Original article

Bayesian inference in threshold models using Gibbs sampling

DA Sorensen [1], S Andersen [2], D Gianola [3], I Korsgaard [1]

[1] *National Institute of Animal Science, Research Centre Foulum, PO Box 39, DK-8830 Tjele;*
[2] *National Committee for Pig Breeding, Health and Production, Axeltorv 3, Copenhagen V, Denmark;*
[3] *University of Wisconsin-Madison, Department of Meat and Animal Sciences, Madison, WI 53706-1284, USA*

Summary – A Bayesian analysis of a threshold model with multiple ordered categories is presented. Marginalizations are achieved by means of the Gibbs sampler. It is shown that use of data augmentation leads to conditional posterior distributions which are easy to sample from. The conditional posterior distributions of thresholds and liabilities are independent uniforms and independent truncated normals, respectively. The remaining parameters of the model have conditional posterior distributions which are identical to those in the Gaussian linear model. The methodology is illustrated using a sire model, with an analysis of hip dysplasia in dogs, and the results are compared with those obtained in a previous study, based on approximate maximum likelihood. Two independent Gibbs chains of length 620 000 each were run, and the Monte-Carlo sampling error of moments of posterior densities were assessed using time series methods. Differences between results obtained from both chains were within the range of the Monte-Carlo sampling error. With the exception of the sire variance and heritability, marginal posterior distributions seemed normal. Hence inferences using the present method were in good agreement with those based on approximate maximum likelihood. Threshold estimates were strongly autocorrelated in the Gibbs sequence, but this can be alleviated using an alternative parameterization.

threshold model / Bayesian analysis / Gibbs sampling / dog

Résumé – Inférence bayésienne dans les modèles à seuil avec échantillonnage de Gibbs. *Une analyse bayésienne du modèle à seuil avec des catégories multiples ordonnées est présentée ici. Les marginalisations nécessaires sont obtenues par échantillonnage de Gibbs. On montre que l'utilisation de données augmentées – la variable continue sousjacente non observée étant alors considérée comme une inconnue dans le modèle – conduit à des distributions conditionnelles* a posteriori *faciles à échantillonner. Celles-ci sont des distributions uniformes indépendantes pour les seuils et des distributions normales*

tronquées indépendantes pour les sensibilités (les variables sous-jacentes). Les paramètres restants du modèle ont des distributions conditionnelles a posteriori identiques à celles qu'on trouve en modèle linéaire gaussien. La méthodologie est illustrée sur un modèle paternel appliquée à une dysplasie de la hanche chez le chien, et les résultats sont comparés à ceux d'une étude précédente basée sur un maximum de vraisemblance approché. Deux séquences de Gibbs indépendantes, longues chacune de 620 000 échantillons, ont été réalisées. Les erreurs d'échantillonnage de type Monte Carlo des moments des densités a posteriori ont été obtenues par des méthodes de séries temporelles. Les résultats obtenus avec les 2 séquences indépendantes sont dans la limite des erreurs d'échantillonnage de Monte-Carlo. À l'exception de la variance paternelle et de l'héritabilité, les distributions marginales a posteriori semblent normales. De ce fait, les inférences basées sur la présente méthode sont en bon accord avec celles du maximum de vraisemblance approché. Pour l'estimation des seuils, les séquences de Gibbs révèlent de fortes autocorrélations, auxquelles il est cependant possible de remédier en utilisant un autre paramétrage.

modèle à seuil / analyse bayésienne / échantillonnage de Gibbs / chien

INTRODUCTION

Many traits in animal and plant breeding that are postulated to be continuously inherited are categorically scored, such as survival and conformation scores, degree of calving difficulty, number of piglets born dead and resistance to disease. An appealing model for genetic analysis of categorical data is based on the threshold liability concept, first used by Wright (1934) in studies of the number of digits in guinea pigs, and by Bliss (1935) in toxicology experiments. In the threshold model, it is postulated that there exists a latent or underlying variable (liability) which has a continuous distribution. A response in a given category is observed, if the actual value of liability falls between the thresholds defining the appropriate category. The probability distribution of responses in a given population depends on the position of its mean liability with respect to the fixed thresholds. Applications of this model in animal breeding can be found in Robertson and Lerner (1949), Dempster and Lerner (1950) and Gianola (1982), and in Falconer (1965), Morton and McLean (1974) and Curnow and Smith (1975), in human genetics and susceptibility to disease. Important issues in quantitative genetics and animal breeding include drawing inferences about (i) genetic and environmental variances and covariances in populations; (ii) liability values of groups of individuals and candidates for genetic selection; and (iii) prediction and evaluation of response to selection. Gianola and Foulley (1983) used Bayesian methods to derive estimating equations for (ii) above, assuming known variances. Harville and Mee (1984) proposed an approximate method for variance component estimation, and generalizations to several polygenic binary traits having a joint distribution were presented by Foulley *et al* (1987). In these methods inferences about dispersion parameters were based on the mode of their joint posterior distribution, after integration of location parameters. This involved the use of a normal approximation which, seemingly, does not behave well in sparse contingency tables (Höschele *et al*, 1987). These authors found that estimates of genetic parameters were biased when the number of observations per

combination of fixed and random levels in the model was smaller than 2, and suggested that this may be caused by inadequacy of the normal approximation. This problem can render the method less useful for situations where the number of rows in a contingency table is equal to the number of individuals. A data structure such as this often arises in animal breeding, and is referred to as the 'animal model' (Quaas and Pollak, 1980). Anderson and Aitkin (1985) proposed a maximum likelihood estimator of variance component for a binary threshold model. In order to construct the likelihood, integration of the random effects was achieved using univariate Gaussian quadrature. This procedure cannot be used when the random effects are correlated, such as in genetics. Here, multiple integrals of high dimension would need to be calculated, which is unfeasible even in data sets with only 50 genetically related individuals. In animal breeding, a data set may contain thousands of individuals that are correlated to different degrees, and some of these may be inbred.

Recent reviews of statistical issues arising in the analysis of discrete data in animal breeding can be found in Foulley *et al* (1990) and Foulley and Manfredi (1991). Foulley (1993) gave approximate formulae for one-generation predictions of response to selection by truncation for binary traits based on a simple threshold model. However, there are no methods described in the literature for drawing inferences about genetic change due to selection for categorical traits in the context of threshold models. Phenotypic trends due to selection can be reported in terms of changes in the frequency of affected individuals. Unfortunately, due to the nonlinear relationship between phenotype and genotype, phenotypic changes do not translate directly into additive genetic changes, or, in other words, to response to selection. Here we point out that inferences about realized selection response for categorical traits can be drawn by extending results for the linear model described in Sorensen *et al* (1994).

With the advent of Monte-Carlo methods for numerical integration such as Gibbs sampling (Geman and Geman, 1984; Gelfand *et al*, 1990), analytical approximations to posterior distributions can be avoided, and a simulation-based approach to Bayesian inference about quantitative genetic parameters is now possible. In animal breeding, Bayesian methods using the Gibbs sampler were applied in Gaussian models by Wang *et al* (1993, 1994a) and Jensen *et al* (1994) for (co)variance component estimation and by Sorensen *et al* (1994) and Wang *et al* (1994b) for assessing response to selection. Recently, a Gibbs sampler was implemented for binary data (Zeger and Karim, 1991) and an analysis of multiple threshold models was described by Albert and Chib (1993). Zeger and Karim (1991) constructed the Gibbs sampler using rejection sampling techniques (Ripley, 1987), while Albert and Chib (1993) used it in conjunction with data augmentation, which leads to a computationally simpler strategy. The purpose of this paper is to describe a Gibbs sample for inferences in threshold models in a quantitative genetic context. First, the Bayesian threshold model is presented, and all conditional posterior distributions needed for running the Gibbs sampler are given in closed form. Secondly, a quantitative genetic analysis of hip dysplasia in German shepherds is presented as an illustration, and 2 different parameterizations of the model leading to alternative Gibbs sampling schemes are described.

