Theor Appl Genet (1995) 91:1137–1147

L. L. G. Janss · R. Thompson · J. A. M Van Arendonk

Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations

Received: 11 April 1995 / Accepted: 21 April 1995

Abstract The application of Gibbs sampling is considered for inference in a mixed inheritance model in animal populations. Implementation of the Gibbs sampler on scalar components, as used for human populations, appeared not to be efficient, and an approach with blockwise sampling of genotypes was proposed for use in animal populations. The blockwise sampling of genotypes was proposed for use in animal populations. The blockwise sampling by which genotypes of a sire and its final progeny were sampled jointly was effective in improving mixing, although further improvements could be looked for. Posterior densities of parameters were visualised from Gibbs samples; from the former highly marginalised Bayesian point- and interval estimates can be obtained.

Key words Animal breeding · Mixed inheritance · Major gene · Inference · Gibbs sampling

Introduction

Gibbs sampling has been proposed for making inferences in a mixed inheritance model in human populations (Guo and Thompson 1992). The Gibbs sampler is a sampling-based computational to that is used to perform marginalisations without analytical approximation (Geman and Geman 1984; Gelfand and Smith 1990). As such, it can marginalise the joint density of unknowns from a mixed inheritance model with respect to both polygenic effects and genotypes, which is an impossible task in general pedigrees using analytical approaches (e.g. Le Roy et al. 1989; Knott et al. 1992; Kinghorn et al. 1993). Due to its potential, Gibbs sampling, or related techniques, may soon dominate other computational methods for making genetic inferences, in particular when modelling single loci, such as in major gene detection and in quantitative trait loci (QTL) and marker mapping. For a review on recent applications of Gibbs sampling in animal breeding, see Sorensen et al. (1994).

Use of the Gibbs sampler implementation for human populations (Guo and Thompson 1992) in animal breeding may show very slow mixing of genotype states, resulting in difficulty in achieving convergence. Large progeny groups in animal breeding are responsible for this effect. The aim of the study presented here was to describe the construction of a markov chain using a modified sampling scheme more suited for inference in animal populations. Because this study is the first report of using Gibbs sampling in a mixed inheritance model in animal breeding, we will describe in detail the construction of the required markov chain. The effect of the modified sampling scheme on mixing will be demonstrated. A small simulation study will be presented showing the types of marginal posterior densities that can be obtained accompanied by a discussion of possible methods of inference based on these marginal densities.

Mixed inheritance model

In a mixed inheritance model a trait is influenced by the genotype at a single locus and by a polygenic effect, which is the aggregate effect of a large number of loci unrelated to the single locus. The single locus is assumed to be an additive, biallelic, autosomal locus with Mendelian transmission probabilities. Alleles at the single locus are A_1 and A_2 with genotypes A_1A_1, A_1A_2, A_2A_1 and A_2A_2. The heterozygotes A_1A_2 and A_2A_1 are distinguished to provide a simple and yet flexible notation for their covariance structure. In an alternative notation,

Communicated by D. Van Vleck

L. L. G. Janss (✉) · J. A. M. Van Arendonk
Department of Animal Breeding, Wageningen Agricultural University, PO Box 338, 6700 AH Wageningen, The Netherlands

R. Thompson
Department of Biometrical Genetics, AFRC Roslin Institute, Roslin, Midlothian EH25 9PS, UK

the genotype of individual i is denoted $\mathbf{w}_i$ with four possible realisations $\boldsymbol{\omega}_{ef}$, a row vector, corresponding to genotype A_eA_f: $\boldsymbol{\omega}_{11} = (1000)$, $\boldsymbol{\omega}_{12} = (0100)$, $\boldsymbol{\omega}_{21} = (0010)$ and $\boldsymbol{\omega}_{22} = (0001)$. We assume a homogeneous population of base individuals with genotypes in Hardy-Weinberg proportions. Relaxation of these assumptions is feasible by increasing the number of parameters to be estimated, which poses no particular difficulty. We also assume that each individual has one observation for the trait. Inbreeding will be accounted for in the computations. The statistical model for the observations is:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{ZWm} + \mathbf{e} \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of fixed nongenetic effects, $\mathbf{X}$ is a design matrix relating nongenetic effects to observations, $\mathbf{u}$ is a vector of random polygenic effects for all individuals in the pedigree, $\mathbf{Z}$ is a design matrix relating polygenic and single locus effects to observations, $\mathbf{Wm}$ is a vector of random effects at the single locus for all individuals and $\mathbf{e}$ is a vector with errors. The effects at the single locus are expressed using $\mathbf{W} = \{\mathbf{w}_i\}$, a matrix containing information on the genotype of each individual, and $\mathbf{m}$, a vector with genotype means, where $\mathbf{m}' = (-a\,0\,0\,a)$. Hence, the A_2 allele is assumed to increase the trait value, hereafter called the favourable allele, no dominance is assumed and no distinction is made between the effects of the two heterozygotes A_1A_2 and A_2A_1.

The distribution of $\mathbf{e}$ is $N(\mathbf{0},\mathbf{I}\sigma_e^2)$ where N denotes the normal distribution. The covariance structures for polygenic effects can be expressed as $\mathbf{u} \sim N(\mathbf{0},\mathbf{A}\sigma_u^2)$, where $\mathbf{A}$ is the numerator relationship matrix and σ_u^2 the polygenic variance. The covariance structure for genotypes, however, cannot be expressed in matrix notation. To show the parallels between polygenic and monogenic effects, therefore, we will specify both in scalar notation. For individual i, the polygenic effect u_i is:

$u_i \sim N(0,\sigma_u^2)$, when i is an individual in the base population (2a)

$u_i \sim N(1/2u_{S,i} + \frac{1}{2}u_{D,i},\ \phi_i\sigma_u^2)$, when i is an individual not in the base population (2b)

where $u_{S,i}$ and $u_{D,i}$ in Eq. 2b are polygenic effects of the sire and dam of i, and $\phi_i = (\frac{1}{2} - \frac{1}{4}F_{S,i} - \frac{1}{4}F_{D,i})$ is the Mendelian sampling term for individual i, where $F_{S,i}$ and $F_{D,i}$ are inbreeding coefficients of the sire and dam of i. An analogous scalar notation for the covariance structure for the genotype $\mathbf{w}_i$ of individual i is:

$$P(\mathbf{w}_i = \boldsymbol{\omega}_{ef}) = p_e p_f \tag{3a}$$

$$P(\mathbf{w}_i = \boldsymbol{\omega}_{ef} | \mathbf{w}_{S,i} = \boldsymbol{\omega}_{gh}, \mathbf{w}_{D,i} = \boldsymbol{\omega}_{g'h'}) = \tau_{e,gh}\tau_{f,g'h'} \tag{3b}$$

where P denotes probability, p_1 and $p_2 (= 1 - p_1)$ are frequencies of alleles A_1 and A_2 in the base population, $\tau_{1,gh}$ is the probability of transmission to an offspring of an A_1 allele from a parent with genotype A_gA_h and $\tau_{2,gh} = 1 - \tau_{1,gh}$. In Eq. 3b $\mathbf{w}_{S,i}$ and $\mathbf{w}_{D,i}$ are genotypes of sire and dam of i. Assuming Mendelian probabilities of transmission, $\tau_{1,11} = 1, \tau_{1,12} = \tau_{1,21} = \frac{1}{2}$ and $\tau_{1,22} = 0$.

