Original article

Estimation of variance components of threshold characters by marginal posterior modes and means *via* Gibbs sampling

I Hoeschele, B Tier *

*Virginia Polytechnic Institute and State University,
Department of Dairy Science, Blacksburg, VA 24061-0315, USA*

Summary – A Gibbs sampling scheme for Bayesian analysis of binary threshold data was derived. A simulation study was conducted to evaluate the accuracy of 3 variance component estimators, deterministic approximate marginal maximum likelihood (AMML), Monte-Carlo marginal posterior mode (MCMML), and Monte-Carlo marginal posterior mean (MCMPM). Several designs with different numbers of genetic groups, herd-year-seasons (HYS), sires and progeny per sire were simulated. HYS were generated as fixed, normally distributed or drawn from a proper uniform distribution. The downward bias of the AMML estimator for small family sizes (50 sires, average of 40 progeny) was eliminated with the MCMML estimator. For designs with many HYS, 0.9 incidence, 50 sires and 40 progeny on average, the marginal posterior distribution of heritability was non-normal; MCMML and MCMPM significantly overestimated heritability under the sire mode, while under the animal model the Gibbs sampler did not converge. For designs with 100 sires and 200 progeny per sire, the marginal posterior distribution of heritability became more normal and the discrepancy among MCMML and MCMPM estimates vanished. Heritability estimates under the animal model were less accurate than those under the sire model. For the smaller designs, the MCMML estimates were very close to the true value when using a normal prior for HYS effects, irrespective of the true state of nature of the HYS effects. For extreme incidence, small data sets and many HYS, observations within an HYS will frequently fall into the same category of response. With flat priors for the HYS effects, the posterior density is likely improper, supported by an analytical proof for a simplified model and analyses from Gibbs output. In analyses of limited binary data with extreme incidence, effects of a factor with many levels should be given a normal prior. Assigning a proper uniform prior or fixing values of such levels was

* On leave from Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia

not useful. Most accurate estimation of genetic parameters requires very large data sets. Further work is needed on diagnosis of improperness and on alternative proper priors.

Bayesian estimation / Gibbs sampling / categorical data / marginal maximum likelihood / variance component estimation

Résumé – Estimation des composantes de variance de caractères à seuil par les modes et les moyennes marginales *a posteriori* à l'aide de l'échantillonnage de Gibbs. *Un plan d'échantillonnage de Gibbs pour l'analyse bayésienne de caractères binaires à seuil a été établi. Par simulation, on a pu comparer la précision de 3 estimateurs de composante de variance, un maximum de vraisemblance marginale déterministe et approximatif (MVMA), le mode marginal* a posteriori *de Monte Carlo (MVMMC) et la moyenne marginale* a posteriori *de Monte Carlo (MMPMC). Plusieurs plans d'expérience avec des nombres différents de groupes génétiques, de cellules troupeau-année-saison (TAS), de pères et de descendants par père ont été simulés. Les TAS ont été établis comme des effets fixes, ou tirés d'une distribution normale, ou tirés d'une distribution uniforme. L'erreur par défaut du MVMA pour de petites tailles de famille (50 pères, 40 descendants par père) a été éliminée par le MVMMC. Pour des dispositifs incluant de nombreux TAS, avec une incidence de 0,9, 50 pères et 40 descendants par père en moyenne, la distribution marginale* a posteriori *de l'héritabilité n'est pas normale ; MVMMC et MMPMC surestiment significativement l'héritabilité dans un modèle paternel, alors qu'avec le modèle individuel la procédure de Gibbs ne converge pas. Avec 100 pères et 200 descendants par père, la distribution marginale* a posteriori *de l'héritabilité se rapproche de la normale et les discordances entre les estimées MVMMC et MMPMC disparaissent. Les estimées d'héritabilités avec le modèle individuel sont moins précises qu'avec le modèle paternel. Pour les petits dispositifs, les estimées MVMMC sont très proches de leur vraie valeur quand la distribution* a priori *des effets TAS est normale, quelle que soit la réalité des TAS. Pour des incidences extrêmes, de petits échantillons et un grand nombre de TAS, les observations à l'intérieur d'une cellule TAS tombent souvent dans la même catégorie de réponse. Avec des* a priori *uniformes pour les effets TAS, la densité* a posteriori *est probablement impropre, comme tend à l'indiquer l'analyse des résultats d'une procédure Gibbs appliquée à un modèle simplifié. Dans l'analyse de données binaires en nombre limité et avec une incidence extrême, une distribution normale* a priori *devrait être assignée aux effets des facteurs ayant de nombreux niveaux, plutôt qu'une distribution uniforme ou des valeurs fixées. Des estimations précises des paramètres génétiques requièrent dans ce cas de très grands ensembles de données. Il reste encore à étudier la manière de déceler l'impropriété des distributions* a priori *et de choisir de meilleurs* a priori.

estimation bayésienne / échantillonnage de Gibbs / données catégorielles / maximum de vraisemblance marginale / composante de variance

INTRODUCTION

Bayesian analysis of binary or polychotomous threshold traits *via* Gibbs sampling has recently been described by Albert and Chib (1993), Sorensen *et al* (1995) and Jensen (1994). In all 3 papers, the Gibbs sampler was implemented in combination with data augmentation, *ie* parameters and missing data were sampled from their fully conditional distributions derived from the joint posterior density of the parameters and missing data. The parameter vector included fixed and random

effects and variance components, while the missing data consisted of the liabilities or latent continuous variables in the threshold model. In contrast with the aforementioned papers, Zeger and Karim (1991) implemented Bayesian analysis with a Gibbs sampler based on the posterior density of the parameters rather than the joint posterior of parameters and missing data. Zeger and Karim (1991) also used a different prior for the dispersion parameters. McCulloch (1994) derived maximum likelihood rather than Bayesian estimation *via* an expectation–maximization (EM) algorithm with Gibbs sampling of the liabilities within each E step.

In this contribution, a Gibbs sampling scheme applied to parameters and liabilities was implemented for Bayesian analysis of a binary trait. A simulation study was conducted to evaluate the accuracy of 3 estimators of variance components, deterministic approximate marginal maximum likelihood (AMML) (Foulley *et al*, 1987; Hoeschele *et al*, 1987), the Monte-Carlo evaluated marginal posterior mode (MCMML), and the Monte-Carlo evaluated marginal posterior mean (MCMPM). The analysis was restricted to a threshold model with one variance component and was performed under sire and animal models. In MML, the marginal posterior density of the variance parameters is maximized with respect to these parameters. 'Fixed' effects and variance parameters have improper, flat prior distributions. Thus, the marginal posterior density of the variance components is proportional to their marginal likelihood. In Bayesian analysis implemented *via* Gibbs sampling applied to all parameters (or all parameters and missing data), the parameter samples provide inferences about any parameter from its marginal posterior density. For a single variance component, the marginal posterior mode is equivalent to the MML estimator. Therefore, the Monte-Carlo evaluated marginal posterior mode was termed the MCMML estimator.

With conventional deterministic algorithms, MML estimates are computed using a normal approximation, and severe biases of variance and covariance estimates have been reported (Gilmour *et al*, 1985; Hoeschele *et al*, 1987; Hoeschele and Gianola, 1989; Simianer and Schaeffer, 1989). While Gilmour *et al* (1985) observed an underestimation of heritability with small family sizes, Hoeschele *et al* (1987) and Hoeschele and Gianola (1989) found an overstimation of heritability for a binary trait with extreme incidence and in the presence of a fixed factor with many levels.

The objective of this study was to compare the 3 variance component estimators AMML, MCMML, and MCMPM in terms of their frequentist properties, and, in particular, to investigate whether the biases observed for the AMML estimator can be eliminated by computing exact MML estimates *via* MCMC algorithms. A related side objective was to investigate potential causes of the biases.

MATERIALS AND METHODS

Methodology

The Bayesian analysis described below is identical to that of Sorensen *et al* (1995), when applied to a binary trait, but differs from Albert and Chib (1993) in the sampling of the variance components. The Gibbs sampling scheme differs from the sampler of Sorensen *et al* (1995) only when run under an animal model where breeding values of sires and offspring were sampled jointly as in Janss *et al* (1994).

Let y represent the observed dichotomous variable and w the liability variable. In the threshold model for 2 categories of response, $y_i = 1$ if $w_i > 0.0$ and $y_i = 0$ otherwise. Conditionally on the fixed (β) and random effects ($\mathbf{u}$), the w_i are independent $N(\mathbf{x}_i'\beta + \mathbf{z}_i'\mathbf{u}, 1)$, and the y_i are independent Bernouilli with $\text{Prob}(y_i = 1) = \Phi(\mathbf{x}_i\beta + \mathbf{z}_i'\mathbf{u})$, where $\Phi(.)$ denotes the standard normal cumulative distribution function. Matrices $\mathbf{X}$ and $\mathbf{Z}$ are the usual incidence matrices with row i denoted by $\mathbf{x}_i'$ ($\mathbf{z}_i'$). The parameter vector (θ) includes β, $\mathbf{u} = [\mathbf{u}_1' \ldots, \mathbf{u}_i', \ldots, \mathbf{u}_q']'$, and the $\sigma_{ij}^2(i, j = 1, \ldots, q)$, with $\text{Cov}(\mathbf{u}_i, \mathbf{u}_j') = \mathbf{A}_{ij}\sigma_{ij}$ and the $\mathbf{A}$ matrices known.