MODEL FOR BINARY RESPONSES

At the phenotypic level, a Bernoulli random variable Y_i is observed for each individual i $(i = 1, 2, \ldots, n)$ taking values $y_i = 1$ or $y_i = 0$ (*eg*, alive or dead). The variable Y_i is the expression of an underlying continuous random variable U_i, the liability of individual i. When U_i exceeds an unknown fixed threshold t, then $Y_i = 1$, and $Y_i = 0$ otherwise. We assume that liability is normally distributed, with the mean value indexed by a parameter $\boldsymbol{\theta}$, and, without loss of generality, that it has unit variance (Curnow and Smith, 1975). Hence:

$$U_i|\boldsymbol{\theta} \sim N(\mathbf{w}_i'\boldsymbol{\theta}, 1) \qquad [1]$$

where $\boldsymbol{\theta}' = (\mathbf{b}', \mathbf{a}')$ is a vector of parameters with p fixed effects ($\mathbf{b}$) and q random additive genetic values ($\mathbf{a}$), and $\mathbf{w}_i'$ is a row incidence vector linking $\boldsymbol{\theta}$ to the ith observation.

It is important to note that conditionally on $\boldsymbol{\theta}$, the U_i are independent, so for the vector $\mathbf{U} = \{U_i\}$ given $\boldsymbol{\theta}$, we have as joint density:

$$p(\mathbf{U}|\boldsymbol{\theta}) = \prod_{i=1}^{n} \phi_{U_i}(\mathbf{w}_i'\boldsymbol{\theta}, 1) = \phi_U(\mathbf{W}\boldsymbol{\theta}, \mathbf{I}) \qquad [2]$$

where $\phi_U(.)$ is a normal density with parameters as indicated in the argument. In [2], put $\mathbf{W}\boldsymbol{\theta} = \mathbf{Xb} + \mathbf{Za}$, where $\mathbf{X}$ and $\mathbf{Z}$ are known incidence matrices of order n by p and n by q, respectively, and, without loss of generality, $\mathbf{X}$ is assumed to have full column rank. Given the model, we have:

$$P(Y_i = 1|\boldsymbol{\theta}, t) = P(U_i > t|\boldsymbol{\theta}, t) = \int_{t-\mathbf{w}_i'\boldsymbol{\theta}}^{\infty} \phi(x)dx = \Phi\left(-(t - \mathbf{w}_i'\boldsymbol{\theta})\right) \qquad [3]$$

where $\Phi(.)$ is the cumulative distribution function of a standardized normal variate. Without loss of generality, and provided that there is a constant term in the model, t can be set to 0, and [3] reduces to

$$P(Y_i = 1|\boldsymbol{\theta}) = \Phi(\mathbf{w}_i'\boldsymbol{\theta}) \qquad [4]$$

Conditionally on both $\boldsymbol{\theta}$ and on $Y_i = y_i$, U_i follows a truncated normal distribution. That is, for $y_i = 1$:

$$p(U_i|\boldsymbol{\theta}, Y_i = 1) = \frac{\phi_{U_i}(\mathbf{w}_i'\boldsymbol{\theta}, 1)}{\Phi(\mathbf{w}_i'\boldsymbol{\theta})}1(U_i > 0) \qquad [5]$$

where $1(X \in A)$ is the indicator function that takes the value 1 if the random variable X is contained in the set A, and 0 otherwise. For $y_i = 0$, the density is $\phi_{U_i}(\mathbf{w}_i'\boldsymbol{\theta}, 1)/\Phi(-\mathbf{w}_i'\boldsymbol{\theta})1(U_i \leqslant 0)$.

Invoking an infinitesimal model (Bulmer, 1971), the conditional distribution of additive genetic values given the additive genetic variance in the conceptual base population (σ_a^2) is multivariate normal:

$$\mathbf{a}|\sigma_a^2 \sim N(0, \mathbf{A}\sigma_a^2) \qquad [6]$$

where $\mathbf{A}$ is a q by q matrix of additive genetic relationships. Note that $\mathbf{a}$ can include animals without phenotypic scores.

We discuss next the Bayesian inputs of the model. The vector of fixed effects $\mathbf{b}$ will be assumed to follow *a priori* the improper uniform distribution:

$$p(\mathbf{b}) \propto \text{ constant} \qquad [7]$$

For a description of uncertainty about the additive genetic variance, σ_a^2, an inverted gamma distribution can be invoked, with density:

$$p(\sigma_a^2|v, S^2) \propto (\sigma_a^2)^{-((v/2)+1)} \exp\left[-\frac{vS^2}{2\sigma_a^2}\right] \qquad [8]$$

where v and S^2 are parameters. When $v = -2$ and $S^2 = 0$, [8] reduces to the improper uniform prior distribution. A proper uniform prior distribution for σ_a^2 is:

$$p(\sigma_a^2) \propto \begin{cases} k_a, & 0 \leqslant \sigma_a^2 \leqslant \sigma_{a\max}^2 \\ 0, & \text{otherwise} \end{cases} \qquad [9]$$

where k_a is a constant and $\sigma_{a\max}^2$ is the maximum value which σ_a^2 can take *a priori*.

To facilitate the development of the Gibbs sampler, the unobserved liability $\mathbf{U}$ is included as an unknown parameter in the model. This approach, known as data augmentation (Tanner and Wong, 1987; Gelfand *et al*, 1992; Albert and Chib, 1993; Smith and Roberts, 1993) leads to identifiable conditional posterior distributions, as shown in the next section.

Bayes theorem gives as joint posterior distribution of the parameters:

$$p(\boldsymbol{\theta}, \mathbf{U}, \sigma_a^2|\mathbf{y}, v, S^2) \propto p(\mathbf{U}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\sigma_a^2)p(\sigma_a^2|v, S^2)p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{U}, \sigma_a^2, v, S^2) \qquad [10]$$

The last term is the conditional distribution of the data given the parameters. We notice that, for $Y_i = 1$, say, we have

$$p(Y_i = 1|\boldsymbol{\theta}, U_i, \sigma_a^2) = p(Y_i = 1|U_i) = \begin{cases} 1 \text{ if } U_i > 0 \\ 0 \text{ otherwise} \end{cases} \qquad [11a]$$

For $Y_i = 0$, we have:

$$p(Y_i = 0|\boldsymbol{\theta}, U_i, \sigma_a^2) = p(Y_i = 0|U_i) = \begin{cases} 1 \text{ if } U_i \leqslant 0 \\ 0 \text{ otherwise} \end{cases} \qquad [11b]$$

This distribution is degenerate, as noted by Gelfand *et al* (1992) because knowledge of U_i implies exact knowlege of Y_i. This can be written (*eg*, Albert and Chib, 1993) as:

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{U}, \sigma_a^2) = \prod_{i=1}^{n} \{1(U_i > 0)1(Y_i = 1) + 1(U_i \leqslant 0)1(Y_i = 0)\} \qquad [12]$$

The joint posterior distribution [10] can then be written as:

$$p(\mathbf{b}, \mathbf{a}, \mathbf{U}, \sigma_a^2 | \mathbf{y}, v, S^2)$$

$$\propto p(\sigma_a^2 | v, S^2) p(\mathbf{a} | \sigma_a^2) \, p(\mathbf{U} | \mathbf{b}, \mathbf{a}) \prod_{i=1}^{n} \{1(U_i > 0)1(Y_i = 1) + 1(U_i \leqslant 0)1(Y_i = 0)\} \quad [13]$$

where the conditioning on hyperparameters v and S^2 is replaced by $\sigma_{a\max}^2$ when the uniform prior [9] for the additive genetic variance is employed.