Flat priors are assigned for nongenetic effects $\boldsymbol{\beta}$, for variance components σ_e^2 and σ_u^2, and for allele effect a and allele frequency p_1, i.e. $f(\boldsymbol{\beta}, \sigma_e^2, \sigma_u^2, a, p_1) \propto$ constant. Variance components are a priori positive, i.e. excluding zero, and the allele frequency is bounded between zero and one, including the bounds. The joint density of all unknowns, given data $\mathbf{y}$, is symbolically denoted:

$$f(\boldsymbol{\beta}, \mathbf{u}, \mathbf{W}, \sigma_e^2, \sigma_u^2, a, p_1 | \mathbf{y}) \tag{4}$$

Gibbs sampling

The Gibbs sampler is based on a markov chain that is primarily used to generate samples from a joint density (Geman and Geman 1984; Gelfand and Smith 1990). These samples enable the study of all marginal densities from that joint density. For statistical problems, the joint density is for the set of unknowns from a statistical model, given observed data, e.g. as in Eq. 4. In statistical problems, the Gibbs sampler is generally used to study marginal posterior densities of parameters; other parameters are considered to be nuisances. Using the primary joint structure in the samples, we can obtain sampling correlations between parameter estimates and for example, two-parameter countour plots. In analytical approaches, the study of marginal densities would require integrations or summations, often not feasible to compute, but which can be circumvented when using Gibbs sampling.

Validity

The use of Gibbs sampling is valid when the joint density considered has a non-zero probability over its entire domain (Tanner 1993), which is similar to the requirement of irreducibility of the Gibbs markov chain. An irreducible chain can be characterised as a chain which, from any state, has a positive probability of transition to each other state. Irreducibility is not always straightforward, a model with a single locus with more than two alleles or a discrete penetrance function leads to a reducible chain (Sheehan and Thomas 1993; Lin et al. 1993). Also, chains may be "practically" reduced, i.e. transition probabilities to certain states are so low that, in practice, these states are never reached. For the model described here, the Gibbs markov chain is theoretically irreducible; possible practical reducibility will be discussed later.

The use of an improper joint density, a non-integrable function, is also invalid for the application of Gibbs sampling. Hobert and Casella (1994) showed that priors $(\sigma^2)^{-(b+1)}$ for variance component estimation in linear models lead to a proper posterior density when

$b < 0$. Hence, a flat prior, corresponding to $b = -1$ yields a proper posterior (see also Besag et al. 1991; Wang et al. 1993, 1994). We assume that the result of Hobert and Casella (1994) obtained for linear models is also valid for the mixed inheritance model.

General construction

The Gibbs markov chain is a continuing series of realisations for the unknowns $\boldsymbol{\beta}, \mathbf{u}, \mathbf{W}, \sigma_e^2, \sigma_u^2, a$ and p_1. Let $\boldsymbol{\theta}^{[t]} = (\boldsymbol{\beta}^{[t]}, \mathbf{u}^{[t]}, \mathbf{W}^{[t]}, \sigma_e^{2[t]}, \sigma_u^{2[t]}, a^{[t]}, p_1^{[t]})$ denote the set of realisations for the unknowns at state or cycle t in the Gibbs chain. Construction of the Gibbs chain requires a set of realisations $\boldsymbol{\theta}^{[t+1]}$, given the current set of realisations $\boldsymbol{\theta}^{[t]}$. To initiate the chain, a set of starting realisations $\boldsymbol{\theta}^{[0]}$ is required, for which we used zeros for $\boldsymbol{\beta}$ and $\mathbf{u}$ and initial guesses for $\sigma_e^2, \sigma_u^2, a$ and p_1. Genotypes $\mathbf{W}$ were initiated as all heterozygotes A_1A_2.

In the most straightforward implementation of the Gibbs sampler, $\boldsymbol{\theta}^{[t+1]}$ is obtained by sampling for each $\theta_i (i = 1,r)$ a new realisation from the conditional distribution of θ_i, given the available realisations $\theta_1^{[t+1]}, \ldots, \theta_{i-1}^{[t+1]}, \theta_{i+1}^{[t]}, \ldots, \theta_r^{[t]}$ and given the data $\mathbf{y}$ (Gelfand and Smith 1990). The form of the conditional densities required often appear to be simple. For model 1, conditional densities are normal for β_i's, u_i's and a, discrete for $\mathbf{w}_i$'s, inverted chi-square for σ_e^2 and σ_u^2 and beta for p_1. The simple form of the conditional densities allows implementation of a Gibbs chain for a mixed inheritance model based on sampling from the exact small sample distributions in each step. Further, for computations on pedigrees, the "neighbourhood set" of an individual (Sheehan and Thomas 1993) plays an important role. This neighbourhood set consists of the polygenic values or genotypes of the parents, progeny and mates of an individual, together with the data on itself. To compute conditional densities for sampling the polygenic effect or genotype of an individual, only the elements in this neighbourhood set are required because of redundancies arising in the conditional densities. One side-effect in particular is that exact computations of conditional genotype probabilities are automatically made in looped pedigrees, whereas analytical approaches become intractable in pedigrees containing loops. Animal pedigrees generally contain many loops due to the common occurrence of multiple matings and inbreeding.

Mixing and blocking

The most straightforward implementation of the Gibbs chain, sampling single parameters, may not be an efficient way to obtain Gibbs samples because strongly dependent parameters may show slow mixing (Smith and Roberts 1993; Tanner 1993). By mixing we generally refer to the speed of movement of the chain in the parameter space. In the initial phase of a markov chain, mixing is important for convergence to the equilibrium distribution and for burn-in time. In the later phase, mixing determines the serial correlations in the chain that affect efficiency by which accurate inferences can be made from the chain. In a mixed inheritance model, genotypes often show slow mixing due to the dependence between genotypes of parents and progeny. This dependence is stronger and mixing is poorer when progeny groups are larger.

As described by Smith and Roberts (1993) and Tanner (1993), for example, mixing can be improved by applying Gibbs sampling to subvectors, thereby treating components as a "block", rather than by using a complete breakdown of the parameter vector in its scalar components. In practice, blocking can be implemented using one or more reduced conditional densities. Use of reduced conditionals densities, as in substitution sampling, was also considered by Gelfand and Smith (1990) for improving convergence. Efficiency of the Gibbs sampler was improved here by blocking genotypes of each sire and its final progeny. Final progeny are progeny that are not parents themselves. By considering only final progeny, the number of individuals involved in computing conditional densities is not increased and remains based on parents, progeny and mates. But, for final progeny, phenotypes are used instead of genotypes. Efficiency was improved further by updating genotypes starting with the youngest families. In this manner, changes appearing in younger families can cause changes in older families within the same update cycle of the Gibbs chain. The blockwise treatment of genotypes of each sire and its final progeny was also applied to polygenic effects. In the Results section the effect of blocking on the changes of genotypes in a Gibbs chain will be demonstrated.