The joint posterior density of θ and $\mathbf{w} = \{w_i\}$ is

$$f(\theta, \mathbf{w}|\mathbf{y}) = c\prod_{i=1}^{N}[I(y_i = 0)I(w_i \leqslant 0.0) + I(y_i = 1)I(w_i > 0.0)]\phi(w_i;\ \mathbf{w}_i'\beta + \mathbf{z}_i'\mathbf{u};\ 1)$$

$$f(\mathbf{u}_1, \ldots, \mathbf{u}_q|\sigma_1^2, \sigma_{12}, \ldots, \sigma_q^2)f(\sigma_1^2, \sigma_{12}, \ldots, \sigma_q^2) \qquad [1]$$

where c is a constant, $\phi(\ ; \mu; \sigma^2)$ is the density function of $N(\mu; \sigma^2)$, and $I(x\varepsilon S)$ is the indicator function equal to 1 if variable x is contained in set S and zero otherwise. For independent $\mathbf{u}_j s'$, the prior densities of the $\mathbf{u}_j$ and the σ_{ij} simplify to products of the $f(\mathbf{u}_j|\sigma_j^2)$, the density of the MVN ($\mathbf{0}; \mathbf{A}_j\sigma_j^2$), and products of the $f(\sigma_j^2)$ or prior densities of the σ_j^2. Samples from the joint posterior distribution can be obtained by sampling in turn from $f(\theta|\mathbf{w}, \mathbf{y}) = f(\theta|\mathbf{w})$ and $f(\mathbf{w}|\theta, \mathbf{y})$. These 2 conditional distributions are of standard forms, and the fully conditional parameter densities, derived from $f(\theta|\mathbf{w})$, are identical to those in the standard mixed linear model (eg, Gelfand et al, 1990; Wang et al, 1993). Then, from standard mixed linear model results (eg, Gianola et al, 1990), and with $q = 1$, $\mathbf{u}_j = \mathbf{u}$ and $\sigma_j^2 = \sigma_u^2$,

$$\beta_i|\beta_{-i}\mathbf{u}, \sigma_u^2, \mathbf{w} \sim N(\mathbf{x}_i'(\mathbf{w} - \mathbf{X}_{-i}\beta_{-i} - \mathbf{Z}\mathbf{u})/\mathbf{x}_i'\mathbf{x}_i, 1/(\mathbf{x}_i'\mathbf{x}_i)| \qquad [2]$$

$$u_i|\beta, \mathbf{u}_{-i}, \sigma_u^2, \mathbf{w} \sim N(\mathbf{z}_i'(\mathbf{w} - \mathbf{X}\beta - \mathbf{Z}_{-i}\mathbf{u}_{-i} - \mathbf{A}_{-i}^i\sigma_u^{-2}\mathbf{u}_{-i})/$$
$$(\mathbf{z}_i'\mathbf{z}_i + a^{ii}\sigma_u^{-2}), 1/(\mathbf{z}_i'\mathbf{z}_i + a^{ii}\sigma_u^{-2})| \qquad [3]$$

where β_i is element i of vector β, β_{-i} is this vector with element i omitted, $\mathbf{x}_i$ is column i of matrix $\mathbf{x}$, $\mathbf{x}_{-i}$ is $\mathbf{x}$ with column i deleted, a^{ii} is element (i, i) of the inverse additive genetic relationship matrix $\mathbf{A}^{-1}$, and $\mathbf{A}_{-i}^i$ is row i of $\mathbf{A}^{-1}$ without element i.

Under the animal model, breeding values (u) of a sire and his offspring were sampled jointly by sampling a sire's u from its marginal normal distribution while sampling the u of each offspring from its full conditional distribution in [3]. Mean and variance of the marginal distribution were obtained as the BLUP of the sire's u and its prediction error variance after absorbing the offspring u into the sire's u in MME for this sire and his offspring.

For $q = 1$ and $f(\sigma_u^2) = \text{constant}$,

$$\sigma_u^2 \sim \mathbf{u}'\mathbf{A}^{-1}\mathbf{u}\chi_{n-2}^{-2} \qquad [4]$$

where the inverse chi-square distribution has $n - 2$ df with n equal to the number of elements in $\mathbf{u}$. The prior in [4] differs from that of Zeger and Karim (1991), who

reported problems in estimating variance components due to the Gibbs sampler "being trapped at zero". Their prior resulted from an inverse chi-square (or inverse Wishart) distribution with zero prior degrees of freedom, yielding the prior density $f(\sigma_u^2) = (\sigma_u^2)^{-1}$ and changing the *df* in [4] to *n*. This problem vanishes when using the flat prior, as also observed by other researchers (D Sorensen, personal communication). Note also that with an improper prior distribution the resulting posterior distribution can be proper or improper (*eg*, Berger and Bernardo, 1992a, b; Hobert and Casella, 1994). Hober and Casella (1994) established necessary and sufficient conditions for the joint posterior density of the fixed and random effects and the variance components in a hierarchical Bayes LMM to be proper, *ie* integrable. One condition was that for any variance component *j*, besides the residual, with prior $f(\sigma_j^2) = (\sigma_j^2)^{-(a_j+1)}$, we must have $a_j < 0$. Setting $a_j = -1$ yields the flat prior $f(\sigma_j^2) = 1$ used in this paper, while $a_j = 0$ produces the prior used by Zeger and Karim (1991). More specifically, in this study, a bounded flat prior for σ_u^2 was used under a sire model as in Sorensen *et al* (1995), while an improper flat prior was used under the animal model.

The marginal posterior mode of heritability was computed by a grid search of the Rao–Blackwell estimate (*eg*, Gelfand *et al*, 1990) of the marginal posterior density of σ_u^2 and a change of variable to h^2, or

$$f(h^2|\mathbf{y}) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{\mathbf{u}^{k'}\mathbf{A}^{-1}\mathbf{u}^k}\left[\frac{h^2/(\delta-h^2)}{\mathbf{u}^{k'}\mathbf{A}^{-1}\mathbf{u}^k}\right]^{-\frac{n}{2}}\exp\left[-\frac{1}{2}\frac{\mathbf{u}^{k'}\mathbf{A}^{-1}\mathbf{u}^k}{h^2/(\delta-h^2)}\right]J \quad [5]$$

where $\delta = 4$ or $\delta = 1$ for the sire or animal model, respectively, *k* denotes the Gibbs sample, and *J* is the Jacobian of the transformation $\sigma_u^2 \to h^2$,

$$J = |\delta/(\delta-h^2)^2| \quad \text{for fixed HYS}$$
$$J = |\delta(1+\sigma_h^2)/(\delta-h^2)^2| \quad \text{for random HYS}$$

Conditional on the parameters, the latent data were sampled from truncated normal distributions, or

$$\mathbf{w}_i|\boldsymbol{\beta}, \mathbf{u}, \mathbf{y} \sim N(\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{u}, 1) \quad [6]$$

with

$$w_i > 0.0 \text{ if } y_i = 1$$
$$w_i \leqslant 0.0 \text{ if } y_i = 0$$

To implement the Gibbs sampler, starting values for the parameters were obtained by computing the maximum *a priori* (MAP) estimates of β and $\mathbf{u}$ (*eg*, Gianola and Foulley, 1983) evaluated at the approximate MML estimates (Foulley *et al*, 1987) of the σ_j^2. Given initial parameter values, a first sample of the latent data was drawn from [5], followed by the sampling of new parameter values from [2], [3] and [4], and further Gibbs cycles.

Because the prior distributions for the fixed effects and for the variance component under the animal model are improper, the posterior distribution might also be improper. Hobert and Casella (1994) gave conditions for the integrability of the

posterior distribution which only hold for hierarchical linear mixed models. Furthermore, a reviewer pointed out that for a simple fixed model with one factor and the logit link function, the joint posterior distribution of the fixed effects is improper, if in at least one factor level all observations fall into the same category of response. An analytical investigation of whether $f(\theta|\mathbf{y})$ is proper for the nonlinear mixed model considered here, however, is very difficult or intractable. Therefore, it is desirable to be able to detect an improper posterior from Gibbs output. A candidate approach is the MC evaluation of the marginal likelihood $f(\mathbf{y})$. With θ representing the parameter vector, the marginal likelihood is

$$f(\mathbf{y}) = \int f(\mathbf{y}|\mathbf{\theta})f(\mathbf{\theta})d\mathbf{\theta} \qquad [7]$$

The reciprocal of $f(\mathbf{y})$ is a normalizing constant ensuring that the posterior density $f(\theta|\mathbf{y})$ integrates to one, if it is proper. If it is not, [7] is infinite.

