [22]$$

Note that $t_1 < t_2$. If only three categories of CE scores are available, there is no need to estimate thresholds under this parameterization. If the fourth category was rare, it would be tempting to combine scores into three categories to avoid estimating thresholds.

GIBBS SAMPLING AND POST-GIBBS ANALYSIS

Densities [15]–[18] (or [19] or [20]), and [21] (or [21.1]) constitute the Gibbs sampler for our model. Gibbs sampling repeatedly draws samples from this set of full conditional posterior distributions. After burning-in, such drawn numbers are random samples, though dependent, from the joint posterior density [9]. Let the Gibbs samples of length m for a particular parameter, say for the direct genetic variance component g_{11} for BW, be $\mathbf{x} = \{x_i\}, i = 1, 2, \dots, m$. An estimate of the mean of the marginal posterior density, $p(g_{11} | \mathbf{Y})$, is:

$$\hat{\mu} = \sum_{i=1}^m x_i \quad [23]$$

and the posterior variance can be estimated by:

$$\hat{V}(g_{11} | \mathbf{Y}) = \frac{1}{m-1} \sum_{i=1}^m (x_i - \hat{\mu})^2 \quad [24]$$

Modes and medians can also be used to estimate location parameter of a posterior density (Wang et al, 1993), though usually requiring more Gibbs samples because the density needs to be estimated first. Both estimators of [23] and [24] are subject to

Monte Carlo errors. Because the Gibbs samples are correlated, one way to estimate Monte Carlo errors is to adopt standard time series analysis techniques as suggested by Geyer (1992) and used by Sorensen et al (1995).

ROUTINE GENETIC EVALUATION

The preceding sections describe a Bayesian analysis via Gibbs sampling for inferences about all the parameters in the model including fixed effects, (functions of) breeding values, genetic and residual covariance matrices, and thresholds based on marginal posterior densities. This is sensible because all uncertainties in estimating other parameters are taken into account when inference is made about a particular parameter of interest, say for the breeding value of a sire. However, it is computationally expensive to carry out large scale analyses routinely. A practical compromise is to estimate covariance matrices and thresholds using a full Bayesian analysis via Gibbs sampling once and subsequently to estimate location parameters based on conditional densities of location parameters given the estimated dispersion and threshold parameters. As data accumulate, covariance matrices and thresholds are reestimated. Explicitly, we suggest a two-stage procedure that might be useful for a large scale routine genetic evaluation program:

1) Estimate \mathbf{G}_0 , \mathbf{R}_0 and \mathbf{t} using mean (mode or median), based on their respective marginal posterior densities, $p(\mathbf{G}_0|\mathbf{Y})$, $p(\mathbf{R}_0|\mathbf{Y})$ and $p(\mathbf{t}|\mathbf{Y})$, via Gibbs sampling, dropping prior parameters \mathbf{S}_g , \mathbf{S}_e , v_g and v_e in notation for convenience; and

2) Estimate location parameters based on the following conditional density:

$$p(\boldsymbol{\theta}|\mathbf{Y}, \hat{\mathbf{t}}, \hat{\mathbf{G}}_0, \hat{\mathbf{R}}_0) \quad [25]$$

which is an approximation to the corresponding marginal density:

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \int_{\mathbf{t}, \mathbf{G}_0, \mathbf{R}_0} p(\boldsymbol{\theta}, \mathbf{t}, \mathbf{G}_0, \mathbf{R}_0|\mathbf{Y}) d\mathbf{t} d\mathbf{G}_0 d\mathbf{R}_0 \quad [26]$$

ie,

$$p(\boldsymbol{\theta}|\mathbf{Y}) \approx p(\boldsymbol{\theta}|\mathbf{Y}, \hat{\mathbf{t}}, \hat{\mathbf{G}}_0, \hat{\mathbf{R}}_0) \quad [26.1]$$

if the marginal density, $p(\mathbf{t}, \mathbf{G}_0, \mathbf{R}_0|\mathbf{Y})$, is symmetric or peaked (Box and Tiao, 1973; Gianola and Fernando, 1986). In other words, if there is sufficient information in the data to estimate \mathbf{G}_0 , \mathbf{R}_0 and \mathbf{t} well, then [25] is a good approximation to [26]. This would be the case if the data set is one of the national data bases, for example, the American Simmental data set. Note that the underlying variable for CE (\mathbf{U}_{20}) and the residual vector associated with the missing data (\mathbf{e}_{1m} and \mathbf{e}_{2m}) have been integrated out of the joint posterior density [9]; that \mathbf{W} matrix no longer contains blocks of $\mathbf{0}$ s corresponding to the missing data, however, the same notation is kept below. Note also that o and m denoting observed and missing data are dropped from the notation because missing data no longer play a role.

The joint posterior mode of [25] can be considered as point estimates for $\boldsymbol{\theta}$, which is also known as maximum a posteriori, or MAP for short (Gianola and

Foulley, 1983). There are at least two ways to compute the MAP estimates: expectation-maximization (EM) (Zhao, 1987; Quaas, 1994, 1996) and Newton-Raphson (Gianola and Foulley, 1983; Foulley et al, 1983).

The EM iteration equation (Quaas, 1996) is:

$$[\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{\Omega}] [\hat{\boldsymbol{\theta}}]^{[i+1]} = \mathbf{W}'\mathbf{R}^{-1}\hat{\mathbf{U}}^{[i]} \quad [27]$$

where the coefficient matrix is exactly the same as that in the usual MME with $\mathbf{\Omega}$ containing $\mathbf{G}_0^{-1} \otimes \mathbf{A}^{-1}$ and blocks of $\mathbf{0}$ s corresponding to the fixed effects, the superscript in [27] indicates iteration number, and $\hat{\mathbf{U}}' = [\mathbf{U}'_1 \ \hat{\mathbf{U}}'_2]$. For a particular record, with $Y_{2i} = k$,

$$\hat{U}_{2i} = \mu + \sigma\lambda \quad [28]$$

where

$$\mu = \begin{cases} \mathbf{w}'_{2i}\boldsymbol{\theta}_2 + b(Y_{1i} - \mathbf{w}'_{1i}\boldsymbol{\theta}_1) & \text{if both BW and CE observed} \\ \mathbf{w}'_{2i}\boldsymbol{\theta}_2 & \text{if only CE observed} \end{cases} \quad [28.1]$$

$$\sigma^2 = \begin{cases} r_{22} - r_{12}^2/r_{11} & \text{if both BW and CE observed} \\ r_{22} & \text{if only CE observed} \end{cases} \quad [28.2]$$

with b as in [18],

$$\lambda = \frac{-\phi(Z_k) + \phi(Z_{k-1})}{P_k} \quad [28.3]$$

and $P_k = \Phi(Z_k) - \Phi(Z_{k-1})$; $Z_\ell = (t_\ell - \mu)/\sigma$

The Newton-Raphson iteration equation (Janss and Foulley, 1993; Quaas, 1994, 1996; Hoeschele et al, 1995), following closely the notation of Quaas (1996), is:

