

# On biased inferences about variance components in the binary threshold model

C Moreno<sup>1</sup>, D Sorensen<sup>2\*</sup>, LA García-Cortés<sup>1</sup>,  
L Varona<sup>1</sup>, J Altarriba<sup>1</sup>

<sup>1</sup> *Unidad de Genética Cuantitativa y Mejora Animal, Facultad de Veterinaria, Universidad de Zaragoza, Miguel Servet 177, 50013 Zaragoza, Spain;*

<sup>2</sup> *Danish Institute of Animal Science, Research Center Foulum, PO Box 39, DK-8830 Tjele, Denmark*

(Received 2 September 1996; accepted 6 February 1997)

**Summary** – A simulation study was conducted to study frequentist properties of three estimators of the variance component in a mixed effect binary threshold model. The three estimators were: the mode of a normal approximation to the marginal posterior distribution of the component, which is denoted in the literature as marginal maximum likelihood (MML); the mean of the marginal posterior distribution of the component, using the Gibbs sampler to perform the marginalisations (GSR); and third, the mode of the joint posterior distributions of location and the variance parameter, used in conjunction with the iterative bootstrap bias correction (MJP-IBC). The latter was recently proposed in the literature as a method to obtain nearly unbiased estimators. The results of this study confirm that MML can yield biased inferences about the variance component, and that the sign of the bias depends on the amount of information associated with either fixed effects or with random effects. GSR can produce positively biased inferences when the amount of data per fixed effect is small. When fixed effects are poorly estimated, the bias persists, despite the fact that posterior distributions are guaranteed to be proper, and that the amount of information about the variance component is large. In this case, the marginal posterior distribution of the variance component is highly peaked and symmetric, but it shows a shift towards the right with respect to the true (simulated) value. This bias can be reduced by assigning a Gaussian probability density function to the prior distribution of the fixed effects, but this strategy does not work with very sparse data structures. The method based on MJP-IBC yielded unbiased inferences about the variance component in all the cases studied. This estimator is computationally simple, but contrary to GSR with normal priors for the fixed effects, can lead to estimates that fall outside the parameter space.

**binary traits / Gibbs sampling / biased estimators / bootstrap**

---

\* Correspondence and reprints

**Résumé – À propos des biais d'estimation de composantes de variance dans le modèle binaire à seuil.** Une étude de simulation a été conduite pour étudier les propriétés fréquentistes de trois estimateurs de la composante de variance dans un modèle mixte à seuils pour variable binaire. Les trois estimateurs ont été : le mode de l'approximation normale de la distribution marginale a posteriori de la composante qui est appelée dans la littérature « maximum de vraisemblance marginale » (MML), l'espérance de distribution marginale a posteriori de la composante, en utilisant l'échantillonnage de Gibbs (GSR) et finalement, le mode de la distribution conjointe a posteriori des paramètres de position et de variance, utilisé avec une méthode itérative de correction des biais par bootstrap (MJP-IBC). Cette dernière a été récemment proposée dans la littérature comme une méthode d'obtention d'estimateurs approximativement non biaisés. Les résultats de l'étude confirment que le MML peut donner des inférences biaisées à propos de la composante de variance et que le signe du biais dépend de la quantité d'information attachée aux effets fixes ou aux effets aléatoires. Le GSR peut produire des estimées à biais positif quand il y a peu d'information par niveau d'effet fixe. Dans ce cas, le biais persiste même quand l'information sur la composante de variance est importante et quand la distribution postérieure est propre. La distribution marginale a posteriori est hautement concentrée et symétrique mais montre un décalage à droite de la vraie valeur (simulée). Ce biais peut être réduit en considérant que la distribution a priori des effets fixes est de type gaussien mais cette stratégie n'est pas efficace pour les structures de données où l'information est très dispersée. La méthode MJB-IBC a produit des estimées non biaisées de la composante de variance dans tous les cas étudiés. Cet estimateur est simple à calculer mais contrairement à GSR avec des effets fixes à distribution gaussienne, elle peut mener à des estimées situées en dehors de l'espace des paramètres.

caractères binaires / échantillonnage de Gibbs / estimateurs biaisés / bootstrap

## INTRODUCTION

Ordered categorical traits are often analysed by means of the threshold liability model, 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. Here we will consider the situation where the liability is modelled as a linear function of fixed effects and possibly correlated random effects. In a typical likelihood setting, one may be interested in making joint inferences about estimable functions of fixed effects and on variance components associated with the random effects. Within a Bayesian framework, one may concentrate inferences on the marginal posterior distributions of contrasts of 'fixed' effects and of the variance components, and possibly on the marginal posterior distributions of the random effects. In a genetic context, the latter involve typically additive genetic values and knowledge about these is important to carry out selection decisions and to describe response to selection. Other inferences of interest may require computing probabilities that observations fall in a given category for given individuals or combinations of parameters.

All these inferential problems, whether in a likelihood or in a Bayesian setting, require marginalisations. For example, computation of the likelihood or of the joint posterior distribution of fixed effects and variance components requires integration

with respect to the random effects. Although these integrals do not have a closed form solution, numerical integration using quadrature is feasible when the random effects are uncorrelated (Anderson and Aitkin, 1985). When random effects are correlated, as is the case with animal models, these numerical integration techniques cannot be used. To circumvent this problem, a large number of approximations and different approaches has been suggested in the literature (see, for example, McCullagh and Nelder (1989) for a discussion of some of these, and Foulley et al (1990) and Foulley and Manfredi (1991) for a review of the analysis of categorical traits in an animal breeding context).

The introduction of Markov chain Monte Carlo methods allows computation of multidimensional integrals and analytic approximations can therefore be avoided. The use of the Gibbs sampler to analyse ordered categorical traits has been reported by Zeger and Karim (1991) and by Albert and Chib (1993), using a Bayesian perspective. McCulloch (1994) applies the Gibbs sampler to estimate variance components for binary data in a likelihood setting. In an animal breeding context, the Gibbs sampler was applied to the analysis of categorical traits by Moreno (1993) and by Sorensen et al (1995), both from a Bayesian perspective.

