# MCMC STRATEGIES FOR A BAYESIAN ANALYSIS OF REACTION NORM MODELS WITH UNKNOWN COVARIATES

**M. M. Shariati and D. Sorensen**

Department of Genetics and Biotechnology, Danish Institute of Agricultural Sciences
PB 50, DK-8830 Tjele, Denmark

## INTRODUCTION

Recently, Su *et al*. 2006 proposed a Bayesian-McMC implementation for the analysis of reaction norm models with unknown covariates. Because conditional posterior distributions have standard forms, Gibbs sampling is an obvious and simple McMC strategy to use. Typically genetic random effects (breeding values for level and slope of each individual in the pedigree) are highly correlated in their posterior distributions. As a consequence, single-site Gibbs updates can lead to mixing problems. Possible remedies are to reparameterize (Gelfand *et al.,* 1995) and/or to sample the highly dimensional parameters jointly (Liu *et al.,* 1994). Relatively little work has been done to compare the benefits of alternative McMC strategies (but see Gustafson, 2004). The present study is aimed at illustrating the implementation of five McMC strategies for fitting the Bayesian reaction norm model to simulated data. The five strategies are single-site Gibbs updates (SG), pairwise (within individual) Gibbs updates (PG), blocked (all location parameters updated jointly) Gibbs updates (BG), Langevin-Hasting proposal for updating genetic random effects (LH), and finally Langevin-Hasting proposal for updating the transformed genetic random effects (TLH).

## MATERIALS AND METHODS

**The model.** The sampling model for the data **y** (order *n*) is of the form

$$\mathbf{y} \mid \mu, \mathbf{h}, \mathbf{a}, \mathbf{b}, \sigma_e^2 \sim N\left(\mathbf{1}\mu + \mathbf{E}\mathbf{h} + \mathbf{Z}\mathbf{a} + \mathbf{Z}_h\mathbf{b}, \mathbf{I}\sigma_e^2\right)$$

where $\mu$ is the general mean, **h** is the vector of environmental values or herd-year effects, **a** and **b** are vectors of genetic levels and slopes and $\sigma_e^2$ is the residual variance, which is assumed to have a scaled inverted Chi-square prior distribution. **I** is the identity matrix, **E** and **Z** are known incidence matrices and $\mathbf{Z}_h$ is an incident matrix containing unknown herd-year effects. The two vectors of genetic effects are normally distributed a priori; that is

$$\left.\begin{matrix}\mathbf{a}\\ \mathbf{b}\end{matrix}\right| \mathbf{G} \sim N\left(\mathbf{0}, \mathbf{G} \otimes \mathbf{A}\right)$$

where $\mathbf{G}_{2\times 2}$ is the covariance matrix of genetic level and slope and is assumed to have an inverse Wishart distribution, a priori. **A** is the additive genetic relationship matrix. The prior distribution of herd-year effects and variance of herd year effects are assumed to be normal and scaled inverted Chi-squared, respectively. In the transformed model (see below), all variance components are assumed to have scaled inverted chi-square distributions a priori, and the correlation coefficient is a priori uniformly distributed in [-1,1].

**Posterior distribution and brief description of algorithms.** The joint posterior distribution in SG, PG, BG and LH strategies is written as

$$p\left(\mu, \mathbf{h}, \mathbf{a}, \mathbf{b}, \mathbf{G}, \sigma_h^2, \sigma_e^2 \,|\, \mathbf{y}\right)$$
$$\propto p\left(\mathbf{y} \mid \mu, \mathbf{h}, \mathbf{a}, \mathbf{b}, \sigma_e^2\right) p\left(\mathbf{a}, \mathbf{b} \,|\, \mathbf{G}\right) p\left(\mathbf{h} \,|\, \sigma_h^2\right) p\left(\mathbf{G}\right) p\left(\sigma_h^2\right) p\left(\sigma_e^2\right).$$