Random number generator

Construction of the Gibbs markov chain requires the sampling of many random deviates, which are based on pseudo-random number generators. Because parameters in the markov chain are updated repeatedly in the same order, the absence of serial correlations in the deviates is important. We used the RAN1, GASDEV and GAMDEV routines (Press et al. 1986), which seemed to meet that requirement. The GAMDEV routine was used to generate chi-square deviates with even-numbered degrees of freedom. Deviates with odd-numbered degrees of freedom were generated by adding one squared random normal deviate.

Sampling of realisations in the Gibbs chain

In this section obtaining $\boldsymbol{\theta}^{[t+1]}$ given $\boldsymbol{\theta}^{[t]}$ will be described. This represents the computation of one "Gibbs cycle". Repetition of this procedure constructs the Gibbs markov chain. The described blockwise treatment for the genotypes and polygenic effects of sires and final

progeny is incorporated, as well as the order of sampling starting with the youngest individuals. In the Gibbs chain, sampling is applied to all unknowns, including genetic parameters, to allow for Bayesian inferences. Parameters are updated in the order given in the joint density (Eq. 4).

Nongenetic effects

Assume first that nongenetic effects $\boldsymbol{\beta}$ are levels of one factor. Then step S1 in the construction of state $t+1$ from t is:

sample $\beta_i^{[t+1]}$ from $N(\check{y}_{i\cdot}/n_i, \sigma_e^{2[t]}/n_i)$. (S1)

Step S1 is based on conditional solutions to the linear model and on conditional standard errors for nongenetic effects. Conditioning on polygenic effects, genotypes and allele effect results in the use of corrected data $\check{\mathbf{y}} = (\mathbf{y} - \mathbf{Zu}^{[t]} - \mathbf{ZW}^{[t]}\mathbf{m}^{[t]})$, with $\mathbf{m}^{[t]} = (-a^{[t]}, 0, 0, a^{[t]})$, where $\check{y}_{i\cdot}$ is the total of observations from $\check{\mathbf{y}}$ pertaining to level i and n_i is the number of observations in level i. More effects would be handled one at a time, correcting $\check{\mathbf{y}}$ also for other nongenetic effects. For two effects, $\boldsymbol{\beta}$ is partitioned as $[\boldsymbol{\beta}_1\boldsymbol{\beta}_2]$ and $\mathbf{X}$ as $[\mathbf{X}_1\mathbf{X}_2]$, and $\boldsymbol{\beta}_1$ is updated to state $t+1$ as above using $\check{\mathbf{y}} = (\mathbf{y} - \mathbf{X}_2\boldsymbol{\beta}_2^{[t]} - \mathbf{Zu}^{[t]} - \mathbf{ZW}^{[t]}\mathbf{m}^{[t]})$, after which $\boldsymbol{\beta}_2$ is updated in the same manner using $\check{\mathbf{y}} = (\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta}_1^{[t+1]} - \mathbf{Zu}^{[t]} - \mathbf{ZW}^{[t]}\mathbf{m}^{[t]})$. Note the direct use of $\boldsymbol{\beta}_1^{[t+1]}$.

Polygenic effects

Steps to update polygenic effects are based on BLUP equations for the linear "animal model" (Henderson 1988) and on conditional standard errors for polygenic effects. The neighbourhood set of polygenic effects (e.g. Sheehan and Thomas 1993) to be considered is represented exactly in BLUP equations. The updating of polygenic effects is based on using step S2.1 for dam j and step S2.2 for sire i with its final progeny l:

sample $u_j^{[t+1]}$ from $N(c_j/d_j, \sigma_e^{2[t]}/d_j)$, (S2.1)

sample $u_i^{[t+1]}$ from $N(c_i/d_i, \sigma_e^{2[t]}/d_i)$ and (S2.2)

sample $u_l^{[t+1]}$ from $N(c_l/d_l, \sigma_e^{2[t]}/d_l)$ for each final progeny l of sire i,

where the BLUP equations are $d_j u_j = c_j$ to solve for the polygenic effect u_j of dam j; $d_i u_i = c_i$ to solve for the polygenic effect u_i of sire i after absorption of all final progeny of i; and $d_l u_l = c_l$ to solve for the polygenic effect u_l of final progeny l. Step S2.2 is the sampling of new realisations for a sire and its final progeny jointly as a block, done in two steps. The first step draws a new realisation for the sire effect from the reduced conditional density, after absorption of final progeny. The second step finalises the joint sampling by obtaining new realisations for final progeny, conditional on the new value for the sire. On the basis of the BLUP equations, elements in S2.1 and S2.2 are:

$$c_j = \check{y}_j + \tfrac{1}{2}\alpha\delta_j(u_{S,j}^{[t]} + u_{D,j}^{[t]}) - \alpha\textstyle\sum_k(\tfrac{1}{4}\delta_k u_{S,k}^{[\cdot]} - \tfrac{1}{2}\delta_k u_k^{[t+1]})$$

$$d_j = 1 + \alpha(\delta_j + \tfrac{1}{4}\textstyle\sum_k\delta_k)$$

$$c_i = \check{y}_i + \tfrac{1}{2}\alpha\delta_i(u_{S,i}^{[t]} + u_{D,i}^{[t]}) - \alpha\textstyle\sum_m(\tfrac{1}{4}\delta_m u_{D,m}^{[\cdot]} - \tfrac{1}{2}\delta_m u_m^{[t+1]}) - \textstyle\sum_l\{\tfrac{1}{4}\alpha\delta_l u_{D,l}^{[\cdot]} - (\check{y}_l + \tfrac{1}{2}\alpha\delta_l u_{D,l}^{[\cdot]})/(1+\alpha\delta_l)\}$$

$$d_i = 1 + \alpha(\delta_i + \tfrac{1}{4}\textstyle\sum_k\delta_k) - \tfrac{1}{4}\textstyle\sum_l(\alpha\delta_l)^2/(1+\alpha\delta_l)$$

$$c_l = \check{y}_l + \tfrac{1}{2}\alpha\delta_l(u_i^{[t+1]} + u_{D,l}^{[\cdot]})$$

$$d_l = 1 + \alpha\delta_l$$

where for individual i, $u_{S,i}$ and $u_{D,i}$ denote polygenic effects of the sire and dam of i, $\check{y}$ is the element pertaining to i from the corrected data $\check{\mathbf{y}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{[t+1]} - \mathbf{ZW}^{[t]}\mathbf{m}^{[t]})$, δ_i is the reciprocal of the Mendelian sampling term ϕ_i from (Eq. 2b); the premises are similar for other individuals j, k, l or m. When individual i or j has unknown parents, zeros are substituted for $u_{S,i}$ and $u_{D,i}$, or $u_{S,j}$ and $u_{D,j}$. In the equation for c_j, $\sum_k$ is evaluated for each progeny k of j; in the equation for c_i, $\sum_l$ is evaluated for each final progeny l of i and $\sum_m$ is evaluated for each nonfinal progeny m of i. In the equation for c_l, i is the sire of l. Finally, α is the variance ratio $\sigma_u^{2[t]}/\sigma_e^{2[t]}$. When it is unclear whether a polygenic value used is from state t or $t+1$, the state is not specifically indicated.

