Strategies for MCMC computation inquantitative genetics

Waagepetersen, Rasmus; Ibánēz-Escriche, Noelia; Sorensen, Daniel

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

AALBORG UNIVERSITY

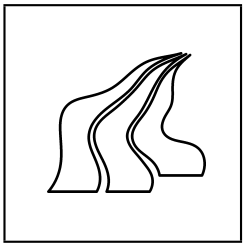
Strategies for MCMC computation in quantitative genetics

by

Rasmus Waagepetersen, Noelia Ibánẽz-Escriche and
Daniel Sorensen

R-2007-07 February 2007

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: www.math.auc.dk/research/reports/reports.htm



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Strategies for MCMC computation in quantitative genetics

Rasmus Waagepetersen*
Department of Mathematical Sciences
Aalborg University, DK-9220 Aalborg
rw@math.aau.dk

Noelia Ibánẽz-Escriche
IRTA
Avda. Rovira Roure, E-25198 Lleida

Daniel Sorensen
Department of Genetics and Biotechnology
Danish Institute of Agricultural Sciences
P.O. Box 50, DK-8830 Tjele

Abstract

In quantitative genetics, Markov chain Monte Carlo (MCMC) methods are indispensable for statistical inference in non-standard models like generalized linear models with genetic random effects or models with genetically structured variance heterogeneity. A particular challenge for MCMC applications in quantitative genetics is to obtain efficient updates of the highdimensional vectors of genetic random effects. We discuss various strategies to approach this problem including reparametrization, Langevin-Hastings updates, and updates based on normal approximations. The methods are compared in applications to Bayesian inference for three data sets using a model with genetically structured variance heterogeneity.

Key words: Langevin-Hastings, Markov chain Monte Carlo, normal approximation, proposal distributions, reparametrization.

Short title: Strategies for MCMC computation.

1 Introduction

Given observations of a trait and a pedigree for a group of animals, the basic model in quantitative genetics is a linear mixed model with genetic random effects. The correlation matrix of the genetic random effects is determined by the pedigree and is typically very high-dimensional but with a sparse inverse. Maximum likelihood inference and Bayesian inference for the linear mixed model are well-studied topics (Sorensen and Gianola, 2002). Regarding Bayesian inference, with appropriate choice of priors, the full conditional distributions are standard distributions and Gibbs sampling can be implemented relatively straightforwardly.

The assumptions of normality, linearity, and variance homogeneity are in many cases not valid. One may then consider generalized linear mixed models where the genetic random effects enter at the level of the linear predictor. San Cristobal-Gaudy *et al.* (1998) proposed another extension of the linear mixed model introducing genetic random effects influencing the log residual variances of the observations thereby producing a genetically structured variance heterogeneity. Considerable computational problems arise when abandoning the standard linear mixed model. Maximum likelihood inference is complicated since it is not possible to evaluate explicitly the likelihood function and conventional Gibbs sampling is difficult since the full conditional distributions are not anymore of standard forms.

The aim of this paper is to discuss strategies to obtain efficient Markov chain Monte Carlo (MCMC) algorithms for non-standard models of the kind mentioned in the previous paragraph. In particular we focus on the problem of constructing efficient updating schemes for the high-dimensional vectors of genetic random effects. We review the methodological background and discuss the various algorithms in the context of the heterogeneous variance model. Apart from being a model of great interest in its own right, this model has proven to be a hard test for MCMC methods. We compare the performances of the different algorithms when applied to three real datasets which differ markedly both in size and regarding the inferences concerning the genetic covariance parameters.

Section 2 discusses general strategies for obtaining efficient MCMC algorithms while Section 3 considers these strategies in the specific context of the San Cristobal-Gaudy *et al.* (1998) model. Section 4 presents results of applying two MCMC schemes to data sets with pig litter sizes, rabbit litter sizes, and snail weights. Some concluding remarks are given in Section 5.

2 MCMC strategies for high-dimensional problems

We initially discuss MCMC strategies in a rather general framework where given the vector of random effects $\mathbf{a}$ and a parameter vector $\boldsymbol{\beta}$, the vector $\mathbf{y}$ of observed traits follows some density $f(\mathbf{y}|\mathbf{a}, \boldsymbol{\beta})$. As usual in quantitative genetics, $\mathbf{a}$ is assumed to be zero mean normal with covariance matrix $\sigma_a^2\mathbf{A}$ where $\mathbf{A}$ is the additive genetic relationship matrix and σ_a^2 is the additive genetic variance. In the following we assume known σ_a^2 and $\boldsymbol{\beta}$ and focus on MCMC strategies for sampling from the posterior

$$p(\mathbf{a}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{a}, \boldsymbol{\beta})p(\mathbf{a}|\sigma_a^2)$$

where we for notational convenience omit $\boldsymbol{\beta}$ and σ_a^2 on the left hand side. An algorithm for sampling $\mathbf{a}$ can typically easily be extended with updates of the lower dimensional quantities $\boldsymbol{\beta}$ and σ_a^2 in order to sample the full posterior distribution of $(\mathbf{a}, \sigma_a^2, \boldsymbol{\beta})$. We assume that the reader has some familiarity with MCMC methods. An introduction to MCMC can be found e.g. in Sorensen and Gianola (2002).