While the computation of Bayesian point estimates and related inferences (*eg*, marginal posterior density plots, highest posterior density regions) from Gibbs output is straightforward, computation of marginal likelihoods or Bayes factors has proved to be very challenging (*eg*, Chib, 1994; Newton and Raftery, 1994). Because all conditional distributions (*ie* [2], [3] and [4]) are standard, the approach of Chib (1994) for marginal likelihood estimation from Gibbs output was adopted here. The marginal likelihood can be written as

$$f(\mathbf{y}) = \frac{f(\mathbf{y}|\mathbf{\theta})f(\mathbf{\theta})}{f(\mathbf{\theta}|\mathbf{y})} \qquad [8]$$

and is estimated by evaluating [8] at a particular point, *eg*, the posterior mean $\widehat{\theta} = E(\theta|\mathbf{y})$. With data augmentation, the denominator of [8] can be written as

$$f(\widehat{\mathbf{\theta}}|\mathbf{y}) = \int f(\widehat{\mathbf{\theta}}|\mathbf{w},\mathbf{y})f(\mathbf{w}|\mathbf{y})d\mathbf{w} \qquad [9]$$

where $\mathbf{w}$ is the vector of missing data. The Monte-Carlo estimate of [9] is

$$\widehat{f}(\widehat{\mathbf{\theta}}|\mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N}f(\widehat{\mathbf{\theta}}|\mathbf{w}^i,\mathbf{y}) \qquad [10]$$

where N is Gibbs sample size and the $\mathbf{w}^i$ are samples from $f(\theta,\mathbf{w}|\mathbf{y})$ and hence are also samples from $f(\mathbf{w}|\mathbf{y})$. To evaluate the densities in the right-hand side of [10], let $\theta = [\mathbf{\beta}',\mathbf{u}',\sigma_u^2]'$. Then, one way of factoring the joint posterior density evaluated at the vector of posterior means of the parameters is

$$f(\widehat{\mathbf{\beta}},\widehat{\mathbf{u}},\widehat{\mathbf{\sigma}}_u^2|\mathbf{y}) = f(\widehat{\mathbf{\sigma}}_u^2|\mathbf{y})f(\widehat{\mathbf{u}}|\widehat{\mathbf{\sigma}}_u^2,\mathbf{y})f(\widehat{\mathbf{\beta}}|\widehat{\mathbf{u}},\widehat{\mathbf{\sigma}}_u^2,\mathbf{y}) \qquad [11]$$

where

$$f(\widehat{\mathbf{\sigma}}_u^2|\mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N}f(\widehat{\mathbf{\sigma}}_u^2|\mathbf{u}^i) \qquad [12a]$$

$$f(\widehat{\mathbf{u}}|\widehat{\boldsymbol{\sigma}}_u^2, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} f(\widehat{\mathbf{u}}|\boldsymbol{\beta}^i, \mathbf{w}^i, \widehat{\boldsymbol{\sigma}}_u^2) \qquad [12b]$$

$$f(\widehat{\boldsymbol{\beta}}|\widehat{\mathbf{u}}, \widehat{\boldsymbol{\sigma}}_u^2, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} f(\widehat{\boldsymbol{\beta}}|\mathbf{w}^i, \widehat{\mathbf{u}}) \qquad [12c]$$

Note that with the parameter vector partitioned into 3 subvectors, [12a, b, c] represents one of 6 possible factorizations of the posterior density. Furthermore note that density [12a] can be evaluated immediately at the termination of the Gibbs chain, while the evaluation of [12b] requires output from a second Gibbs chain with σ_u^2 fixed at its posterior mean estimate obtained from the first chain, and evaluation of [12c] is from a third Gibbs chain with $\mathbf{u}$ and σ_u^2 fixed at their posterior mean estimates from the first Gibbs chain.

Simulated data and analyses

Data were simulated on a binary threshold trait. Eight different designs were investigated, which firstly differed in the numbers of genetic groups, herd-year-season (HYS) effects, sires, and progeny per sire. Secondly, designs different in the way HYS effects were simulated: as fixed (generated once from a normal distribution and held constant in all replicates), as random and normally distributed, or as random and drawn from a proper uniform distribution. Genetic groups were always fixed and sires or animals were random. The designs are defined in table I.

Table I. Description of designs in terms of number of genetic groups, number of HYS levels, number of sires, number of progeny per sire, and status of HYS effects (fixed = simulated once from $N(0.046)$ and constant across replicates, $U(-1.175, 1.175)$ = simulated from this uniform distribution in each replicate, $N(0.046)$ = simulated from this normal distribution in each replicate).

Design	Number of groups	Number of HYS	Number of sires	Number of progeny	HYS status
I-F	4	135	50	40	Fixed
II-F	1	135	50	40	Fixed
II-U	1	135	50	40	$U(-1.175, 1.175)$
II-N	1	135	50	40	$N(0.046)$
III-F	1	32	50	40	Fixed
IV-F	1	320	100	200	Fixed
V-F	1	1	50	40	Fixed
VI-F	1	1	100	200	Fixed

Genetic group means for liability were -0.4, -0.15, 0.15, and 0.4. Sire or animal effects on liability were generated from $N(0, .\sigma_u^2)$ for a heritability of $h^2 = 0.25$ ($\sigma_u^2 = h^2/(\delta - h^2)$ for HYS fixed and $\sigma_u^2 = h^2(1 + \sigma_{\text{HYS}}^2)/(\delta - h^2)$ for HYS random, with $\delta = 4$ ($\delta = 1$) under the sire (animal) model), and residual liabilities were

generated from $N(0, 1)$. HYS effects were generated from $N(0, \sigma^2_{\text{HYS}} = 0.46)$ or from $U[a, b]$, where $a = -0.5(12\sigma^2_{\text{HYS}})^{0.5} = -b$ was set such that HYS effects had the same variance under both distributions. The truncation point used to dichotomize the liabilities was $\mathbf{\Phi}^{-1}(p)*(1+\sigma^2_{\text{HYS}}+\sigma^2_u)^{0.5}$, where p was the desired incidence equal to 0.9 for all designs except VI-F with $p = 0.5$. With 135 HYS, the probabilities of having 0, 1, or 2 offspring in any HYS were 0.8, 0.1, and 0.1, respectively, for each sire; with 32 HYS, sires had 0, 6 or 7 offspring per HYS with the same probabilities; and with 320 HYS, sires had 0, 2 or 4 progeny per HYS with probabilities 0.8, 0.09, and 0.11, yielding approximately the average progeny group sizes given in table I. The numbers of sires in the 4 genetic groups were 12, 14, 13, and 11. Designs with average progeny group size of 40 (200) were replicated 40 (20) times.

All data sets were analyzed with the sire model, and data sets II-F, V-F and VI-F were also analyzed with the animal model. Note that for these designs and if a linear mixed model were used, sire and animal, models would be (linearly) equivalent (Henderson, 1985), ie yield the same estimates of fixed and sire effects. HYS effects were treated as fixed (f) in the analysis, ie had improper uniform priors, for all designs where HYS effects were fixed (F). Additionally, HYS were treated as random with normal prior (n) in the analyses of data sets for designs II-F, II-N and II-U where HYS were fixed, random and normally distributed, and random and drawn from the proper uniform distribution, respectively. Treating HYS as random with normal prior required to also estimate HYS variance σ^2_{HYS}.

For some of the designs, data sets contained HYS levels with the so-called extreme category problem (ECP) (Misztal and Gianola, 1989). Extreme categories are the first and last category, hence for a binary trait both categories are extreme. In an HYS exhibiting the ECP, all records are in the same extreme category. On average, frequency of HYS with the ECP was 45% for designs I-F and II-F/N/U, 22% for design III-F, 17% for design IV-F, and 0% for all other designs. Solutions for such HYS classes do not converge but tend toward infinity in deterministic AMML algorithms. Therefore, in AMML solutions for these HYS were fixed at ± 10 (Misztal and Gianola, 1989). For the Gibbs sampler, options were considered for the treatment of HYS in the presence of the ECP: (i) to use a proper normal prior in the analysis; (ii) to use a proper uniform distribution as prior, eg, $U(-10., 10.)$; and (iii) to fix HYS effects with the ECP at ± 10 or at ± 3 in all Gibbs cycles, denoted by f10 and f3, respectively.

RESULTS AND DISCUSSION

Estimates of heritability (h^2) were obtained under the sire model for all designs and under the animal model for designs II-F, V-F, and VI-F. For designs II-F and IV-F, HYS variance was also estimated when HYS effects had a prior normal distribution in the analysis. Estimators were deterministic, AMML, MCMML and MCMPM.