Genotypes

Obtaining new realisations for genotypes is done as for polygenic effects except that discrete distributions are sampled. Conditional probabilities for genotypes are obtained by peeling (Cannings et al. 1978), but genotypes of the individuals in the neighbourhood set, i.e. parents, progeny and mates, are taken as known. Analogous to polygenic effects, the updating of all genotypes is based on step S3.1 for dam j and step S3.2 for sire i and its final progeny l:

sample $\mathbf{w}_j$ according to the probabilities: (S3.1)

$$P(\mathbf{w}_j = \boldsymbol{\omega}_{ef}) \propto f(\check{y}_j|\mathbf{w}_j = \boldsymbol{\omega}_{ef})P(\mathbf{w}_j = \boldsymbol{\omega}_{ef}|\mathbf{w}_{S,j}^{[t]}, \mathbf{w}_{D,j}^{[t]}) \cdot \textstyle\prod_k P(\mathbf{w}_k^{[t+1]}|\mathbf{w}_{S,k}^{[\cdot]}, \mathbf{w}_j = \boldsymbol{\omega}_{ef})$$

sample $\mathbf{w}_i$ according to the probabilities (S3.2)

$$P(\mathbf{w}_i = \boldsymbol{\omega}_{ef}) \propto f(\check{y}_i|\mathbf{w}_i = \boldsymbol{\omega}_{ef}) P(\mathbf{w}_i = \boldsymbol{\omega}_{ef}|\mathbf{w}_{S,i}^{[t]}, \mathbf{w}_{D,i}^{[t]})$$
$$\cdot \Pi_m P(\mathbf{w}_m^{[t+1]}|\mathbf{w}_i = \boldsymbol{\omega}_{ef}, \mathbf{w}_{D,m}^{[\cdot]})$$
$$\cdot \Pi_l \textstyle\sum_{g,h} P(\mathbf{w}_l = \boldsymbol{\omega}_{gh}|\mathbf{w}_i = \boldsymbol{\omega}_{ef}, \mathbf{w}_{D,l}^{[\cdot]}) f(\check{y}_l|\mathbf{w}_l = \boldsymbol{\omega}_{gh})$$

and for each final progeny l of sire i, sample $\mathbf{w}_l$ according to:

$$P(\mathbf{w}_l = \boldsymbol{\omega}_{gh}) \propto f(\check{y}_l|\mathbf{w}_l = \boldsymbol{\omega}_{gh}) P(\mathbf{w}_l = \boldsymbol{\omega}_{gh}|\mathbf{w}_i^{[t+1]}, \mathbf{w}_{D,l}^{[\cdot]})$$

where the notation is analogous to that for polygenic effects and P denotes probability. Here $\check{\mathbf{y}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{[t+1]} - \mathbf{Z}\mathbf{u}^{[t+1]})$, $f(\check{y}_i|\mathbf{w}_i = \boldsymbol{\omega}_{ef}) \propto \exp\{-\frac{1}{2}(\check{y}_i - \boldsymbol{\omega}_{ef}\mathbf{m}^{[t]})^2/\sigma_e^{2[t]}\}$ is the normal penetrance function for i, and $P(\mathbf{w}_i = \boldsymbol{\omega}_{ef}|\mathbf{w}_{S,i}^{[t]}, \mathbf{w}_{D,i}^{[t]})$ is a transmission probability for i, available from Eq. 3b. When parents of i are unknown, the transmission probability is replaced by $p_e^{[t]}p_f^{[t]}$. The products over k, l and m are evaluated for the same individuals as the sums over k, l and m for polygenic effects, and the sum within Π_l is evaluated over the possible genotypes of progeny l, for $g = 1, 2$ and $h = 1, 2$. Step S3.2 is the sampling of the genotypes of a sire and its final progeny where in the first part a new genotype for the sire is sampled from a reduced conditional density. In the reduced conditional density for a sire, the phenotypes of final progeny are used. The actual sampling of genotypes is done by evaluating the above probabilities for all possible realisations $\boldsymbol{\omega}_{11}$, $\boldsymbol{\omega}_{12}$, $\boldsymbol{\omega}_{21}$ and $\boldsymbol{\omega}_{22}$, and the sampling of a new genotype according to these probabilities. Probabilities are given to proportionality, and so need to be normalised.

Residual and polygenic variance

Variance components follow inverted chi-square distributions, with new realisations for σ_e^2 and σ_u^2 obtained as:

sample $\sigma_e^{2[t+1]}$ as $\mathbf{e}'\mathbf{e}/\chi^2(n-2)$ (S4)

sample $\sigma_u^{2[t+1]}$ as $\mathbf{u}^{[t+1]\prime}\mathbf{A}^{-1}\mathbf{u}^{[t+1]}/\chi^2(q-2)$ (S5)

where $\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{[t+1]} - \mathbf{Z}\mathbf{u}^{[t+1]} - \mathbf{Z}\mathbf{W}^{[t+1]}\mathbf{m}^{[t]})$, $\mathbf{A}$ is the numerator relationship matrix, n is the number of observations, q is the number of individuals, and $\chi^2(n-2)$ and $\chi^2(q-2)$ are random deviates from chi-squared distributions with $n-2$ and $q-2$ degrees of freedom. When degrees of freedom $n-2$ and $q-2$ are used, a flat prior for variance components is used (Wang et al. 1994). The quadratic $\mathbf{u}^{[t+1]\prime}\mathbf{A}^{-1}\mathbf{u}^{[t+1]}$ is computed as $\sum_i u_i^2 + \sum_j d_j(u_j - \frac{1}{2}u_{S,j} - \frac{1}{2}u_{D,j})^2$, a scalar computation due to the factorisation of $\mathbf{A}$ (Quaas 1976). The first summation is over all base animals and the second summation is over all non-base animals. Further notation is as in sampling steps S2.1 and S2.2. To prevent accidental rounding-off of variance components to zero, variances were not allowed to be smaller than 10^{-12}. Whenever a realised value fell below 10^{-12}, the sampling, i.e. step S4 or S5, was repeated.

Allele effect

If genotypes are used as a known classification factor, the effect of an allele is estimated as the deviation of homozygotes from an assumed mean of zero, yielding a linear model equation $(n_1 + n_4)\hat{a} = (\check{y}_4 \cdot - \check{y}_1 \cdot)$. This leads to:

sample $a^{[t+1]}$ from $N((\check{y}_{4.} - \check{y}_{1.})/(n_1 + n_4), \sigma_e^{2[t+1]}/(n_1 + n_4))$. (S6)

where n_i is the diagonal element i of $\mathbf{W}^{[t+1]\prime}\mathbf{Z}'\mathbf{Z}\mathbf{W}^{[t+1]}$, giving the number of genotypes of each type; $\check{y}_{i.}$ is element i of $\mathbf{W}^{[t+1]\prime}\mathbf{Z}'\check{\mathbf{y}}$, containing sums of corrected data per genotype with $\check{\mathbf{y}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{[t+1]} - \mathbf{Z}\mathbf{u}^{[t+1]})$. When all genotypes are $\mathbf{A}_1\mathbf{A}_1$ or $\mathbf{A}_2\mathbf{A}_2$, i.e. n_1 or n_4 is equal to the total number of animals, the effect of the allele is nonestimable and the new realisation for a is taken as zero.