2.1 The Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm generates a Markov chain $\mathbf{a}^1, \mathbf{a}^2, \ldots$ as follows. Given the current value $\mathbf{a}^i$ of the Markov chain, a proposal $\mathbf{a}^{\text{prop}}$ is generated from a proposal density $q(\mathbf{a}^{\text{prop}}|\mathbf{a}^i)$. With probability

$$\min\{1, \frac{p(\mathbf{a}^{\text{prop}}|\mathbf{y})q(\mathbf{a}^i|\mathbf{a}^{\text{prop}})}{p(\mathbf{a}^i|\mathbf{y})q(\mathbf{a}^{\text{prop}}|\mathbf{a}^i)}\} \tag{1}$$

the new state $\mathbf{a}^{i+1}$ is given by $\mathbf{a}^{\text{prop}}$; otherwise $\mathbf{a}^{i+1} = \mathbf{a}^i$. Under weak conditions of regularity and after a suitable 'burn-in', the generated Markov chain provides a dependent sample from the posterior distribution of $\mathbf{a}$. The question is now how to choose a suitable proposal density q.

A simple and often used proposal density is a multivariate normal density centered in the current value $\mathbf{a}^i$ of the chain and with covariance matrix $h\mathbf{I}$ where h is a user-specified proposal variance and $\mathbf{I}$ is the identity matrix, i.e. $q(\mathbf{a}^{\text{prop}}|\mathbf{a}^i)$ is the density of $N(\mathbf{a}^i, h\mathbf{I})$. The resulting Metropolis-Hastings algorithm is known as a *random-walk Metropolis* algorithm. In high-dimensional problems, the random-walk Metropolis algorithm may converge very slowly and produce highly auto-correlated samples.

A simple step forward is to use gradient information in the proposal density. The proposal distribution of a *Langevin-Hastings* algorithm is given by

$N(\mathbf{a}^i + h\nabla \log p(\mathbf{a}|\mathbf{y})/2, h\mathbf{I})$ where $\nabla \log p(\mathbf{a}|\mathbf{y})$ is the gradient of the posterior density. Intuitively, the use of gradient information helps to direct the algorithm towards regions of high posterior density. In applications in spatial statistics (Christensen and Waagepetersen, 2002) the Langevin-Hastings algorithm has proven superior to the random walk Metropolis algorithm. In the context of quantitative genetics, Langevin-Hastings has been successfully applied to implement Bayesian inference in Sorensen and Waagepetersen (2003), Ros *et al.* (2004), Ibáñez-Escriche *et al.* (2006) and Damgaard (2007).

When choosing the proposal variance h, rules of thumb suggest that one should aim at acceptance rates of about 25% for random walk and 60% for Langevin-Hastings updates. Single-site schemes where the components in $\mathbf{a}$ are updated in turn may lead to poorly mixing Markov chains due to high correlation between the components of $\mathbf{a}$.

2.2 Reparametrization

Simulation studies in Gustafson *et al.* (2004) show that Langevin-Hastings updates may not work well if the components of $\mathbf{a}$ have very different posterior variances. In applications in quantitative genetics, the individuals may contribute with different numbers of observations and may have different numbers of relatives with records. Hence the posterior variances may be very different. The correlation structure of the Langevin-Hastings proposal described in the previous section moreover typically differs markedly from the posterior correlation structure where the components are not independent. It may therefore be useful to transform $\mathbf{a}$ into a quantity whose components are less correlated a posteriori. Using the factorisation $\mathbf{A} = \mathbf{T}\mathbf{D}\mathbf{T}^\mathsf{T}$ (Henderson, 1976), one may let $\mathbf{a} = \sigma_a \boldsymbol{\gamma}\mathbf{B}^\mathsf{T}$ where $\mathbf{B} = \mathbf{D}^{1/2}\mathbf{T}$ and $\boldsymbol{\gamma}$ is a priori standard normal $N(0, \mathbf{I})$ (note that we regard vectors as row vectors). The posterior correlation matrix of $\boldsymbol{\gamma}$ given $\mathbf{y}$ is then closer to the correlation matrix $\mathbf{I}$ of the Langevin-Hastings proposal. The joint posterior of $(\boldsymbol{\gamma}, \sigma_a^2)$ given $(\mathbf{y}, \boldsymbol{\beta})$ is of the form

$$p(\boldsymbol{\gamma}, \sigma_a^2|\mathbf{y}, \boldsymbol{\beta}) \propto f(\mathbf{y}|\sigma_a\boldsymbol{\gamma}\mathbf{B}, \boldsymbol{\beta})p(\boldsymbol{\gamma})p(\sigma_a^2) \tag{2}$$

where $p(\boldsymbol{\gamma})$ and $p(\sigma_a^2)$ denote respectively the multivariate standard normal density of $\boldsymbol{\gamma}$ and the prior for σ_a^2. Given a current value $\boldsymbol{\gamma}^i$, the Langevin-Hastings proposal is of the form

$$\boldsymbol{\gamma}^{\text{prop}} = \boldsymbol{\gamma}^i(1 - h/2) + (h/2)\nabla \log p(\boldsymbol{\gamma}^i|\mathbf{y}, \sigma_a^2, \boldsymbol{\beta}) + \boldsymbol{\epsilon}^i \tag{3}$$

where $\boldsymbol{\epsilon}^i$ is $N(0, h\mathbf{I})$ distributed. Posterior samples $\mathbf{a}^i$ are straightforwardly obtained by back-transforming samples $\boldsymbol{\gamma}^i$ from (2).