Conditional posterior distributions

In order to implement the Gibbs sampler, all conditional posterior distributions of the parameters of the model are needed. The starting point is the full posterior distribution [13]. Among the 4 terms in [13], the third is the only one that is a function of $\mathbf{b}$ and we therefore have for the fixed effects:

$$p(\mathbf{b} | \mathbf{a}, \mathbf{U}, \sigma_a^2, \mathbf{y}) \propto p(\mathbf{U} | \mathbf{b}, \mathbf{a}) \quad [14]$$

which is proportional to $\phi_U(\mathbf{Xb} + \mathbf{Za}, \mathbf{I})$. As shown in Wang *et al* (1994a), the scalar form of the Gibbs sampler for the ith fixed effect consists of sampling from:

$$b_i | \mathbf{b}_{-i}, \mathbf{a}, \mathbf{U}, \sigma_a^2, \mathbf{y} \sim N\left(\widehat{b}_i, (\mathbf{x}_i'\mathbf{x}_i)^{-1}\right) \quad [15]$$

where $\mathbf{x}_i$ is the ith column of the matrix $\mathbf{X}$, and $\widehat{b}_i$ satisfies:

$$\mathbf{x}_i'\mathbf{x}_i\widehat{b}_i = \mathbf{x}_i'(\mathbf{U} - \mathbf{X}_{-i}\mathbf{b}_{-i} - \mathbf{Za}) \quad [16]$$

In [16], $\mathbf{X}_{-i}$ is the matrix $\mathbf{X}$ with the column associated with i deleted, and $\mathbf{b}_{-i}$ is $\mathbf{b}$ with the ith element deleted. The conditional posterior distribution of the vector of breeding values is proportional to the product of the second and third terms in [13]:

$$p(\mathbf{a} | \mathbf{b}, \mathbf{U}, \sigma_a^2, \mathbf{y}) \propto p(\mathbf{a} | \sigma_a^2) p(\mathbf{U} | \mathbf{b}, \mathbf{a}) \quad [17]$$

which has the form $\phi_a(\mathbf{0}, \mathbf{A}\sigma_a^2)\phi_U(\mathbf{u} | \mathbf{b}, \mathbf{a})$. Wang *et al* (1994a) showed that the scalar Gibbs sampler draws samples from:

$$a_i | \mathbf{b}, \mathbf{a}_{-i}, \mathbf{U}, \sigma_a^2, \mathbf{y} \sim N\left(\widehat{a}_i, (\mathbf{z}_i'\mathbf{z}_i + c_{ii}\lambda)^{-1}\right) \quad [18]$$

where $\mathbf{z}_i$ is the ith column of $\mathbf{Z}$, c_{ii} is the element in the ith row and column of $\mathbf{A}^{-1}$, $\lambda = (\sigma_a^2)^{-1}$, and $\widehat{a}_i$ satisfies:

$$(\mathbf{z}_i'\mathbf{z}_i + c_{ii}\lambda)\widehat{a}_i = \mathbf{z}_i'(\mathbf{U} - \mathbf{Xb}) - \lambda\mathbf{c}_{i,-i}\mathbf{a}_{-i} \quad [19]$$

In [19], $\mathbf{c}_{i,-i}$ is the row of $\mathbf{A}^{-1}$ corresponding to the ith individual with the ith element excluded. We notice from [14] and [17], that augmenting with the underlying variable U, leads to an implementation of the Gibbs sampler which is the same as for the linear model, with the underlying variable replacing the observed data.

For the variance component, we have from [13]:

$$p(\sigma_a^2|\mathbf{b},\mathbf{a},\mathbf{U},\mathbf{y},v,S^2) \propto p(\sigma_a^2|v,S^2)p(\mathbf{a}|\sigma_a^2) \qquad [20]$$

Assuming that the prior for σ_a^2 is the inverted gamma given in [8], this becomes:

$$p(\sigma_a^2|\mathbf{b},\mathbf{a},\mathbf{U},\mathbf{y},v,S^2) \propto (\sigma_a^2)^{-\left(\frac{q+v}{2}+1\right)}\exp\left(-\frac{(\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}+vS^2)}{2\sigma_a^2}\right) \qquad [21a]$$

and assuming the uniform prior [9], it becomes:

$$p(\sigma_a^2|\mathbf{b},\mathbf{a},\mathbf{U},\mathbf{y}) \propto (\sigma_a^2)^{-\left(\frac{q}{2}\right)}\exp\left(-\frac{\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}}{2\sigma_a^2}\right) 1(0,\sigma_{a\max}^2) \qquad [21b]$$

Expression [21a] is in the form of a scaled inverted gamma density, and [21b] in the form of a truncated scaled inverted gamma density.

The conditional posterior distribution of the underlying variable U_i is proportional to the last 2 terms in [13]. This can be seen to be a truncated normal distribution, on the left if $Y_i = 1$ and on the right otherwise. The density function of this truncated normal distribution is given in [5]. Thus, depending on the observed Y_i, we have:

$$p(U_i|\mathbf{b},\mathbf{a},U_{-i},\sigma_a^2,Y_i=1) = \frac{\phi_{U_i}(\mathbf{w}_i'\boldsymbol{\theta},1)}{\Phi(\mathbf{w}_i'\boldsymbol{\theta})}1(U_i>0) \qquad [22a]$$

or

$$p(U_i|\mathbf{b},\mathbf{a},\mathbf{U}_{-i},\sigma_a^2,Y_i=0) = \frac{\phi_{U_i}(\mathbf{w}_i'\boldsymbol{\theta},1)}{1-\Phi(\mathbf{w}_i'\boldsymbol{\theta})}1(U_i\leqslant 0) \qquad [22b]$$

Sampling from the truncated distribution can be done by generating from the untruncated distribution and retaining those values which fall in the constraint region. Alternatively and more efficiently, suppose that U is truncated and defined in the interval $[i,j]$ only, where i and j are the lower and upper bounds, respectively. Let the distribution function of U be F, and let ν be a uniform $[0,1]$ variate. Then $U = F^{-1}[F(i) + \nu(F(j) - F(i))]$ is a drawing from the truncated random variable (Devroye, 1986).

Albert and Chib (1993) also constructed the mixed model in terms of a hierarchical model, but proposed a block sampling strategy for the parameters in the underlying scale, instead. Essentially, they suggest sampling from the distributions $(\sigma_a^2,\mathbf{a},\mathbf{b}|\mathbf{U})$ as $(\sigma_a^2|\mathbf{U})(\mathbf{a},\mathbf{b}|\mathbf{U},\sigma_a^2)$, instead of from the full conditional posterior distributions [15], [18] and [21], and they assumed a uniform prior for $\log(\sigma_a^2)$ in a finite interval. To facilitate sampling from $p(\sigma_a^2|\mathbf{U})$, they use an approximation which consists of placing all prior probabilities on a grid of σ_a^2 values, thus making the prior and the posterior discrete. The need for this approximation is questionable, since the full conditional posterior distribution of σ_a^2 has a simple form as noted in [21] above. In addition, in animal breeding, the distribution $(\mathbf{a},\mathbf{b}|\mathbf{U},\sigma_a^2)$ is a high dimensional multivariate normal and it would not be simple computationally to draw a large number of samples.

MULTIPLE ORDERED CATEGORIES

Suppose now that the observed random variable Y can take values in one of C mutually exclusive ordered categories delimited by $C + 1$ thresholds. Let $t_0 = -\infty$, $t_C = +\infty$, with the remaining thresholds satisfying $t_1 \leqslant t_2 \ldots \leqslant t_{C-1}$. Generalizing [3]:

$$P(Y_i = j|\boldsymbol{\theta}, \mathbf{t}) = P(t_{j-1} < U_i \leqslant t_j|\boldsymbol{\theta}, \mathbf{t}) = \Phi(t_j - \mathbf{w}_i'\boldsymbol{\theta}) - \Phi(t_{j-1} - \mathbf{w}_i'\boldsymbol{\theta}) \qquad [23]$$

$j = 1, 2, \ldots, C$

Conditionally on $\boldsymbol{\theta}$, $Y_i = j$, t_{j-1} and t_j, the underlying variable associated with the ith observation follows a truncated normal distribution with density:

$$p(U_i|\boldsymbol{\theta}, Y_i = j, t_{j-1}, t_j) = \frac{\phi_{U_i}(\mathbf{w}_i'\boldsymbol{\theta}, 1)}{\Phi(t_j - \mathbf{w}_i'\boldsymbol{\theta}) - \Phi(t_{j-1} - \mathbf{w}_i'\boldsymbol{\theta})}, t_{j-1} < U_i \leqslant t_j \qquad [24]$$

Assuming that σ_a^2, $\mathbf{b}$ and $\mathbf{t}$ are independently distributed *a priori*, the joint posterior density is written as:

$$p(\mathbf{b}, \mathbf{a}, \mathbf{U}, \mathbf{t}, \sigma_a^2|v, S^2, \mathbf{y}) \propto p(\mathbf{b})p(\mathbf{t})p(\mathbf{a}|\sigma_a^2)p(\sigma_a^2|v, S^2)p(\mathbf{U}|\mathbf{b}, \mathbf{a}, \mathbf{t})p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{U}, \mathbf{t}, \sigma_a^2)$$

$$[25]$$

where $p(\mathbf{U}|\mathbf{b}, \mathbf{a}, \mathbf{t}) = p(\mathbf{U}|\mathbf{b}, \mathbf{a})$. Generalizing [12], the last term in [25] can be expressed as (Albert and Chib, 1993):

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{U}, \mathbf{t}, \sigma_a^2) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{C} 1(t_{j-1} < U_i \leqslant t_j)1(Y_i = j) \right\} \qquad [26]$$

All the conditional posterior distributions needed to implement the Gibbs sampler can be derived from [25]. It is clear that the conditional posterior distributions of b_i, a_i and σ_a^2 are the same as for the binary response model and given in [15], [18] and [21]. For the underlying variable associated with the ith observation we have from [25]:

$$p(U_i|\mathbf{b}, \mathbf{a}, \mathbf{t}, \mathbf{U}_{-i}, \sigma_a^2, \mathbf{y}) \propto p(U_i|\mathbf{b}, \mathbf{a}) \left(\sum_{j=1}^{C} 1(t_{j-1} < U_i \leqslant t_j)1(Y_i = j) \right)$$

This is a truncated normal, with density function as in [24].