Allele frequency

Given genotypes of base individuals, allele frequency in the base generation has a beta distribution. This leads to:

sample $p_1^{[t+1]}$ from $f(p_1) \propto p_1^{B_1}(1 - p_1)^{B_2}$ (S7)

where B_1 is the number of A_1 alleles and B_2 the number of A_2 alleles in genotypes of base individuals. An acception-rejection technique is used to sample a new allele frequency. A "suggested" sample p_1^* is generated from a uniform density. This p_1^* is accepted as the new sample for p_1 with probability $f_{\max}(p_1)/f(p_1^*)$, where $f_{\max}(p_1)$ is the maximum value of $f(p_1)$, attained for $p_1 = B_1/(B_1 + B_2)$. When p_1^* is rejected, the procedure is repeated.

Statistical inference

In the following we will describe a straightforward use of a Gibbs chain for making statistical inferences. In the Discussion section, we will elaborate on alternative approaches. For statistical inference, a long markov chain is produced, repeating the update scheme described in the previous section to obtain vectors with subsequent realisations for parameters. The subsequent realisations, or states in the markov chain, will show serial correlations, so that not every state is used to obtain Gibbs samples. Instead, virtually independent

samples are obtained by "thinning the chain". From the original chain, every K^{th} sample is taken, which is referred to as thinning "by K". Determining a suitable K-value or thinning parameter will be described first.

Thinning parameter

An initial run of the Gibbs sampler is required to determine a suitable K value. Following Raftery and Lewis (1992), thinning is based on a transformation of the original output into a binary process, for which transition probabilities are studied. Let $\theta^{[t]}$ be the value for a certain parameter at state t in the test run. The binary process is defined as $Z^{[t]} = \delta(\theta^{[t]} \leqslant c)$, where δ is the indicator function and c, in our application, is the mean of $\theta^{[t]}$'s. Thus, $Z^{[t]}$ indicates whether the realisation at state t was below or above the mean. The mean was taken because we are primarily interested in a central location parameter for the posterior densities. A suitable thinning parameter is obtained as follows, using for computations the binary process $Z^{[t]}$:

1) a thinning parameter k_1 is determined such that $Z^{[t]}$, thinned by k_1, is approximately a first-order markov (Raftery and Lewis 1992);
2) $Z^{[t]}$ thinned by k_1, being a first-order markov, can be described by a simple transition mechanism with transition probabilities α and β, which are estimated from $Z^{[t]}$ thinned by k_1;
3) an additional thinning parameter k_2 is determined, such that the transition probabilities in $Z^{[t]}$, thinned by $K = k_1 k_2$, differ only ε from the transition probabilities for $k_2 \to \infty$, i.e., for $K \to \infty$. Based on estimated transition probabilities from (2) and powers of the corresponding transition probability matrix, $k_2 = \ln(\varepsilon)/\ln(1 - \alpha - \beta)$. In our application, we took $\varepsilon = 0.001$.

Step (3) differs from the approach suggested by Raftery and Lewis (1992), who thinned only by k_1, yielding serially correlated realisations. Raftery and Lewis (1992) also determined the number of "burn in" cycles to be $\ln(\varepsilon(\alpha + \beta)/\max(\alpha,\beta))/\ln(1 - \alpha - \beta)$, which, for small ε and α and β approximately equal, is close to K. Therefore, taking the first Gibbs sample at state K generally allows for a sufficient burn in as well. In practice this was indeed observed.

The determination of K can be repeated for various parameters, or for functions of parameters in the Gibbs chain, for example, a heritability as a ratio of variance components. Different parameters or different functions of parameters may yield different K's. The approach we used is to determine K for various parameters and functions and to choose the largest K to be applied to all. Hence, the Gibbs chain can be constructed and, at every K^{th} cycle, realisations for parameters at that cycle are saved as being a "Gibbs sample" for the set of model parameters. When the same thinning for all parameters is used, the primary joint structure in the samples is retained, allowing, for example, computation of sampling correlations between parameter estimates.

Inference from marginal densities

So far we have considered sets of realisations $\boldsymbol{\theta}^{[t]}$ arising in the Gibbs chain as coherent units, being joint samples. Marginal densities of parameters are studied by observing realisations of a single parameter in these samples, irrespective of realisations for other parameters. We will focus on the genetic hyper parameters, i.e. variance components and effect and frequency of the major gene. However, nongenetic effects, polygenic effects and genotypes could be studied as well from the Gibbs chains.

A very general inference is made by visualising the marginal posterior densities in a density estimate. In this study, we supply nonparametric density estimates in the form of average shifted histograms (Scott 1992). At boundaries of parameter spaces, a reflection boundary technique (e.g. Scott 1992, p 149) was used to smooth the histogram up to the boundary. The posterior density can be summarised by one or more statistics. Straightforward cases are, approximately, symmetric densities where the mean and standard deviation are appropriate for describing the density. To describe more complicated densities, the mode is often a valuable third statistic. For symmetric densities, the mean will correspond to a maximum likelihood or maximum a-posteriori point estimate, and the standard deviation will correspond to the small sample standard error of this parameter estimate. Parameter estimates based on Gibbs samples are subject to Monte Carlo (MC) error. Because our analysis is based on nearly independent Gibbs samples, empirical MC error on the posterior mean simply can be assessed from the estimated standard deviation of the posterior density and the number of Gibbs samples generated.

In the Gibbs chain, allele effect a may appear positive as well as negative. The sign of a, however, is not relevant, being based on the arbitrary assignment of A_2A_2 as the genotype with value $+a$. From the Gibbs samples, therefore, we studied the absolute values of a. For consistency, we also studied the frequency of the favourable allele, denoted p_h. The favourable allele is A_2 when a is positive and A_1 when a is negative.

Simulated data

A population was simulated in which ten males were mated with four dams each, producing five progeny per female, yielding 200 offspring per generation. A sex was assigned to each progeny at random on a 1:1 ratio, but with the requirement of at least one male and one female in each full-sibship. For each subsequent generation, each sire was replaced by a son and each dam was replaced by a daughter. Generations were non-overlapping, and no intentional selection was practiced. Mating

Table 1 Parameter values used in simulation

Parameter	Data set 0	Data set 1
σ_e^2	100	100
σ_u^2	50	50
a	0	12.5
p_h	–	0.7
$\sigma_m^2 = 2p_1p_2a^2$	0	65.6

was at random; unintentional inbreeding could be present from the second generation onwards because of finite population size. The theoretical rate of inbreeding was approximately 0.8% per generation. The population was simulated for five generations, which resulted in a population of 1050 individuals, including the 50 base generation individuals.