It is important to note that with the reparametrized posterior (2) one looses conjugacy in the sense that the conditional distribution of σ_a^2 given $(\mathbf{y}, \boldsymbol{\gamma}, \boldsymbol{\beta})$ is not χ^{-2} when the usual χ^{-2} prior is used for σ_a^2. To maintain conjugancy one may stick to the posterior of $\mathbf{a}$ but use proposals $\mathbf{a}^{\text{prop}} = \sigma_a \boldsymbol{\gamma}^{\text{prop}} \mathbf{B}^\mathsf{T}$ obtained by transforming Langevin-Hastings proposals (3) with $\boldsymbol{\gamma}^i = \mathbf{a}^i (\sigma_a \mathbf{B}^\mathsf{T})^{-1}$.

2.3 Normal approximation of the posterior

Suppose for a moment that $q(\mathbf{a}^{\text{prop}}|\mathbf{a}^i)$ is equal to the target density $p(\mathbf{a}^{\text{prop}}|\mathbf{y})$. The Metropolis-Hastings algorithm then produces independent draws from the posterior. This indicates that an efficient Metropolis-Hastings algorithm might be obtained by constructing a proposal density which is a good approximation of the posterior density.

Consider the second-order Taylor expansion

$$\log p(\mathbf{a}^{\text{prop}}|\mathbf{y}) \approx$$
$$\log p(\hat{\mathbf{a}}|\mathbf{y}) + (\mathbf{a}^{\text{prop}} - \hat{\mathbf{a}}) \nabla \log p(\hat{\mathbf{a}}|\mathbf{y})^\mathsf{T} - \frac{1}{2}(\mathbf{a}^{\text{prop}} - \hat{\mathbf{a}}) \mathbf{H}(\hat{\mathbf{a}}) (\mathbf{a}^{\text{prop}} - \hat{\mathbf{a}})^\mathsf{T} \quad (4)$$

around a value $\hat{\mathbf{a}}$ where $\mathbf{H}(\hat{\mathbf{a}}) = \mathbf{A}^{-1}/\sigma_a^2 - \frac{\mathrm{d}^2}{\mathrm{d}\mathbf{a}^\mathsf{T}\mathrm{d}\mathbf{a}} \log f(\mathbf{y}|\mathbf{a})|_{\mathbf{a}=\hat{\mathbf{a}}}$ is minus the Hessian matrix of second derivatives. Provided $\mathbf{H}(\hat{\mathbf{a}})$ is positive definite, the exponential of the right hand side of (4) is proportional to the density

$$q(\mathbf{a}^{\text{prop}}|\mathbf{a}^i) \propto |\mathbf{H}(\hat{\mathbf{a}})|^{1/2} \exp\Big(-\frac{1}{2}\big(\mathbf{a}^{\text{prop}} - \mu(\hat{\mathbf{a}})\big) \mathbf{H}(\hat{\mathbf{a}}) \big(\mathbf{a}^{\text{prop}} - \mu(\hat{\mathbf{a}})\big)^\mathsf{T} \Big) \quad (5)$$

of a multivariate normal distribution with mean $\mu(\hat{\mathbf{a}}) = \hat{\mathbf{a}} + \nabla \log p(\hat{\mathbf{a}}|\mathbf{y}) \mathbf{H}(\hat{\mathbf{a}})^{-1}$ and precision matrix $\mathbf{H}(\hat{\mathbf{a}})$. Choosing $\hat{\mathbf{a}}$ to be the mode of the posterior, $\mu(\hat{\mathbf{a}}) = \hat{\mathbf{a}}$ and the determinant $|\mathbf{H}(\hat{\mathbf{a}})|$ conveniently cancels out when evaluating the Metropolis-Hastings ratio (1). However, if finding the mode is very time-consuming, other possibilities are to let $\hat{\mathbf{a}}$ be equal to the current value $\mathbf{a}^i$ or the result of one Newton-Raphson step starting from $\mathbf{a}^i$. Sampling from a normal approximation is discussed in Section 2.4.

2.4 Implementation of the normal approximation

Typically, the genetic random effects enter the sampling density $f(\mathbf{y}|\mathbf{a})$ via a linear predictor $\boldsymbol{\eta} = \mathbf{a}\mathbf{Z}$ where $\mathbf{Z}$ is an incidence matrix relating the observed traits to the random effects. The precision matrix $\mathbf{H}(\hat{\mathbf{a}})$ in the normal approximation proposal density (5) then takes the form

$$\mathbf{A}^{-1}/\sigma_a^2 + \mathbf{Z}\boldsymbol{\Sigma}^{-1}\mathbf{Z}^\mathsf{T}$$

where $\boldsymbol{\Sigma}^{-1} = \frac{\mathrm{d}^2}{\mathrm{d}\boldsymbol{\eta}^\mathsf{T}\mathrm{d}\boldsymbol{\eta}} \log f(\mathbf{y}|\mathbf{a})|_{\mathbf{a}=\hat{\mathbf{a}}}$. The normal approximation proposal distribution is thus formally equivalent to the conditional distribution of random effects in a 'virtual' linear normal model

