Markov chain Monte Carlo and Rao–Blackwellization

Ian W. McKeague*
Florida State University

Wolfgang Wefelmeyer
University of Siegen

Abstract

We introduce a form of Rao–Blackwellization for Markov chains which uses the transition distribution for conditioning. We show that for reversible Markov chains, this form of Rao–Blackwellization always reduces the asymptotic variance, and derive two explicit forms of the variance reduction obtained through repeated Rao–Blackwellization. The result applies to many Markov chain Monte Carlo methods used in practice. In particular, we discuss an application to data augmentation and give some simulation results for Ising model samplers.

1 Introduction

The Rao–Blackwell theorem is considered to be one of the fundamental paradigms of modern statistics [see Pathak (1992)]. In its simplest form, this theorem provides an appealing method of improving the efficiency of an unbiased estimator $\hat{\theta}$ by conditioning with respect to some statistic X. The method works provided $E(\hat{\theta}|X)$ is easy to compute and does not depend on any unknown parameters (e.g., when X is a sufficient statistic). Recently, there has been considerable interest in developing versions of the Rao–Blackwell theorem in the context of stochastic simulation, and for Markov chain Monte Carlo (MCMC) in particular—see Casella and Robert (1996a) and the references cited therein. MCMC now plays an important role in many areas of statistics;

*Research partially supported by NSF Grant ATM-9417528.

[1] *AMS 1991 subject classifications.* Primary: 62M05, 62G05; secondary: 62M40, 65C05.

[2] *Key words and Phrases.* Empirical estimator, conditioning, data augmentation, Ising model, Gibbs sampler, Metropolis algorithm.

in Bayesian inference, for instance, it is widely used for the exploration of posterior distributions, see Tierney (1994).

The present paper introduces a form of Rao–Blackwellization for the estimation of features in the invariant law of a Markov chain when certain information concerning the transition distribution of the chain is available. The primary application will be to MCMC, where the aim is to estimate the expectation $\pi f = E_\pi f$ of a given function f under some finite-dimensional 'target' distribution π; this is done by calculating long run averages of f at values given by the path of a simulated Markov chain $\{X^i,\, i \geq 0\}$ that has π as its stationary distribution.

Many types of updating mechanism are available for the specification of MCMC samplers (Gibbs, Metropolis, Metropolis–Hastings, etc.). Often the choice of updating mechanism is guided by how quickly the sampler converges to stationarity, an efficient sampler being one that rapidly explores the support of π. For numerous routine applications of MCMC, however, mixing is quite rapid. Our concern here will be with estimator efficiency (or, more precisely, asymptotic variance reduction), rather than sampler efficiency. The choice of updating mechanism determines the transition distribution $Q(x, dy)$ of the chain, which is therefore known in advance and can in principle be used to reduce estimator variance by conditioning.

We propose a Rao–Blackwellized procedure that makes use of the conditional expectation $Q(x, f) = E(f(X^i)|X^{i-1} = x)$, assuming that this function is sufficiently tractable to be calculated explicitly from the given f and Q. We introduce the following estimator of πf:

$$E_n Q f = \frac{1}{n} \sum_{i=1}^{n} Q(X^i, f),$$

which can be thought of as a Rao–Blackwellized version of the usual empirical estimator $E_n f = n^{-1} \sum_{i=1}^{n} f(X^i)$.

Our main result, proved in Section 2, is that for reversible Markov chains, $E_n Q f$ has smaller asymptotic variance than $E_n f$. Although MCMC samplers are not in general reversible, many of them have simple modifications that are reversible and to which our result applies. Applications of the proposed estimator and some alternative types of Rao–Blackwellization for MCMC are discussed in Section 3. Two specific examples are considered in some detail: data augmentation and estimation of correlation functions in the Ising model.

2 Variance reduction

In this section we show that the Rao–Blackwellized empirical estimator $E_n Qf$ has smaller asymptotic variance than the ordinary empirical estimator $E_n f$ when the chain is reversible. Repeated Rao–Blackwellizations $E_n Q^k f$, $k > 1$, of course give further variance reductions. In fact, we show that as k goes to infinity, the asymptotic variance of $E_n Q^k f$ goes to zero. In practice, however, more than one Rao–Blackwellization may not be feasible because of the difficulty of finding $Q^k f$, except in simple cases.

We need conditions under which the empirical estimator and its Rao–Blackwellizations are consistent and asymptotically normal. To introduce such conditions, write $\|f\|$ and (f, g) for the norm and inner product of $L_2(\pi)$, and regard Q as the operator defined by $(Qf)(x) = Q(x, f)$. The following ergodicity assumption will be in force throughout.