For all individuals observations were simulated according to the model of analysis. This simulation included polygenic effects from the normal densities (Eq. 2a) for base animals and from Eq. 2b for non-base animals, genotypes according to probabilities from Eqs. 3a and 3b and sampling of normally distributed random errors. Two data sets were simulated, for which the genetic parameters are shown in Table 1. In data set 0 no effect at the single locus was simulated. This set was used to demonstrate results found when no single gene effect in present. In data set 1 a single gene effect was simulated with a difference (25) between extreme genotypes of approximately 2 standard deviations of the variation within single genes. The effect of the single gene in set 1 was expected to be clearly detectable. Average inbreeding coefficients in generation 5 were 4.1% in data set 1 and 4.2% in data set 2, matching the theoretical predicted rate. Numbers of individuals with non-zero inbreeding coefficients were 450 in data set 0 and 430 in data set 1. This indicates a large number of pedigree loops in these data sets already due to inbreeding alone. Multiple matings applied in this simulated breeding structure resulted in an additional large number of loops. Inbreeding was taken into account in the simulation of polygenic effects and in the analysis at steps S2 and S4. An effect of sex was simulated that favoured males by + 2 units, and sex was used in the analysis as an explanatory nongenetic effect.

Results

Mixing and the effect of blocked Gibbs sampling

For data set 1, with a simulated effect of a major gene, changes of genotypes were studied for three classes of individuals: final progeny, dams and sires. In Table 2 the average number of genotype changes per cycle is given for each class of individuals. Without blocking, virtually none of the sire-genotypes changed in the majority of the Gibbs cycles. Results were such that about once in 100 cycles, one sire-genotype was changed. Hence, the genotype configuration for sires remains practically the same over many hundreds of cycles, and movement of the markov chain is restricted to a small subspace. In this case, changes appearing for final progeny and dams are relatively meaningless because these changes are limited due to the near fixation of all sire genotypes. With the blocking technique, mixing is improved, changing about 5% of the sire-genotypes each cycle. The increased changes in sire genotypes resulted in a general increase of changes in the entire pedigree, which can be seen in particular for dam genotypes.

Table 2 Average number of genotype changes per Gibbs cycle for three groups of individuals with a scalar updating of genotypes ("scalar") and with a block updating of genotypes of sires and final progeny ("block") in data set 1 (average of 10 000 Gibbs cycles).

	Average number of changes per Gibbs cycle	
Group (total number)	Scalar	Block
Finals (800)	234 (29%)	258 (32%)
Dams (200)	12.7 (6.4%)	39.8(20%)
Sires (50)	0.008 (0.02%)	2.62 (5.2%)

Data without a major gene

Determination of a thinning parameter K for data set 0 was based on an initial run of the Gibbs chain of 10 000 cycles. Starting realisations for genetic parameters were taken as the simulated values (Table 1), which represented a pure polygenic mechanism. Allele frequency p_1 was initiated as 0.5. Thinning parameters were determined for the variance components for errors, σ_e^2, for polygenic effects, σ_u^2, and for major gene effects, $\sigma_m^2 = 2p_1p_2a^2$, for the absolute value of the effect of the allele $|a|$, and for the frequency of the favourable allele, p_h. Allele frequency p_h showed the strongest dependencies, requiring $K \approx 890$ to yield independent samples.

For data set 0, two Gibbs chains were run, each initiated with different seeds for the random number generator. From each chain, 250 Gibbs samples were obtained using $K = 890$. Results from each set of Gibbs samples are given in Table 3, which presents the es-

Table 3 Estimated means and standard deviations of posterior densities for genetic parameters in data set 0 (no major gene) in two runs of the Gibbs sampler, based on 250 samples per run

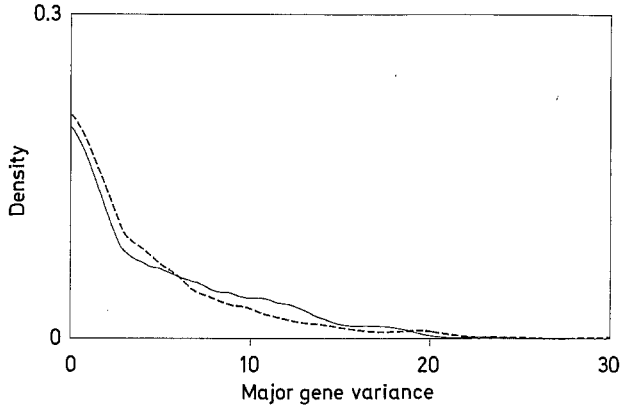
Parameter	Mean ± standard deviation	
	Run 1	Run 2
σ_e^2	98.0 ± 7.4	95.7 ± 6.8
σ_u^2	38.1 ± 14.3	48.5 ± 11.4
$\sigma_m^2 = 2p_1p_2a^2$	5.3 ± 4.9[a]	4.8 ± 5.3[a]

[a] Mode is zero

timated contributions of the two genetic components and error in terms of variances. Runs were consistent in the estimate for major gene variance. In each case, a unimodal density for major gene variance was found with a mode at zero (Fig. 1). From such densities, we infer the variance component to be zero, which means an absence of the major gene effect.

Estimates for variance components were different in the two runs, especially for polygenic variance. Posterior means for σ_u^2 differed about 10 units, which cannot be explained by Monte Carlo (MC) error. The empirical MC error on the means for σ_u^2 was estimated as 0.9 for run 1 and 0.7 for run 2. Differences in these estimates must be caused by a near reducibility of the chain, with allele frequency moving in a few subspaces between which mixing is relatively bad. When the probability of moving to a different subspace is low, this type of behaviour is unlikely to be spotted in the tuning phase, which we based on 10000 cycles only.

Fig. 1 Estimated posterior densities (averaged histogram frequencies) for major gene variance in data set 0 for run 1 (*solid line*) and run 2 (*dashed line*) of the Gibbs sampler, based on 250 samples per run



Data with a major gene

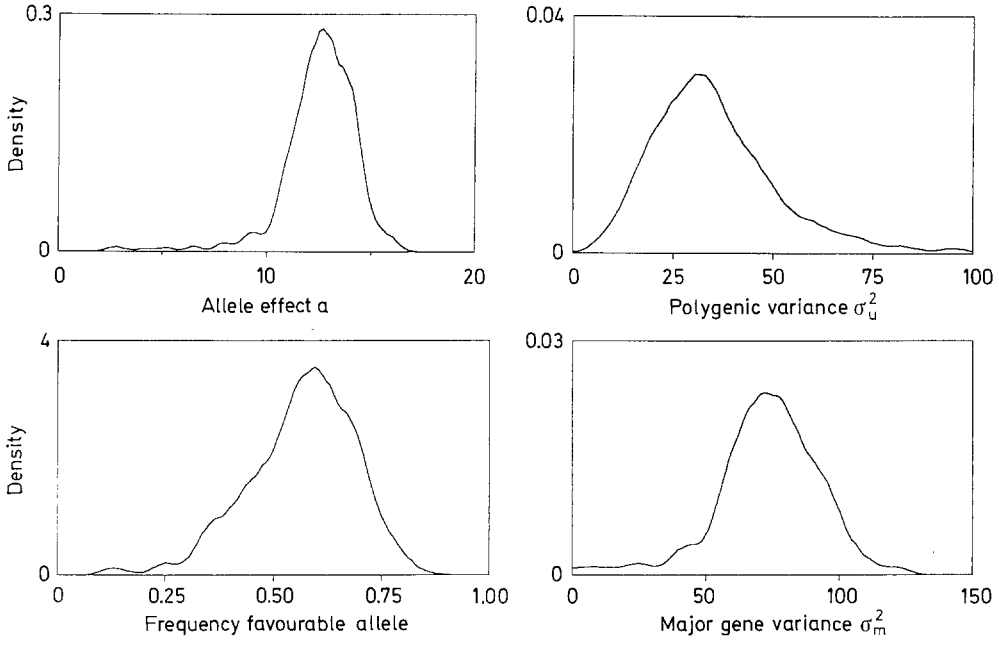
Posterior means and standard errors for parameter estimates in data set 1 are in Table 4. These estimates are based on 500 Gibbs samples from a single Gibbs chain using $K = 400$. Starting values for data set 1 represented a pure polygenic model, i.e. $\sigma_u^2 \approx 116$ and $a = 0$, which does not correspond to the simulated parameters. From this polygenic starting point, the Gibbs chain was observed to move to a mixed inheritance model in a few hundred cycles. Density estimates for σ_u^2, $|a|$, p_h and $\sigma_m^2 = 2p_1p_2a^2$ are presented in Fig. 2. The density estimate for σ_u^2 shows a unimodal density with a mode for $\sigma_u^2 > 0$, indicating the significance of the polygenic component in the model. The density estimate for σ_m^2 shows a local mode for $\sigma_m^2 = 0$ and a global mode for $\sigma_m^2 > 0$. The odds are 1:26 between the estimated density for $\sigma_m^2 = 0$ and for $\sigma_m^2 > 0$, which is taken as evidence for a significant single gene component. Neither posterior means nor modes agree perfectly with the simulated values, but in each case the simulated value was well within a 90% highest posterior density region of the estimate.