Zeger et al (1988) studied non-linear models for the analysis of longitudinal data, and drew attention to the dependence between the marginal expectation of the data and the variance of the random effects. In the context of a generalized estimating equations approach, they showed that estimators of the fixed effects and of the variance component associated with the random effects are not even asymptotically orthogonal, as they are in the Gaussian model. A similar dependency in the case of the logistic mixed model was shown by Drum and McCullagh (1993) in the context of restricted maximum likelihood, where the latter was constructed using error contrasts. This approach to constructing the restricted likelihood generates an equation that is not a function of the fixed effects in the case of the Gaussian model, but not in the case of non-linear models. The association between the fixed effects and the variance component indicates that inference about the latter is dependent upon the amount of information in the data with which fixed effects are estimated.

The study of Moreno (1993) reported frequentist properties of the mean of the marginal posterior distribution of the variance component in a binary mixed threshold model. In this study, a sire model was assumed. Moreno (1993) noted that when there is little information about fixed effects, the mean of the marginal posterior distribution of the variance component is biased upwards. This is so even in cases where there are very large numbers of sires and offspring per sire, and consequently this marginal posterior distribution is peaked and symmetric. It is well known that when the data associated with a particular fixed effect fall all within the same category (known as the extreme category problem (ECP), the likelihood is underidentified (eg, Misztal et al, 1989; Gelman et al, 1995). In this situation, in a Bayesian setting, the assignment of improper prior distributions to the 'fixed effects' can lead to improper posterior distributions. However, we show that the bias related to inference about the variance component when there is little information about fixed effects persists, when all the parameters of the model are assigned proper priors. In this case, the posterior distribution is guaranteed to be proper.

This bias problem has been the subject of an extensive simulation study by Hoeschele and Tier (1995). These authors, who used the Gibbs sampler with data augmentation (Tanner and Wong, 1987), confirm the results of Moreno (1993), and indicate that part of the problem can be alleviated by assuming that 'fixed effects' follow a priori a Gaussian distribution. As we show below, treating 'fixed effects' as if they were 'random' can reduce the bias associated with the mean of the marginal posterior distribution of the variance component, but this is not a strategy that always works.

An alternative approach to inferences in the mixed threshold model was presented by Tempelman (1994). The method is quite general and is based on an approximation to the marginal posterior distribution of the variance component, where the marginalisation is obtained using Laplacian integration. The computational burden of this method amounts to obtaining repeatedly the determinant of the Hessian of the joint posterior distribution, conditional on the current value of the variance component. The dimension of the Hessian is equal to the number of 'fixed' and random effects in the model. In a small simulation study, Tempelman (1994) showed that the method performed well in a sire model, but yielded biased estimates of the variance component in an animal model.

Breslow and Lin (1995) have recently developed methods to adjust for the bias in approximate estimators of the variance component and regression coefficients for generalised linear mixed models having one source of random variation. The bias correction applies to approximations based on penalised quasilielihood and holds for generalised linear mixed models having canonical link functions, the latter being a condition that is not satisfied in the case of the probit mixed model studied in this work.

Kuk (1995) proposed an alternative iterative method using the bootstrap to obtain asymptotically unbiased estimators. The appealing aspect of this approach is that in principle it can be applied to any estimator. Though computationally intensive, the method can be implemented in a straightforward manner.

The objective of this work is to study frequentist properties of three approaches to making inferences about variance components in mixed binary threshold models. One of these approaches, which does not rely on analytic approximations, is based on the mean of the marginal posterior distribution of the variance components. We study its properties using different data structures and assuming either proper or improper prior distributions for the 'fixed' effects and for the variance component. We report also frequentist properties of a computationally simple estimator based on the mode of the joint posterior distribution of location and scale parameters, used in conjunction with the iterative bias correction using the bootstrap proposed by Kuk (1995).

## MATERIAL AND METHODS

### *Model for the analysis of binary responses*

A Bernoulli random variable  $y_{ij}$  is recorded for observation  $j$  ( $j = 1, \dots, n_i$ ) in subclass  $i$  ( $i = 1, \dots, s$ ), and takes the value  $y_{ij} = 1$  if  $\ell_{ij} > t$  or  $y_{ij} = 0$  otherwise, where  $\ell_{ij}$  represents the underlying variable (liability) and  $t$  is the threshold.

Conditionally on fixed effects (herds) ( $\mathbf{b}$ , of order  $p$ ) and on random (sire) effects ( $\mathbf{u}$ , of order  $q$ ), the  $\ell_{ij}$  are independently and normally distributed, ie,  $N(\mathbf{x}'_i\mathbf{b} + \mathbf{z}'_i\mathbf{u}, 1)$ , and the  $y_{ij}$  are independent with  $\Pr(y_{ij} = 0|\mathbf{b}, \mathbf{u}) = \Phi(t - \mathbf{x}'_i\mathbf{b} - \mathbf{z}'_i\mathbf{u})$ . The symbol  $\Phi(\cdot)$  denotes the standard normal cumulative distribution function and  $\mathbf{x}'_i(\mathbf{z}'_i)$  denotes the  $i$ th row of the known design matrix  $\mathbf{X}(\mathbf{Z})$ . The subclass  $i$  is defined as the herd-sire combination  $i$ . We will denote the prior distribution for the fixed effects as  $p(\mathbf{b})$ . Invoking an infinitesimal model, the vector of sire effects is multivariate normally distributed:  $\mathbf{u}|\mathbf{A}, \sigma_u^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_u^2)$ , where  $\mathbf{A}$  is a known  $q$  by  $q$  matrix of additive genetic relationships among sires, and  $\sigma_u^2$  is the unknown variance between sires. The prior probability density for  $\sigma_u^2$  will be denoted  $p(\sigma_u^2)$ . Under the model, the conditional distribution of the data  $\mathbf{y}$ , given  $\mathbf{b}$  and  $\mathbf{u}$  is:

$$p(\mathbf{y}|\mathbf{b}, \mathbf{u}) \propto \prod_{i=1}^s [\Phi(t - \mathbf{x}'_i\mathbf{b} - \mathbf{z}'_i\mathbf{u})]^{n_{i1}} [1 - \Phi(t - \mathbf{x}'_i\mathbf{b} - \mathbf{z}'_i\mathbf{u})]^{n_{i2}} \quad [1]$$

where  $n_{i1}$  is the number of observations in class  $i$  with  $y_{ij} = 0$  and  $n_{i2}$  is the number of observations in class  $i$  with  $y_{ij} = 1$ . The joint posterior distribution is given by:

$$p(\mathbf{b}, \mathbf{u}, \sigma_u^2|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{b}, \mathbf{u})p(\mathbf{b})p(\mathbf{u}|\sigma_u^2)p(\sigma_u^2) \quad [2]$$