$$[\mathbf{W}'\tilde{\mathbf{R}}^{-1}\mathbf{W} + \mathbf{\Omega}][\hat{\boldsymbol{\theta}}]^{[i+1]} = \mathbf{W}'\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{U}}^{[i]} \quad [29]$$

where, for a particular record,

$$\tilde{U}_{2i} = \mu - \gamma^{-1}\delta \quad [30]$$

with μ as in [28.1],

$$\delta = \frac{\phi(Z_k) - \phi(Z_{k-1})}{P_k\sigma} \quad [31]$$

σ as in [28.2], P_k as in [28.3], and

$$\gamma = \frac{\phi(Z_k)Z_k}{P_k\sigma^2} + \frac{-\phi(Z_{k-1})Z_{k-1}}{P_k\sigma^2} + \left[\frac{\phi(Z_k)}{P_k\sigma} + \frac{-\phi(Z_{k-1})}{P_k\sigma} \right]^2 \quad [32]$$

Structurally, $\tilde{\mathbf{R}}$ in [29] is the same as \mathbf{R} in [2] but consisting of $\tilde{\mathbf{R}}_0$ matrices as specified by:

$$\tilde{\mathbf{R}}_0^{-1} = \begin{bmatrix} \tilde{r}_{11} & \tilde{r}_{12} \\ \tilde{r}_{21} & \tilde{r}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} b^2\gamma + r_{11}^{-1} & -b\gamma \\ -b\gamma & \gamma \end{bmatrix} \quad [33]$$

with b as in [18] and r_{11} the residual variance of BW. It is clear that $\tilde{\mathbf{R}}_0$ depends on $\boldsymbol{\theta}$, \mathbf{t} , \mathbf{R}_0 and \mathbf{Y} ; thus it is record specific. If an animal is missing BW or CE, $\tilde{\mathbf{R}}_0^{-1}$ is a scalar: γ or r_{11}^{-1} , respectively.

A Bayesian analog of the Beef Improvement Federation measure of ‘accuracy’ of a bull’s genetic prediction (BIF, 1990) is:

$$\text{BIF accuracy} = 1 - \sqrt{\frac{\text{posterior var}(a_i)}{\text{prior var}(a_i)}} \quad [34]$$

If information contained in the data conflicts with prior belief, then posterior variance could be larger than prior variance resulting in a negative accuracy. This is peculiar to a frequentist, particularly to a producer, why after collecting data on his animal, the uncertainty about his animal’s BV has increased! Posterior variances of breeding values are usually approximated, based on large sample theory, by the inverse of Fisher’s (expected) information matrix or by the inverse of negative Hessian matrix. The latter approximation is:

$$\text{Var}(\boldsymbol{\theta}|\mathbf{Y}) = \left[-\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} = \left[\mathbf{W}' \tilde{\mathbf{R}}^{-1} \mathbf{W} + \boldsymbol{\Omega} \right]^{-1} \quad [35]$$

where $l = \log\{[25]\}$. This is the inverse of the coefficient matrix of [29]. Note that [27], the inverse of the coefficient matrix used for EM, is not a large sample approximation to the posterior variance matrix of [25], or, at least, not a very good one. We shall return to this point in the numerical example section below.

NUMERICAL EXAMPLE

A data set representing continuous BW and discrete CE scores with ordered categories, 1–4, of ‘Simmental calves’ was simulated and analyzed to illustrate the methodology.

Model

The records were simulated under the sire–maternal grandsire model [1]. Model equations for the k th calf with i th bull as sire and j th bull as maternal grandsire were

$$U_{1ijk} = \mu_1 + a_{1i} + 1/2a_{1j} + e_{1ijk}$$

$$U_{2ijk} = \mu_2 + a_{2i} + m_{2j} + e_{2ijk}$$

for BW with only direct effects and CE underlying variable with both direct and maternal effects, respectively. The random vectors $[a_{1k}, a_{2k}, m_{2k}]$ were drawn from a $N(\mathbf{0}, \mathbf{G}_0)$ distribution, ie, bulls were unrelated. Similarly $[e_{1ijk}, e_{2ijk}] \sim N(\mathbf{0}, \mathbf{R}_0)$. Parameter values were similar to previous estimates from the Simmental data (Dong et al, 1991). A residual correlation of 0.6 and a genetic correlation matrix,

$$\begin{bmatrix} 1.00 & -0.40 & 0.15 \\ & 1.00 & -0.10 \\ & & 1.00 \end{bmatrix}$$

were scaled to obtain \mathbf{R}_0 and \mathbf{G}_0 . Assuming the phenotypic SD for BW was nine units and $h_{bw}^2 = 0.4 \Rightarrow \sigma_{a_1}^2 = 1/4 \times 0.4 \times 81$. The CE components, σ_{a_2} and σ_{m_2} , were assumed equal and computed by solving $h_{ce}^2 = \frac{4\sigma_x^2}{2\sigma_x^2 + \sigma_{e_2}^2}$ for $\sigma_x^2 = \sigma_{a_2}^2 = \sigma_{m_2}^2$. The residual SD which establishes the scale of \mathbf{U}_2 was set to ten to be comparable with BW; $h_{ce}^2 = 0.2$. The dispersion matrices were

$$\mathbf{G}_0 = \begin{bmatrix} 8.100 & -2.6833 & 1.0062 \\ & 5.5556 & -0.5556 \\ & & 5.5556 \end{bmatrix} \text{ and } \mathbf{R}_0 = \begin{bmatrix} 70.8750 & 50.5124 \\ & 100.0000 \end{bmatrix}$$

As found in previous analyses of Simmental data (Zhao, 1986, 1987; Quaas et al, 1988), thresholds were equally spaced at 1 SD intervals: $t_1 = 0$, $t_2 = 10$ and $t_3 = 20$. Corresponding to $\sim 65\%$ unassisted births of Simmental calves out of 2-year-old dams (unpublished data):

$\mu_2 = -\sigma_{e_2} \Phi^{-1}(0.65) = -3.8532$. Thus, $P_k = \Phi\{t_k - \mu_2\}/\sigma_{e_2} - \Phi\{t_{k-1} - \mu_2\}/\sigma_{e_2} = 0.6500, 0.2670, 0.0744$ and 0.0085 , for $k = 1, 2, 3, 4$, respectively. For BW, $\mu_1 = 85$. Calving ease scores were assigned such that $Y_{2ijl} = k$ if $t_{k-1} < U_{2ijl} \leq t_k$.