MC estimates were computed from 20 000 consecutive Gibbs samples for the sire model and 200 000 samples for the animal model, with a burn-in period of an additional 2 000 cycles. The burn-in period was determined based on plots of sample values for heritability *versus* cycle number showing consistently that 2 000 cycles were more than needed. Typical plots for design-analysis combinations II-F-f and IV-F-f are shown in figures 1 and 2, respectively. The short burn-in period was
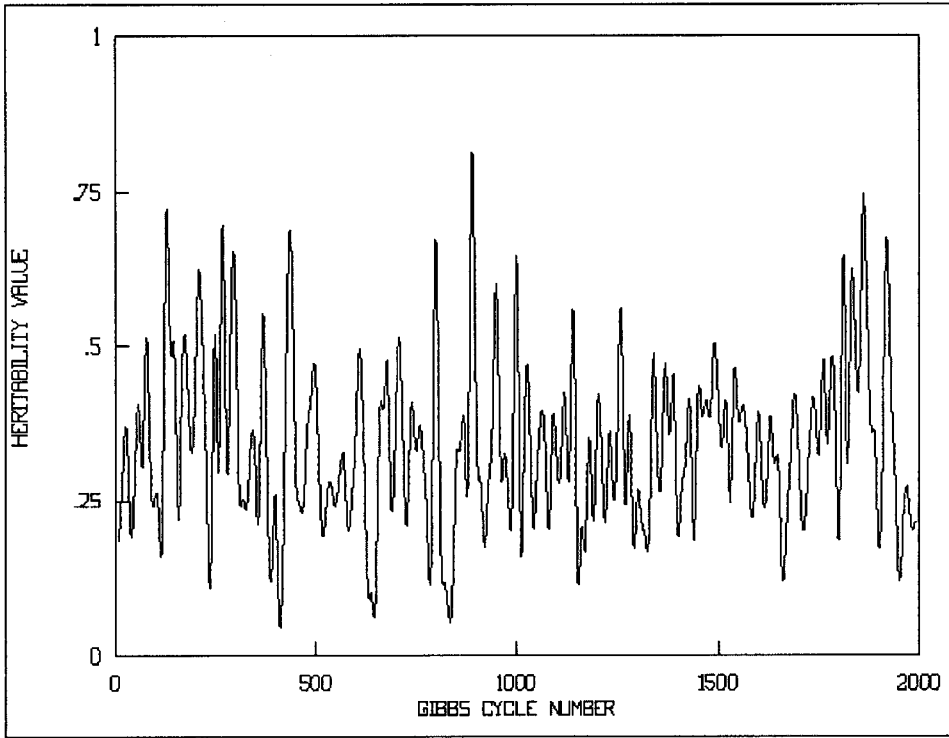
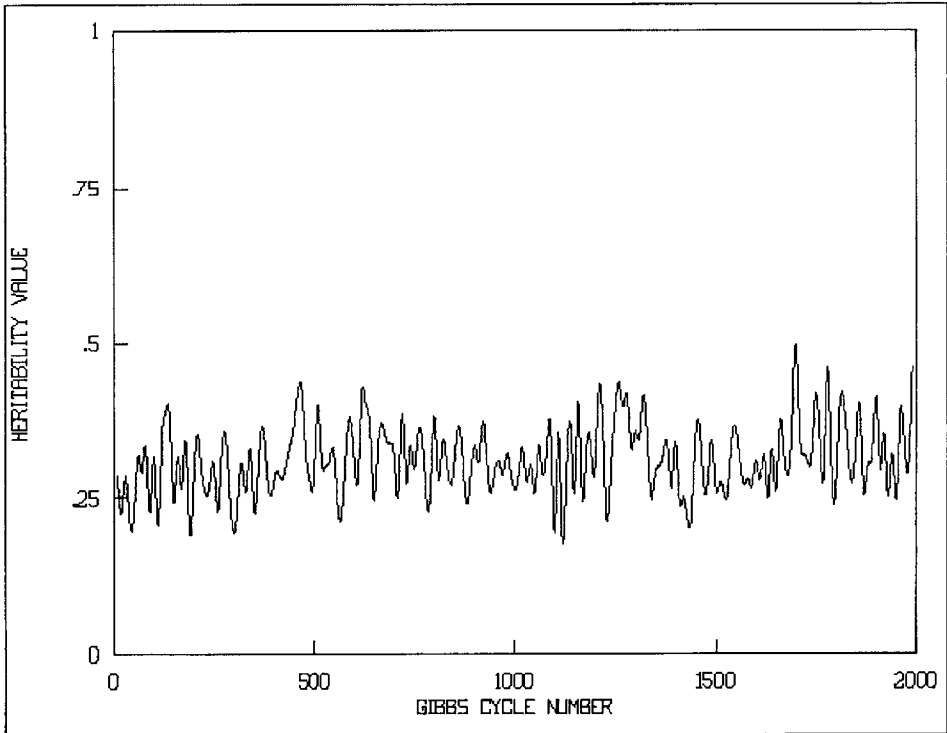**Fig 1.** Plot of sample value of heritability *versus* Gibbs cycle within the burn-in period for design-analysis II-F-f.

due to the use of good starting values for both location and variance parameters, which were the estimates obtained from the AMML procedure.

The number of Gibbs cycles of 20 000 (200 000) was determined as sufficient based on the autocovariance structure of the sample of heritabilities. This entailed calculating an effective sample size as the ratio of variance of samples to the variance of the sample mean computed from the estimated autocorrelations (Sorensen *et al*, 1995). Autocovariance for lag t was estimated as

$$\widehat{\rho}(t) = \frac{1}{K} \sum_{i=1}^{K-t}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_K)(\mathbf{x}_{i+t} - \widehat{\boldsymbol{\mu}}_K) \qquad \widehat{\boldsymbol{\mu}}_K = \frac{1}{K} \sum_{i=1}^{K}\mathbf{x}_i$$

Variance of sample mean given the estimated autocovariances was estimated with the initial positive sequence estimator of Geyer (1992), or

$$\text{V}\widehat{\text{a}}\text{r}(\widehat{\boldsymbol{\mu}}_K) = \frac{\widehat{\boldsymbol{\rho}}(0)}{K} + \frac{2}{K} \sum_{i=1}^{2m+1} \widehat{\boldsymbol{\rho}}(i)$$

Fig 2. Plot of sample value of heritability *versus* Gibbs cycle within the burn-in period for design-analysis IV-F-f.

where m is the largest integer satisfying the condition

$$\widehat{\rho}(2i) + \widehat{\rho}(2i + 1) > 0, \quad i = 1, 2, \ldots, m$$

and effective sample size ESS was

$$\text{ESS} = \frac{\widehat{\rho}(0)}{\text{Vâr}(\mu_K)}$$

For the sire model with 20 000 cycles, the minimum ESS was 580 and was found in one of the design I-F data sets analyzed with HYS treated as fixed (f). For a typical replicate with ESS = 1 435, autocorrelations at lags 1, 10, 20 and 50 were 0.66, 0.34, 0.13 and 0.007, respectively. For most data sets of this design, ESS exceeded 1 000. For other designs (II–VI), ESS was in the 2 000 to 6 000 range, and, for example, autocorrelations at the above lags were 0.39, 0.14, 0.03, and 0.006 with ESS = 3 705 for a design IV-F data set with HYS treated as fixed. For the animal model and analysis of design V-F and VI-F data sets, 200 000 cycles were required to achieve ESS values in the order of 500 to 1 000.

Mean estimates and empirical SE across 40 replicates are presented in table II. The first column of table II specifies the design (*eg*, I-F) in combination with the type of analysis (*eg*, I-F-f) according to the treatment of HYS effects. Progeny group sizes typically ranged from 25 to 55 for average progeny group size of 40 and from 120 to 280 for an average of 200.

Table II. Estimates of heritability (h^2; true value = 0.25) and of HYS variance (σ^2_{HYS}; true value = 0.46) with a deterministic AMML, MCMML and MCMPM.

Design analysis	Model/parameter	AMML	MCMML	MCMPM
I-F-f	Sire/h^2	0.304 (0.022)	0.431 (0.027)	0.514 (0.024)
II-F-f	Sire/h^2	0.288 (0.021)	0.424 (0.025)	0.502 (0.023)
II-F-n	Sire/h^2	0.191 (0.011)	0.251 (0.013)	0.291 (0.013)
II-F-n	Sire/σ^2_{HYS}	0.264 (0.007)	0.398 (0.012)	0.433 (0.013)
II-N-n	Sire/h^2	0.186 (0.013)	0.247 (0.015)	0.285 (0.015)
II-N-n	Sire/σ^2_{HYS}	0.322 (0.010)	0.469 (0.016)	0.511 (0.018)
II-U-n	Sire/h^2	0.195 (0.012)	0.255 (0.013)	0.295 (0.013)
II-U-n	Sire/σ^2_{HYS}	0.281 (0.008)	0.458 (0.015)	0.504 (0.017)
II-F-f3[a]	Sire/h^2	0.286 (0.023)	0.394 (0.026)	0.464 (0.024)
II-F-f10[b]	Sire/h^2	0.298 (0.022)	0.422 (0.025)	0.498 (0.023)
III-F-f	Sire/h^2	0.239 (0.021)	0.288 (0.025)	0.360 (0.024)
IV-F-f	Sire/h^2	0.257 (0.009)	0.269 (0.009)	0.276 (0.009)
IV-F-f3[a]	Sire/h^2	0.256 (0.009)	0.267 (0.009)	0.274 (0.009)
IV-F-n	Sire/h^2	0.244 (0.009)	0.251 (0.009)	0.259 (0.009)
V-F-f	Sire/h^2	0.237 (0.014)	0.259 (0.015)	0.307 (0.016)
V-F-f[c]	Animal/h^2	0.239 (0.015)	0.334 (0.025)	0.357 (0.024)
VI-F-f	Sire/h^2	0.248 (0.008)	0.251 (0.008)	0.259 (0.008)
VI-F-f	Animal/h^2	0.248 (0.008)	0.252 (0.008)	0.261 (0.008)

[a]Estimates are averages of 20–40 replicates (see text) with empirical SE in parentheses. Estimates of effects of ECP herd-year-seasons (all observations in the category coded 0) were set to −3. [b]Estimates of effects of ECP herd-year-seasons were set to −10. [c] The AMML estimate was computed under the sire model for this design.

