

On a multivariate implementation of the Gibbs sampler

LA García-Cortés, D Sorensen *

*National Institute of Animal Science, Research Center Foulum, PB 39,
DK-8830 Tjele, Denmark*

(Received 2 August 1995; accepted 30 October 1995)

Summary – It is well established that when the parameters in a model are correlated, the rate of convergence of Gibbs chains to the appropriate stationary distributions is faster and Monte-Carlo variances of features of these distributions are lower for a given chain length, when the Gibbs sampler is implemented by blocking the correlated parameters and sampling from the respective conditional posterior distributions takes place in a multivariate rather than in a scalar fashion. This block sampling strategy often requires knowledge of the inverse of large matrices. In this note a block sampling strategy is implemented which circumvents the use of these inverses. The algorithm applies in the context of the Gaussian model and is illustrated with a small simulated data set.

Gibbs sampling / block sampling / Bayesian analysis

Résumé – Une mise en œuvre multivariante de l'échantillonnage de Gibbs. Il est bien établi que, lorsque les paramètres d'un modèle sont corrélés, l'estimation de ces paramètres par échantillonnage de Gibbs converge lentement lorsque les composantes du modèle sont traitées séparément. Mais, si l'échantillonnage de Gibbs est conduit en fixant des valeurs pour les paramètres corrélés et en échantillonnant dans les distributions conditionnelles respectives, la convergence est plus rapide et les variances de Monte-Carlo des caractéristiques des distributions sont diminuées pour une chaîne de longueur donnée. Cet échantillonnage multidimensionnel et non plus scalaire requiert souvent l'inversion de matrices de grande taille. Cette note présente une méthode d'échantillonnage en bloc de ce type qui évite le passage par ces inverses. L'algorithme s'applique dans le contexte d'un modèle gaussien et est illustré numériquement sur un petit échantillon de données simulées.

échantillonnage de Gibbs / échantillonnage en bloc / analyse bayésienne

* Correspondence and reprints

INTRODUCTION

The Gibbs sampler is a numerical technique that has received considerable attention in statistics and animal breeding. It is particularly useful in the solution of high dimensional integrations and has therefore been applied in likelihood and Bayesian inference problems in a wide variety of models.

The Gibbs sampler produces realizations from a joint posterior distribution by sampling repeatedly from the full conditional posterior distributions of the parameters of the model. In theory, drawing from the joint posterior density takes place only in the limit, as the number of drawings becomes infinite. The study of convergence to the appropriate distributions is still an active area of research, but it is well established that convergence can be slow when highly correlated scalar components are treated individually (Smith and Roberts, 1993). In such cases it is preferable to block the scalars and to perform sampling from multivariate conditional distributions. Liu et al (1994) show that a block sampling strategy can lead to considerably smaller Monte-Carlo variances of estimates of features of posterior distributions.

In animal breeding applications this block sampling strategy can be difficult to implement because repeated inversions and factorizations of very large matrices are required in order to perform the multivariate sampling. The purpose of this note is to describe a block sampling computer strategy which does not require knowledge of the inverse of these matrices. The results of applying the method are illustrated using a small simulated data set.

THE MODEL

Let \mathbf{y} , \mathbf{a} and \mathbf{b} represent vectors of data (order n), of additive genetic values (order q) and of fixed effects (order p), respectively, and let \mathbf{X} and \mathbf{Z} be design matrices associating the data with the additive genetic values and fixed effects, respectively. We will assume that the data are conditionally normally distributed, that is:

$$\mathbf{y}|\mathbf{a}, \mathbf{b}, \sigma_e^2 \sim N(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a}, \mathbf{I}\sigma_e^2) \quad [1]$$

where σ_e^2 is the residual variance. Invoking an infinitesimal additive genetic model, the distribution of additive genetic values is also normal:

$$\mathbf{a}|\sigma_a^2 \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2) \quad [2]$$

where \mathbf{A} is the known numerator relationship matrix and σ_a^2 is the additive genetic variance. For illustration purposes we will assume that the vector of fixed effects \mathbf{b} , and of the variance components σ_a^2 and σ_e^2 are all a priori independent and that they follow proper uniform distributions. Under the model, the joint posterior distribution of the parameters is:

$$p(\mathbf{b}, \mathbf{a}, \sigma_a^2, \sigma_e^2|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{a}, \mathbf{b}, \sigma_e^2) p(\mathbf{a}|\sigma_a^2) \quad [3]$$