Design

There were 75 bulls in three batches of 25. Bulls in the first batch were MGS of the calves sired by bulls in the second batch; bulls 26–50 had two progeny from daughters of each of the first 25 bulls. Likewise, bulls 51–75 (third batch) had two progeny from daughters of each of the second 25 bulls (second batch). Thus bulls 1–50 were each MGS of 50 calves; bulls 26–75 were each sires of 50 calves; only bulls 26–50 (second batch) were both sires and MGS. There were a total of 2500 calves, each with a BW and a CE score.

Gibbs sampling

Priors for $\boldsymbol{\mu}$, \mathbf{t} and \mathbf{R}_0 were uniform while that for \mathbf{G}_0 was the inverted Wishart in [5] with \mathbf{S}_g a diagonal matrix of the genetic variances used for simulation and $v_g = 5$. The latter ‘slightly informative’ prior was adopted because during initial testing of the programs, a nearly singular \mathbf{SS}_g would cause the sampler to crash. The last \mathbf{G}_0 would look reasonable but would have an effectively zero eigenvalue. The informative prior was incorporated to ensure $(\mathbf{SS}_g + \mathbf{S}_g)$ was positive definite. Subsequently, the problem was found in the simulation program where both h_{ce}^2 were almost zero; \mathbf{G}_0 was nearly singular! The safeguard, however, was left in the program.

The initial dispersion matrices were diagonal matrices of the variances used for simulation, bull effects were null and the average BW was assigned to μ_1 . Thresholds and μ_2 were computed from the observed frequencies: $\mu_2 = 10 \times [-\Phi^{-1}(C_1)]$ and $t_k = 10 \times [\Phi^{-1}(C_k) - \Phi^{-1}(C_1)]$ where C_k is the observed cumulative frequency of scores 1 through k , $k = 1, 2, \dots, 3$. The values for $t_1 (= 0)$ and t_2 were fixed; t_3 was estimated. Initial values are displayed in table I.

Table I. Parameter values.

	t_3	μ_1	μ_2	ξ_{11}	ξ_{12}	ξ_{13}	ξ_{22}	ξ_{23}	ξ_{33}	Γ_{11}	Γ_{12}	Γ_{22}
True ^a	20.0000	85.0000	-3.8532	8.1000	-2.6833	1.0062	5.5556	-0.5556	5.5556	70.8750	50.5124	100.0000
Initial values ^b	18.4162	85.0015	-3.4459	8.1000	0.0000	0.0000	5.5556	0.0000	5.5556	70.8750	0.0000	100.0000
Posterior mean ^c	18.3440	85.0865	-3.3049	7.8940	-3.2421	1.7627	4.3004	-0.1706	4.4224	72.4930	44.3328	88.4873
Posterior SD ^c	0.7443	0.3491	0.2993	1.7323	1.1965	1.2960	1.4106	1.1137	1.4300	2.0758	2.1597	5.6807

^a Values of parameters used to simulate data; $t_1 = 0$, $t_2 = 10$. ^b Initial values of Gibbs chain, see text for how initial values were chosen; $t_1 = 0$ (fixed), $t_2 = 9.3696$ (fixed). ^c Average (SD) of 10 000 Gibbs samples after a burn-in of 500 samples.

From these starting values the parameters were repeatedly sampled in the following order: U_{2ijk} from [19], t_3 from [21], $(\mu_1 \mu_2)$ jointly from [15], $(a_{1i} a_{2i} m_i)$ jointly from [15], \mathbf{G}_0 from $IW_3(\mathbf{SS}_g + \mathbf{S}_g, 75+5)$, [16], and \mathbf{R}_0 from $IW_2(\mathbf{SS}_e, 2\,500-3)$, [17]. Each bull's three effects were sampled jointly as were the two μ s. Most of the results to be presented came from a single chain of 10 500 samples; all summary statistics were computed from the last 10 000. We feel a single chain of this length is sufficient to demonstrate the procedures but would probably not suffice for a real application. A few references will be made to results from replicated Gibbs chains and replicated data. Our purpose here, however, was not to study Monte Carlo error, single long chain versus multiple short chains nor the small sample properties of posterior means; no systematic study was carried out on these replications. Comparisons of frequentist properties of estimators of certain parameters are found in Jensen (1994) and Hoeschele and Tier (1995).

The most time consuming step was sampling the CE underlying variables. This was due partly to the number of these but also because of the way the truncated normals in [19] were generated, ie, normals were sampled until one was obtained within the range determined by the CE score. The time for a complete cycle could almost triple; this variability was entirely due to generating the truncated normals. A more efficient truncated normal generator might be necessary for a large data set in which the probability of some categories is very small, eg, a score of four for a heifer calf out of a mature dam. In several million records, there will be a handful of these and they will cause problems.

Joint conditional posterior modes

Joint posterior modes of $\boldsymbol{\mu}$ and the bull effects given \mathbf{t} , \mathbf{R}_0 and \mathbf{G}_0 [25] were computed by both the EM [27] and Newton–Raphson [29] algorithms. The values of \mathbf{t} , \mathbf{R}_0 , and \mathbf{G}_0 were fixed at the estimates of their posterior means from the Gibbs sampler. Starting values were the same as in the Gibbs sampler. Iteration was continued until \log_{10} of the maximum absolute change of any bull effect was < -10 at which point the modal estimates of $\boldsymbol{\mu}$ and bull effects computed by the alternative algorithms differed by $< 10^{-10}$.

NUMERICAL RESULTS

Visual inspection of plots of parameters (or functions of parameters) against sample number suggests that a burn-in of 500 cycles was probably unnecessary (eg, fig 1). The parameters in figure 1 were chosen to illustrate the markedly different patterns observed as a result of correlations among sequential Gibbs samples. There were marked differences among parameters in these serial correlations (table II) which were particularly high for the threshold, t_3 , and also for the CE residual variance, r_{22} . In contrast, mixing for the other r_{ij} , all g_{ij} and the bull effects was much more rapid. However, overall convergence of a Gibbs chain is tied to parameters with slow mixing properties. In another words, one cannot declare that a Gibbs chain is converged for some parameters but not others.