The thresholds $\mathbf{t} = (t_1, t_2, \ldots, t_{C-1})$ are clearly dependent *a priori*, since the model postulates that these are distributed as order statistics from a uniform distribution in the interval $[t_{\min}, t_{\max}]$. However, the full conditional posterior distributions of the thresholds are independent. That is, $p(t_j|\mathbf{t}_{-j}, \mathbf{b}, \mathbf{a}, \mathbf{U}, \sigma_a^2, \mathbf{y}) = p(t_j|\mathbf{U}, \mathbf{y})$, as the following argument shows. The joint prior density of $\mathbf{t}$ is:

$$p(\mathbf{t}) = (C - 1)! \left(\frac{1}{t_{\max} - t_{\min}} \right)^{C-1} 1(\mathbf{t} \in T) \qquad [27]$$

where $T = \{(t_1, t_2, \ldots, t_{C-1}) | t_{\min} \leqslant t_1 \leqslant t_2 \leqslant \ldots \leqslant t_{C-1} \leqslant t_{\max}\}$ (Mood *et al*, 1974). Note that the thresholds enter only in defining the support of $p(\mathbf{t})$. The conditional posterior distribution of t_j is given by:

$$p(t_j | \mathbf{t}_{-j}, \mathbf{b}, \mathbf{a}, \mathbf{U}, \sigma_a^2, \mathbf{y}) \propto p(t_j, \mathbf{t}_{-j} \mathbf{b}, \mathbf{a}, \mathbf{U}, \sigma_a^2, \mathbf{y})$$
$$\propto p(\mathbf{t}) p(\mathbf{b}, \mathbf{a}, \mathbf{U}, \sigma_a^2, \mathbf{y} | \mathbf{t})$$
$$\propto p(\mathbf{t}) p(\mathbf{y} | \mathbf{b}, \mathbf{a}, \mathbf{U}, \mathbf{t}, \sigma_a^2)$$
$$\propto p(\mathbf{y} | \mathbf{b}, \mathbf{a}, \mathbf{U}, \mathbf{t}, \sigma_a^2) 1(\mathbf{t} \in T),$$

which has the same form as [26]. Regarded as a function of t, [26] shows that, given $\mathbf{U}$ and $\mathbf{y}$, the upper bound of threshold t_j is $\min(\mathbf{U}|Y = j+1)$ and the lower bound is $\max(\mathbf{U}|Y = j)$. The *a priori* condition $\mathbf{t} \in T$ is automatically fulfilled, and the bounds are unaffected by knowledge of the remaining thresholds. Thus t_j has a uniform distribution in this interval given by:

$$p(t_j | \mathbf{t}_{-j}, \mathbf{b}, \mathbf{a}, \mathbf{U}, \sigma_a^2, \mathbf{y}) = \frac{1}{\min(U|Y = j+1) - \max(U|Y = j)} \qquad [28]$$

This argument assumes that there are no categories with missing observations. To accommodate for the possibility of missing observations in 1 or more categories, Albert and Chib (1993) define the upper and lower bounds of threshold j, as $\min\{\min(\mathbf{U}|Y = j+1), \mathbf{t}_{j+1}\}$ and as $\max\{\max(\mathbf{U}|Y = j), \mathbf{t}_{j-1}\}$, respectively. In this case, the thresholds are not conditionally independent. The Gibbs sampler is implemented by sampling repeatedly from [15], [18], [21], [24] and [28].

Alternative parameterization of the multiple threshold model

The multiple threshold model can also be parameterized such that the conditional distribution of the underlying variable U, given $\boldsymbol{\theta}$, has unknown variance σ_e^2 instead of unit variance. The equivalence of the 2 parameterizations is shown in the *Appendix*. This parameterization requires that records fall in at least 3 mutually exclusive ordered categories; for C categories, only C-3 thresholds are identifiable. In this new parameterization, one must sample from the conditional posterior distribution of σ_e^2. Under the priors [8] or [9], the conditional posterior distribution of σ_e^2 can be shown to be in the form of a scaled inverted gamma. The parameters of this distribution depend on the prior used for σ_e^2. If this is in the form [8], then

$$p(\sigma_e^2 | \mathbf{b}, \mathbf{a}, \mathbf{U}, \sigma_a^2, \mathbf{y}, v_e, S_e^2) \propto (\sigma_e^2)^{-\left(\frac{n+v_e}{2}+1\right)} \exp\left(-\frac{SSE + v_e S_e^2}{2\sigma_e^2}\right) \qquad [29]$$

where, $SSE = (\mathbf{U} - \mathbf{Xb} - \mathbf{Za})'(\mathbf{U} - \mathbf{Xb} - \mathbf{Za})$, and v_e and S_e^2 are parameters of the prior distribution. If a uniform prior of the form [9] is assumed to describe the prior uncertainty about σ_e^2, the conditional posterior distribution is a truncated version of [29] (*ie* [21b]), with $v_e = -2$ and $S_e^2 = 0$. With exactly 3 categories, the Gibbs sampler requires generating random variates from [15], [18], [21], [24] and [29], and no drawings need to be made from [28].

EXAMPLE

We illustrate the methodology with an analysis of data on hip dysplasia in German shepherd dogs. Results of an early analysis and a full description of the data can be found in Andersen *et al* (1988). Briefly, the records consisted of radiographs of 2 674 offspring from 82 sires. These radiographs had been classified according to guidelines approved by FCI (Fédération Cynologique Internationale, 1983), each offspring record was allocated to 1 of 7 mutually exclusive ordered categories.

The model for the underlying variable was:

$$U_i|\mu, a_i \sim N(\mu + a_i, 1) \tag{30}$$

where a_i is the effect of sire i $(i = 1, 2, \ldots, 82;\ j = 1, 2, \ldots, n_i)$. The prior distribution of μ was as in [7] and sire effects were assumed to follow the normal distribution:

$$\mathbf{a}|\sigma_a^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_a^2) \tag{31}$$

The prior distribution of the sire variance (σ_a^2) was in the form given in [8], with $v = 1$ and $S^2 = 0.05$. The prior for $t_2, \ldots, t_6$ was chosen to be uniform on the ordered subset of $[f_1 = -1.365,\ f_7 = +\infty]^5$ for which $t_1 < t_2 < \ldots < t_7$, where f_1 was the value at which t_1 was set, and f_7 is the value of the 7th threshold. The value for f_1 was obtained from Andersen *et al* (1988), in order to facilitate comparisons with the present analysis. The analysis was also carried out under the parameterization where the conditional distribution of U given $\boldsymbol{\theta}$ has variance σ_e^2. Here, σ_e^2 was assumed to follow a prior of the form of [8], with $v = 1$ and $S^2 = 0.05$ and t_6 was set to 0.429. Results of the 2 analyses were similar, so only those from the second parameterization are presented here.

Gibbs sampler and post Gibbs analysis

The Gibbs sampler was run as a single chain. Two independent chains of length 620 000 each were run, and in both cases, the first 20 000 samples were discarded. Thereafter, samples were saved every 20 iterations, so that the total number of samples kept was 30 000 from each chain. Start values for the parameters were, for the case of chain 1, $\sigma_e^2 = 2.0$, $\sigma_a^2 = 0.5$, $t_2 = -0.8$, $t_3 = -0.5$, $t_4 = -0.2$, $t_5 = 0.1$. For chain 2, estimates from Andersen *et al* (1988) were used, and these were $\sigma_e^2 = 1.0$, $\sigma_a^2 = 0.1$, $t_2 = -1.05$, $t_3 = -0.92$, $t_4 = -0.62$, $t_5 = -0.34$. In both runs, starting values for sire effects were set to zero.