The scalar implementation of the Gibbs sampler consists of deriving from [3] the full conditional posterior distributions of the scalar parameters pertaining to the model (eg, Gelfand et al, 1990; Wang et al, 1994).

METHODS

Let $\mathbf{W} = (\mathbf{XZ})$; $\boldsymbol{\theta}' = (\mathbf{b}'\mathbf{a}')$;

$$\Omega = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1}k \end{bmatrix}$$

and $\mathbf{C} = \mathbf{W}'\mathbf{W} + \Omega$, where $k = \sigma_e^2/\sigma_a^2$. Then the mixed-model equations associated with [1] and [2] are:

$$\mathbf{C}\hat{\boldsymbol{\theta}} = \mathbf{W}'\mathbf{y} \tag{4}$$

The implementation of the Gibbs sampler requires sampling $\boldsymbol{\theta}$ from:

$$\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y} \sim N\left(\hat{\boldsymbol{\theta}}, \mathbf{C}^{-1}\sigma_e^2\right) \tag{5}$$

and σ_i^2 from:

$$\sigma_i^2|\boldsymbol{\theta}, \mathbf{y} \sim S_i\chi_{\nu_i}^{-2}, \quad i = a, e \tag{6}$$

where $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y})$ satisfies the linear system [4], and $\mathbf{C}^{-1}\sigma_e^2 = \text{Var}(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y})$.

In [6], $S_a = \mathbf{a}'\mathbf{A}^{-1}\mathbf{a}$; $\nu_a = q - 2$; $S_e = (\mathbf{y} - \mathbf{Xb} - \mathbf{Za})'(\mathbf{y} - \mathbf{Xb} - \mathbf{Za})$; $\nu_e = n - 2$; and $\chi_{\nu_i}^{-2}$ is an inverse chi square variate with ν_i degrees of freedom. In order to sample the whole vector $\boldsymbol{\theta}$ simultaneously from [5] without involving the inverse of the coefficient matrix \mathbf{C} , we draw from ideas in García-Cortés et al (1992, 1995). Given starting values for σ_a^2, σ_e^2 , apply the method of composition (eg, Tanner, 1993) to solve the following equation:

$$p(\mathbf{y}|\sigma_a^2, \sigma_e^2) = \int p(\mathbf{y}|\boldsymbol{\theta}, \sigma_a^2, \sigma_e^2) p(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2) d\boldsymbol{\theta} \tag{7}$$

To obtain samples from $p(\mathbf{y}|\sigma_a^2, \sigma_e^2)$, draw $\boldsymbol{\theta}^*$ from $p(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2)$ and \mathbf{y}^* from $p(\mathbf{y}|\boldsymbol{\theta}^*, \sigma_a^2, \sigma_e^2)$. Then $(\mathbf{y}^*, \boldsymbol{\theta}^*)$ is a drawing from $p(\mathbf{y}, \boldsymbol{\theta}|\sigma_a^2, \sigma_e^2)$ and (\mathbf{y}^*) from $p(\mathbf{y}|\sigma_a^2, \sigma_e^2)$. Further, the pair $(\mathbf{y}^*, \boldsymbol{\theta}^*)$ can also be viewed as a realized value from $p(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y}^*)$ and from $p(\mathbf{y}|\boldsymbol{\theta}^*, \sigma_a^2, \sigma_e^2)$.