Estimates of posterior means and SDs are presented in table I. The parameters involving the continuous trait, BW, were reasonably close to the values used to

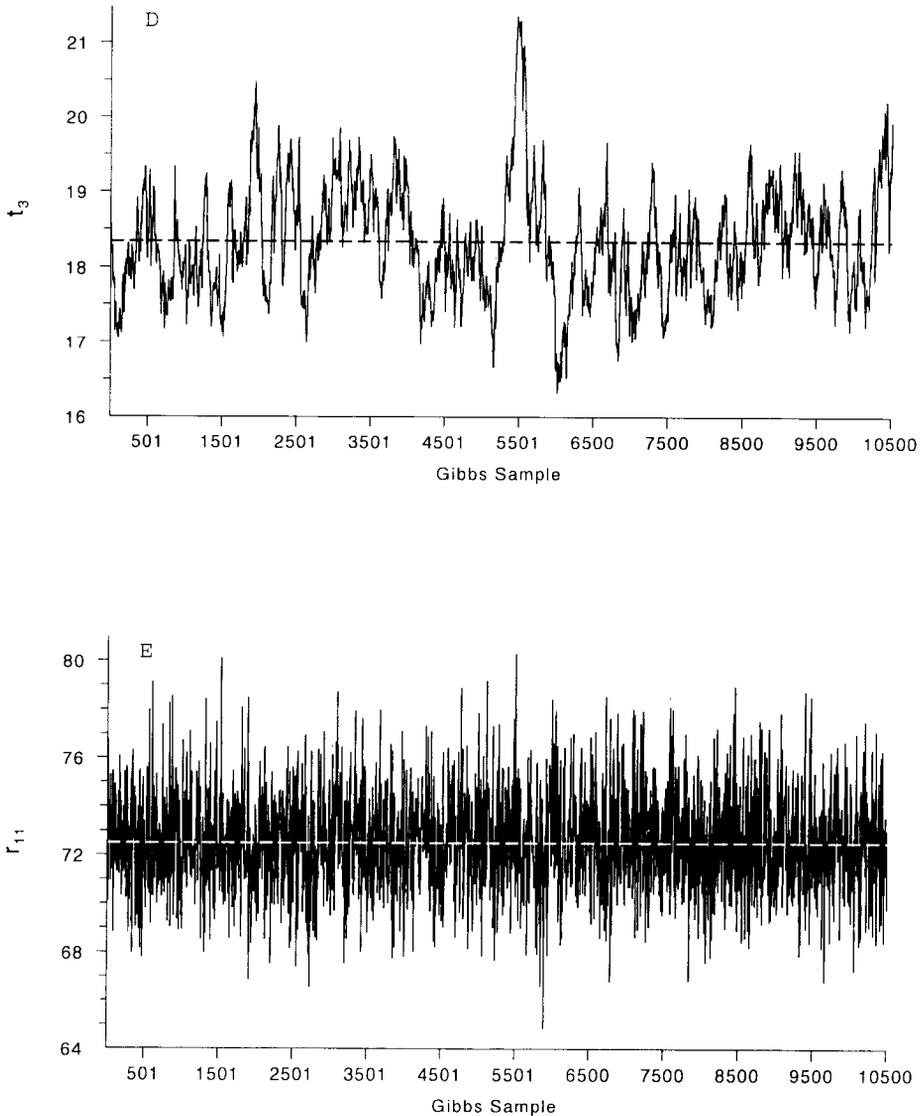


Fig 1. Gibbs sample values versus Gibbs cycle. (Every fifth sample for t_3 and r_{11} is plotted.)

simulate the data. This was not the case, however, for the parameters affecting CE. The dispersion parameters, especially, were markedly smaller than the 'true' values. Too much cannot be made of this from a single sample and we cannot conclude that such a pattern is a consequence of using the marginal posterior mean as an estimator. It is quite possible that it is a matter of identifiability. For example,

Table II. Serial correlations: t_3 , μ , \mathbf{G}_0 and \mathbf{R}_0 .

Lag	Parameter											
	t_3	μ_1	μ_2	g_{11}	g_{12}	g_{13}	g_{22}	g_{23}	g_{33}	r_{11}	r_{12}	r_{22}
1	0.990	0.892	0.830	0.406	0.560	0.687	0.722	0.777	0.718	0.033	0.230	0.800
11	0.894	0.382	0.224	-0.014	0.048	0.049	0.121	0.145	0.074	-	-0.004	0.392
21	0.806	0.207	0.108	-	-	-	0.054	0.039	-	-	-	0.311
31	0.723	0.113	0.088	-	-	-	-	-	-	-	-	0.284
41	0.656	0.049	-	-	-	-	-	-	-	-	-	0.275
51	0.602	-	-	-	-	-	-	-	-	-	-	0.270
101	0.370	-	-	-	-	-	-	-	-	-	-	0.133
151	0.193	-	-	-	-	-	-	-	-	-	-	0.077
201	0.103	-	-	-	-	-	-	-	-	-	-	-

neither t_k nor σ_{ce} is identifiable but $(t_k - t_{k-1})/\sigma_{ce}$ is identifiable. The ‘true’ thresholds were equally spaced 1 residual SD apart; analogous functions of t_1, t_2 (fixed) and the posterior means of t_3 and r_{22} are 0.996 and 0.954, quite close to the ‘true’ values. Likewise, estimated heritabilities and genetic correlations, table III, differed considerably less from the ‘true’ values than did the (co)variances.

Table III. Heritabilities and genetic correlations.

	h_1^2	$r_{g_{12}}$	$r_{g_{13}}$	h_2^2	$r_{g_{23}}$	h_3^2
From true g_{ij} and r_{ij}	0.4000	-0.4000	0.1500	0.2000	-0.1000	0.2000
From \hat{g}_{ij} and \hat{r}_{ij} ^a	0.3834	-0.5564	0.2983	0.1770	-0.0391	0.1820
Posterior mean ^b	0.3818	-0.5648	0.2923	0.1765	-0.0405	0.1816
Posterior SD ^b	0.0738	0.1591	0.1927	0.0549	0.2515	0.0559

^a \hat{g}_{ij} and \hat{r}_{ij} are posterior means of g_{ij} and r_{ij} , respectively. ^b Average (SD) of 10 000 transformed Gibbs samples after a burn-in of 500 samples.

For this small example precise estimates of the marginal posterior densities were not attempted. Coarse histograms were examined; most were slightly skewed away from zero; in figure 2 are typical examples. None were ‘peaked’ in the eyes of the authors.

Though the EM and Newton–Raphson algorithms both gave the same values for the joint conditional modes of [25], there was a marked difference in the pattern of convergence: linear versus quadratic (fig 3). The EM algorithm took six times as many rounds, but elapsed times were close: EM < two times longer. For a large, field data application it is questionable as to which algorithm will be faster; we do not expect a large difference. While EM is easier to code, approximations of SDs are not a by-product as they are if the Hessian is computed.