$$\tilde{\mathbf{y}} = \mathbf{a}^{\text{prop}}\mathbf{Z} + \tilde{\boldsymbol{\epsilon}}$$

where $\mathbf{a}^{\text{prop}}$ is $N(0, \sigma_a^2\mathbf{A})$, $\mathrm{E}[\mathbf{a}^{\text{prop}}|\tilde{\mathbf{y}}] = \mu(\hat{\mathbf{a}})$, and $\tilde{\boldsymbol{\epsilon}} \sim N(0, \boldsymbol{\Sigma})$ and $\tilde{\mathbf{y}}$ represents 'virtual' noise and data. Hence we may sample from the normal approximation by applying the García-Cortés and Sorensen (1996) algorithm based on the decomposition

$$\mathbf{a}^{\text{prop}} = \left(\mathbf{a}^{\text{prop}} - \mathrm{E}[\mathbf{a}^{\text{prop}}|\tilde{\mathbf{y}}]\right) + \mathrm{E}[\mathbf{a}^{\text{prop}}|\tilde{\mathbf{y}}] = \mathbf{e} + \mu(\hat{\mathbf{a}})$$

where the 'prediction error' $\mathbf{e} = (\mathbf{a}^{\text{prop}} - \mathrm{E}[\mathbf{a}^{\text{prop}}|\tilde{\mathbf{y}}])$ and the 'prediction' $\mu(\hat{\mathbf{a}}) = \mathrm{E}[\mathbf{a}^{\text{prop}}|\tilde{\mathbf{y}}]$ are independent. Hence if $\mathbf{e}_{\text{sim}}$ is a simulation of $\mathbf{e}$ then

$$\mathbf{a}_{\text{sim}} = \mathbf{e}_{\text{sim}} + \mu(\hat{\mathbf{a}})$$

is a conditional simulation of $\mathbf{a}^{\text{prop}}$ given $\tilde{\mathbf{y}}$. The simulated prediction error $\mathbf{e}_{\text{sim}}$ may be generated as follows

1. simulate $(\mathbf{a}_{\text{sim}}, \tilde{\mathbf{y}}_{\text{sim}})$ from the joint distribution of $(\mathbf{a}^{\text{prop}}, \tilde{\mathbf{y}})$ (using the Henderson factorization $\mathbf{A} = \mathbf{T}\mathbf{D}\mathbf{T}^\mathsf{T}$)

2. compute $\mu(\hat{\mathbf{a}}_{\text{sim}}) = \mathrm{E}[\mathbf{a}^{\text{prop}}|\tilde{\mathbf{y}}_{\text{sim}}]$ by solving the standard mixed model equations $\mu(\hat{\mathbf{a}}_{\text{sim}})[\mathbf{A}^{-1}/\sigma_a^2 + \mathbf{Z}\boldsymbol{\Sigma}^{-1}\mathbf{Z}^\mathsf{T}] = \mathbf{y}_{\text{sim}}\boldsymbol{\Sigma}^{-1}\mathbf{Z}^\mathsf{T}$ for the virtual linear model

3. return $\mathbf{e}_{\text{sim}} = \mathbf{a}_{\text{sim}} - \mu(\hat{\mathbf{a}}_{\text{sim}})$.

Alternatively, one may exploit the sparseness of $\mathbf{H}(\hat{\mathbf{a}})$ which enables fast computation of the Cholesky factorization of $\mathbf{H}(\hat{\mathbf{a}})$, (see Rue (2001) and Rue and Knorr-Held (2005)). For the latter approach the `c` library GMRFLib (`www.math.ntnu.no/~hrue/GMRFLib/` and Appendix B in Rue and Knorr-Held, 2005) provides an extensive suite of procedures for computation of and sampling from normal approximations. Using this library, sophisticated MCMC algorithms can be constructed with little programming effort. `GMRFlib` is used in Steinsland and Jensen (2005) to implement Bayesian inference for a multiple trait model.

After sampling a proposal from the normal approximation one needs to evaluate the proposal density appearing in the Metropolis-Hastings ratio (1). The proposal density involves the determinant of the precision matrix $\mathbf{H}(\hat{\mathbf{a}})$ which cancels out if $\hat{\mathbf{a}}$ is the mode of the posterior density or another value which does not depend on the current state of the Markov chain. If $\hat{\mathbf{a}}$ is given by e.g. the current value $\mathbf{a}^i$ or the result of one Newton-Raphson step, sparse matrix methods as those used in `GMRFLib` are needed to evaluate the determinant.