Unless otherwise stated,  $p(\mathbf{b})$  is assumed uniform in the interval  $[t - 5, t + 5]$ , and  $p(\sigma_u^2)$  is assumed uniform in the interval  $[0, 1/3]$ . Consequently, in these cases, posterior distributions are proper.

As is well established (Cox and Snell, 1989), in the binary threshold model neither the residual variance nor the threshold can be estimated. In this work, the simulated data are analysed using a model in which the residual variance is set equal to 1, and the value of the threshold is arbitrarily set equal to the simulated value.

### Methods of inference

In this section we briefly describe the three methods of inference studied in this work. These are, first, marginal maximum likelihood (MML) (Foulley et al, 1987). The second method is based on an estimator of the mean of the marginal posterior distribution of the sire variance. Here, marginalisations were obtained using the Gibbs sampler applying an adaptive rejection sampling algorithm due to Gilks and Wild (1992) (GSR). The third method is based on the mode of the joint posterior distribution [2], used in conjunction with the iterative bootstrap bias correction (MJP-IBC).

The method based on MML was chosen because it is one of the most common methods used in the analysis of categorical traits. Method GSR does not involve analytic approximations and therefore it is interesting to compare it to MML under a variety of data structures. As we show below, both MML and GSR in the form implemented in this work, can lead to biased inferences, even in cases with a large number of sires and many offspring per sire. Method MJP-IBC is an ad hoc method which yields inferences with no detectable bias.

### Marginal maximum likelihood (MML)

This method is based on locating the modal value of an approximation to  $\ln p(\sigma_u^2|\mathbf{y})$ , the logarithm of the marginal posterior distribution of the variance component. As shown in Foulley et al (1987) and applying their result for the case  $\sigma_u^2 \sim \text{Uniform}[0, 1/3]$ :

$$\frac{\partial \ln \left[ \int_{\mathbf{b}, \mathbf{u}} p(\mathbf{b}, \mathbf{u}, \sigma_u^2 | \mathbf{y}) d\mathbf{b} d\mathbf{u} \right]}{\partial \sigma_u^2} = E_{\mathbf{u} | \mathbf{y}, \sigma_u^2} \left[ \frac{\partial \ln p(\mathbf{u} | \sigma_u^2)}{\partial \sigma_u^2} \right] \quad [3]$$

The expression on the right hand side of [3] involves finding the expected value of  $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}$ , where this expectation is taken with respect to the distribution of  $\mathbf{u} | \mathbf{y}, \sigma_u^2$ . That is:

$$E_{\mathbf{u} | \mathbf{y}, \sigma_u^2} [\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}] = [E(\mathbf{u} | \mathbf{y}, \sigma_u^2)]' \mathbf{A}^{-1} [E(\mathbf{u} | \mathbf{y}, \sigma_u^2)] + \text{tr} [\mathbf{A}^{-1} \text{Var}(\mathbf{u} | \mathbf{y}, \sigma_u^2)] \quad [4]$$

In the Gaussian model,  $\mathbf{u} | \mathbf{y}, \sigma_u^2$  is normal, and the expectation in [4] involves the solution of Henderson's (1973) mixed model equations. This yields  $E(\mathbf{u} | \mathbf{y}, \sigma_u^2)$ , and the inverse of the coefficient matrix of the mixed model equations, yields  $\text{Var}(\mathbf{u} | \mathbf{y}, \sigma_u^2)$ . In the context of the present binary model, the form of the distribution of  $\mathbf{u} | \mathbf{y}, \sigma_u^2$  is not known. MML consists of approximating  $E(\mathbf{u} | \mathbf{y}, \sigma_u^2)$  by the mode of  $p(\mathbf{b}, \mathbf{u} | \mathbf{y}, \sigma_u^2)$  and  $\text{Var}(\mathbf{u} | \mathbf{y}, \sigma_u^2)$  by the inverse of minus the expected value of the second derivatives of  $p(\mathbf{b}, \mathbf{u} | \mathbf{y}, \sigma_u^2)$  with respect to  $\mathbf{b}$  and  $\mathbf{u}$ . Setting [3] equal to zero and solving for the unknown, leads to an iterative scheme which involves the solution of a non-linear system reminiscent of Henderson's (1973) mixed model equations. The equation for  $\sigma_u^2$  requires the computation of:

$$\sigma_u^2 = [\hat{\mathbf{u}}' \mathbf{A}^{-1} \hat{\mathbf{u}} + t(\mathbf{A}^{-1} \mathbf{C}_{uu})] / q \quad [5]$$

where  $\mathbf{C}_{uu}$  is the part of the complete inverse of the coefficient matrix corresponding to the  $\mathbf{u}$  by  $\mathbf{u}$  block.

This method is known to yield biased inference in small samples with respect to the variance component (Hoeschele et al, 1987; Dellaportas and Smith, 1993; Engel et al, 1995). The sign of the bias depends on the data structure (Hoeschele and Tier, 1995; Engel et al, 1995). With a large amount of information on fixed effects and little information on random effects, the bias is negative. When the amount of information on the random effects is large, but fixed effects are poorly estimated, the bias is positive. This behaviour of the estimator can lead to situations where no bias is detected.