Correlations among bull effects are presented in table IV. The joint modes given the posterior mean estimates of t_3 , \mathbf{G}_0 and \mathbf{R}_0 , [25], were very highly correlated

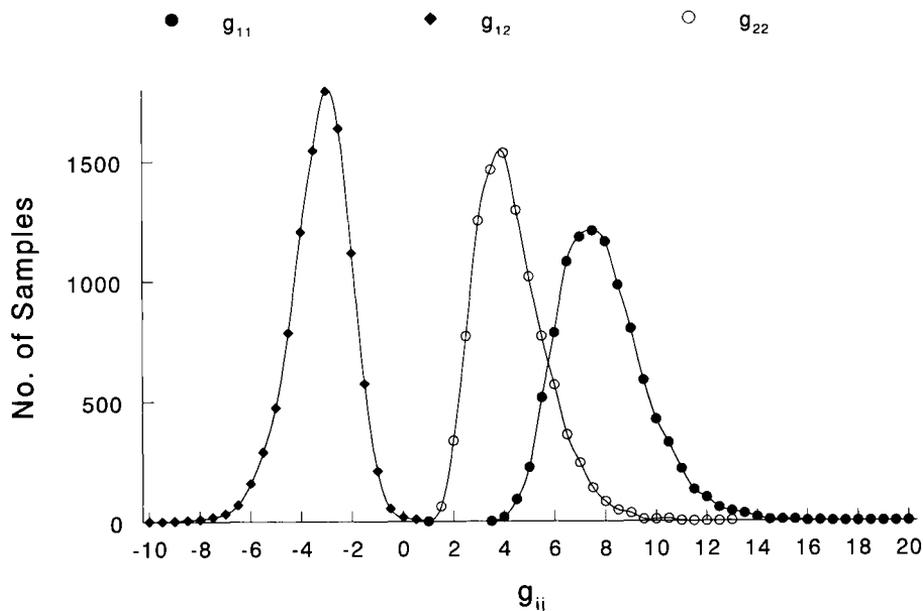


Fig 2. Approximate shape of marginal posterior densities for three elements of G_0 . (Spline curve fit to rough histogram.)

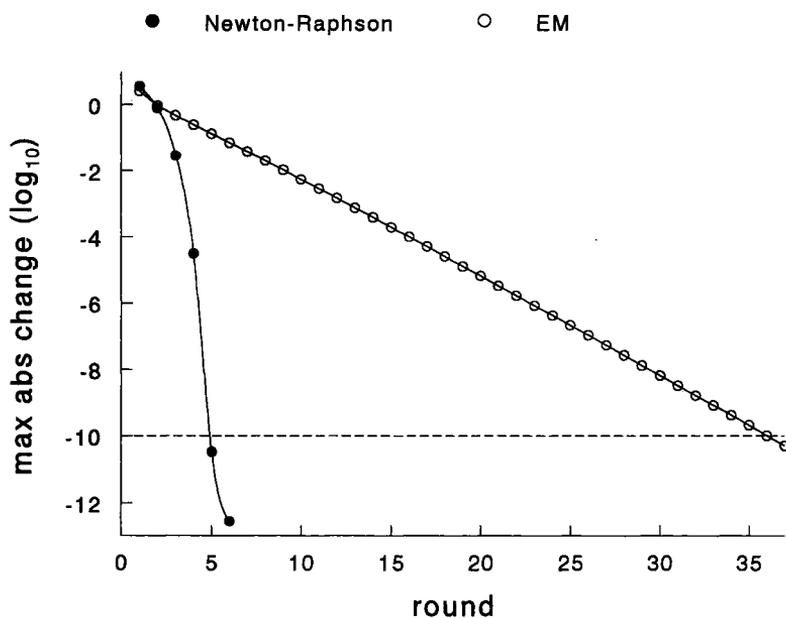


Fig 3. Convergences of Newton-Raphson and EM algorithms in computing the joint conditional posterior mode of [25].

Table IV. Correlations between true and/or estimated bull effects for three simulated data sets. (Data set 3 is that summarized elsewhere.)

Trait	Data set	True versus mode ^a			True versus mean ^b			Mode versus mean		
		bulls 1-25	bulls 26-50	bulls 51-75	bulls 1-25	bulls 26-50	bulls 51-75	bulls 1-25	bulls 26-50	bulls 51-75
bw	1	0.76	0.92	0.94	0.76	0.92	0.94	0.9999	0.9999	0.9999
	2	0.79	0.93	0.95	0.78	0.93	0.95	0.9997	0.9998	0.9999
	3	0.81	0.95	0.95	0.81	0.95	0.95	0.9999	0.9998	1.0000
ce-d	1	0.05	0.81	0.89	0.06	0.81	0.88	0.9997	0.9994	0.9998
	2	0.43	0.80	0.89	0.43	0.79	0.89	0.9994	0.9994	0.9997
	3	0.35	0.80	0.84	0.35	0.80	0.83	0.9997	0.9989	0.9991
ce-mgs	1	0.73	0.88	-0.26	0.73	0.88	-0.26	0.9999	0.9997	0.9996
	2	0.78	0.85	0.03	0.77	0.85	0.03	0.9999	0.9996	0.9988
	3	0.77	0.85	0.30	0.74	0.85	0.31	0.9993	0.9994	0.9993

^a Joint posterior mode given posterior mean estimates of t_3 , \mathbf{G}_0 and \mathbf{R}_0 . ^b Marginal posterior mean.

to the marginal posterior means of [26] despite the posterior densities for t_3 and dispersion parameters being neither symmetric nor particularly peaked. The similarity of marginal mean and joint conditional mode is graphically illustrated for the CE-D effect in figure 4. The regression of mode on mean is slightly greater than unity, 1.08, reflecting the greater spread of modes when there is no uncertainty in thresholds nor dispersion parameters. This was seen for all three bull effects.

Correlations, within bull batch, were also computed between true and estimated (mode or mean) bull effects. These correlations caused debate among the authors. The Bayesian amongst us was not sure how to interpret them whereas the lapsed frequentists had no difficulty whatsoever: a big number is good; a small number is not so good. Except for CE-D for batch 1 bulls and CE-MGS for batch 3 bulls where the estimation came entirely from correlated information, these correlations seem satisfyingly large. Of some interest was the observation that the correlations between true and estimated direct effects, both BW and CE-D, were a bit higher for the batch 3 bulls than for the batch 2 bulls. Both had 50 progeny but the latter also had 50 grandprogeny. What seems to be adding information from grandprogeny, apparently, is noise. This could be due to sampling, though the same pattern was seen in three independent samples of simulated data.