In the SG strategy, elements of (**a**, **b**) are sampled individually; in the PG strategy the genetic level *a* and the slope *b* of each individual are sampled from their bivariate normal fully conditional distribution. In the BG strategy, the vector of location parameters $\boldsymbol{\theta} = (\mu, \mathbf{a}, \mathbf{b})$ is sampled in one step (Garcia-Cortes and Sorensen, 1996; Sorensen and Gianola, 2002). In the LH and TLH strategies, all elements of the genetic random effects are sampled and updated in one pass. In the TLH algorithm, the genetic random vector (**a**, **b**) is transformed in its prior distribution into independent standard normal variables as follows:

$$\begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix} = \mathbf{L}^{-1} \otimes \mathbf{D}^{-0.5}\mathbf{T}^{-1} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}$$

where $\mathbf{G} = \mathbf{LL}'$ and $\mathbf{A} = \mathbf{TDT}'$ (*i.e.* Henderson, 1976). The joint posterior distribution of the parameters after transformation is:

$$p\left(\mu, \mathbf{h}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \sigma_a^2, \sigma_b^2, \rho, \sigma_h^2, \sigma_e^2 \,|\, \mathbf{y}\right)$$
$$\propto p\left(\mathbf{y} \mid \mu, \mathbf{h}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \sigma_a^2, \sigma_b^2, \rho, \sigma_e^2\right) p\left(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2\right) p\left(\mathbf{h} \,|\, \sigma_h^2\right) p\left(\sigma_a^2\right) p\left(\sigma_b^2\right) p\left(\sigma_h^2\right) p\left(\sigma_e^2\right).$$

The idea here is that the elements of the transformed random variables will be less correlated a posteriori and that this would induce better mixing of the chain.

**Simulation.** Two sets of data were simulated, one with higher proportion of G×E interaction variance (18% of the total phenotypic variance) and one with a smaller proportion of G×E interaction variance (6% of the total phenotypic variance). The variance of G×E interaction is defined as Var($bh$) = $\sigma_b^2\,\sigma_h^2$ (Su *et al*., 2006). For both sets of data the following values were used: $\mu = 10$, $\sigma_b^2 = 1$, $\rho = 0.3$ and $\sigma_h^2 = 10$. In dataset 1 (higher interaction) it was assumed that $\sigma_a^2 = 6$ and $\sigma_e^2 = 30$ whereas in dataset 2 (lower interaction), $\sigma_a^2 = 24$ and $\sigma_e^2 = 120$. The phenotypic variances across herd-years of the datasets 1 and 2 were 56 and 164 respectively. Two generations were simulated; the first generation, which did not contain records, consisted of 1000 sires and 4000 dams. Each sire mated with 4 dams and each dam produced 2 offspring leading to 8000 animals with records in generation 2. The number of herds in both sets of data was 50 and each sire had progeny with records in 2 or 3 herds.

**Evaluation criterion.** Suppose that realizations $X_1$, $X_2$ ...$X_n$ are simulated from a target distribution ($X_i$ is the $i$th draw of the vector of parameters $X$ of the model) using a particular algorithm and that one computes an estimate of the mean of some function $g$ of the parameters

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g\left(X_i\right).$$

Then $\sqrt{n}(\hat{\mu} - \mu)$ converges to $N(0, \sigma^2)$ and $\sigma / \sqrt{n}$ is the Monte Carlo error associated with the estimator $\hat{\mu}$. The asymptotic variance $\sigma^2$, which is peculiar for a given algorithm, can be estimated using methods proposed in (Geyer, 1992). Here we use what Geyer (1992) calls the initial positive sequence estimator. Suppose that $n$ samples are generated using an algorithm with a cost per sample (CPU-time per sample, for example) equal to $m$, and variance $\sigma^2$. For a

Monte Carlo error $\sigma / \sqrt{n}$ equal to $k$, it can be shown that the total cost is $m\sigma^2 / k^2$. Then for a given value of $k$, the performance of an algorithm can be measured in terms of $m\sigma^2$.

**RESULTS AND DISCUSSION**

Although the product $m\sigma^2$ is the relevant quantity to use as a measure of the performance of an algorithm, we have chosen to present values of $m$ and of $\sigma^2$ separately. Estimates of the asymptotic variance $\sigma^2$ of chosen parameters for the five algorithms are shown in tables 1 and 2. Table 3 shows computing costs $m$ in terms of CPU-time. For dataset 1, BG leads to the smallest values of $\sigma^2$ over the whole set of parameters studied. This is followed by PG, SG, TLH and LH. The transformation of (**a**, **b**) into *iid* variables in the prior has a very significant positive effect in terms of $\sigma^2$. However, TLH cannot compete with any of the versions of the Gibbs sampler. The results are a little different for dataset 2; BG performs well but does not lead to the smallest asymptotic variances across the whole range of parameters. When computing cost is taken into account, the picture is less transparent. For variance components and the correlation coefficient, SG is the strategy that performs best, followed by BG and PG, that do not differ by much. However BG outperforms the other strategies in the case of the genetic random effects of the model. This holds for both datasets. The two versions of the Langevin-Hastings algorithm are outperformed by the Gibbs samplers in terms of the product $m\sigma^2$.