Two important issues are the assessment of convergence of the Gibbs sampler, and the Monte-Carlo error of estimates of features of posterior distributions. Both issues are related to the question of whether the chain, or chains, have been run long enough. This is an area of active research in which some guidelines based on theoretical work (Roberts, 1992; Besag and Green, 1993; Smith and Roberts, 1993; Roberts and Polson, 1994) and on practical considerations (Gelfand *et al*, 1990; Gelman and Rubin, 1992; Geweke, 1992; Raftery and Lewis, 1992) have been suggested. The approach chosen here is based on Geyer (1992), who used time series methods to estimate the Monte-Carlo error of moments estimated from the Gibbs chain. Other approaches include, for example, batching (Ripley, 1987), and Raftery

and Lewis (1992) proposed a method based on 2-state Markov chains to calculate the number of iterations needed to estimate posterior quantiles.

Let $X_1, \ldots, X_m$ be elements of a single Gibbs chain of length m. The m sampled values, which are generally correlated, can be used to compute features of the posterior distribution. In our model, the Xs could represent samples from the posterior distribution of a sire value, or of the sire variance, or of functions of these. Invoking the ergodic theorem, for large m, the posterior mean can be estimated by

$$\widehat{\mu}_m = \frac{1}{m} \sum_{i=1}^{m} X_i$$

and the variance of the posterior distribution can be estimated by

$$\mathrm{Var}(X) = \frac{1}{m} \sum_{i=1}^{m} (X_i - \widehat{\mu}_m)^2$$

(Cox and Miller, 1965; Geyer, 1992). Similarly, a marginal posterior distribution function, $F(Z) = P(X < Z)$, can be estimated by

$$\widehat{F}_m(Z) = \frac{1}{m} \sum_{i=1}^{m} 1_{(X_i < Z)}$$

which means that $F(Z)$ is estimated by the empirical distribution function. All these estimators are subject to Monte-Carlo sampling error, which is reduced by prolongation of the chain.

Consider the sequence $g(X_1), \ldots, g(X_m)$, where $g(.)$ is a suitable function, eg, $g(X) = X$ or $g(X) = 1_{(X<Z)}$, and let $E(g(X)) = \mu$. The lag-time auto-covariance of the sequence is estimated as:

$$\widehat{\gamma}_m(t) = \frac{1}{m} \sum_{i=1}^{i=m-t} (g(X_i) - \widehat{\mu}_m)(g(X_{i+t}) - \widehat{\mu}_m) \qquad [32]$$

where

$$\widehat{\mu}_m = (1/m) \sum_{i=1}^{i=m} g(X_i)$$

is the sample mean for a chain of length m, and t is the lag. The auto-correlation is then estimated as $\widehat{\gamma}_m(t)/\widehat{\gamma}_m(0)$. The variance of the sample mean of the chain is given by

$$\mathrm{Var}(\widehat{\mu}_m) = \frac{\gamma(0)}{m} + \frac{1}{m} 2 \sum_{t=1}^{t=m-1} \left(1 - \frac{t}{m}\right) \gamma(t) \qquad [33]$$

which will exceed $\gamma(0)/m$ if $\gamma(t) > 0$ for all t, as is usual. Several estimators of the variance of the sample mean have been proposed (Priestley, 1981), but we chose one suggested by Geyer (1992), which he calls the initial positive sequence estimator.

Let $\widehat{\Gamma}_m(t) = \widehat{\gamma}_m(2t) + \widehat{\gamma}_m(2t+1)$, $t = 0, 1, \ldots$. The estimator can then be written as

$$m\text{V}\hat{\text{a}}\text{r}(\widehat{\mu}_m) = \widehat{\gamma}_m(0) + 2\sum_{i=1}^{2t+1}\widehat{\gamma}_m(i) = -\widehat{\gamma}_m(0) + 2\sum_{i=0}^{t}\widehat{\Gamma}_m(i) \qquad [34]$$

where t is chosen such that it is the largest integer satisfying $\widehat{\Gamma}_m(i) > 0$, $i = 1, \ldots, t$. The justification for this choice is that $\Gamma(i)$ is a strictly positive, strictly decreasing function of i. If $X_1, X_2, \ldots X_m$ are independent, then $\text{Var}(\widehat{\mu}_m) = \gamma(0)/m$. To obtain an indication of the effect of the correlation on $\text{Var}(\widehat{\mu}_m)$, an 'effective number' of independent observations can be assessed as $\widehat{\psi}_m = \widehat{\gamma}_m(0)/\text{V}\hat{\text{a}}\text{r}(\widehat{\mu}_m)$. When the elements of the Gibbs chain are independent, $\widehat{\psi}_m = m$.

RESULTS

Estimates of the empirical distribution function of various parameters of the model for each of the 2 chains of the Gibbs sampler are shown in figures 1–5. For example, figure 2 shows that there is a 90% posterior probability that the sire variance lies between 0.065 and 0.14, and the median of this posterior distribution is slightly
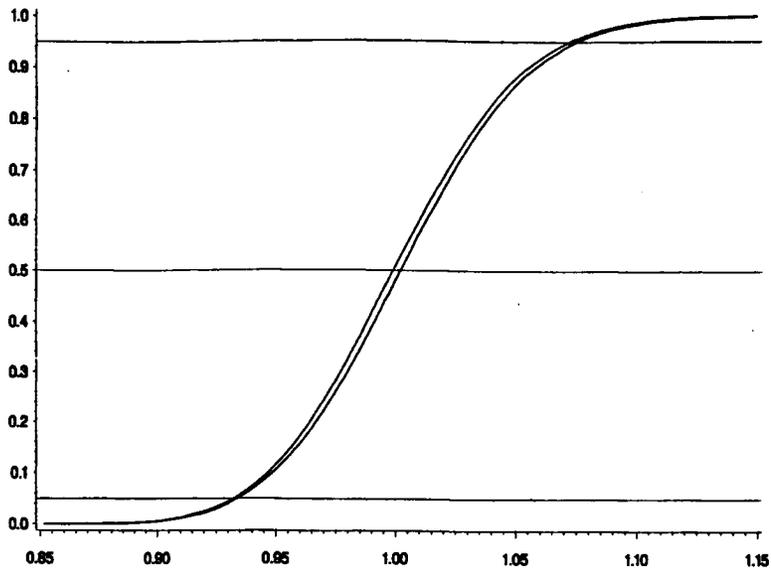

Fig 1. Estimated marginal posterior cumulative distribution functions of the residual variance (chains 1 and 2).
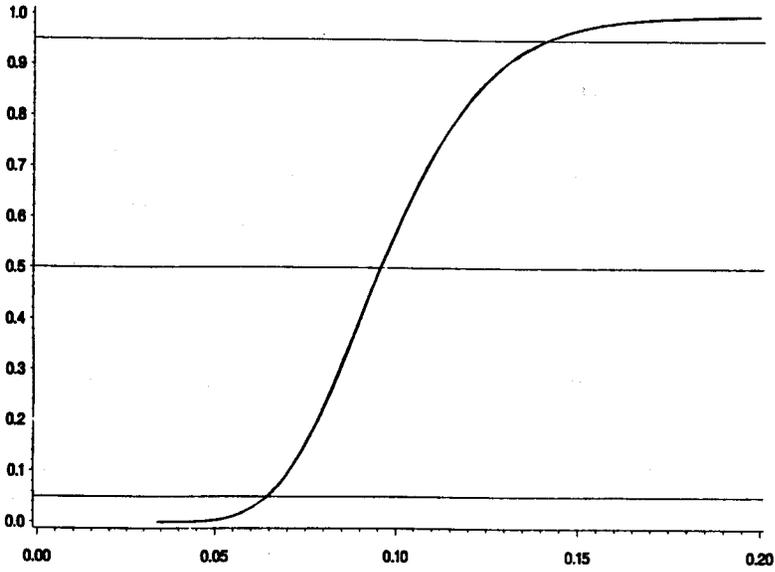
Fig 2. Estimated marginal posterior cumulative distribution functions of the sire variance (chains 1 and 2).
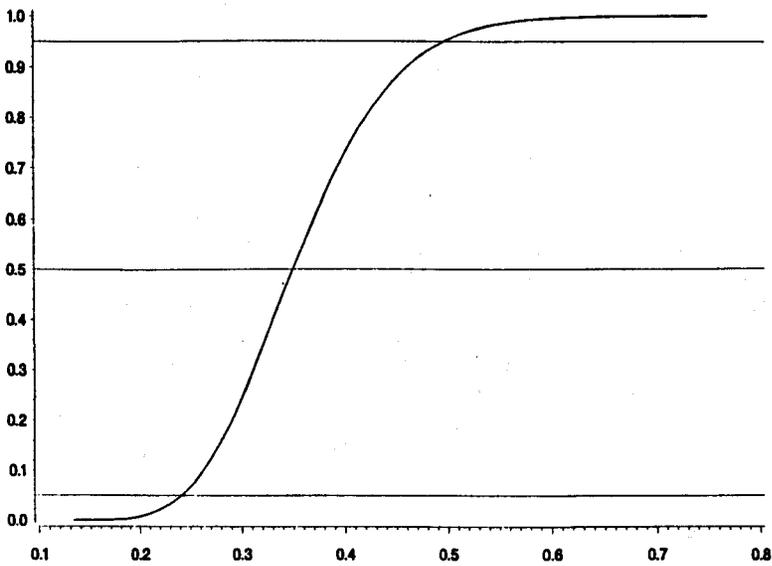
Fig 3. Estimated marginal posterior cumulative distribution functions of the heritability (chains 1 and 2).