### Mean of the marginal posterior distribution of sire variance, using the Gibbs sampler (GSR)

The Gibbs sampler is an iterative numerical technique for the approximate generation of (possibly dependent) samples from multivariate distributions. The iterative scheme consists of obtaining draws from the fully conditional posterior distributions

of all the parameters of the model. The literature on Markov chain Monte Carlo methods is very large and many theoretical and applied issues can be found in the recent publication of Gilks et al (1996).

In many applications, the fully conditional posterior distributions are known, and the implementation of the Gibbs sampler is straightforward. This is the case of the binary threshold mixed model, when the Gibbs sampler is used in conjunction with data augmentation, as in Albert and Chib (1993). If the technique of data augmentation is not used, the present model leads to fully conditional posterior distributions of the location parameters that are not of standard form. In this case, other sampling schemes can be implemented, such as the adaptive rejection sampling proposed by Gilks and Wild (1992). This is the approach chosen in this paper. Details of the fully conditional posterior distributions can be found in the *Appendix*. We follow closely the adaptive rejection sampling algorithm as described in Dellaportas and Smith (1993).

The implementation of GSR requires draws from the fully conditional posterior distributions of the parameters. One of these is the fully conditional posterior distribution of the sire variance,  $p(\sigma_u^2 | \mathbf{b}, \mathbf{u}, \mathbf{y})$ , which is in the form of a scaled inverted chi-square. The heritability,  $h^2 = 4\sigma_u^2 / (\sigma_u^2 + 1)$ , was constructed for each value drawn from  $p(\sigma_u^2 | \mathbf{b}, \mathbf{u}, \mathbf{y})$ . The Monte Carlo estimate of the mean of the marginal posterior distribution of heritability was computed as the mean  $h^2$  over the Gibbs samples. In the present study, differences between mean, mode and median of the marginal posterior distribution of heritability were negligible and results based on the mean only are reported. The Gibbs sampler was implemented using several chains, and all the samples in each were kept. Burn in periods were determined on a pragmatic basis. The criterion of Gelman and Rubin (1992) was adopted to determine convergence.

The Gibbs sampler allows the drawing of inferences using a fully Bayesian approach, without resorting to analytical approximations. Asymptotically, posterior distributions are normal, with mean equal to the maximum likelihood estimator. It seems relevant to study the behaviour of this approach under different data structures, in view of the fact that analytic approximations are not involved.

### Mode of the joint posterior distribution followed by the iterative boost-rap bias correction (MJP-IBC)

The first step of this method is to find the mode of the joint posterior distribution of location and dispersion parameters [2], under the assumption that the prior of  $p(\sigma_u^2)$  follows a scaled inverted chi-square distribution. The latter is of the form:

$$p(\sigma_u^2) \propto (\sigma_u^2)^{-\left(\frac{q^*}{2} + 1\right)} \exp\left(-\frac{q^* \sigma_u^{*2}}{2}\right) \quad [6]$$

where  $q^*$  and  $\sigma_u^{*2}$  are parameters of the prior distribution. In the present study,  $q^*$  was set equal to  $q$ , and  $\sigma_u^{*2}$  was derived assuming  $h^2 = 0.9$ . Other starting values of  $h^2$  were also tried, but convergence was faster using this value. Differentiating

the resulting joint posterior with respect to  $\sigma_u^2$ , setting to zero and solving for  $\sigma_u^2$ , produces the iterative scheme:

$$\hat{\sigma}_u^2 = [\hat{\mathbf{u}}' \mathbf{A}^{-1} \hat{\mathbf{u}} + q^* \sigma_u^{*2}] / [q + q^* + 2] \tag{7}$$

as shown by Hoeschele et al (1987), where  $\hat{\mathbf{u}}$  is the mode of  $p(\mathbf{b}, \mathbf{u} | \sigma_u^2, \mathbf{y})$  as in MML. Notice that in contrast to [5], [7] does not require knowledge of the inverse of any matrix and is very easy to compute. Use of a proper prior distribution for  $\sigma_u^2$  is important, because with an improper uniform prior, this method is known to yield zero estimates for  $\sigma_u^2$  (Lindley and Smith, 1972).

The second step, consists of applying the iterative bias correction that was proposed by Kuk (1995). This procedure can be described as follows. Let  $\theta$  be the parameter of interest, and let  $\tilde{\theta}$  represent an initial estimate obtained as a solution to a system of estimating equations  $\psi(\theta | \mathbf{y})$ , where  $\mathbf{y}$  are the data that are assumed to follow a distribution  $f(\mathbf{y} | \theta)$ . The system of equations  $\psi(\theta | \mathbf{y})$  could be the score function, for example, or some approximation to it. In the present case,  $\psi(\theta | \mathbf{y})$  is the system of equations that results from the joint maximization of the posterior distribution with respect to location and dispersion parameters. Let  $\theta^*$ , the expected value of the estimator of  $\theta$ , satisfy:

$$\int \psi(\theta | \mathbf{y}) \Big|_{\theta=\theta^*} f(\mathbf{y} | \theta) d\mathbf{y} = \mathbf{0} \tag{8}$$

If the estimator of  $\theta$  is biased, the asymptotic bias is given by:

$$\boldsymbol{\beta}(\theta) = \theta^* - \theta = \mathbf{g}(\theta) - \theta \tag{9}$$

where  $g(\theta) = \theta^*$ . Kuk (1995) proposes the following correction for this bias. Let  $\boldsymbol{\beta}^{(0)}$  be an initial estimate of the bias of  $\tilde{\theta}$ . Then define an updated estimate of the bias as:

$$\boldsymbol{\beta}^{(k+1)} = \mathbf{g}(\tilde{\theta} - \boldsymbol{\beta}^{(k)}) - (\tilde{\theta} - \boldsymbol{\beta}^{(k)}) \tag{10}$$

and define an updated bias-corrected estimate of  $\theta$  as:

$$\hat{\theta}^{(k+1)} = \tilde{\theta} - \boldsymbol{\beta}^{(k+1)} \tag{11}$$

Kuk (1995) shows that this estimator is asymptotically consistent. Since the function  $\mathbf{g}(\theta)$  is not usually known, Kuk (1995) proposes to approximate  $\mathbf{g}(\theta)$  by  $\mathbf{g}_M(\theta)$ , which is the average of  $\tilde{\theta}$  over simulated samples. That is:

$$\mathbf{g}_M(\theta) = \frac{1}{m} \sum_{i=1}^{i=m} \tilde{\theta}(\mathbf{y}_i) \tag{12}$$

where  $\mathbf{y}_1, \dots, \mathbf{y}_m$  are simulated samples (ie, data) from  $f(\mathbf{y} | \theta)$ , and  $\tilde{\theta}(\mathbf{y}_i)$  is the estimate of  $\theta$  obtained using the simulated sample  $i$ ,  $\mathbf{y}_i$ . Replacing  $g$  by  $\mathbf{g}_M$  in [10], Kuk (1995) obtains:

$$\boldsymbol{\beta}_M^{(k+1)} = \mathbf{g}_M(\tilde{\theta} - \boldsymbol{\beta}_M^{(k)}) - (\tilde{\theta} - \boldsymbol{\beta}_M^{(k)}) \tag{13}$$

and

$$\hat{\boldsymbol{\theta}}^{(k+1)} = \tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\beta}}_M^{(k+1)} \quad [14]$$

The simulated values  $\mathbf{y}_1, \dots, \mathbf{y}_m$  are obtained from  $f(\mathbf{y}|\boldsymbol{\theta})$ , where in the  $k$ th round of iteration,  $\boldsymbol{\theta}$  is set to  $\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\beta}}_M^{(k)}$ .

The implementation of the above procedure using method (MJP-IBC) was as follows. First, estimates of location and scale parameters,  $\tilde{\boldsymbol{\theta}}$ , are obtained from the joint maximization of the posterior distribution [2], conditional on the original data. Second, simulated samples are generated from  $f(\mathbf{y}|\boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  is set to the value of these initial estimates  $\tilde{\boldsymbol{\theta}}$ . Third,  $\tilde{\boldsymbol{\theta}}(\mathbf{y}_i)$  are obtained from the joint maximization of the posterior distribution [2], conditional on the  $i$ th simulated sample ( $i = 1, \dots, m$ ). Fourth, an average of  $\tilde{\boldsymbol{\theta}}(\mathbf{y}_i)$  over the  $m$  samples is computed, which is denoted by  $\mathbf{g}_M(\boldsymbol{\theta})$ , and the estimate of the bias is obtained from [13]. Finally,  $\tilde{\boldsymbol{\theta}}$  is corrected using [14], which is used to generate a new set of simulated samples. This iterative scheme is repeated until convergence is achieved.

In the presence of the extreme category problem (ECP), it was required that the proportion of ECPs in the simulated samples was the same as in the original data. This was achieved by adjusting the values of the fixed effects associated with ECPs. Another point to highlight in the application of MJP-IBC in models with ECPs is that the correction in expression [14] involves the component of variance and the non-ECP fixed effects only. Moreover, if in a simulated sample  $\mathbf{y}_i$ , an ECP is generated for a fixed effect that was a non-ECP fixed effect in the original data, then the element in  $\tilde{\boldsymbol{\theta}}(\mathbf{y}_i)$  associated with this fixed effect, is set to the value that was used to simulate this fixed effect in the sample  $\mathbf{y}_i$ , and not to its estimated value. This ad hoc strategy was arrived at on a trial and error basis, and it was followed because it avoided problems of lack of convergence. Other strategies could be envisaged, one such being based on assigning mildly informative priors to the 'fixed' effects.

### Simulation study

A Monte Carlo simulation study was carried out in order to evaluate the three methods of inference under a variety of data structures. The criteria used to assess the performance of the methods were bias and mean square error. The data were simulated as follows. Liability was drawn from a normal distribution, with mean equal to the fixed effect corresponding to the individual in question and variance given by  $\sigma_u^2 + \sigma_e^2$ , where  $\sigma_u^2$  is the sire variance and  $\sigma_e^2$  was set equal to 1. Only one set of fixed effects were simulated (herds), and these were drawn from uniform distributions within the interval  $(-2, 2)$ . Sires were drawn from a normal distribution, with mean  $(1/2)u_p$  and variance  $(3/4)\sigma_u^2$ , where  $u_p$  is the realized value of the sire of the individual. Base population sires were drawn assuming a mean of zero and variance  $\sigma_u^2$ . Three non-overlapping generations of data (with no selection) were created in this way, and relationships among sires were included in the analyses of these data. In a few cases, the simulated data comprised unrelated base population sire families only, and the data were analysed accordingly.

The data at the phenotypic scale were generated as follows:

$$y_{ij} = \begin{cases} 0 & \text{if } l_{ij} \leq t \\ 1 & \text{if } l_{ij} > t \end{cases}$$

where the value of the known threshold  $t$ , was set depending on the desired level of incidence.

The simulations of the original data were carried out such that the resulting structures led to cases with different amounts of information associated with the parameters of the model. The simulation results were all based on 50 replicates.

## RESULTS

Estimates of heritability using MML and GSR (mean of the marginal posterior distribution of heritability) for three data structures are shown in table I. The analyses based on GSR assumed the proper uniform priors for the 'fixed' effect and the variance component, as described in *Model for the analysis of binary responses*. The structure A1 is one in which there is adequate information in the data to make inferences about all the parameters of the model. There is no ECP in this structure and the incidence is set at 60%. Both MML and GSR yield estimates that on average are in good agreement with the true value. In structure A2 there is adequate information related to fixed effects, but there is relatively less information associated with the random effects. In A2 there are no ECPs. In structure A3 there are ECPs, but there is a large amount of data associated with the non-ECP fixed effects and there is little information about the random effects. As expected, the MML yields estimates of heritability that are biased downwards. GSR does not show signs of bias. Structures A4 and A5 represent cases where there is little information about fixed effects. The percentage of ECPs is higher under structure A5 (65% in A5 versus 25% in A4). Both methods produce positively biased results; the bias using GSR is larger than that using MML.