The Markov chain is positive Harris recurrent, and $\sup\{\|Q^r f - \pi f\| : \|f\| = 1\} \to 0$ *for* $r \to \infty$.

Gordin and Lifšic (1978) show that if $f \in L_2(\pi)$, then $n^{1/2}(E_n f - \pi f)$ is asymptotically normal with mean zero and variance

$$\|(I - Q)^{-1}(f - \pi f)\|^2 - \|(I - Q)^{-1}Q(f - \pi f)\|^2. \tag{1}$$

The variance is easily rewritten as

$$\begin{aligned}&\Big(f - \pi f, \Big(I + 2\sum_{r=1}^{\infty} Q^r\Big)(f - \pi f)\Big) \\ &= \|f - \pi f\|^2 + 2\sum_{r=1}^{\infty} \Big(f - \pi f, Q^r(f - \pi f)\Big). \end{aligned} \tag{2}$$

For the above central limit theorem and representations of the asymptotic variance we refer also to Meyn and Tweedie (1993, Chapter 17, in particular Section 17.4.3) and Chan and Geyer (1994).

The Markov chain is reversible under the stationary law if Q is in detailed balance with π,

$$\pi(dx)Q(x,dy) = \pi(dy)Q(y,dx). \tag{3}$$

Detailed balance is equivalent to selfadjointness of Q as an operator on $L_2(\pi)$,

$$(f,Qg) = (Qf,g) \quad \text{for } f,g \in L_2(\pi). \tag{4}$$

Theorem *If the above ergodicity assumption holds and the Markov chain is reversible, then for $f \in L_2(\pi)$ the asymptotic variance of E_nQ^kf is less than that of E_nf, and the reduction is*

$$\sum_{j=0}^{k-1} \|(I+Q)Q^j(f-\pi f)\|^2.$$

The asymptotic variance of E_nQ^kf tends to zero as k goes to infinity.

Proof The asymptotic variance of E_nQ^kf is calculated from (2) with Q^kf in place of f, and using selfadjointness (4), as

$$\begin{aligned}
&\Big(Q^k(f-\pi f), \Big(I + 2\sum_{r=1}^{\infty} Q^r\Big)Q^k(f-\pi f)\Big) \\
&= \Big(f-\pi f, \Big(Q^{2k} + 2\sum_{r=1}^{\infty} Q^{r+2k}\Big)(f-\pi f)\Big) \\
&= \Big(f-\pi f, Q^{2k}(f-\pi f)\Big) + 2\sum_{r=2k+1}^{\infty} \Big(f-\pi f, Q^r(f-\pi f)\Big).
\end{aligned} \tag{5}$$

The asymptotic variance (5) is obtained from the asymptotic variance (2) by omitting the term $\|f-\pi f\|^2$, the terms of order $r = 1,\dots,2k-1$, and half the term of order $2k$. This implies the second part of the Theorem.

The difference between (2) and (5) can be written as

$$\begin{aligned}
&\big(f-\pi f, \big(I+2(Q+\cdots+Q^{2k-1})+Q^{2k}\big)(f-\pi f)\big)\\
&=\big(f-\pi f, \sum_{j=0}^{k-1}(I+Q)^2 Q^{2j}(f-\pi f)\big)\\
&=\sum_{j=0}^{k-1}\|(I+Q)Q^j(f-\pi f)\|^2,
\end{aligned}$$

which completes the proof. □

Remark Even for $k=1$, the variance reduction can be substantial. As seen in the proof, the asymptotic variance of $E_n Qf$ is obtained from the asymptotic variance (2) of $E_n f$ by omitting the terms

$$\|f-\pi f\|^2 + 2(f-\pi f, Q(f-\pi f)) + \|Q(f-\pi f)\|^2 = \mathrm{Var}_\pi(f+Qf).$$

If the chain is weakly correlated, $\|f-\pi f\|^2 = \mathrm{Var}_\pi f$ dominates the asymptotic variance of $E_n f$, and the asymptotic variance of $E_n Qf$ is relatively small. We will see this effect in our Ising model simulations at high temperature.