under 0.10. Similarly, figure 3 indicates that there is 90% posterior probability that heritability in the underlying scale ($h^2 = 4\sigma_a^2/(\sigma_a^2 + \sigma_e^2)$) lies between 0.24 and 0.49, and the median of the posterior distribution is 0.35. Although this distribution is slightly skewed, the estimate of the median agrees well with the ML type estimate of heritability of 0.35, reported in Andersen *et al* (1988). Figure 4 depicts estimates of distribution functions for the mean (μ) and for each of 3 sire effects (a_1, a_2, a_3). Figure 5 gives corresponding distributions for 2 threshold parameters (t_2, and t_3). The figures fall in 3 categories. The distribution functions obtained from chains 1 and 2 coincide for each of the variables σ_a^2, h^2, a_1, a_2 and a_3, where the sire effects a_1, a_2 and a_3 pertain to 3 males with 31, 5 and 158 offspring, respectively. A small deviation between chains 1 and 2 is observed for σ_e^2 and μ, and a larger deviation is observed for the threshold parameters (fig 5). The Gibbs sequence for the threshold parameters showed very slow mixing properties. For example, for threshold 2, the autocorrelations between sampled values were 0.785, 0.663 and 0.315, for lags between 5, 10 and 50 samples, respectively. The reason is that the sampled value for a given threshold is bounded by the values of the neighbouring underlying variables U. If these are very close, the value of the threshold in subsequent samples is likely to change very slowly. Under the parameterization where 1 of the thresholds is substituted by the residual variance, the autocorrelations associated with the lags above, between samples from the marginal posterior distribution of σ_e^2, were 0.078, 0.064 and 0.032, respectively. Another scheme that may accelerate mixing is to sample jointly from the threshold and the liability. For sire effects, lag autocorrelations were close to zero.

A comparison between marginal posterior means (average of the 30 000 samples) estimated from the 2 chains is shown in table I. The difference between chains 1 and 2 is in all cases within Monte-Carlo sampling error, which was estimated within chains, using [34]. The 'effective number' of observations $\widehat{\psi}_m$ for the means of marginal posterior distributions is close to 30 000 for σ_a^2, a_1, a_2 and a_3, but is about 3 000 for σ_e^2 and μ and between 200 and 400 for $t_2, \ldots, t_5$.

The marginal posterior distributions for $\mu, t_2, \ldots, t_5, a_1, a_2, a_3$ are well approximated by normal distributions. The posterior means and standard deviations of these marginal distributions can be compared to estimates reported in Andersen *et al* (1988), who used a 2-stage procedure. The authors first estimated variances using an REML-type estimator (Harville and Mee, 1984). Secondly, assuming that the estimated variances were the true ones, fixed effects and sire effects were estimated as suggested by Gianola and Foulley (1983). For example, for the 3 sires with 158, 31 and 5 offspring respectively, the 2-stage procedure yielded estimates of sire effects (approximate posterior standard deviations) of 0.30 (0.09), 0.17 (0.17) and −0.093 (0.26), respectively. The present Bayesian approach with the Gibbs sampler, yielded estimates of marginal posterior means and standard deviations for these sires of 0.304 (0.088), 0.177 (0.166), and −0.092 (0.263), respectively (table I).

DISCUSSION

We have described a Gibbs sampler for making inferences from threshold models for discrete data. The method was illustrated using a model with unrelated sires;
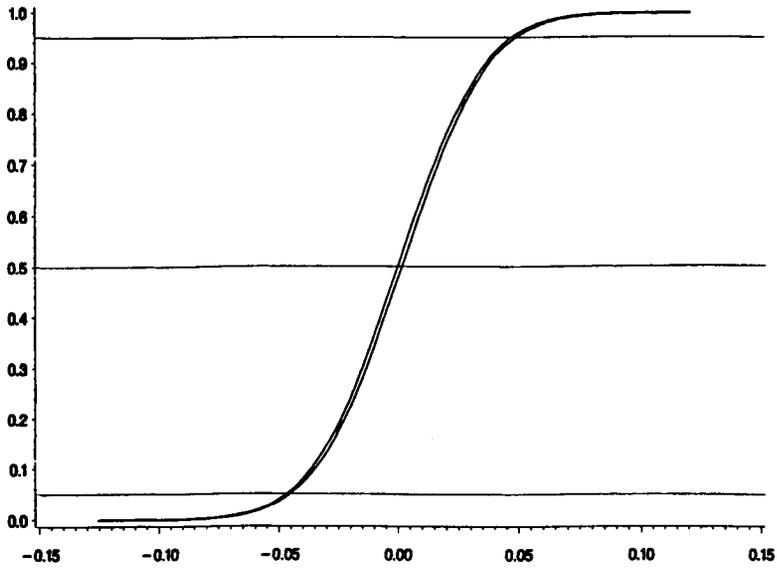
a
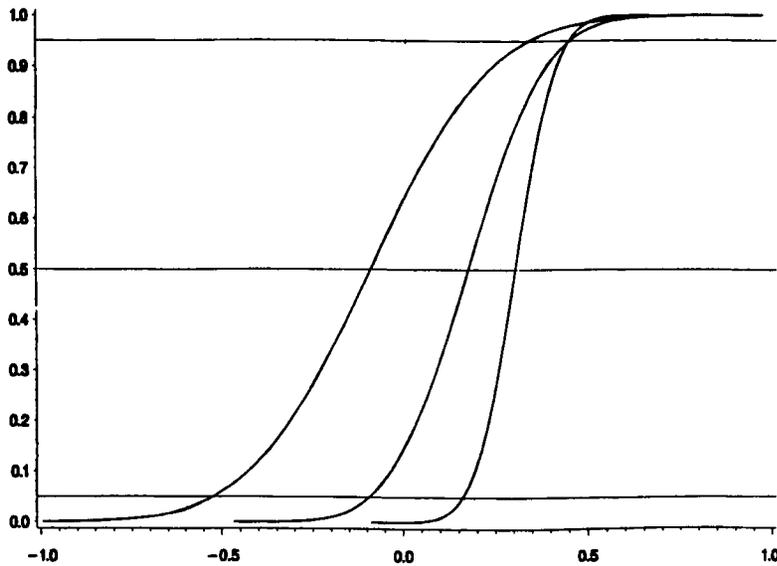
b