The posterior SD were also examined. These were approximated in three ways: from the coefficient matrix of the MME used in EM [27], from the Hessian used in Newton-Raphson [29] and by Gibbs sampling. In applications these are sometimes used interchangeably but they are not, of course, the same. The first is the posterior SD of $p(\boldsymbol{\theta}|\mathbf{U}_1, \hat{\mathbf{U}}_2, \hat{\mathbf{G}}_0, \hat{\mathbf{R}}_0)$ or the frequentist's standard error of prediction (SEP) for BLUP $p(\boldsymbol{\theta}|\mathbf{U}_1, \hat{\mathbf{U}}_2, \hat{\mathbf{G}}_0, \hat{\mathbf{R}}_0)$ note that this assumes the CE underlying variable is observed. The second is the large sample approximation of the joint posterior SD given t_3 , \mathbf{G}_0 and \mathbf{R}_0 . The last is the marginal posterior SD, subject to Monte

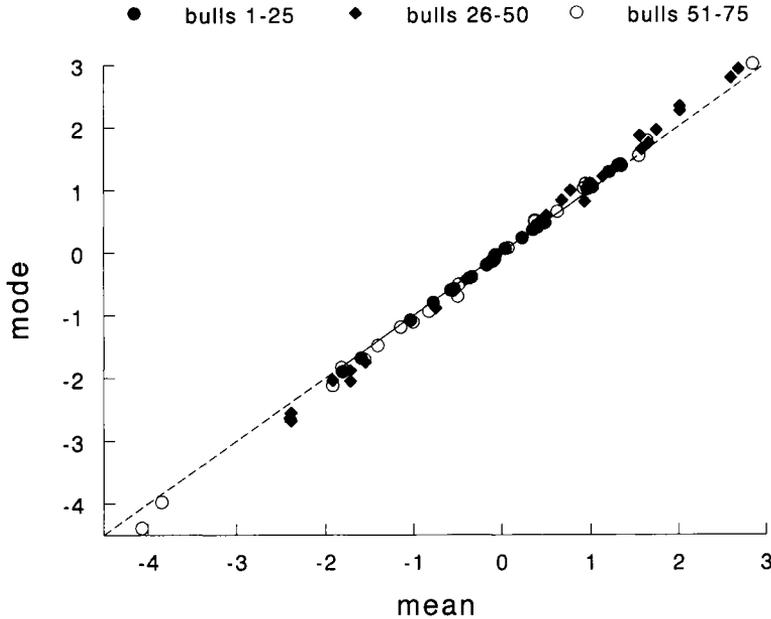


Fig 4. Joint conditional posterior mode of [25] versus marginal posterior mean of [26] for the bulls' CE-D effects. —, Unit slope.

Carlo error. These are plotted for the bulls' BW effects in figure 5. It is immediately obvious that the first behaves quite differently to either of the latter two. The former depends only on \mathbf{G}_0 , \mathbf{R}_0 and the information that is available, hence a different value for each of the three batches of bulls. The second also depends on \mathbf{G}_0 and \mathbf{R}_0 (and t_3) but also on $\hat{\boldsymbol{\mu}}$, $\hat{\mathbf{a}}_1$, $\hat{\mathbf{a}}_2$ and \mathbf{m}_2 minimally, hence the values vary a bit but noticeably so only for the bulls with progeny. (Considerably more variation was seen for the CE effects and the SDs were larger, in general, reflecting the fact that \mathbf{U}_2 was not actually observed.) In marked contrast, the marginal posterior SD are all over the place. This is not just due to Monte Carlo error, see figure 6 where results from two independent Gibbs samples are plotted. It is due to the dependence of the SD on the posterior means. This is shown graphically in figure 7. Presumably this pattern is due primarily to the uncertainty of the dispersion parameters. Heuristically, a bull with average progeny BW will have zero \hat{a}_{1k} regardless of the value of (or uncertainty about) h_{bw}^2 whereas the uncertainty about (SD) \hat{a}_{1k} for a bull whose progeny BW depart markedly from average will depend on uncertainty about h_{bw}^2 .

DISCUSSION AND CONCLUSIONS

Following closely the work of Sorensen et al (1995), we have extended the Gibbs sampling scheme to make full Bayesian inferences about location, dispersion, and thresholds from modeling one multiple ordered categorical trait (CE) and one continuous variable (BW) with the possibility of missing patterns of data. Inferences

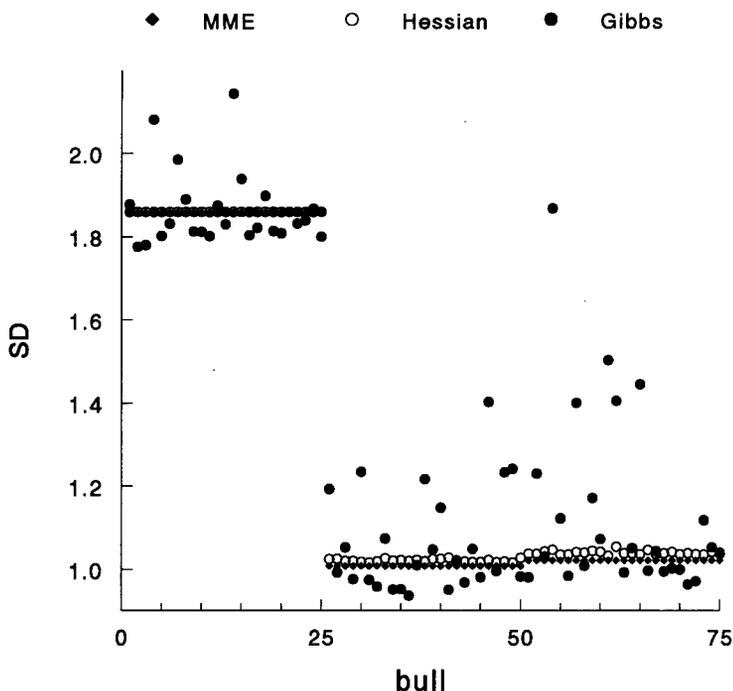


Fig 5. Posterior SD for bulls' BW effects approximated from the coefficient matrix of the MME used in EM [27], the Hessian used in Newton–Raphson [29] and Gibbs samples.

about genetic and residual covariance matrices and thresholds are based on their respective marginal posterior densities in a unified fashion without analytical approximations, in contrast to traditional methods based on approximations (Foulley et al, 1987; Harville and Mee, 1984).

Our model can be expanded to include heterogeneous variances in a similar way as for linear models (Gianola et al, 1992). Foulley and Gianola (1996) expanded a log-linear structural model to describe heterogeneous variances (Foulley et al, 1990; San Cristobal et al, 1993) to a single-trait threshold model. This idea can be extended to model heterogeneous variances in our situation as well.

It is advantageous to estimate location parameters including fixed effects and breeding values jointly with dispersion parameters because uncertainties in estimation of dispersion parameters can be taken into account, particularly in small populations. However, in real genetic evaluation systems with data sets of millions of records, joint modeling may be neither possible nor necessary. With good estimates for dispersion parameters and thresholds from the joint posterior density [9] from large data sets, we argue that the conditional posterior density of location parameters [25] with such estimated dispersion parameters as fixed values is a good approximation to the marginal posterior density of location parameters [26]. In our numerical example the joint mode of [25] approximated well the marginal posterior means of [26] even though the dispersion parameters were not very well estimated.