2.5 Comparison of samplers in terms of Monte Carlo error and computational cost

Given an MCMC sample $(\mathbf{a}^1, \theta^1)$,$(\mathbf{a}^2, \theta^2)$,...,$(\mathbf{a}^n, \theta^n)$ from the posterior distribution of $(\mathbf{a}, \theta)$, $\theta = (\boldsymbol{\beta}, \sigma_a^2)$, and some function $h(\mathbf{a}, \theta)$, the posterior expectation

$$\mathrm{E}[h(\mathbf{a}, \theta)|\mathbf{y}] = \int h(\mathbf{a}, \theta) p(\mathbf{a}, \theta|\mathbf{y}) \mathrm{d}\mathbf{a}\mathrm{d}\theta$$

is estimated by the average $\bar{h}_n = \sum_{i=1}^n h(\mathbf{a}^i, \theta^i)/n$. The Monte Carlo variance of $\bar{h}_n$ is given by V_{asymp}/n where

$$V_{\mathrm{asymp}} = \lim_{n\to\infty} \mathrm{Var}\, \sqrt{n}\bar{h}_n = \mathrm{Var}[h(\mathbf{a})|y](1 + 2\sum_{m=1}^{\infty} \rho_m)$$

is the so-called asymptotic variance given in terms of the posterior variance $\mathrm{Var}[h(\mathbf{a})|y]$ and the Markov chain lag-m autocorrelations

$$\rho_m = \mathrm{Corr}[h(\mathbf{a}^k, \theta^k), h(\mathbf{a}^{k+m}, \theta^{k+m})].$$

To attain a given Monte Carlo variance of size V, we need a sample size of $n_V = V_{\mathrm{asymp}}/V$ and if the cost of generating one sample is c then the total cost becomes $cn_V = c\tau\, \mathrm{Var}[h(\mathbf{a})|y]/V$ where $\tau = 1 + 2\sum_{m=1}^{\infty} \rho_m$ is the integrated autocorrelation. Thus $c\tau$ is an appropriate performance measure for an MCMC algorithm. The so-called effective sample size is given by n/τ. The integrated autocorrelation can be estimated as suggested e.g. in Geyer (1992).

Note that the ratio τ_2/τ_1 of integrated autocorrelations for two MCMC samplers is equal to the ratio n_2/n_1 of numbers of iterations n_2 and n_1 required to obtain the same MCMC variance V with the two samplers.

3 A model with genetically structured variance heterogeneity

We now discuss the methods of the previous sections within the context of the San Cristobal-Gaudy *et al.* (1998) model for genetically structured variance heterogeneity. For ease of presentation, we here omit systematic and environmental effects. Let $\mathbf{a}^*$ denote random effects affecting the residual variance of $\mathbf{y}$. Given $\mathbf{a}$ and $\mathbf{a}^*$, the components y_i of $\mathbf{y}$ are independent $N(\mu + \mathbf{a}\mathbf{z}_i^\mathsf{T}, \exp(\mu^* + \mathbf{a}^*\mathbf{z}_i^\mathsf{T}))$ where $\mathbf{z}_i$ is an incidence vector with components

equal to zero or one. The joint distribution of $\mathbf{a}$ and $\mathbf{a}^*$ is zero mean normal with covariance matrix $\mathbf{G} \otimes \mathbf{A}$ where

$$\mathbf{G} = \begin{bmatrix} \sigma_a^2 & \rho\sigma_a\sigma_{a^*} \\ \rho\sigma_a\sigma_{a^*} & \sigma_{a^*}^2 \end{bmatrix}.$$

The correlation between the two types of random effects is given by ρ, and σ_a^2 and $\sigma_{a^*}^2$ are the variances for the genetic random effects.

3.1 Reparametrizations

Various reparametrization strategies are possible for this model. Letting $\mathbf{U_G}$ denote the upper triangular Cholesky factor of $\mathbf{G}$ (i.e. $\mathbf{G} = \mathbf{U_G}^\mathsf{T}\mathbf{U_G}$), $(\mathbf{a}, \mathbf{a}^*) = (\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)\mathbf{U_G} \otimes \mathbf{B}^\mathsf{T}$ where $(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$ is a standard normal vector and $\mathbf{B}$ is defined in Section 2.2. Langevin-Hastings updates for the vector of a priori uncorrelated random effects $(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$ are used in Sorensen and Waagepetersen (2003), Ros *et al.* (2004), and Ibáñez-Escriche *et al.* (2006). An alternative reparametrization is based on $(\mathbf{a}, \mathbf{a}^*) = (\mathbf{a}, \alpha\mathbf{a} + \mathbf{u})$ where $\alpha = \rho\sigma_{\tilde{a}}/\sigma_a$ and $\mathbf{u} = \mathbf{a}^* - \mathrm{E}[\mathbf{a}^*|\mathbf{a}]$. Then $\mathbf{u}$ is $N(0, \sigma_{\tilde{a}}^2(1-\rho^2)\mathbf{A})$ and a priori independent of $\mathbf{a}$. Hence one might update $\mathbf{a}$ and $\mathbf{u}$ in turn hoping that these quantities are only weakly correlated a posteriori. Note that it is not guaranteed that the Hessian matrix with respect to $\mathbf{u}$ is positive definite.