All estimators strongly overestimated heritability for the design-analysis combinations I-F-f and II-F-f- under the sire model. The upward bias was largest for MCMPM, followed by MCMML and AMML. The designs were characterized by a small number of sires with average progeny group size of 40, a large number of HYS effects, a high incidence of 0.9, and as a consequence a high percentage of ECP HYS levels. The discrepancy between the posterior mean and mode of h^2 indicate that the posterior distribution was not normal, with the mode always closer to the

true value than the mean. A plot of the marginal posterior density of h^2 based on [5] for II-F-f can be found in figure 3.
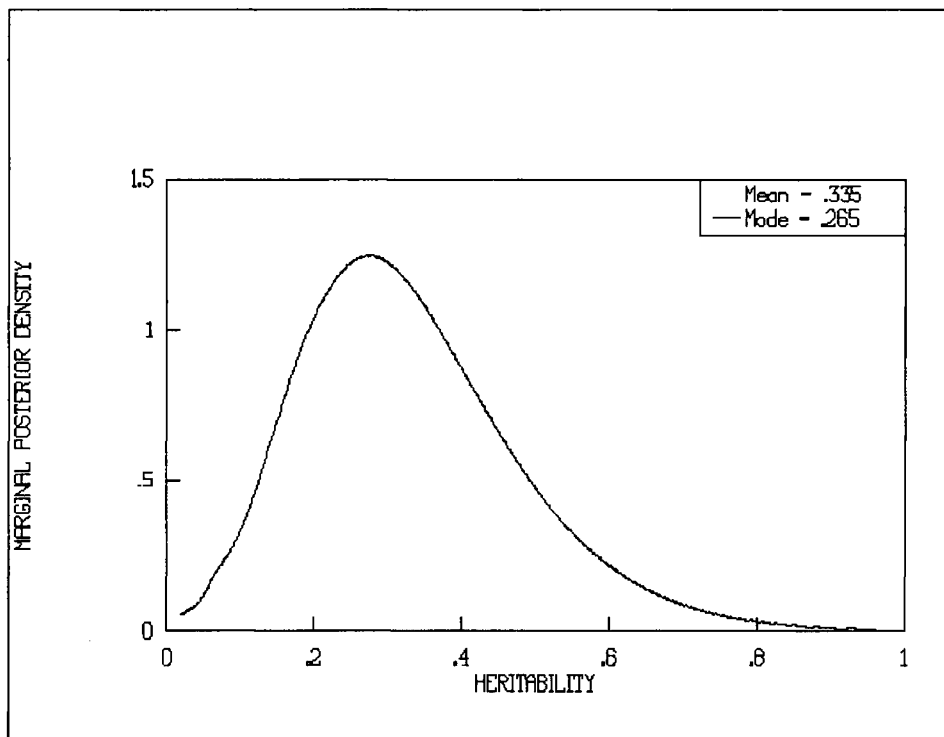

Fig 3. Marginal posterior density of heritability estimated from Gibbs output for design-analysis II-F-f.

For design-analysis combination III-F-f, with the design containing the same number of sires and records as, but a smaller number of HYS levels than, I-F and II-F, the upward biases of the MCMPM and MCMML estimators decreased while the AMML estimator tended to underestimate h^2. Overestimation (Hoeschele *et al*, 1987; Hoeschele and Gianola, 1989; Simianer and Schaeffer, 1989) and underestimation (Gilmour *et al*, 1985; Thompson, 1990) with AMML are in good agreement with the literature. It appears that the AMML estimates are closest to the true value of h^2. However, a comparison between the results for I-F-f and II-F-f *versus* III-F-f strongly suggests that this finding is due to a partial counterbalancing of upward and downward biases of the AMML estimator for data sets with a small number of records and progeny per sire.

For the larger designs with 100 sires and average progeny group size of 200, biases of and discrepancies among estimators were strongly reduced (design IV-F) and negligible for design VI-F with the mean as the only fixed effect and an incidence of 0.5. For design-analysis IV-F-f, the number of records and the number of HYS

levels were increased by a factor of 10 relative to III-F-f. As a result, the AMML, estimator no longer underestimated h^2, while the overstimations obtained with and the discrepancy between the MCMPM and MCMML estimators strongly decreased. A plot of the marginal posterior distribution of h^2 for design-analysis combination IV-F-f is presented in figure 4, which shows a more symmetric distribution with smaller variance relative to figure 3.
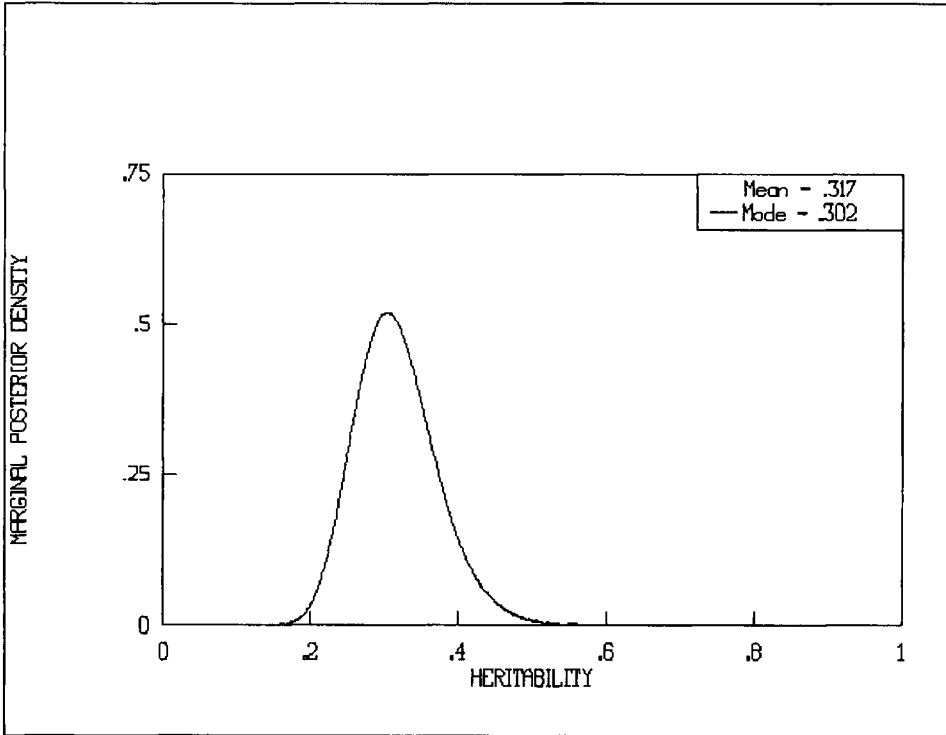

Fig 4. Marginal posterior density of heritability estimated from Gibbs output for design-analysis IV-F-f.