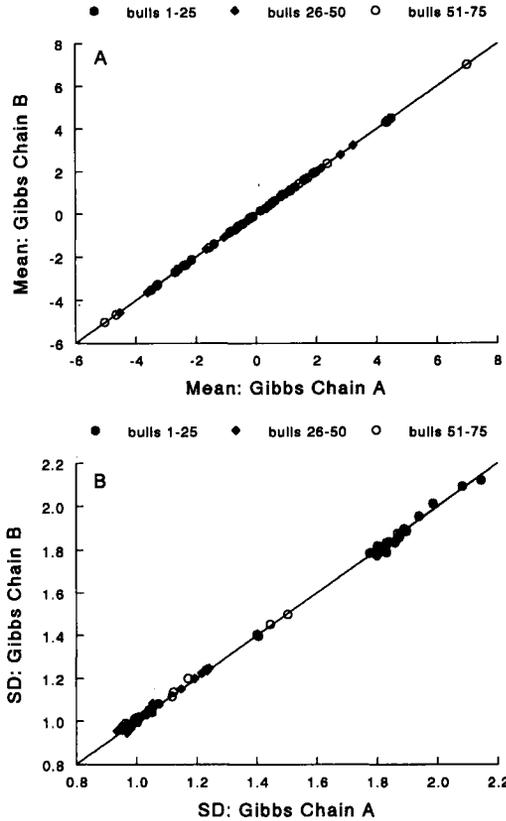


Fig 6. Marginal posterior means and SD of [26] for bulls' BW effects from replicate Gibbs chains. –, Unit slope.

The posterior SD of [26], however, was not well approximated by the inverse of the coefficient matrix of [27] or [29]. The latter are large sample approximations; our data were not numerous.

A main purpose of the data augmentation used in the paper was to result in fully conditional posterior distributions of parameters in standard recognizable forms such that samples were easily drawn. We chose to augment the underlying variables of observed CE (\mathbf{U}_{2o}) and the residuals associated with missing BW and CE (\mathbf{e}_{1m} and \mathbf{e}_{2m}); the whole augmented data vector was ($\mathbf{e}'_{1m} \mathbf{U}'_{2o} \mathbf{e}'_{2m}$). Another way is to augment the underlying variables of observed CE (\mathbf{U}_{2o}) and that of missing CE (\mathbf{U}_{2m}), and continuous data corresponding to missing BW (\mathbf{Y}_{1m}); the whole augmented data vector is ($\mathbf{Y}'_{1m} \mathbf{U}'_{2o} \mathbf{U}'_{2m}$). Sorensen (1996) gave a parallel treatment of the problem under the latter data augmentation. In general, the whole design matrix \mathbf{W} is considered to be fixed. In other words, the design vector \mathbf{w}_i associated with an observed or a missing (either BW or CE) calf record is assumed to be either fully or partially known and fixed. Our approach to data

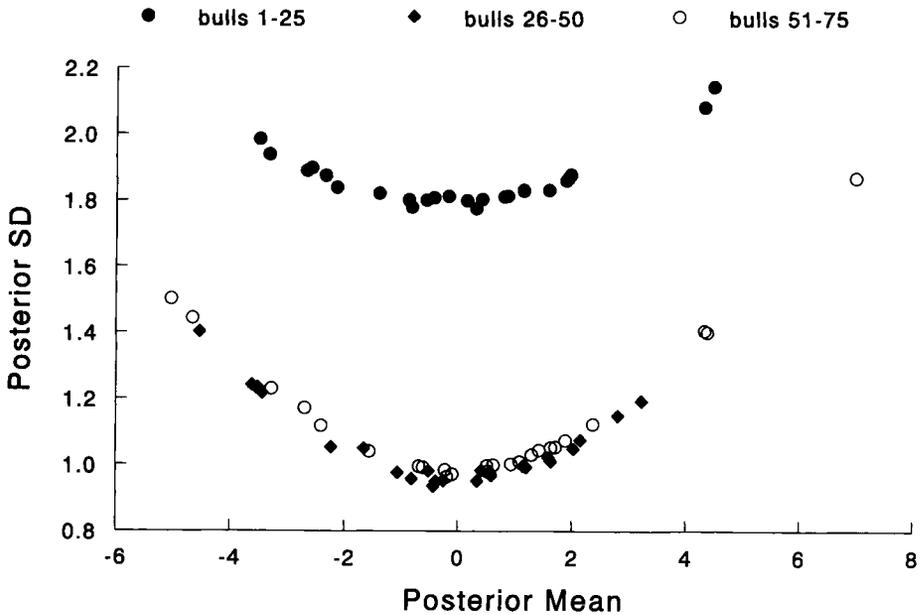


Fig 7. Marginal posterior SD versus means of [26] for bulls' BW effects. —, Unit slope.

augmentation was, implicitly, equivalent to treating \mathbf{w}_i s associated with missing BW and CE as random and to assigning uniform priors to them, and integrating them out of the joint posterior density, as opposed to Sorensen (1996) in which the joint posterior density was conditioned on those \mathbf{w}_i s of missing BW and CE. We conjecture that our approach has computational advantages while not sacrificing theoretical simplicity, particularly when the missing data rate is high. For example, for American Simmental, about 1/3 of the records have either BW or CE missing. It is clear also that other combinations of data augmentation are possible, such as $(\mathbf{e}'_{1m} \mathbf{U}'_{2o} \mathbf{U}'_{2m})$ ($\mathbf{Y}'_{1m} \mathbf{U}'_{2o} \mathbf{e}'_{2m}$) or $(\mathbf{e}'_{1m} \mathbf{e}'_{2o} \mathbf{e}'_{2m})$. If no data augmentation is used, fully conditional distributions of certain parameters may not be in recognizable forms and alternative sampling procedures such as rejection sampling need to be used. Zeger and Karim (1991) presented algorithms for a single trait threshold model without data augmentation.

There is a danger in using improper priors in a Bayesian analysis. In a linear model setting, some improper priors induce improper joint posterior densities even though the fully conditional posterior densities are well defined (Hobert and Casella, 1996). We do not know whether or not the uniform priors we used in the numerical example for $\boldsymbol{\mu}$, \mathbf{t} and \mathbf{R}_0 will induce a proper joint posterior density [9]. The fact that no difficulties were encountered in analysis does not necessarily mean that [9] was suitable. The safe way may be to use a noninformative but proper prior in an analysis.

ACKNOWLEDGMENTS

We thank the American Simmental Association, Bozeman MT, for partially supporting this work. We thank D Sorensen for sending us a copy of his lecture notes (Sorensen, 1996) in which an alternative data augmentation is detailed, and are indebted to him along with an anonymous reviewer for numerous constructive criticisms which improved the presentation of the paper considerably.