3.2 Normal approximations

Let $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_i^2)$ and $\mathbf{R} = \mathrm{diag}(r_i)$ denote diagonal matrices where $\sigma_i^2 = \exp(\mu^* + \mathbf{a}^*\mathbf{z}_i^\mathsf{T})$ is the conditional variance given $\mathbf{a}^*$ for the ith observation and $r_i = (y_i - \mu - \mathbf{a}\mathbf{z}_i^\mathsf{T})$ is the ith residual. The precision matrix in the normal approximation of the posterior for $(\mathbf{a}, \mathbf{a}^*)$ is then

$$\begin{bmatrix} \mathbf{Z}\boldsymbol{\Sigma}^{-1}\mathbf{Z}^\mathsf{T} + \mathbf{A}^{-1}g^{11} & \mathbf{Z}\boldsymbol{\Sigma}^{-1}\mathbf{W}^\mathsf{T} + \mathbf{A}^{-1}g^{12} \\ \mathbf{W}\boldsymbol{\Sigma}^{-1}\mathbf{Z}^\mathsf{T} + \mathbf{A}^{-1}g^{12} & \frac{1}{2}\mathbf{W}\boldsymbol{\Sigma}^{-1}\mathbf{W}^\mathsf{T} + \mathbf{A}^{-1}g^{22} \end{bmatrix} \tag{6}$$

where $\mathbf{W} = \mathbf{ZR}$ and g^{ij} are the entries of $\mathbf{G}^{-1}$. Due to the factor $1/2$ in the lower right block, this matrix cannot be recognized as the covariance matrix of a conditional normal distribution and there is in fact no guarantee that it is positive definite. This is further illustrated in the toy example in Section 3.3 which shows that the joint posterior of $(\mathbf{a}, \mathbf{a}^*)$ can be far from multivariate normal. In Section 4 we use the normal approximation in turn for $\mathbf{a}$ and $\mathbf{a}^*$ separately. In this case the covariance matrices are given by the diagonal blocks in (6).

3.3 Toy example

We illustrate the various reparametrization strategies in the very simple case where $\mathbf{y} = (-2.62, -2.42)$ consists of two observations, and $\mathbf{a}$ and $\mathbf{a}^*$ are one-dimensional. Figure 1 shows the posterior densities of $(\mathbf{a}, \mathbf{a}^*)$, $(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$ and $(\mathbf{a}, \mathbf{u})$ in the case where $\mu = 0$, $\mu^* = -1$, $\sigma_a^2 = 1$, $\sigma_{a^*}^2 = 0.25$, and $\rho = 0.75$. The plots demonstrate for the given parameter settings that $(\mathbf{a}, \mathbf{a}^*)$ are highly correlated a posteriori and that the joint posterior distribution of $(\mathbf{a}, \mathbf{a}^*)$ is not well approximated by a normal distribution. The transformed random effects $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^*$ seem approximately uncorrelated but have different posterior variances. As expected, $\mathbf{a}$ and $\mathbf{u}$ are less correlated a posteriori than $\mathbf{a}$ and $\mathbf{a}^*$ but the joint distribution is far from normal.

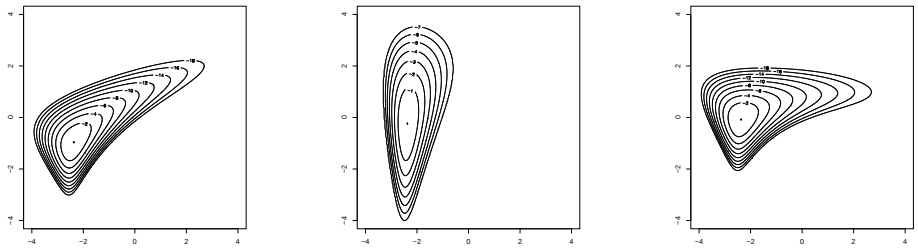


Figure 1: Posterior densities for $(\mathbf{a}, \mathbf{a}^*)$ (left), $(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$ (middle), and $(\mathbf{a}, \mathbf{u})$ (right). Note that the modal value of the posterior density is subtracted in all four plots and that contour curves are omitted for very low values of the posterior densities.

The plots in Figure 2 illustrate the random walk and Langevin-Hastings proposals. The Langevin-Hastings proposal mean lies in a region of higher posterior density than the current value. This means that proposals in a rather large region around the proposal mean have a good chance of being accepted. Figure 3 shows that the normal approximation is poor for the joint posterior of $(\mathbf{a}, \mathbf{a}^*)$ while it works well for $\mathbf{a}$ and $\mathbf{a}^*$ separately.

4 Examples

In this section we compare the performance of Langevin-Hastings and normal approximation MCMC algorithms applied to three data sets which have been previously analyzed in Sorensen and Waagepetersen (2003), Ros *et al.* (2004), and Ibáñez-Escriche *et al.* (2006). The first data set originates from a selection experiment for pig litter size and contains 10,060 litter size records

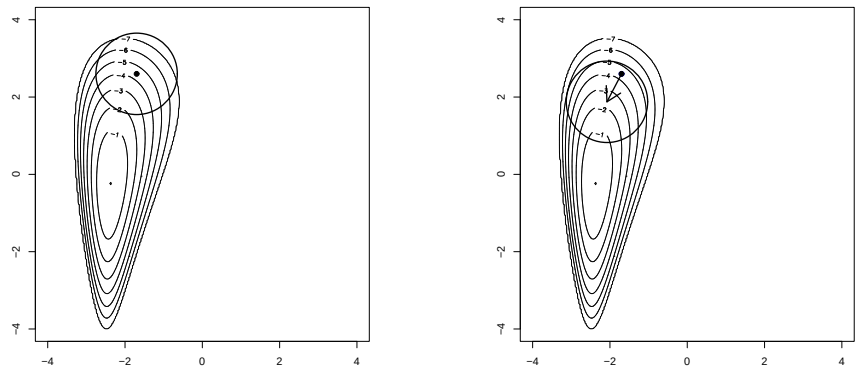


Figure 2: Illustration of random walk and Langevin-Hastings for $(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$ with $h = 0.4$. In both plots, the bold dot represents a current state $\mathbf{a}^i$ and 75 % of the proposals fall within the circle. In the right plot the proposal mean is given by the current value plus $h\nabla \log p(\mathbf{a}^i|\mathbf{y})/2$ indicated by the arrow.