**Table I.** Heritability estimates obtained with MML (marginal maximum likelihood) and GSR<sup>a</sup> (mean of the marginal posterior distribution using the Gibbs sampler) in a sire mixed latent model (50 replications in each case).

	FL	RL	OS	MF	MR	I	E	$h^2$	MML		GSR	
									$h_e^2$	MSE	$h_e^2$	MSE
A1	100	500	2	275	55	60	0	0.2	0.200	0.0006	0.208	0.0007
A2	15	500	2	275	8	60	0	0.2	0.184	0.0025	0.211	0.0030
A3	15	500	2	275	8	95	46	0.5	0.305	0.0527	0.512	0.0304
A4	900	250	10	15	55	60	25	0.2	0.251	0.0051	0.270	0.0086
A5	900	250	10	15	55	95	65	0.5	0.652	0.0599	0.772	0.1060

FL and RL: number of fixed and random levels. OS: maximum number of observations per subclass. MF and MR: mean number of observations per fixed and random level. I: incidence. E: percentage of ECP.  $h^2$  and  $h_e^2$ : true and estimated heritability. MSE: mean square error. <sup>a</sup> GSR implemented with proper prior distributions for all parameters.

Table II shows estimates of the mean of the marginal posterior distribution of the heritability using GSR under a variety of data structures, and performing the Bayesian analysis on the basis of proper and improper posterior distributions. The point of the results in the table is to draw attention to the type of data structures that can cause GSR to yield biased inferences about the variance component and to illustrate that the bias is not necessarily related to the impropriety of the posterior distribution. In cases B1 and B3, the data comprise unrelated sire families and an improper uniform prior was assigned to the fixed effects and to the variance component in such a way that the conditions required to guarantee that the posterior distribution is proper (Natarajan and McCulloch, 1995) are not met. Although the ECP is present at a level of 27% in B1, with a large number of families and a large number of observations per family and per fixed effect, the estimate of the mean of the marginal posterior distribution of heritability over the 50 Monte Carlo replicates shows good agreement with the true value. This result is disturbing since the sampler retrieved a meaningful estimate from a posterior distribution which is not guaranteed to be proper! This point is discussed at length in Casella (1996).

**Table II.** Heritability estimates obtained with GSR<sup>a</sup> (mean of the marginal posterior distribution using the Gibbs sampler) in a sire mixed latent model (50 replications in each case).

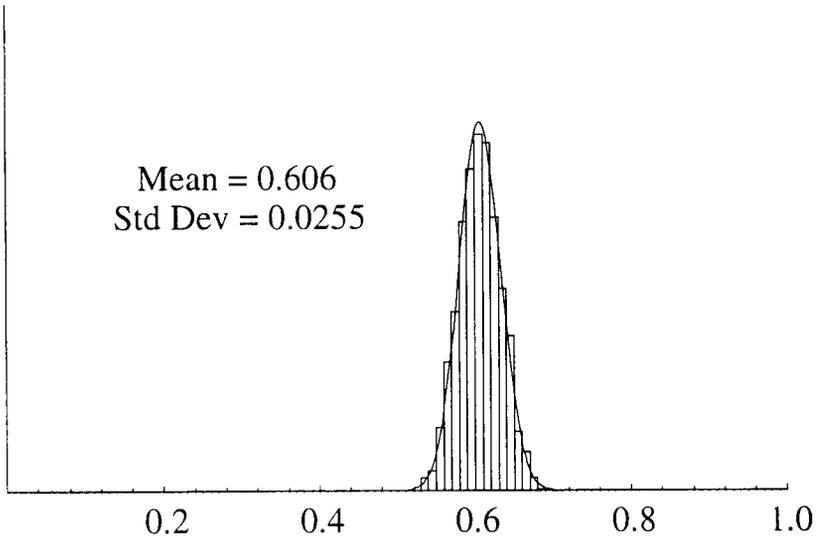
	<i>FL</i>	<i>RL</i>	<i>OS</i>	<i>MF</i>	<i>MR</i>	<i>I</i>	<i>E</i>	$h^2$	$h_e^2$	<i>MSE</i>
B1	100	500	10	275	55	90	27	0.5	0.519	0.0051
B2	100	500	2	38	8	65	0	0.5	0.598	0.0265
B3	100	500	2	38	8	65	0	0.5	0.605	0.0234
B4	800	4 000	2	38	8	65	0	0.5	0.593	0.0093

FL and RL: number of fixed and random levels. OS: maximum number of observations per subclass. MF and MR: mean number of observations per fixed and random level. *I*: incidence. *E*: percentage of ECP.  $h^2$  and  $h_e^2$ : true and estimated heritability. MSE: mean square error. <sup>a</sup> GSR implemented with proper prior distributions for all parameters in cases B2 and B4; and with improper priors for 'fixed' effects and for the variance component in cases B1 and B3.

In cases B2 and B3 there is a limited amount of information per random effect and per fixed effect. The ECP is not present. The difference between these two cases is that under B2, GSR was implemented with proper priors for all the parameters, which leads to a proper joint posterior distribution. However, the estimates using GSR are biased in both cases and the size of the bias is similar. Thus the presence of bias does not seem to be necessarily related to the fact that the posterior distribution is improper.

In B4, the analysis was carried out assigning proper priors to all the parameters and there are no ECPs (the marginal posterior distribution of heritability is guaranteed to be proper). In common with cases B2 and B3, there are relatively few observations per fixed and random effect. Relative to cases B2 and B3, case B4 comprises a large number of sire families (4 000). Indeed, the mean square error

of the heritability is almost three times smaller than in cases B2 and B3. Here, the marginal posterior distribution of heritability is highly peaked and symmetric, but is shifted to the right relative to the true value of 0.5. One of the 50 replicates was arbitrarily chosen and this distribution is shown in figure 1. Despite the large number of sire families, the absence of ECPs and the fact that the marginal posterior distribution is proper, the bias using GSR persists.