It may be concluded that the discrepancies found between the true h^2, the MCMPM and the MCMML estimates are due to a lack of information in the data causing the posterior distribution to be non-normal and its mean to be a biased estimator of h^2. Another explanation might be that the joint posterior distribution of the parameters is improper. As mentioned earlier, for the logit link and a simple fixed model, improperness can be shown analytically when levels of the fixed factor exhibit the ECP. Consequently, use of a proper prior distribution for the fixed effects, resulting in a proper posterior, should eliminate this problem. Therefore, 2 proper uniform distributions were employed as priors, $U(-3, 3)$ and $U(-10, 10)$ in the analysis of II-F. However, this leads to rejection of all sampled values for HYS effects with the ECP. Figure 5 provides an explanation by presenting a plot

of sample value *versus* Gibbs cycle for an HYS with the ECP. In this analysis, an improper uniform prior was used, and sample values drifted toward extremely small numbers far below the lower limits of -3 and -10 in the proper uniform priors.
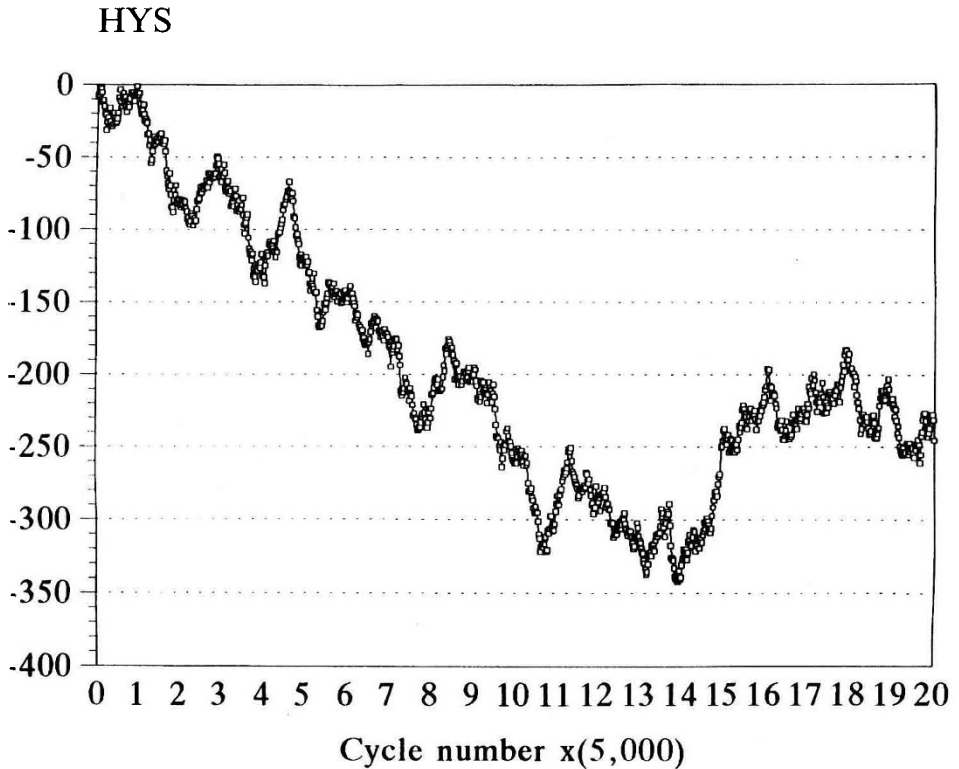
HYS

Fig 5. Plot of sample value of HYS effect *versus* Gibbs cycle for an HYS with the ECP.

Next, values of HYS effects with the ECP (all records equal to 0) were fixed at -3 or at -10 across all Gibbs cycles while the other HYS effects were sampled and had an improper prior. Results from these analyses are in table II in the rows for design-analysis combinations II-F-f3 and II-F-f10. Upward biases were still substantial and only slightly smaller than those for II-F-f. The posterior distribution for the parameters sampled should have been proper and, if so, the results indicate that biases were (mostly) due to limited information in the data.

Because a proper uniform prior could not be used for technical reasons, a normal prior distribution was postulated for the HYS. The designs analyzed with a normal prior were II-F, II-N, and II-U, where the true state of nature of the HYS effects was fixed, normally distributed and drawn from a proper uniform, respectively. The results for all 3 designs were very similar. The AMML estimator significantly underestimated h^2, MCMPM overestimated h^2 with a smaller absolute bias, and MCMML estimates were very close to the true value.

The empirical SE of the MC estimators tended to be slightly higher than the SE of the deterministic AMML estimator for the smaller designs. Doubling the number of Gibbs cycles had virtually no effect, suggesting that the cause was not the Monte-Carlo error. Possibly the AMML estimator had smaller variance in cases where it exhibited a large downward bias.

When an animal model was used for design-analysis combination II-F-f, the Gibbs sampler 'blew up' in each of several replicates analyzed, *ie* the additive genetic variance continued to increase in subsequent Gibbs cycles, soon reaching unreasonably high values. When design V-F with only one fixed effect was analyzed under the animal model, the Gibbs sampler 'converged', but estimates had larger biases and were more variable compared to those obtained under the sire model. 'Convergence' of the Gibbs sampler refers to the sampler reaching stationarity, where sample values fluctuate around a constant (see figure 6 for presence of and figure 5 for lack of 'convergence'; these figures are discussed further below). To verify this result and ensure the correctness of the software under the animal model option, the larger design VI-F was also analyzed under the animal model. Then, the average estimates were much closer to the true values and almost identical to the sire model estimates.
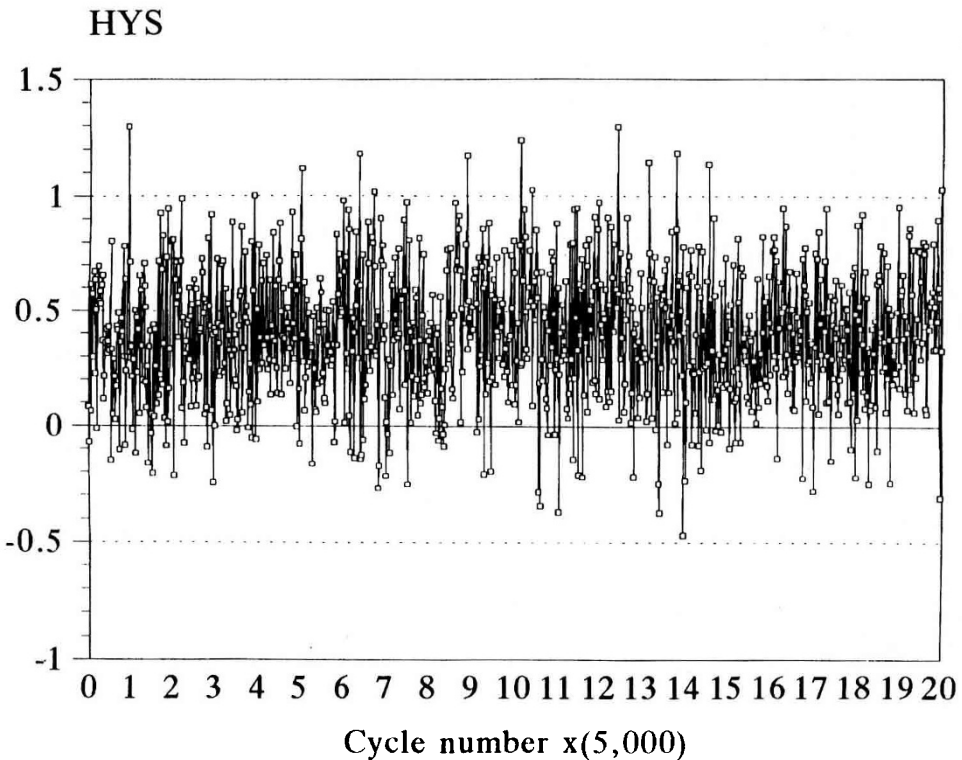

Fig 6. Plot of sample value of HYS effect *versus* Gibbs cycle for an HYS without the ECP.

The results discussed so far strongly suggest that the cause of the upward bias in the MCMPM and MCMML estimators is a highly non-symmetrical marginal posterior distribution of heritability in situations where the data contain little information. For the larger designs (IV-F, VI-F), the biases strongly decreased. However, it appears to be likely that the posterior density of the parameters is also improper in the presence of HYS levels with the ECP. As mentioned earlier, improperness can be shown analytically for a simple fixed model and the logit link for binary data when at least one level of the fixed factor exhibits the ECP. The fact that the Gibbs sampler 'blew up' for the smaller designs under the animal model also supports the hypothesis of an improper posterior. Moreover, when a joint posterior density of the parameters is improper, 'marginal posterior mean or mode' estimates or 'marginal posterior density plots' from Gibbs output may look 'reasonable', although these were not obtained from a Markov chain with known stationarity properties (Hobert and Casella, 1994). In the logit link example, effects of fixed levels not showing ECP are still 'reasonably well estimated' in the presence of other levels with the ECP causing the posterior to be improper.

Because an analytical proof of an improper posterior under the nonlinear mixed model for binary data employed in this investigation appeared to be very difficult or intractable, the marginal likelihood in [7], or the reciprocal of the integration constant of the posterior density, was considered as a potential criterion for detecting improperness from Gibbs output. It was estimated from several modified design II-F data sets using expression [11] and required 3 consecutive Gibbs chains which were run at lengths of 10 000 and 20 000 (after burn-in) cycles. Two situations were considered in which the posterior density was probably improper: in the presence of the ECP and when using the prior $1/\sigma_u^2$ for the variance component.

To examine the ECP, 4 data sets were created. Data set 1 had 70 HYS with the ECP, data set 2 had only 1 HYS with the ECP and data set 3 had none. Data sets 2 and 3 were obtained by reducing the incidence of the trait from 0.9 to 0.5 and by reducing the size of the differences among HYS differences. Data set 4 was from design II-N with random HYS effects.