Allele frequency in data set 1 was poorly estimable, showing values in a range between 0.15 and 0.9 (Fig. 2).

Table 4 Estimated means and standard deviations of posterior densities for genetic parameters in data set 1, based on 500 Gibbs samples

Parameter	Estimate
σ_e^2	104.6 ± 10.4
σ_u^2	35.6 ± 16.6
$\|a\|$	12.5 ± 1.82
p_h	0.56 ± 0.12
$\sigma_m^2 = 2p_1p_2a^2$	73.5 ± 19.5

Fig. 2 Estimated posterior densities (averaged histogram frequencies) for genetic parameters in data set 1, based on 500 Gibbs samples



The influence of allele frequency on estimated polygenic variance and major gene variance is large. Two more analyses of data set 1 were performed, fixing the allele frequency of the favourable allele at 0.74 or at 0.60 (Table 5). The value of 0.74 was the true realised value in the simulation of data set 1, and the value of 0.60 was around the mode of the marginal posterior of allele frequency. Each value, therefore, can be taken as a plausible estimate which, based on the posterior from Fig. 2, is not dramatically different from the other. The use of an estimated value, treated as a true value without error in a further estimation step, is a procedure common to classical inference from a joint likelihood function. When allele frequency was fixed, the portion of polygenic variance in the total genetic variance ranged from 38% for $p_h = 0.74$ to 28% for $p_h = 0.60$. The MC error on the posterior means for polygenic variance is about 0.5% of the estimated total genetic variance and, therefore, is too small to account for these differences. Hence, fixing the unknown allele frequency at some value substantially affects estimates for the two genetic variances. In contrast, the "marginal" estimates (Table 4), which are averaged over all possible allele frequencies, are not affected by the arbitrary choice of a point estimate.

Discussion

Mixing in the Gibbs chain

In this study we described the construction of a Gibbs markov chain for inference in a mixed inheritance model. Efficiency of the Gibbs sampler depends on the parameterisation used and on the sampling scheme applied. A Gibbs sampling approach for a mixed inheritance model applied to human populations (Guo and Thompson 1992) is inefficient when applied to animal populations. We suggested a blockwise treatment for genotypes, which would yield faster changes in the Gibbs chain without considerable complications in computing. The blocking is typically applied to parents with large progeny groups. We applied this for sires, but the technique can also be applied to dams. Without blocking, markov chains remain stuck in a subspace of the parameter space, making a proper inference impossible. With blocking, mixing was improved, although inference in data set 0 remained difficult. Here, two Gibbs chains did not yield exactly similar results for all parameters, possibly the result of a more subtle type of bad mixing. Multiple runs of the Gibbs sampler, preferably with various starting values, can be used to spot, but not to solve, such problems of mixing. The blocking technique, therefore, is possibly only a first step to improve mixing, and more methods could be developed and added. Note further that the efficiency of blocking will depend on the data structure; in particular, on the progeny group sizes and on the allele effect at the major locus. In animal breeding practice, progeny groups are generally sufficiently large to recommend the use of blocking.

Table 5 Estimated means and standard deviations of posterior densities for genetic parameters in data set 1 fixing allele frequency at two different values, based on 500 Gibbs sampler per case

Parameter	Mean ± standard deviation using	
	$p_h = 0.74$	$p_h = 0.60$
σ_e^2	102.0 ± 10.5	104.1 ± 10.0
σ_u^2	40.1 ± 13.9	31.0 ± 12.4
$\lvert a \rvert$	13.1 ± 1.39	12.6 ± 1.25
$\sigma_m^2 = 2p_1p_2a^2$	66.7 ± 13.5	81.3 ± 15.1

Alternative uses of Gibbs chains

The efficiency in using realisations from a markov chain for statistical inference can possibly be improved. For instance, the use of independent samples is not required. Posterior means and other density features, including the density itself, can be estimated directly using serially correlated states in the chain (Geyer 1992; Wang et al. 1994). The advantages of our approach of using independent samples is that the accuracy of output from a Gibbs chain can be appreciated directly, simply by the number of samples. Independent samples also allow the comparison of output from multiple chains by standard analysis-of-variance methods. A further measure to increase the accuracy of the estimate of a mean is the use of Rao-Blackwell estimates (Gelfand and Smith 1990). This procedure uses from every state the expected value for a certain parameter rather than the realised value in the chain. Expected values are often directly available from the intermediate computations in the Gibbs chain and vary less because the disturbance from the conditional variance is eliminated.

Statistical inference

In the mixed major gene-polygenic inheritance model, maximum likelihood (ML) inference is classically employed (Elston and Stewart 1971; Morton and MacLean 1974). Gibbs sampling can also be used to obtain such ML estimates (Guo and Thompson 1992). The specification of prior densities is then circumvented by updating a parameter, for example, a variance component, not with samples from the specified densities but with the expectation for that parameter given realisations of other parameters. This technique is known as Monte Carlo EM (Tanner 1993). In our model, a REML inference could be made by omitting the sampling steps for $\sigma_e^2, \sigma_u^2, a$ and p_1, and by updating these parameters as their expectation. A ML inference could be implemented by also updating elements of $\boldsymbol{\beta}$ with their expectation. In this manner, based on the Gibbs sampler, a hierarchy of inferential methods can be obtained by suppressing

certain sampling steps in the construction of the chain. Note that when using this Monte Carlo EM technique, fluctuations in the chain will not correspond to standard errors of parameter estimates, and density estimates of posteriors cannot be made.

ML inference and associated hypothesis testing in major gene models, however, have several shortcomings. For instance, REML (Patterson and Thompson 1971) was developed to overcome biases in ML point estimates for variance components, and ML standard errors and likelihood ratio tests are based on asymptotic normal approximations. For application of the likelihood ratio test, moreover, assumed asymptotic distribution of the test statistic is questionable when dealing with mixture distributions (Titterington et al. 1985). When Gibbs sampling is used, alternatives for ML inference are available.