By analogy with the identity:

$$\boldsymbol{\theta} = E(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y}) + (\boldsymbol{\theta} - E(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y})) \tag{8}$$

we define the random variable:

$$\tilde{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y}) + (\boldsymbol{\theta} - E(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y}^*)) \tag{9}$$

where the random variable $\boldsymbol{\theta}$ in [9] has density $p(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y}^*)$. The expected value of $\tilde{\boldsymbol{\theta}}$ in [9] with respect to $p(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y}^*)$ is equal to $E(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y})$ and the variance is $\mathbf{C}^{-1}\sigma_e^2$, independent of \mathbf{y}^* . In addition, the random variable $\boldsymbol{\theta}$ in [9] is normally distributed; therefore $\tilde{\boldsymbol{\theta}}$ in [9] and $\boldsymbol{\theta}$ in [8] have the same density $p(\boldsymbol{\theta}|\sigma_a^2, \sigma_e^2, \mathbf{y})$,

which is the conditional posterior distribution of interest. Using [4] and [5], and replacing the random variable $\boldsymbol{\theta}$ in [9] by its realized values $\boldsymbol{\theta}^*$, it follows that drawings from this conditional posterior distribution, which we denote $\tilde{\boldsymbol{\theta}}^*$, can be constructed by solving the linear system:

$$\tilde{\boldsymbol{\theta}}^* = \boldsymbol{\theta}^* + \mathbf{C}^{-1}\mathbf{W}'(\mathbf{y} - \mathbf{y}^*) \quad [10]$$

A wide variety of efficient algorithms are available which do not require \mathbf{C}^{-1} to solve the mixed-model equations [10]. We note in passing that under the assumption of either proper or improper uniform priors for \mathbf{b} , a simple manipulation of [10] shows that this expression is not a function of \mathbf{b}^* , where $\boldsymbol{\theta}^* = (\mathbf{a}^* \mathbf{b}^*)'$. Therefore the drawing of $\boldsymbol{\theta}^*$ from $p(\boldsymbol{\theta} | \sigma_a^2 \sigma_e^2)$ involves \mathbf{a}^* only, and \mathbf{b}^* can be set arbitrarily to zero.

With $\tilde{\boldsymbol{\theta}}^*$ available, draw from:

$$\sigma_i^2 | \tilde{\boldsymbol{\theta}}^*, \mathbf{y} \sim \tilde{S}_i^* \chi_{\nu_i}^{-2}, \quad i = a, e \quad [11]$$

The samples σ_i^2 are now used in the new round of iteration as input in [7].

AN EXAMPLE

As an illustration, the block (multivariate) sampling strategy is compared with the traditional scalar implementation of the Gibbs sampler. A data file based on a univariate animal model with 250 individuals (all with one record, 25 individuals per generation, ten generations) and one fixed-effect factor with ten levels was simulated. For both strategies, a single chain of length 31 000 was run, and the first 1 000 samples were discarded.

Table I shows estimates of the Monte-Carlo variance of the mean of the marginal posterior distributions of the first level of the fixed-effect factor, of the additive genetic value of the last individual and of both variance components. The mean was estimated by summing the samples and dividing by the number of samples (30 000) and the Monte-Carlo variance was computed using the Markov chain estimator (Geyer, 1992). Also shown in table I are estimates of the chains effective length. Briefly, the effective chain length is associated with the amount of information available in a given chain. This parameter becomes smaller as the dependence between the samples of the chain increases. When the samples are in fact independent, the effective and actual chain lengths are equal. Details can be found in Sorensen et al (1995). The figures in table I show that for this model and data structure, there is a reduction in the Monte-Carlo variance by a factor of ten using the block sampling strategy in the case of the fixed effects and breeding values, and a twofold reduction in the case of the variance components.