REFERENCES

- Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J Am Stat Assoc* 88, 669–679
- Berger PJ, Lin EC, Van Arendonk J, Janss L (1995) Properties of genetic parameter estimates from selection experiments by Gibbs sampling. *J Dairy Sci* 78 (Suppl 1), 246
- BIF (1990) *BIF Guidelines for Uniform Beef Improvement Programs*. Beef Improvement Federation, Athens, 6th ed
- Box GEP, Tiao GC (1973) *Bayesian Inference in Statistical Analysis*. Wiley, New York
- Dong MC, Quaas RL, Pollak EJ (1991) Estimation of genetic parameters of calving ease and birth weight by a threshold model. *J Anim Sci* 69 (Suppl 1), 204
- Foulley JL, Gianola D, Thompson R (1983) Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Genet Sel Evol* 15, 407–424
- Foulley JL, Im S, Gianola D, Hoeschele I (1987) Empirical Bayes estimation of parameters for n polygenic binary traits. *Genet Sel Evol* 19, 197–224
- Foulley JL, Gianola D, San Cristobal M, Im S (1990) A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models. *J Dairy Sci* 73, 1612–1624
- Foulley JL, Gianola D (1996) Statistical analysis of ordered categorical data via a structural heteroskedastic threshold model. *Genet Sel Evol* 28, 249–273
- Garcia-Cortes LA, Sorensen DA (1996) On a multivariate implementation of the Gibbs sampler. *Genet Sel Evol* 28, 121–126
- Garrick DJ, Pollak EJ, Quaas RL, Van Vleck LD (1989) Variance heterogeneity in direct and maternal weight traits by sex and percent purebred for Simmental-sired calves. *J Anim Sci* 67, 2515–2528
- Gelfand AE, Hills SE, Racine-Poon A, Smith AFM (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *J Am Stat Assoc* 85, 972–985
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans Pattern Analysis and Machine Intelligence* 6, 721–741
- Geyer CJ (1992) Practical Markov chain Monte Carlo (with discussion). *Stat Sci* 7, 467–511
- Gianola D (1982) Theory and analysis of threshold characters. *J Anim Sci* 54, 1079–1096
- Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet Sel Evol* 15, 201–224
- Gianola G, Fernando RL (1986) Bayesian methods in animal breeding theory. *J Anim Sci* 63, 217–244
- Gianola D, Im S, Macedo FW (1990) A framework for prediction of breeding value. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond eds), Springer-Verlag, Berlin, 210–238

- Gianola D, Foulley JL, Fernando RL, Henderson CR, Weigel KA (1992) Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. *J Dairy Sci* 75, 2805–2823
- Harville DA, Mee RW (1984) A mixed-model procedure for analyzing ordered categorical data. *Biometrics* 40, 393–408
- Henderson CR (1973) Sire evaluation and genetic trends. In: *Proc Animal Breeding and Genetics Symposium in Honor of Dr Jay L Lush*. ASAS and ADSA, Champaign, IL
- Henderson CR (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph
- Hobert JP, Casella G (1996) The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J Am Stat Assoc* 91, 1461–1473
- Hoeschele I, Tier B, Graser HU (1995) Multi-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models. *J Anim Sci* 73, 1609–1627
- Hoeschele I, Tier B (1995) Estimation of variance components of threshold characters by marginal posterior mode and means via Gibbs sampling. *Genet Sel Evol* 27, 519–540
- Janss LLG, Foulley JL (1993) Bivariate analysis for one continuous and one threshold dichotomous trait with unequal design matrices and an application to birth weight and calving difficulty. *Livestock Prod Sci* 33, 183–198
- Jensen J (1994) Bayesian analysis of bivariate mixed models with one continuous and one binary trait using the Gibbs sampler. *Proc 5th World Congress on Genetics Applied to Livestock Production, 7–12 August 1994, University of Guelph, Guelph* 18, 333–336
- Jensen J, Wang CS, Sorensen DA, 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, Sect A, Anim Sci* 44, 193–201
- Johnson NL, Kotz S (1972) *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York
- Koger MJ, Mitchell JS, Widder RW, Warnick WC (1967) Factors influencing survival in beef calves. *J Anim Sci* 26, 205
- Pollak EJ (1975) Dystocia in Holsteins. PhD thesis, Iowa State University, Ames, Iowa
- Quaas RL (1994) Threshold model for calving ease EPDs. *Proc 4th Genetic Prediction Workshop, Kansas City, 21–22 January*, Beef Improvement Federation
- Quaas RL (1996) Threshold model for calving ease EPDs. In: *BIF Guidelines for Uniform Beef Improvement Programs*. Beef Improvement Fed, Athens, 7th ed
- Quaas RL, Zhao Y, Pollak, EJ (1988) Describing interactions in dystocia scores with a threshold model. *J Anim Sci* 66, 396–399.
- San Cristobal M, Foulley JL, Manfredi E (1993) Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genet Sel Evol* 25, 3–30
- Sorensen D, Wang CS, Jensen J, Gianola D (1994) Bayesian analysis of genetic trend due to selection using Gibbs sampling. *Genet Sel Evol* 26, 333–360
- Sorensen DA, Anderson S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. *Genet Sel Evol* 27, 229–249
- Sorensen DA (1996) Gibbs sampling in quantitative genetics. Intern Report No. 82, Danish Institute Of Animal Science, Tjele
- Tanner MA (1993) *Tools for Statistical Inference*. Springer-Verlag, New York
- Van Tassell CP (1994) The use of Gibbs sampling for variance component estimation with simulated and weaning weight data using animal and maternal effects models. PhD thesis, Cornell University, Ithaca, NY
- Van Tassell CP, Van Vleck LD (1996) Multiple trait Gibbs sampler for animal models: flexible programs for Bayesian and likelihood-based (co)variance component inference. *J Anim Sci* 74, 2586–2597

- Van Tassell CP, Van Vleck LD, Gregory KE (1996) Bayesian analysis of twinning and ovulation rate. *J Anim Sci* 74 (suppl), 115
- Wang CS, Rutledge JJ, Gianola D (1993) Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet Sel Evol* 25, 41-62
- Wang CS, Rutledge JJ, Gianola D (1994a) Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet Sel Evol* 26, 91-115
- Wang CS, Gianola D, Sorensen D, Jensen J, Christensen A, Rutledge JJ (1994b) Response to selection for litter size in Danish Landrace pigs: a Bayesian analysis. *Theor Appl Genet* 88, 220-230
- Wang CS, Quaas RL, Pollak EJ (1995) Bayesian analysis of calving ease scores and birth weights. *Development and Implementation of Statistical Analyses and Computer Strategies for National Cattle Evaluation*, Proceedings of Symposium held at the 1995 WRCC-100 Regional Coordinating Committee Meeting, 20-21 October, 1995, Brainard, Minnesota, 50-79
- Wright S (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19, 506-536
- Zeger SL, Karim MR (1991) Generalized linear models with random effects: a Gibbs sampling approach. *J Am Stat Assoc* 86, 79-86
- Zhao Y (1986) Application of thresholds model to calving ease scores in Simmental cattle. MS thesis, Cornell University, Ithaca, NY
- Zhao Y (1987) Estimation of parameters in a mixed threshold model: its application to dystocia and birth weight in Simmental cattle. PhD thesis, Cornell University, Ithaca, NY