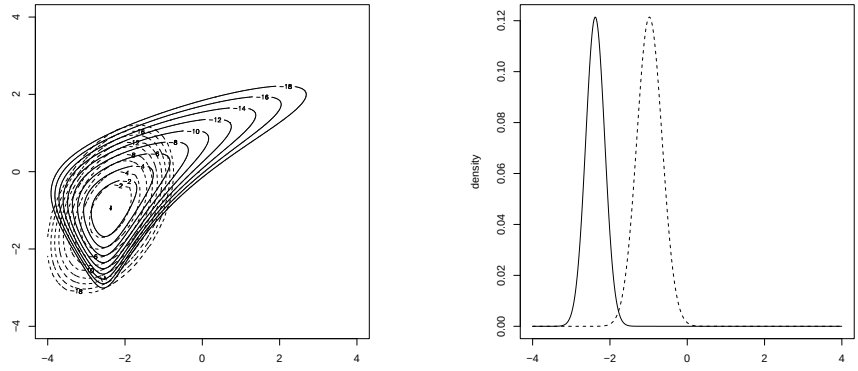


Figure 3: Illustration of normal approximation (dashed contours) evaluated at the mode for the joint posterior (solid contours) of $(\mathbf{a}, \mathbf{a}^*)$ (left) and for the conditional distributions of $\mathbf{a}|\mathbf{a}^*, \mathbf{y}$ (solid curve) and $\mathbf{a}^*|\mathbf{a}, \mathbf{y}$ (dashed curve).

from 4,149 sows. The pedigree file includes 6,437 individuals. The second data set contains 2,996 litter sizes from a divergent selection experiment for rabbit uterine capacity with 1,161 individuals in the pedigree. The third and largest data set consists of weights for each of 22,033 adult snails and the pedigree file includes 22,454 individuals. For all three datasets we consider

the model from Section 3 extended with systematic and environmental effects and impose prior distributions on the unknown location and variance parameters. The posterior means of the genetic covariance parameters are given in the last three columns of Table 2. The genetic correlation parameter ρ is negative for the pig and rabbit litter size data while it is positive for the snail weights. More details on the data, priors, and posterior results can be found in Sorensen and Waagepetersen (2003), Ros *et al.* (2004) and Ibáñez-Escriche *et al.* (2006).

The first algorithm LH is the one employed in Sorensen and Waagepetersen (2003), Ros *et al.* (2004) and Ibáñez-Escriche *et al.* (2006) where the components of the posterior distribution are updated in turn using Langevin-Hastings updates for the reparametrized random effects $(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$ and either Langevin-Hastings or random walk updates for the other components. The second algorithm NX is as LH except that the Langevin-Hastings update for $(\boldsymbol{\gamma}, \boldsymbol{\gamma}^*)$ is replaced with normal approximation updates for $\mathbf{a}$ and $\mathbf{a}^*$ separately. The normal approximation updates are implemented using `GMRFLib`. The proposal variances for the Langevin-Hastings and random walk updates are chosen according to the rules of thumb mentioned in Section 2.1.

Table 1 shows ratios of integrated autocorrelations obtained using respectively LH and NX. The integrated autocorrelations are evaluated for quadratic forms involving $\mathbf{a}$ and $\mathbf{a}^*$ and the first two components of $\mathbf{a}$ and $\mathbf{a}^*$. Considering e.g. the random effect a_1 for the pigs data, the integrated autocorrelation is 818 times larger for LH than for NX. This means that 818 times more iterations are needed with LH to obtain the same precision as for the NX algorithm (see Section 2.5 for details concerning integrated autocorrelation and MCMC precision). Columns 2-4 in Table 2 show ratios of integrated autocorrelations for the genetic covariance parameters using respectively LH and NX. Regarding integrated autocorrelation, NX clearly outperforms LH. However, the improvement is smaller for the genetic covariance parameters. The reason is that with LH and reparametrized genetic random effects, we can use 3-5 times larger proposal standard deviations than for NX in the random walk updates for the genetic covariance parameters while still maintaining acceptance rates around 25%.

Data	$\mathbf{a}\mathbf{A}^{-1}\mathbf{a}^\mathsf{T}$	$\mathbf{a}\mathbf{A}^{-1}\mathbf{a}^{*\mathsf{T}}$	$\mathbf{a}^*\mathbf{A}^{-1}\mathbf{a}^{*\mathsf{T}}$	a_1	$\tilde{a}_1$
Rabbits	44	44	38	107	111
Pigs	102	69	24	818	529
Snails	377	470	212	359	172

Table 1: Ratios of integrated autocorrelations (LH/NX) for quadratic forms and two genetic random effects using LH and NX.

To evaluate the performance of the algorithms, computing time must be taken into account. Column 5 in Table 2 shows that the computing time for one MCMC iteration is between 20 to 100 times higher for the NX algorithm than for the LH algorithm. Hence, based on the product of integrated autocorrelations and computing time, the LH algorithm seems preferable for the pigs data while NX seems superior for the rabbits and snails data.