Fig 4. Estimated marginal posterior cumulative distribution functions of the mean (a) and of 3 sire effects (b) (chains 1 and 2).
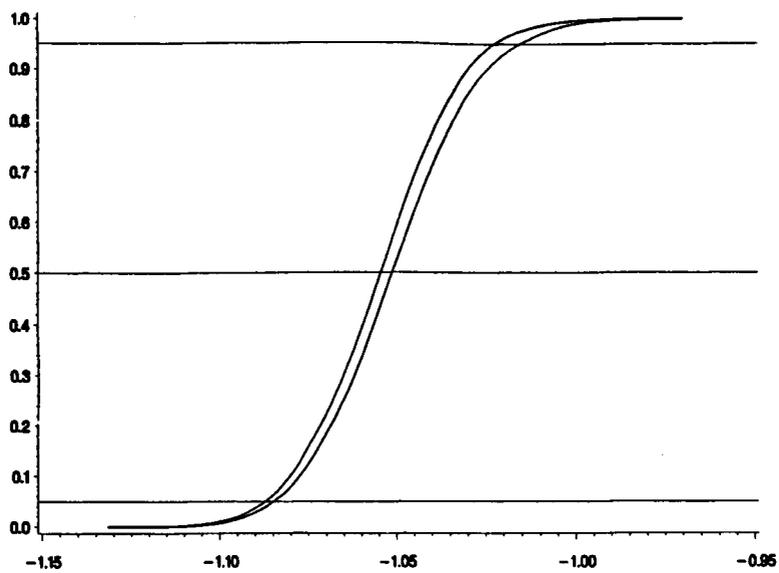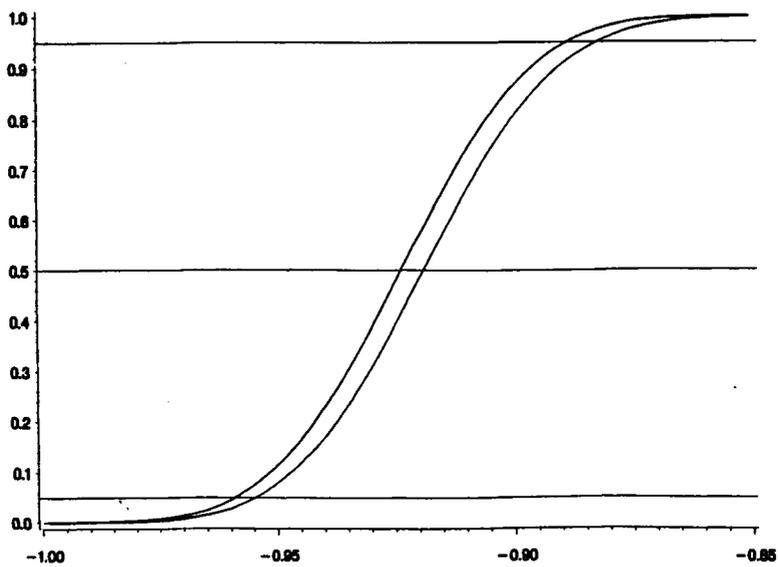
Fig 5. Estimated marginal posterior cumulative distribution functions of the second (a) and third (b) thresholds (chains 1 and 2).

Table I. Marginal posterior means (MPM), Monte-Carlo (MC) standard deviation (sd) of MPM, standard deviation of marginal posterior distribution, and effective number of independent observations $(\widehat{\psi}_m)$ for chains 1 and 2, for sire variance (σ_a^2), variance of liability (σ_e^2), mean (μ), threshold parameters (t_i) and sires (a_i).

	MPM		*MC sd* $100(\hat{Var}(\widehat{\mu}_m)^{1/2})$		*Posterior sd* $(\widehat{\gamma}_m(0)^{1/2})$		$\widehat{\psi}_m = \dfrac{\widehat{\gamma}_m(0)}{\hat{Var}(\widehat{\mu}_m)}$	
	Chain 1	*Chain 2*	*Chain 1*	*Chain 2*	*Chain 1*	*Chain 2*	*Chain 1*	*Chain 2*
σ_a^2	0.0983	0.0986	0.014	0.014	0.0237	0.0237	28 187	29 848
σ_e^2	1.0000	1.0028	0.082	0.076	0.0425	0.0429	2 659	3 205
μ	0.0018	0.0001	0.049	0.048	0.0288	0.0287	3 494	3 668
t_2	−1.051	−1.055	0.121	0.099	0.0208	0.0198	298	399
t_3	−0.019	−0.924	0.143	0.119	0.0218	0.0212	230	321
t_4	−0.628	−0.628	0.180	0.134	0.0222	0.0219	152	267
t_5	−0.348	−0.352	0.179	0.128	0.0212	0.0210	152	270
a_1	0.177	0.176	0.096	0.096	0.1656	0.1655	29 726	29 832
a_2	−0.092	−0.091	0.154	0.153	0.2631	0.2636	29 138	29 839
a_3	0.3037	0.3039	0.051	0.050	0.0878	0.0878	29 661	30 235

here likelihood inference with numerical integration is a competing alternative. For this model and data, marginal posterior distributions of sire effects are well approximated by normal distributions. On the other hand, with little information per random effect, *eg*, animal models, the normality of marginal posterior distributions when variances are not known is unlikely to hold. A strength of the Bayesian approach *via* Gibbs sampling is that inferences can be made from marginal posterior distributions in small sample problems, without resorting to asymptotic approximations. Further, the Gibbs sampler can accommodate multivariately distributed random effects, such as is the case with animal models, and this cannot be implemented with numerical integration techniques.

It seems important to investigate threshold models further, especially with sparse data structures consisting of many fixed effects with few observations per subclass. This case was studied by Moreno *et al* (manuscript submitted for publication) in the binary model, where they investigated frequency properties of Bayesian point estimators in a simulation study. They showed that improper uniform prior distributions for the fixed effects lead to an average of the marginal posterior mean of heritability which was larger than the simulated value. They obtained better agreement when fixed effects were assigned proper normal prior distributions. The use of 'non-informative' improper prior distributions is discouraged on several grounds by, among others, Berger and Bernardo (1992), as this can lead to improper posterior distributions. It seems that the disagreement between simulated and estimated values in Moreno *et al* is due to lack of information and not due to impropriety of posterior distributions. Thus, the bias persists, though smaller, when all parameters of the model are assigned proper prior distributions (Moreno, personal communication). It is clear that as the number of fixed effects increases, for a constant amount of observations, a larger proportion of fixed effect levels will

contain data falling into only one of the 2 dichotomies. There is no information in the data (in the Fisherian sense) to estimate these fixed effects and the likelihood is ill-conditioned. In the case of these sparse data structures, the choice of the prior distribution for the fixed effects may well be the most critical part of the problem. This needs to be studied further.

Data augmentation in the Gibbs sampler led to conditional posterior distributions which are easy to sample from. This facilitates programming. We have noted though, that threshold parameters have very slow mixing properties, and this is probably related to the data augmentation approach used in this study (Liu *et al*, 1994). With our data, the parameterization in terms of the residual variance resulted in smaller autocorrelations between samples of σ_e^2 than between samples of the thresholds. A scheme that is likely to accelerate mixing is to sample jointly from the threshold and liability. This step may necessitate other Monte-Carlo sampling techniques such as a Metropolis algorithm (Tanner, 1993), since sampling is from a non-standard distribution. Alternative computational strategies and parameterizations of the model may be more critical with animal models. Here, there is typically little information on additive genetic effects, and these are correlated. These properties slow down convergence of the Gibbs chain.

The methods described in this paper can be adapted easily to draw inferences about genetic change when selection is for categorical data. Sorensen *et al* (1994) described how to make inferences about response to selection in the context of the Gaussian model. In the threshold model, the only difference is that observed data are replaced by the unobserved underlying variable U (liability). In order to make inferences about response to selection, the parameterization must be in terms of an animal model.