**Fig 1.** Histogram of marginal posterior distribution of heritability (with its mean and standard deviation) for one replicate corresponding to case B4 (see table II), and a normal distribution fit superimposed.

Table III shows comparisons of GSR and MJP-IBC. Method GSR was implemented assigning normally distributed priors for the fixed effects and the proper uniform distribution in the interval  $[0, 1/3]$  for the prior sire variance.

Cases C1 and C2 have the same data structure as B2 in table II, where there are relatively few observations per 'fixed' effect but no ECPs. Using GSR and assigning a normal prior to the 'fixed' effects is effective in removing the bias. Data structures C3 and C4 are identical to A4 and A5. These are characterised by few observations per 'fixed' effect, but with ECP. Again, assigning a normal prior to the 'fixed' effects produces estimates of heritability in better agreement with the true value. In all these four cases, MJP-IBC yields estimates of heritability in good agreement with the true value. The mean square errors of estimates based on GSR and on MJP-IBC are of similar magnitude.

In data structure C5, the frequency of ECP is 80% and the incidence is 90%. The strategy of using GSR treating fixed effects as random yields estimates that are severely biased upwards. MJP-IBC yields estimates that do not deviate significantly from the true value, with a mean square error two and a half times smaller.

**Table III.** Heritability estimates obtained with GSR<sup>a</sup> (mean of the marginal posterior distribution using the Gibbs sampler), where the fixed effects are estimated as if they were random, and MJP-IBC (mode of joint posterior distribution with iterative bias correction) in a sire mixed latent model (50 replications in each case).

	<i>FL</i>	<i>RL</i>	<i>OS</i>	<i>MF</i>	<i>MR</i>	<i>I</i>	<i>E</i>	$h^2$	$h_e^2$	<i>MML</i> MSE	<i>MJP-IBC</i> $h_e^2$ MSE
C1	100	500	2	38	8	65	0	0.2	0.211	0.0041	0.204 0.0051
C2	100	500	2	38	8	65	0	0.5	0.503	0.085	0.512 0.0074
C3	900	250	10	15	55	60	25	0.2	0.221	0.0053	0.177 0.0065
C4	900	250	10	15	55	95	65	0.5	0.475	0.0128	0.479 0.0128
C5	500	250	1	3	6	90	80	0.5	0.743	0.1976	0.496 0.0757

FL and RL: number of fixed and random levels. OS: maximum number of observations per subclass. MF and MR: mean number of observations per fixed and random level. *I*: incidence. *E*: percentage of ECP.  $h^2$  and  $h_e^2$ : true and estimated heritability. MSE: mean square error. <sup>a</sup> GSR implemented with proper prior distributions for all parameters, and 'fixed' effects assigned normally distributed prior distributions.

## DISCUSSION

The purpose of this work has been to draw attention to the poor frequentist properties of the estimator of the mean of the marginal posterior distribution of variance components in mixed threshold models when there is little information associated with the fixed effects. The simulation results indicate that this bias is not necessarily related to the impropriety of posterior distributions. Indeed, we have shown that biased inferences result in cases when the posterior is guaranteed to be proper, and agreement with the true value was obtained when the propriety of the posterior distributions cannot be guaranteed. Further, this is seen as a problem, because the bias persists in situations when this marginal posterior distribution is highly peaked, symmetric and guaranteed to be proper. In other words, this is not the bias to be expected from a non-linear estimator in a small sample setting.

The type of data structures that can lead to this problem is ubiquitous in animal breeding. This is characterised by models with a large number of fixed effects, with relatively few observations per fixed effect. As we have shown, the biased inferences associated with the variance component are not necessarily related to the extreme category problem, although this can be conducive to then.

The inability to obtain a closed form solution to the marginal posterior distribution of the variance component has precluded us from fully understanding the cause of the bias. The lack of asymptotic independence between the estimator of the fixed effect and of the variance component, either in the context of the generalized estimating equation studied by Zeger et al (1988) or in the context of the likelihood of error contrasts reported by Drum and McCullagh (1993), does not necessarily cast light on the present problem. In a Bayesian setting, one is marginalising with respect to the fixed effects, and provided that one does not stumble into problems of impropriety of distributions, it is not clear to us how assertions that are valid from a likelihood viewpoint can help explain the present results.

A strategy that can be interesting to explore is to run the Gibbs sampler after having integrated the fixed effects out analytically from the joint posterior distribution. The relevance of this is based on the observation that running the sampler conditioning on the true values of the fixed effects, yields correct inferences about heritability (Moreno, unpublished).

Using GSR but treating fixed effects as random, removes much of the bias in inference about variance components, as shown by Hoeschele and Tier (1995). However, we have illustrated that this alternative does not always work. In such cases, it is useful for the researcher to have a choice of other methods to solve the problem. The iterative bootstrap bias correction is an alternative, computationally simple method, that shows good frequentist behaviour. In this work, we chose to use it in conjunction with the mode of the joint posterior distribution, due to its computational simplicity. We stress, however, that the iterative bias correction can in principle be applied to any method, but if it is used, for example, together with the Gibbs sampler, computing time can become a limiting factor. An important point to stress is that the price to pay for unbiasedness is that estimates can fall outside the parameter space. This will never be the case using GSR treating fixed effects as random, or using approximations to the marginal posterior distribution of the variance component. It would certainly be interesting to extend the study of Tempelman (1994) and to explore the behaviour of alternative computational forms of the Laplacian approximation further.

The present simulation study was based on the sire model where each sire effect is represented by several observations. A more relevant model in an animal breeding context is the individual animal model, where each additive genetic value is associated at the most with one observation. In the case of the animal model, the bias problems discussed in this paper are accentuated, as illustrated in the simulation study of Hoeschele and Tier (1995). The study of Tempelman (1994) also showed poor behaviour of the Laplacian approximation in the case of the animal model. It is precisely in such a situation that methods which control this bias such as the MJP-IBC implemented here may be useful.