In the above example, the reduction of the Monte-Carlo variance using either the scalar or block sampling strategies was compared on the basis of the simple estimator of the mean of marginal posterior distributions (the raw average of the elements of the Gibbs chain). A more efficient estimator is based on the average of conditional densities (Gelfand and Smith, 1990; Liu et al, 1994). Liu et al (1994)

Table I. Monte-Carlo variances (Var) of the estimated means (computed as the raw average of the relevant Gibbs chain) of the marginal posterior distributions and effective chain lengths (ECL) of the first level of the fixed-effect factor, of the last individual's breeding value, of the additive genetic variance (Va) and of the residual variance (Ve) implementing the Gibbs sampler with scalar or block strategies.

	<i>Fixed effect</i>		<i>Breeding value</i>		<i>Additive genetic variance</i>		<i>Residual variance</i>	
	10^{-3} Var	ECL	10^{-3} Var	ECL	Var	ECL	Var	ECL
Scalar	5.90	2 244	8.90	2 675	2.61	255	1.08	339
Block	0.49	27 022	0.94	26 233	1.32	489	0.49	713

The actual chain length was 30 000 in all cases.

refer to this as the mixture estimator. For example, let X , Y and Z denote three parameters and assume that interest focuses on the estimate of the mean of the marginal posterior distribution of X .

The mixture estimator is given by

$$\hat{E}(X) = \frac{1}{n} \sum_i \sum_j x_j f(x_j | y_i z_i)$$

where the summation over i is over the n elements of the Gibbs chain, and the summation over j is over the chosen values of X . Alternatively, when the integral has a closed-form solution, the mixture estimator can take the form

$$\hat{E}(X) = \frac{1}{n} \sum_i \int x f(x | y_i, z_i) dx$$

The Monte-Carlo variances of the mixture estimator of the mean of the marginal posterior distributions of the first level of the fixed effect factor, of the additive genetic value of the last individual and of both variance components were respectively 5.11×10^{-3} , 7.62×10^{-3} , 2.59 and 1.03 for the scalar sampling strategy and 0.13×10^{-3} , 0.56×10^{-3} , 1.30 and 0.43 for the block sampling strategy. There is a small increase in efficiency relative to the 'raw means estimator' in the case of the location parameters but not in the case of the variance components. The mixture estimator is especially beneficial when the chain mixes quickly (Liu et al, 1994) and this is not the case in animal models.

CONCLUSIONS

We have presented an algorithm which allows us to implement the Gibbs sampler in a multivariate fashion. The development is in terms of the single trait Gaussian model, but extension to a multiple trait analysis with arbitrary pattern of missing data is straightforward, provided the procedure is used in conjunction with data augmentation. The benefit of block sampling relative to scalar sampling in terms of CPU time was not investigated, since the results will be dependent on the model

and data structure. The important feature of the strategy is that it only involves the solution of a linear system. This means that either computer time or storage requirements can be optimized by choice of the appropriate method to solve the linear system. This is in contrast with the scalar Gibbs sampler which has computer requirements analogous to Gauss–Seidel iterative methods.

ACKNOWLEDGMENTS

Our colleagues D Gianola and S Andersen contributed with insightful comments.

REFERENCES

- García-Cortés LA, Moreno C, Varona L, Altarriba J (1992) Variance component estimation by resampling. *J Anim Breed Genet* 109, 358-363
- García-Cortés LA, Moreno C, Varona L, Altarriba J (1995) Estimation of prediction error variances by resampling. *J Anim Breed Genet* 112, 176-182
- Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85, 398-409
- Gelfand AE, Hills SE, Racine-Poon A, Smith AFM (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *J Am Stat Assoc* 85, 972-985
- Geyer CJ (1992) Practical Markov chain Monte-Carlo (with discussion). *Stat Sci* 7, 46-511
- Liu JS, Wong WH, Kong A (1994) Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika* 81, 27-40
- Smith AFM, Roberts GO (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte-Carlo methods. *J R Stat Soc B* 55, 3-23
- Sorensen DA, Andersen S, Gianola D, Korsgaard I (1995) Bayesian inference in threshold models using Gibbs sampling. *Genet Set Evol* 27, 229-249
- Tanner M (1993) *Tools for Statistical Inference*. Springer-Verlag, New York, USA
- 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