REFERENCES

Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J Am Stat Assoc* 88, 669-679

Andersen S, Andresen E, Christensen K (1988) Hip dysplasia selection index exemplified by data from German shepherd dogs. *J Anim Breed Genet* 105, 112-119

Anderson DA, Aitkin M (1985) Variance component models with binary response: interviewer variability. *J R Stat Soc B* 47, 203-210

Berger JO, Bernardo JM (1992) On the development of reference priors. *In: Bayesian Statistics IV* (JM Bernardo, JO Berger, AP Dawid, AFM Smith, eds), Oxford University Press, UK, 35-60

Besag J, Green PJ (1993) Spatial statistics and Bayesian computation. *J R Stat Soc B* 55, 25-37

Bliss CI (1935) The calculation of the dosage–mortality curve. *Ann Appl Biol* 22, 134-167

Bulmer MG (1971) The effect of selection on genetic variability. *Am Nat* 105, 201-211

Cox DR, Miller HD (1965) *The Theory of Stochastic Processes*. Chapman and Hall, London, UK

Curnow R, Smith C (1975) Multifactorial models for familial diseases in man. *J R Statist Soc A* 138, 131-169

Dempster ER, Lerner IM (1950) Heritability of threshold characters. *Genetics* 35, 212-236

Devroye L (1986) *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, USA

Falconer DS (1965) The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann Hum Genet* 29, 51-76

Fédération Cynologique Internationale (1983) Hip dysplasia–international certificate and evaluation of radiographs (W Brass, S Paatsama, eds), Helsinki, Finland

Foulley JL (1993) Prediction of selection response for threshold dichotomous traits. *Genetics* 132, 1187-1194

Foulley JL, Manfredi E (1991) Approches statistiques de l'évaluation génétique des reproducteurs pour des caractères binaires à seuils. *Genet Sel Evol* 23, 309-338

Foulley JL, Im S, Gianola D, Höschele I (1987) Empirical Bayes estimation of parameters for n polygenic binary traits. *Genet Sel Evol* 19, 197-224

Foulley JL, Gianola D, Im S (1990) Genetic evaluation for discrete polygenic traits in animal breeding. *In: Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, Hammond K, eds) Springer-Verlag, NY, USA, 361-409

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

Gelfand A, Smith AFM, Lee TM (1992) Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *J Am Stat Assoc* 87, 523-432

German S, Geman D (1984) Stochastic relaxation, Gibbs distributions and Bayesian restoration of images. *IEEE Trans Pattern Anal Machine Intelligence* 6, 721-741

Gelman A, Rubin DB (1992) A single series from the Gibbs sampler provides a false sense of security. *In: Bayesian Statistics IV* (JM Bernardo, JO Berger, AP Dawid, AFM Smith, eds), Oxford University Press, UK, 625-631

Geweke J (1992) Evaluating the accurracy of sampling-based approaches to the calculation of posterior moments. *In: Bayesian Statistics IV* (JM Bernardo, JO Berger, AP Dawid, AFM Smith, eds), Oxford University Press, UK, 169-193

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

Harville DA, Mee RW (1984) A mixed-model procedure for analyzing ordered categorical data. *Biometrics* 40, 393-408

Höschele I, Gianola D, Foulley JL (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet* 104, 334-349

Jensen J, Wang CS, Sorensen DA, Gianola D (1994) Marginal inferences of variance and covariance components for traits influenced by maternal and direct genetic effects using the Gibbs sampler. *Acta Agric Scand* 44, 193-201

Liu JS, Wong WH, Kong A (1994) Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* 81, 27-40

Mood AM, Graybill FA, Boes DC (1974) *Introduction to the Theory of Statistics.* McGraw-Hill, NY, USA

Morton NE, McLean CJ (1974) Analysis of family resemblance. III. Complex segregation of quantitative traits. *Am J Hum Genet* 26, 489-503

Priestley MB (1981) *Spectral Analysis and Time Series.* Academic Press, London, UK

Quaas RL, Pollak EJ (1980) Mixed-model methodology for farm and ranch beef cattle testing programs. *J Anim Sci* 51, 1277-1287

Raftery AE, Lewis SM (1992) How many iterations in the Gibbs sampler? *In: Bayesian Statistics IV* (JM Bernardo, JO Berger, AP Dawid, AFM Smith, eds), Oxford University Press, UK, 763-773

Ripley BD (1987) *Stochastic Simulation.* John Wiley & Sons, NY, USA

Roberts GO (1992) Convergence diagnostics of the Gibbs sampler. *In: Bayesian Statistics IV* (JM Bernardo, JO Berger, AP Dawid, AFM Smith, eds), Oxford University Press, 775-782

Roberts GO, Polson NG (1994) On the geometric convergence of the Gibbs sampler. *J R Statist Soc B* 56, 377-384

Robertson A, Lerner IM (1949) The heritability of all-or-none traits: viability of poultry. *Genetics* 34, 395-411

Smith AFM, Roberts GO (1993) Bayesian computation *via* the Gibbs sampler and related Markov chain Monte-Carlo methods. *J R Stat Soc B* 55, 3-23

Sorensen DA, Wang CS, Jensen J, Gianola D (1994) Bayesian analysis of genetic change due to selection using Gibbs sampling. *Genet Sel Evol* 26, 333-360

Tanner MA (1993) *Tools for Statistical Inference.* Springer-Verlag, NY, USA

Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. *J Am Stat Assoc* 82, 528-540

Wang CS, Rutledge JJ, Gianola D (1993) Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet Sel Evol* 21, 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 DA, 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

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

APPENDIX

Here it is shown that there is a one-to-one relationship between 2 chosen parameterizations of the threshold model, and that they lead to the same probability distribution.

As a starting point assume that the conditional distribution of liability is:

$$U_i|\boldsymbol{\theta} \sim N(\mathbf{w}_i'\boldsymbol{\theta}, \sigma_e^2) \qquad [1.1]$$

where $\boldsymbol{\theta} = (b_1, \ldots, b_p, a_1, \ldots a_q)$, b_1 is an intercept common to all observations, and associated with C categories there are thresholds t_i, $i = 0, 1, \ldots, C$, with $t_0 = -\infty$ and $t_C = \infty$. We assume that the p fixed effects are estimable. In order to make the parameters identifiable, in what we call the *standard parameterization*, we define:

$$\widetilde{U}_i = \frac{U_i - t_1 + f_1}{\sigma_e}, \text{ such that } \widetilde{b}_1 = \frac{b_1 - t_1 + f_1}{\sigma_e}, \ \widetilde{b}_i = \frac{b_i}{\sigma_e}, \ (i = 2, \ldots, p),$$

$$\widetilde{a}_i = \frac{a_i}{\sigma_e}(i = 1, \ldots, q), \ \widetilde{t}_i = \frac{t_i - t_1 + f_1}{\sigma_e} \text{ and } \widetilde{\sigma}_a^2 = \frac{\sigma_a^2}{\sigma_e^2}$$

Note that by setting $f_1 = 0$, then $\widetilde{t}_1 = 0$. In terms of this parameterization, the conditional distribution of liability is:

$$\widetilde{U}_i|\widetilde{\boldsymbol{\theta}} \sim N(\mathbf{w}_i'\widetilde{\boldsymbol{\theta}}, 1) \qquad [1.2]$$

There are $p + q + C - 1$ identifiable unknown parameters which are:

$$\widetilde{\boldsymbol{\theta}} = (\widetilde{b}_1, \ldots, \widetilde{b}_p, \widetilde{a}_1, \ldots, \widetilde{a}_q), \ \widetilde{t}_i (i = 2, \ldots, C - 1) \text{ and } \widetilde{\sigma}_a^2.$$

In the *alternative parameterization*, one can define $\widehat{U}_i = \sigma_e \widetilde{U}_i$, such that:

$$\widehat{U}_i | \widehat{\boldsymbol{\theta}} \sim N(\mathbf{w}_i' \widehat{\boldsymbol{\theta}}, \sigma_e^2) \qquad [1.3]$$

where $\widehat{\boldsymbol{\theta}} = \sigma_e \widetilde{\boldsymbol{\theta}}, \widehat{\sigma}_a^2 = \widehat{\sigma}_a^2 \sigma_e^2, \sigma_e = \widehat{t}_i / \widetilde{t}_i$, and $\widehat{t}_1 = f_1$. The number of identifiable parameters is of course the same as before. In this paper we chose to set $t_{C-1} = f_2$, where $f_2 > f_1$ is an arbitrary known constant.

The probability distributions are given by:

$P(Y_i = j | \widetilde{\boldsymbol{\theta}}, \widetilde{t}) = P(\widetilde{t}_{j-1} < \widetilde{U}_i \leqslant \widetilde{t}_j | \widetilde{\boldsymbol{\theta}}, \widetilde{t}) = P(\widetilde{t}_{j-1} < \widetilde{U}_i \leqslant \widetilde{t}_j | \widehat{\boldsymbol{\theta}}, \widehat{t})$ (since the relationship between $(\widetilde{\boldsymbol{\theta}}, \widetilde{t})$ and $(\widehat{\boldsymbol{\theta}}, \widehat{t})$ is one-to-one) $= P(\widehat{t}_{j-1} < \widehat{U}_i \leqslant \widehat{t}_j | \widehat{\boldsymbol{\theta}}, \widehat{t}) = P(Y_i = j | \widehat{\boldsymbol{\theta}}, \widehat{t}).$

Finally we notice that $\widetilde{h}^2 = \widehat{h}^2$.