Table III contains the estimated normalizing constants (reciprocals of marginal likelihoods) and the MCMPM estimates of heritability for each of the 5 data sets and the 2 lengths of the sampler. The estimates of the integration constant for data sets 1 and 2 of table III differed strongly between lengths of 10 000 and 20 000 cycles, *ie* did not 'converge'. The estimate of the integration constant, however, did converge for data set 3 not containing HYS with ECP and for data set 4 treating HYS as random. For all 4 data sets, the estimate of heritability appeared to have 'converged', because virtually the same means were obtained after 10 000 and 20 000 cycles. 'Convergence' was also indicated by the plot of sample value for heritability *versus* Gibbs cycle in figure 7 although it showed more variability than a corresponding plot in figure 8 for data set 3 not containing any HYS with ECP. Note that the numbering of the Gibbs cycles in the figures begins after burn-in. Expectedly, a plot of HYS sample value *versus* Gibbs cycle (fig 5) showed non-convergence for an ECP HYS in data set 1 of table III. As a control, figure 6 presents the plot of sample value *versus* Gibbs cycle for an HYS without ECP, which demonstrated convergence. Plots of the marginal posterior density for an HYS effect with and without ECP can be found in figures 9 and 10, respectively.

Table III. Estimates of heritability[a] and of the integration constant[b] and with the Gibbs sampler run at 2 different lengths. Data sets 1–4 are from design II-F and data set 5 is from design IV-F (see text).

Data set	Parameter/length			
	h^2/10 000	h^2/20 000	$[f(y)]^{-1}$/10 000	$[f(y)]^{-1}$/20 000
1 (70 ECP HYS)	0.297	0.290	$e^{-585258}$	$e^{-957785}$
2 (1 ECP HYS)	0.307	0.305	$e^{-3703.0}$	$e^{-13984.5}$
3 (0 ECP HYS)	0.286	0.285	$e^{-2007.8}$	$e^{-2008.6}$
4 (HYS random)	0.349	0.350	$e^{-1713.2}$	$e^{-1718.7}$

[a]Estimator of heritability was MCMPM. [b]Estimation by equations [11] and [12a], [12b], [12c].
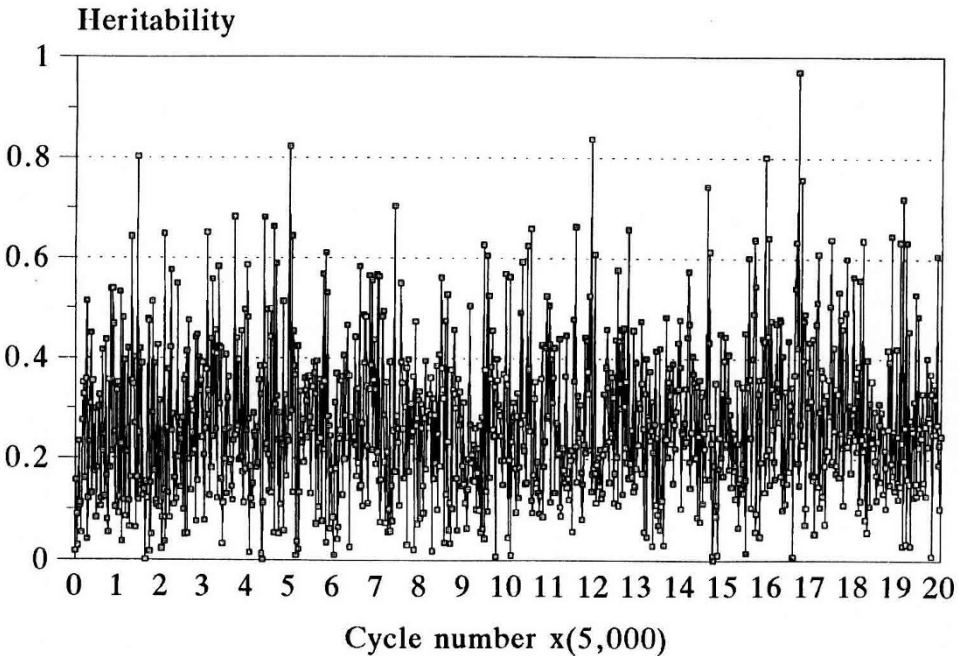
Fig 7. Plot of sample value for heritability *versus* Gibbs cycle for data set 1 of table III.

To verify the effect of the particular factorization of the posterior density evaluated at the vector of posterior means of the parameters in [11], the analysis of data set 1 in table II was repeated for the other 5 factorizations, which all indicated non-convergence of the integration constant (results not presented).

To examine the second case of an improper posterior caused by the $1/\sigma_u^2$ prior, design V-F data sets were analyzed with this prior, which is known to produce
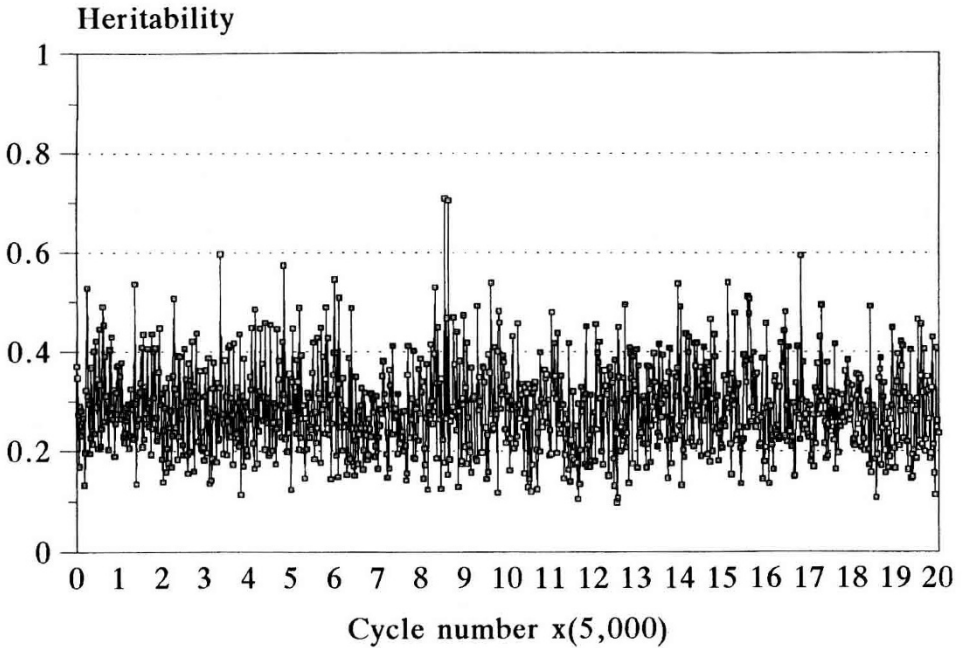
Heritability

Fig 8. Plot of sample value for heritability *versus* Gibbs cycle for data set 3 of table III.

improper posteriors in the linear model (Hobert and Casella, 1994). Use of this prior in estimation under design V-F occasionally yielded zero h^2 estimates or underestimated h^2. Although heritability was underestimated ($h^2 = 0.09$), the Monte-Carlo estimates of the integration constant were almost unchanged after 10 000 and 20 000 Gibbs cycles indicating that this form of improperness, if it exists, was not detected by Monte-Carlo estimation of the integration constant based on [11].

CONCLUSIONS

This study confirmed that the AMML estimator of the heritability of a threshold character has a downward bias if family sizes are small (Gilmour *et al*, 1985), in this case if sire progeny group size is small for a binary trait with high (or low) incidence. This bias is due to the approximation in the AMML estimator (Foulley *et al*, 1987; Hoeschele *et al*, 1987), because it is eliminated by computing exact MML estimates *via* Gibbs sampling.

