Biometrika (1996), **83**, 3, *pp.* 681–682
Printed in Great Britain

Miscellanea

Peskun's theorem and a modified discrete-state Gibbs sampler

BY JUN S. LIU

Department of Statistics, Stanford University, Stanford, California 94305, U.S.A.

SUMMARY

Attention is drawn to the use of Peskun's theorem in improving statistical efficiency of discrete-state Gibbs sampling.

Some key words: Ising model; Metropolis–Hastings algorithm; Markov chain Monte Carlo.

1. PESKUN'S THEOREM

Let X be a discrete random variable following distribution π, and let P be the transition matrix of a Markov chain with π as its invariant distribution. We call P reversible if

$$\pi(x)P(x, y) = \pi(y)P(y, x).$$

Following Peskun (1973), we define $P_2 \geqslant P_1$ for any two transition matrices if each of the off-diagonal elements of P_2 is greater than or equal to the corresponding off-diagonal elements of P_1. The following lemma is Theorem 2.1.1 of Peskun (1973).

LEMMA 1·1. *Suppose each of the irreducible transition matrices P_1 and P_2 is reversible for the same invariant probability distribution π. If $P_2 \geqslant P_1$ then, for any f,*

$$v(f, \pi, P_1) \geqslant v(f, \pi, P_2), \tag{1}$$

where

$$v(f, \pi, P) = \lim_{N\to\infty} N \operatorname{var}(\hat{I}_N),$$

and $\hat{I}_N = \sum_{t=1}^{N} f\{X^{(t)}\}/N$ is an estimator of $I = E_\pi(f)$ using N consecutive samples from the Markov chains. Kemeny & Snell (1969, *p.* 84) *gave an expression for $v(f, \pi, P)$ in terms of f, P and π.*

Whenever (1) holds, we say that P_2 is statistically more efficient than P_1.

2. A MODIFIED GIBBS SAMPLER

Suppose that $X = (X_1, \ldots, X_d)$, where X_i takes m_i possible values, and that $\pi(x)$ is the distribution of interest. In the random scan Gibbs sampler (Geman & Geman, 1984), each successive step chooses a coordinate i independently, according to a probability distribution $\alpha = (\alpha_1, \ldots, \alpha_d)$, and then the current value x_i of X_i is replaced by a value y_i, drawn from the corresponding full conditional distribution. Thus, the nonzero elements of the transition matrix P_1 are $P_1(x, y) = \alpha_i \pi(y_i|x_{-i})$, where $y = x$, except that y_i replaces x_i; and $x_{-i} = x$, except that x_i is omitted.

Here we consider a modification of the above procedure in which a value y_i, different from x_i, is drawn with probability

$$\frac{\pi(y_i|x_{-i})}{1 - \pi(x_i|x_{-i})};$$

then y_i replaces x_i with the Hastings (1970) acceptance probability,

$$\min\left\{1, \frac{1-\pi(x_i|x_{-i})}{1-\pi(y_i|x_{-i})}\right\},$$

else x_i is retained. This provides a time-reversible transition matrix P_2 with nonzero off-diagonal elements

$$P_2(x, y) = \alpha_i \times \min\left\{\frac{\pi(y_i|x_{-i})}{1-\pi(x_i|x_{-i})}, \frac{\pi(y_i|x_{-i})}{1-\pi(y_i|x_{-i})}\right\}$$

for $y_i \neq x_i$, where y is defined as above. It follows that $P_2 \geqslant P_1$ and hence Lemma 1 implies the following.

THEOREM 2·1. *The modified Gibbs sampler for discrete random variables as defined above is statistically more efficient than the random scan Gibbs sampler.*

When $m_i = 2$, the Gibbs sampler is essentially Barker's (1965) method, whereas the modified procedure becomes a Metropolis et al. (1953) algorithm. Peskun (1973) makes some general comparisons between these two samplers. Besag et al. (1995) note that the superiority of Metropolis for binary systems results from its increased mobility around the state space. This rationale applies more generally to the modified Gibbs sampler. Although Theorem 2·1 does not even require m_i to be finite, the modification is likely to be most useful for components with m_i rather small.

It is easily shown from (1) that the second largest eigenvalue of P_1 is greater than or equal to that of P_2. Frigessi et al. (1993) prove that, for the binary Ising model, Metropolis converges faster than Gibbs for strong interaction, and more slowly for weak interaction. This does not conflict with our result, which concerns statistical efficiency in equilibrium, rather than rate of convergence. Whereas the eigenvalues of the Gibbs sampler are necessarily nonnegative (Liu, Wong & Kong, 1995), slow Metropolis convergence under weak interaction is the product of a large negative eigenvalue.

ACKNOWLEDGEMENT

The author is grateful to Professor Persi Diaconis and the Associate Editor for their insightful comments and to the National Science Foundation for financial support.

REFERENCES

BARKER, A. A. (1965). Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. *Aust. J. Phys.* **18**, 119–33.

BESAG, J., GREEN, P., HIGDON, D. & MENGERSEN, K. (1995). Bayesian computation and stochastic systems (with Discussion). *Statist. Sci.* **10**, 3–66.

FRIGESSI, A., HWANG, C.-R., DI STEFANO, P. & SHEU, S.-J. (1993). Convergence rates of the Gibbs sampler, the Metropolis algorithm and other single-site updating dynamics. *J. R. Statist. Soc.* B **55**, 205–19.

GEMAN, S. & GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pat. Anal. Mach. Intel.* **6**, 721–41.

HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

KEMENY, J. G. & SNELL, J. L. (1969). *Finite Markov Chains.* Princeton, NJ: Van Nostrand.

LIU, J. S., WONG, W. H. & KONG, A. (1995). Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. R. Statist. Soc.* B **57**, 157–69.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. & TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–91.

PESKUN, P. H. (1973). Optimal Monte Carlo sampling using Markov chains. *Biometrika* **60**, 607–12.

[*Received April* 1995. *Revised January* 1996]