## ACKNOWLEDGEMENTS

The authors are very grateful to J Ocaña, who suggested the bootstrap to study the problem of bias on inferences about variance components in threshold models. We gratefully acknowledge the support of the Diputación General de Aragón (Grant 2-2-94) and of Caja de Ahorros de la Inmaculada (Grant CA 396). Three anonymous reviewers made useful suggestions.

## REFERENCES

- Albert JH, Chib S (1993) Bayesian analysis of binary and polychotomous response data. *J Am Stat Assoc* 88, 669–679
- Anderson DA, Aitkin M (1985) Variance component models with binary response: interviewer variability. *J R Stat Soc B* 47, 203–210
- Bliss CI (1935) The calculation of the dosage–mortality curve. *Ann Appl Biol* 22, 134–167
- Breslow NE, Lin X (1995) Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* 82, 81–91

- Casella G (1996) Statistical inference and Monte-Carlo algorithms. *Test* 5, 1–30
- Cox DR, Snell EJ (1989) *Analysis of Binary Data*. Chapman and Hall, London
- Dellaportas P, Smith AFM (1993) Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Appl Stat* 42, 443–459
- Drum ML, McCullagh P (1993) REML estimation with exact covariance in the logistic mixed model. *Biometrics* 49, 677–689
- Engel B, Buist W, Visscher A (1995) Inference for threshold models with variance components from the generalized linear mixed model perspective. *Genet Sel Evol* 27, 15–32
- 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, Im S (1990) Genetic evaluation for discrete polygenic traits in animal breeding. In: *Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds), Springer-Verlag, New York, 361–409
- 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
- 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, Oxford, 625–631
- Gelman A, Carlin JB, Stern HS, Rubin DB (1995) *Bayesian Data Analysis*. Chapman and Hall, London, 1st ed
- Gilks WR, Wild P (1992) Adaptive rejection sampling for Gibbs sampling. *Appl Stat* 41, 337–348
- Gilks WR, Richardson S, Spiegelhalter DJ (1996) *Markov Chain Monte-Carlo in Practice*. Chapman and Hall, London, 1st ed
- Henderson CR (1973) Sire evaluation and genetic trends. In: *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr DL Lush*, Blacksburg, Virginia, 10–41
- Hoeschele I, Foulley JL, Gianola D (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet* 104, 334–349
- Hoeschele I, Tier B (1995) Estimation of variance components of threshold characters by marginal posterior modes and means via Gibbs sampling. *Genet Sel Evol* 27, 519–540
- Kuk AYC (1995) Asymptotically unbiased estimation in generalized linear models with random effects. *J R Stat Soc B* 57, 395–407
- Lindley DV, Smith AFM (1972) Bayes estimates for the linear model. *J Roy Stat Soc B* 34, 1–41
- McCullagh P, Nelder JA (1989) *Generalized Linear Models*. Chapman and Hall, London, 2nd ed
- McCulloch CE (1994) Maximum likelihood variance components estimation for binary data. *J Am Stat Assoc* 89, 330–335
- Misztal I, Gianola D, Foulley JL (1989) Computing aspects of a nonlinear method of sire evaluation for categorical data. *J Dairy Sci* 72, 1557–1568
- Moreno C (1993) Estudio de los componentes genéticos del modelo umbral. Tesis doctoral. Universidad de Zaragoza, Spain
- Natarajan R, McCulloch CE (1995) A note on the existence of the posterior distribution for a class of mixed models for binomial responses. *Biometrika* 82, 639–643
- Sorensen D, Andersen S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. *Genet Sel Evol* 27, 229–249
- Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. *J Am Stat Assoc* 82, 528–540

- Tempelman RJ (1994) Evaluation of derivative-free marginal maximum likelihood estimation of variance components in threshold mixed models. In: *Proc of the 5th World Cong Genet Appl to Livestock Prod*, University of Guelph, Guelph, 22, 15–18
- Wright S (1934) An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* 19, 506–536
- Zeger SL, Liang K, Albert PS (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44, 1049–1060
- Zeger SL, Karim MR (1991) Generalized linear models with random effects: a Gibbs sampling approach. *J Am Stat Assoc* 86, 79–86

## APPENDIX

This appendix shows the log-conditional posterior distributional forms for the fixed and random effects and the variance component of the binary threshold model. The adaptive rejection sampling algorithm was implemented using these distributional forms.

The log-fully conditional posterior distributional form for the  $x$ th fixed effect is:

$$\ln f(b_x | \boldsymbol{\theta}, \mathbf{y}) \sum_{i=1,s} c_i \left\{ n_{i1} \ln [\Phi(t - b_{i(j)} - u_{i(k)})] + n_{i2} \ln [1 - \Phi(t - b_{i(j)} - u_{i(k)})] \right\}$$

where  $c_i = \begin{cases} 1 & \text{if } i(j) = x \\ 0 & \text{if } i(j) \neq x \end{cases}$  and  $\boldsymbol{\theta}$  represents the rest of parameters.

Log-conditional distributional form for the  $x$ th random effect is:

$$\ln f(u_x | \boldsymbol{\theta}, \mathbf{y}) = \sum_{i=1,s} c_i \left\{ n_{i1} \ln [\Phi(t - b_{i(j)} - u_{i(k)})] + n_{i2} \ln [1 - \Phi(t - b_{i(j)} - u_{i(k)})] \right\} + \left( -\frac{\mathbf{u}' \mathbf{A}^{-1} \mathbf{u}}{2\sigma_u^2} \right)$$

where  $c_i = \begin{cases} 1 & \text{if } i(k) = x \\ 0 & \text{if } i(k) \neq x \end{cases}$

Log-conditional distributional form for the variance component is:

$$\ln f(\sigma_u^2 | \boldsymbol{\theta}, \mathbf{y}) = \left( -\frac{\mathbf{u}' \mathbf{A}^{-1} \mathbf{u}}{2\sigma_u^2} \right) - \frac{q}{2} \ln(\sigma_u^2)$$

which is proportional to a scaled inverted chi-square distribution.