Data	σ_a^2	$\sigma_{a^*}^2$	ρ	c	$E[\sigma_a^2\|\mathbf{y}]$	$E[\sigma_{a^*}^2\|\mathbf{y}]$	$E[\rho\|\mathbf{y}]$
Rabbits	47	14	18	20	0.82	0.16	-0.74
Pigs	56	5	13	100	1.62	0.10	-0.62
Snails	194	13	31	35	1.71	0.29	0.81

Table 2: Column 2-4: Ratios of integrated autocorrelations (LH/NX) for genetic variance parameters using LH and NX. Column c contains the ratios of computing times (NX/LH) for one MCMC iteration. Last three columns contain posterior means of the genetic covariance parameters

For the data set of snail weights with high dimensional $\mathbf{a}^*$, the acceptance rate for the normal approximation updates of $\mathbf{a}^*$ is quite small – around 5% – while larger acceptance rates 50% and 30 % are obtained for the smaller rabbits and pigs data sets. If the MCMC algorithm is initialized in values far from the posterior mode, the acceptance probability for the normal approximation update of $\mathbf{a}^*$ may be very small in which case a large burn-in is needed.

We also tried out normal approximations for the $(\mathbf{a}, \mathbf{u})$ reparametrization but this did not offer any improvement.

5 Discussion

Normal approximation proposal distributions are intuitively appealing and provide smaller integrated autocorrelations than Langevin-Hastings updates in the examples in Section 4. A distinct advantage of the normal approximation updates is moreover that they do not require user tuning of proposal variances. However, normal approximation is not always optimal when taking into account computing time. The computing time is on the other hand sensitive to the choice of implementation. The normal approximation updates used in Section 4 are implemented using general routines in `GMRFLib` based on numerical methods for sparse matrices. This approach reduces very much the programming effort but one looses the computational advantages offered by the specific structure of the genetic correlation matrix. For the

rabbits data, Ibáñez-Escriche (2006) reduces the computing cost of the NX algorithm by a factor three using the approach described in Section 2.4 where samples from the normal approximation is obtained using the García-Cortés and Sorensen (1996) algorithm and the Henderson factorization.

The results from the present study confirm that it is difficult to suggest globally optimal MCMC strategies. Relative efficiency is not only dependent on the structure of the data, as shown here, but also on the values of the parameters of the model, as demonstrated by Shariati and Sorensen (2007) in a simulation study. Our advice to the practitioner is try out both simple MCMC strategies like Langevin-Hastings and more sophisticated methods like normal approximation - and perhaps mix the different strategies. Mixing Langevin-Hastings and normal approximation updates would e.g. resolve the problem with a potentially long burn-in for normal approximation algorithms mentioned in Section 4 .

References

Christensen, O. F. & Waagepetersen, R. P. (2002). Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics* **58**, 280–286.

Damgaard, L. H. (2007). Cox proportional hazards model with time varying additive genetic effects: a case study on sow longevity. In preparation.

García-Cortés, L. A. & Sorensen, D. (1996). On a multivariate implementation of the Gibbs sampler. *Genetics, Selection, Evolution* **28**, 121–126.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical Science* **7**, 473–511.

Gustafson, P., Macnab, Y. C. & Wen, S. (2004). On the value of derivative evaluations and random walk suppression in Markov chain Monte Carlo algorithms. *Statistics and Computing* **14**, 23–38.

Henderson, C. R. (1976). A simple method for the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**, 69–83.

Ibáñez-Escriche, N. (2006). *Models for residual variances in quantitative genetics. An application to rabbit uterine capacity*. Ph.D. thesis, Departamento de Ciencia Animal, Universidad Politécnica de Valencia, Valencia.

Ibáñez-Escriche, N., Sorensen, D., Waagepetersen, R. & Blasco, A. (2006). A study of canalization and response to selection for uterine capacity in rabbits. Submitted.

Ros, M., Sorensen, D., Waagepetersen, R., Dupont-Nivet, M., SanCristobal, M., Bonnet, J.-C. & Mallard, J. (2004). Evidence for genetic control of adult weight plasticity in the snail helix aspersa. *Genetics* **168**, 2089–2097.

Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society B* **63**, 325–338.

Rue, H. & Knorr-Held, L. (2005). *Gaussian Markov random fields - theory and applications*. Chapman & Hall/CRC.

San Cristobal-Gaudy, M., Elsen, J. M., Bodin, L. & Chevalet, C. (1998). Prediction of the response to a selection for canalisation of a continuous trait in animal breeding. *Genetics, Selection, Evolution* **30**, 423–451.

Shariati, M. & Sorensen, D. (2007). Efficiency of alternative McMC strategies using the reaction norm model. Submitted.

Sorensen, D. & Gianola, D. (2002). *Likelihood, Bayesian, and Markov chain Monte Carlo methods in quantitative genetics*. Springer-Verlag.

Sorensen, D. & Waagepetersen, R. (2003). Normal linear models with genetically structured variance heterogeneity: a case study. *Genetical Research* **82**, 207–222.

Steinsland, I. & Jensen, H. (2005). Making inference from Bayesian animal models utilising Gaussian Markov random field properties. *Statistics Preprint 10*, Norwegian University of Science and Technology.