Inference could possibly be improved by using the Gibbs chain as implemented in this study, including the sampling steps for all genetic hyper-parameters and by making use of the marginal posterior densities of parameters obtained. This approach is generally Bayesian and, for our implementation with flat priors for all hyper-parameters, could be classified as "empirical" Bayesian. With this approach, standard errors of parameter estimates or, in general, interval estimates in any form, are directly available. Interval estimates will be based on small sample distributions and respect the natural bounds on parameter spaces. As point estimates, mean or mode of the posterior density could be used, which would be respectively marginal APE (a-posteriori expectation) or marginal MAP (maximum a-posteriori) estimates. APE is simple to compute from Gibbs chains, and APE estimates are considered more optimal than estimators based on a marginal or joint mode (Henderson 1953; Harville 1977). However, absence of a variance component shows a density with a gobal mode at zero, as we showed, in which case the posterior mode is an appealing point estimate. This favours, from a more practical point of view, use of MAP estimates. Further, when Gibbs chains are used as we presented, highly marginalised densities are used, in which for each parameter all other parameters are considered as nuisances. This provides a richer summary and may improve estimation of the two genetic components in the mixed inheritance model. For instance, we showed when fixing allele frequency, that estimates for genetic variances in the mixed inheritance model depend on the value used for allele frequency. Marginal estimates, however, take into account the error in estimating allele frequency, or any other parameter. This gives a more realistic inference, which represents better uncertainty in the estimates and provides a better disentanglement between, for example, polygenic and major gene variance.

Hypothesis testing

We did not thoroughly consider power to detect single genes or test of significance of the single-gene component. It was shown for major gene variance, σ^2_m, that absence of a single-gene effect leads to a global mode for $\sigma^2_m = 0$. As discussed, the MAP estimate would be zero in this case, correctly indicating absence of a single-gene effect. Presence of a single-gene effect showed a density with a global mode for $\sigma^2_m > 0$ and a local mode for $\sigma^2_m = 0$. We used the odds ratio of the densities at both modes as a criterion, assuming significance at a 5% level when the odds ratio is above 1:20. This criterion, however, may be very severe. An alternative would be to assume a mixed mode of inheritance as soon as the mode for $\sigma^2_m > 0$ dominates the mode for $\sigma^2_m = 0$. When experimenting with smaller effects of the major gene, a gradual increase of the density at $\sigma^2_m = 0$ was indeed observed, indicating the less likely action of a major gene. It would be of interest to further develop hypothesis testing because a test based on small sample distributions obtained from the Gibbs sampler has the potential to improve the likelihood ratio test for presence of a major gene. Gibbs sampling approaches also can handle very large data sets, as shown by Van der Lugt et al. (1994) using a polygenic model, because Gibbs sampling implementations require little memory and do not accumulate round-off errors. This facilitates the use of the generally abundant amount of information in animal populations, which is a simple measure to increase power. Additional simulations showed, for instance, that a major gene with $a = 6$ and other parameters as in data set 1, which explain about one-quarter of all genetic variance, was detected easily in a data set with 5000 individuals.

Acknowledgements Daniel Gianola and Daniel Sorensen are acknowledged for discussions on the use of improper priors for variance components. This research was supported financially by the Dutch Product Board for Livestock and Meat, the Dutch Pig Herdbook Society, Bovar, Euribrid, Fomeva and Nieuw Dalland.

References

Besag J, York J, Mollie A (1991) Bayesian image restoration with two applications in spacial statistics (with discussion). Ann Inst Stat Math 43:1–59

Cannings C, Thompson EA, Skolnick MH (1978) Probability functions on complex pedigrees. Adv Appl Prob 10:26–61

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21:523–542

Gelfand AE, Smith AFM (1990) Sampling based approaches to calculating marginal densities. J Am Stat Assoc 85:398–409

Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans Pattn Anal Mach Intell 6:721–741

Geyer CJ (1992) A practical guide to Markov chain Monte Carlo. Stat Sci 7:467–511

Guo SW, Thompson EA (1992) A Monte Carlo method for combined segregation and linkage analysis. Am J Hum Genet 51:1111–1126

Harville DA (1977) Maximum likelihood approaches to variance component estimation and related problems. J Am Stat Assoc 72:320–339

Henderson CR (1953) Estimation of variance and covariance components. Biometrics 9:226–252

Henderson CR (1988) Theoretical basis and computational methods for a number of different animal models. J Dairy Sci 71 [Suppl 2]: 1–16

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

Kinghorn BP, Kennedy BW, Smith C (1993) A method of screening for genes of major effect. Genetics 134:351–360

Knott SA, Haley CS, Thompson R (1992) Methods of segregation analysis for animal breeding data: a comparison of power. Heredity 68:299–311

Le Roy P, Elsen JM, Knott SA (1989) Comparison of four statistical methods for detection of a major gene in a progeny test design. Genet Sel Evol 21:341–357

Lin S, Thompson E, Wijsman E (1993) Achieving irreducibility of the Markov chain Monte Carlo method applied to pedigree data. IMA J Math Med Biol 10:1–17

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

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1986) Numerical recipes; the art of scientific computing. Cambridge University Press, Cambridge, Mass.

Quaas RL (1976) Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:949–953

Raftery AE, Lewis SM (1992) How many iterates in the Gibbs sampler? In: Bernardo JM, Bergen JO, David AP, Smith AFM (eds) Bayesian statistics, Oxford University Press, Oxford, pp 765–776

Scott DW (1992) Mutivariate density estimation. Wiley and Sons, New York

Sheehan N, Thomas A (1993) On the irreducibility of a Markov chain defined on a space of genotype configurations by a sampling scheme. Biometrics 49:163–175

Smith AFM, Roberts GO (1993) Bayesian computation via the Gibbs sampler and related markov chain Monte Carlo methods. J R Stat Soc Ser B 55:3–24

Sorensen D, Anderson S, Jensen J, Wang CS, Gianola D (1994) Inferences about genetic parameters using the Gibbs sampler. In: Smith C, Gavora JS, Benkel B, Chesnais J, Fairfull W, Gibson JP, Kennedy BW, Burnside EB (eds) Proc 5th World Congr Genet Appl Livest Prod, vol 18. Organizing committee, 5th World Congr Genet Appl Livest Prod Guelph, Canada, pp 321–328

Tanner MA (1993) Tools for statistical inference. Springer, Berlin Heidelberg New York

Titterington DM, Smith AFM, Makov EU (1985) Statistical analysis of finite mixture distributions. Wiley and Sons, New York

Van der Lugt AW, Janss LLG, Van Arendonk JAM (1994) Estimation of variance components in large animal models using Gibbs sampling. In: Smith C, Gavora JS, Benkel B, Chesnais J, Fairfull W, Gibson JP, Kennedy BW, Burnside EB (eds) Proc 5th World Congr Genet Appl Livest Prod, vol 18. Organizing committee, 5th World Congr Genet Appl Livest Prod, Guelph Canada, pp 329–332

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

Wang CS, Rutledge JJ, Gianola D (1994) Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. Genet Sel Evol 26:91–115