For small data sets with extreme incidence and many fixed effects (*eg*, design II-F), *ie* with little information about the heritability, the marginal posterior density of heritability is highly non-normal, its mean and mode differ and overestimate h^2 with the mean being more strongly biased. Even for a data set larger by a factor of 10 (design IV-F), marginal mean and mode were still significantly biased upward. Biases became insignificant when in the large design the mean was the only fixed
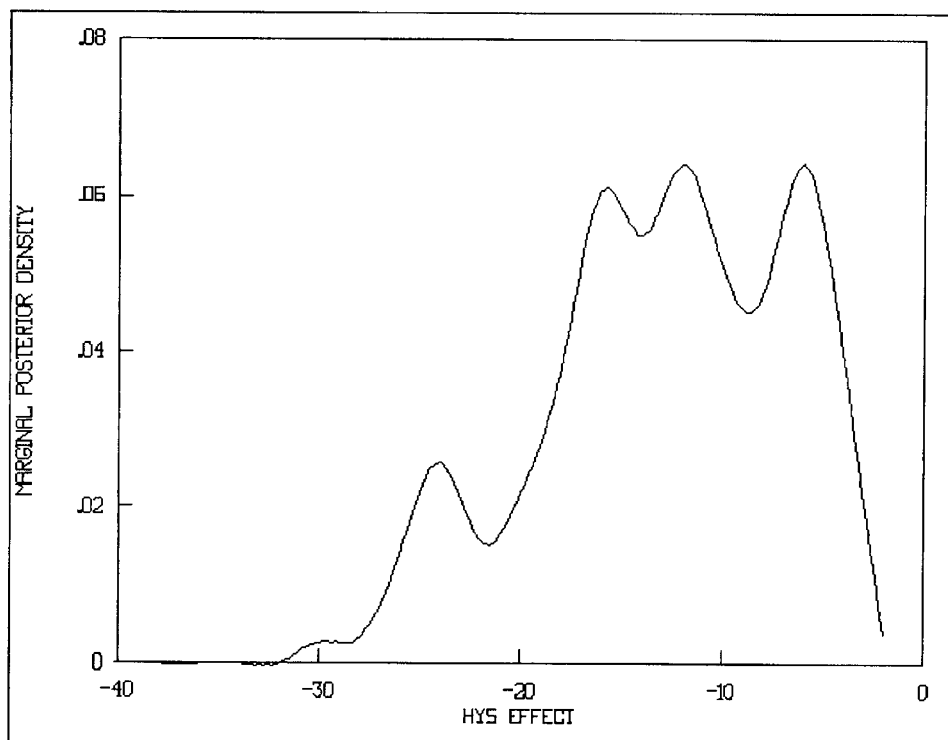
Fig 9. Marginal posterior density of HYS effect estimated from Gibbs output for an HYS with the ECP.

effect and the incidence was 0.5. The phenomenon of obtaining biased estimates when random effects are incorporated into a generalized linear model is quite well known, and an iterative bias correction method for parametric models using the bootstrap has been developed (Kuk, 1995). Moreno (personal communication) obtained promising results when applying Kuk's method to a binary threshold trait, however, he noted that the bootstrap is computationally extremely demanding.

Without bias correction, accurate estimation of genetic parameters for binary threshold traits requires a large amount of data in absolute terms and relative to the number of fixed effects. This is likely even more important for genetic correlations than for heritabilities with the former known to be poorly estimated from binary data (Simianer and Schaeffer, 1989). An encouraging result, however, is that for situations where only a small data set containing many HYS levels is available for analysis, the MCMML estimator of heritability appears to be unbiased if a normal prior is used for the HYS effects, irrespective of the true state of nature of the HYS effects (fixed, normally or bounded uniformly distributed). As categorical traits are usually not under intense selection, a non-random association of sires and herds seems unlikely which justifies treating HYS effects as random in practice. Other approaches to improving heritability estimation for such data, *eg*, fixing values of HYS levels with the ECP at predetermined constants or use of a bounded uniform
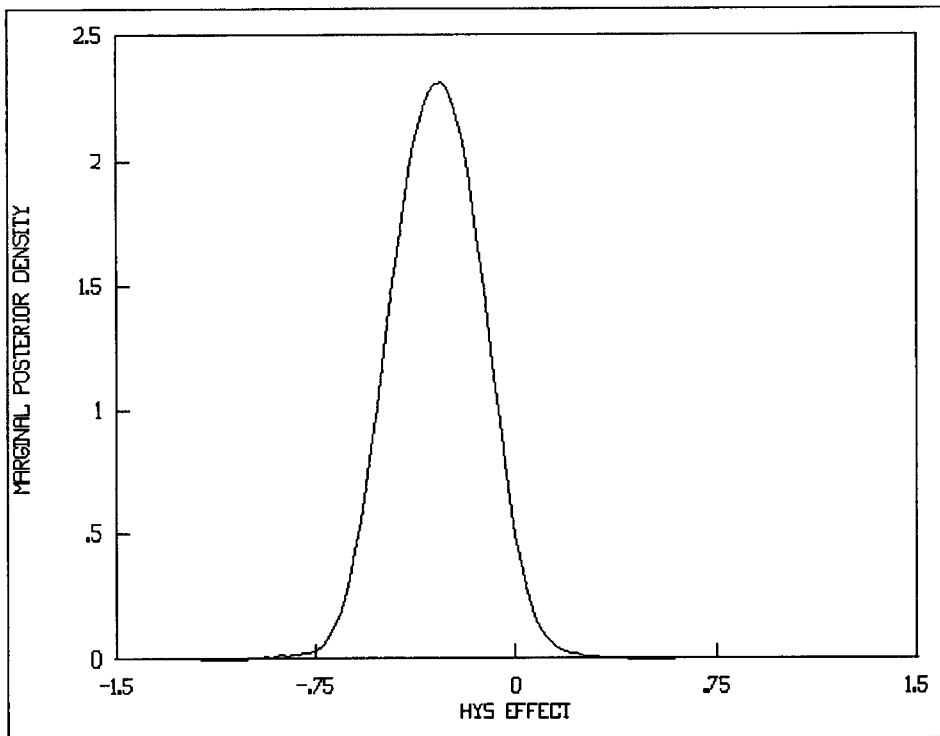
Fig 10. Marginal posterior density of HYS effect estimated from Gibbs output for an HYS with the ECP.

prior, were not successful. The use of alternative prior distributions (*eg*, Berger and Bernardo, 1992a, b) for fixed factors with many levels exhibiting the ECP should be investigated further.

The accuracy of heritability estimation was also found to differ among sire and animal models. For data with extreme incidence, a limited number of sires (50) and small progeny group size (40), estimates obtained with the animal model were less accurate than those from the sire model. When the number of HYS effects (treated as fixed) was additionally large, the Gibbs sampler did not 'converge'. Therefore, the animal model should be used only when there is sufficient information in the data; otherwise the apparently more robust sire model should be preferred.

ACKNOWLEDGMENTS

REFERENCES

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

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

Berger JO, Bernardo JM (1992b) Reference priors in a variance components problem. *In: Proceedings of the Indo-USA Workshop on Bayesian Analysis in Statistics* (P Goel, ed), Springer Verlag, New York, 177-194

Chib S (1994) *Marginal Likelihood from the Gibbs Output*. Technical Report, Olin School of Business, Washington University, Washington, DC

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

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

Geyer CJ (1992) Practical Markov chain Monte-Carlo (with discussion). *Stat Sci* 7, 467-511

Gianola D, Foulley JL (1983) Sire evaluation for ordered categorical data with a threshold model. *Genet Sel Evol* 15, 201-224

Gianola D, Im S, Macedo FW (1990) A framework for prediction of breeding value. *In: Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds), Springer Verlag, Berlin, 210-238

Gilmour AR, Anderson RD, Rae AL (1985) The analysis of binomial data in a generalised linear mixed model. *Biometrika* 72, 593-599

Henderson CR (1985) Equivalent linear models to reduce computations. *J Dairy Sci* 68, 2267-2277

Hobert JP, Casella G (1994) Gibbs sampling with improper prior distributions. Technical Report BU-1221-M, Biometrics Unit, Cornell University, Ithaca, NY

Hoeschele I (1989) A note on local maxima in Maximum Likelihood, Restricted Maximum Likelihood, and Bayesian estimation of variance components. *J Stat Comp Simul* 33, 149-160

Hoeschele I, Foulley JL, Gianola D (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet* 104, 334-339

Hoeschele I, Gianola D (1989) Bayesian *versus* Maximum Quasi Likelihood methods for sire evaluation with categorical data. *J Dairy Sci* 72, 1569-1577

Janss LLG, Thompson R, van Arendonk JAM (1994) Application of Gibbs sampling for inference in a mixed major gene–polygenic inheritance model in animal populations. *Theor Appl Genet* (in press)

Jensen J (1994) Bayesian analysis of bivariate mixed models with one continuous and one binary trait using the Gibbs sampler. *Proc 5th World Congr Genet Appl Livest Prod* Vol 18, 333

Kuk AYC (1995) Asymptotically unbiased estimation in Generalized Linear Models with random effects. *J R Stat Soc* B 57, 395-407

Mc Culloch CE (1994) Maximum likelihood variance components estimation for binary data. *J Am Stat Assoc* 89, 330-335

Misztal I, Gianola D (1989) Computing aspects of a nonlinear method of sire evaluation for categorical data. *J Dairy Sci* 72, 1557-1568

Newton MA, Raftery AE (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *J R Stat Soc B* 56, 3-48

Simianer H, Schaeffer LR (1989) Estimation of covariance components between one continuous and one binary trait. *Genet Sel Evol* 21, 303-315

Sorensen DA, Andersen S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. *Genet Sel Evol* 27, 229-249

Thompson R (1990) Generalized linear models and applications to animal breeding. *In: Advances in Statistical Methods for Genetic Improvement of Livestock* (D Gianola, K Hammond, eds), Springer Verlag, Berlin, 312-328

Wang CS, Rutledge JJ, Gianola D (1993) Marginal inferences about variance components in a mixed linear model using Gibbs sampling. *Genet Sel Evol* 25, 41-62

Zeger SL, Karim MR (1991) Generalized linear models with random effects: a Gibbs sampling approach. *J Am Stat Assoc* 86, 79-86