**Table 1. Estimates of $\sigma^2$ ($\times 10^{-1}$) of chosen parameters**

| | Strategy | $\mu$ | $\sigma_e^2$ | $\sigma_a^2$ | $\sigma_b^2$ | $\rho$ | $\sigma_h^2$ | $h_1$ | $h_2$ |
|---|---|---|---|---|---|---|---|---|---|
| High G×E interaction | SG | 32.7 | 58.2 | 109 | 0.42 | 5.79 | 15.0 | 33.4 | 35.2 |
| | PG | 26.4 | 47.9 | 77.5 | 0.48 | 5.02 | 11.9 | 15.7 | 29.5 |
| | BG | 15.1 | 39.8 | 73.9 | 0.30 | 2.86 | 10.7 | 15.1 | 15.5 |
| | LH | 1190 | 1510 | 2490 | 19.5 | 207 | 545 | 964 | 1310 |
| | TLH | 70.2 | 138 | 208 | 3.21 | 11.1 | 64.4 | 65.6 | 88.4 |
| Low G×E interaction | SG | 8.36 | 1010 | 916 | 4.27 | 1.66 | 14.6 | 11.1 | 18.0 |
| | PG | 8.74 | 864 | 928 | 4.48 | 1.63 | 15.1 | 11.0 | 16.6 |
| | BG | 4.80 | 1300 | 985 | 7.52 | 0.74 | 11.1 | 6.43 | 11.1 |
| | LH | 168 | 53910 | 54600 | 53.1 | 22.6 | 604 | 159 | 293 |
| | TLH | 29.5 | 2030 | 2650 | 22.3 | 2.28 | 47.0 | 35.6 | 38.6 |

**Table 2. Estimates of $\sigma^2$ of three chosen additive genetic values and slopes**

| | Strategy | $a_{s1}$ | $a_{p1}$ | $a_{p2}$ | $b_{s1}$ | $b_{p1}$ | $b_{p2}$ |
|---|---|---|---|---|---|---|---|
| High G×E interaction | SG | 38.5 | 29.5 | 25.1 | 6.97 | 4.94 | 4.87 |
| | PG | 31.8 | 23.7 | 19.2 | 6.53 | 4.94 | 4.39 |
| | BG | 8.55 | 14.2 | 5.42 | 0.91 | 1.05 | 1.05 |
| | LH | 8088 | 6510 | 6126 | 1542 | 2259 | 1971 |
| | TLH | 870 | 887 | 833 | 172 | 173 | 220 |
| Low G×E interaction | SG | 178 | 126 | 132 | 5.60 | 3.97 | 3.96 |
| | PG | 131 | 84.9 | 83.7 | 4.13 | 2.72 | 2.71 |
| | BG | 50.6 | 19.4 | 23.5 | 0.61 | 0.62 | 0.61 |
| | LH | 21018 | 24900 | 17100 | 534 | 489 | 552 |
| | TLH | 4591 | 4987 | 5568 | 106 | 122 | 126 |

**Tabel 3. Computing time (in seconds) required for each iteration on a pc with 3 Ghz cpu speed**

| | SG | PG | BG | LH | TLH |
|---|---|---|---|---|---|
| Computing time | 0.094 | 0.186 | 0.265 | 0.330 | 0.306 |

**CONCLUSION**

The results of the present study confirm that it may be extremely hard to provide advice to the practitioner concerning the choice of a given McMC strategy holding broadly across a range of target distributions. McMC has revolutionized Bayesian analysis but its efficient use requires a considerable degree of tuning and experimentation.

**REFERENCES**

Garcia-Cortes, L. A., and Sorensen, D. (1996) *Genet. Sel. Evol*. **28:** 121-126.
Gelfand, A., Sahu, S.K. and Carlin, B.P. (1995) *Biometrika* **82:** 479-488.
Geyer, C. J. (1992) *Statistical Science* **7:** 473-511.
Gustafson, P., Macnab, Y. and Wen, S. (2004) *Statistics and Computing* **14:** 23-38.
Henderson, C. R. (1976) *Biometrics* **32 :** 69-83.
Liu, J. S., Wong, W.H., and Kong, A. (1994) *Biometrika* **81:** 27-40.
Sorensen, D. and Gianola, D. (2002) «Likelihood, Bayesian and MCMC Methods in Quantitative Genetics». Springer-Verlag, New York, NY.
Su, G., Madsen, P., Lund, M.S., Sorensen, D., Korsgaard, I.R. and Jensen, J. (2006) *J. Anim. Sci.* In press