Remark If the state space has N elements, the decrease in variance through repeated Rao–Blackwellization can also be expressed in terms of the eigenvalues and eigenvectors of Q. Let $1=\lambda_1$ and $1>\lambda_2\geq\lambda_3\geq\cdots\geq\lambda_N>-1$ be the eigenvalues of Q, and $e_1=1, e_2, e_3, \ldots, e_N$ the corresponding eigenvectors. Then

$$(f-\pi f, e_1) = \pi(f-\pi f) = 0$$

and

$$f-\pi f = \sum_{r=2}^{N}(f-\pi f, e_r)e_r.$$

Assume again reversibility. By detailed balance (3), the asymptotic variance (2) of the empirical estimator $E_n f$ can be written in terms of the eigenvalues and eigenvectors of Q as

$$\sum_{r=2}^{N}\frac{1+\lambda_r}{1-\lambda_r}(f-\pi f, e_r)^2. \tag{6}$$

We refer to Frigessi et al. (1992) and Green and Han (1992). The asymptotic variance of $E_n Q^k f$ is (6) with f replaced by $Q^k f$. Using selfadjointness (4), we can write the asymptotic variance of $E_n Q^k f$ as

$$\begin{aligned} &\sum_{r=2}^{N} \frac{1+\lambda_r}{1-\lambda_r} (Q^k(f-\pi f), e_r)^2 \\ &= \sum_{r=2}^{N} \frac{1+\lambda_r}{1-\lambda_r} \lambda_r^{2k} (f-\pi f, e_r)^2. \end{aligned} \tag{7}$$

The difference between the asymptotic variance (6) of the empirical estimator and the asymptotic variance (7) of the k-fold Rao–Blackwellized empirical estimator is therefore

$$\sum_{r=2}^{N} \frac{1+\lambda_r}{1-\lambda_r} (1-\lambda_r^{2k})(f-\pi f, e_r)^2$$

which is again seen to be nonnegative. $\square$

3 Application to Markov chain Monte Carlo

When the distribution π is on a d-dimensional space, MCMC schemes for sampling from π typically involve the updating of one variable at a time. In this section we discuss how our approach can be applied to such samplers.

Recall that a Markov chain with transition distribution Q is reversible if Q is in detailed balance with π, or equivalently if Q is selfadjoint. Usually, for local Gibbs and Metropolis–Hastings updates, in particular, the local transition distribution Q_j corresponding to the update of the j-th component is in detailed balance with π. The components of the state vector are updated according to a given sweep strategy [see Roberts and Sahu (1996) for a comparison of various sweep strategies for the Gibbs sampler]. If a component is chosen *at random* and then updated according to the corresponding local transition distribution, the sampler has transition distribution $Q = \frac{1}{d}\sum_{j=1}^{d} Q_j$ which is again in detailed balance with π, and our result applies.

For a d-step sampler with *deterministic* sweep and the components updated in the order $1, 2, \ldots, d$ say, the transition distribution is $Q = Q_1 \cdots Q_d$. Its adjoint is

$Q^* = Q_d \cdots Q_1 \neq Q$. Hence the sampler is not reversible, and our result does not apply. A simple modification, however, makes the sampler reversible: use a forward sweep $1, 2, \ldots, d$ followed by a backward sweep $d, d-1, \ldots, 1$. The resulting transition distribution QQ^* is reversible. Nevertheless, deterministic sweeps have the disadvantage that Q is a *composition* of local transition distributions, which makes it computationally more costly to evaluate $Qf = Q_1 \cdots Q_d f$ compared with $\frac{1}{d}\sum_{j=1}^{d} Q_j f$. For random fields π with only nearest neighbor dependence, an effective and computationally feasible alternative method for reducing the variance of empirical estimators based on deterministic sweeps is the von Mises type statistic approach of Greenwood, McKeague and Wefelmeyer (1995).

Other types of Rao–Blackwellization have been introduced previously in the MCMC context. They differ from our version in that they condition $f(X^i)$ on a component of X^i rather than on the previous value of the chain. Gelfand and Smith (1990, 1991) consider i.i.d. runs of a Gibbs sampler. In the empirical estimator based on the final value of each run, they replace f by a conditional expectation under π. For long runs, the final values are approximately distributed according to π, so the classical Rao–Blackwell theorem implies that the variance is usually reduced. Liu et al. (1994) consider a single run of a two-step Gibbs sampler (data augmentation) and a function f of one component; they replace f by its conditional expectation given the other component, similar to Gelfand and Smith, and prove that the variance is always reduced. Our version of Rao–Blackwellization can be applied to data augmentation and leads to greater variance reduction than that of Liu et al.; see Section 3.1. Casella and Robert (1995, 1996a, 1996b) propose some types of Rao–Blackwellization for the accept-reject and Metropolis algorithms. Their approach is to integrate out some or all of the uniform random variables involved. They established a variance reduction for one of their accept-reject Rao–Blackwellization methods.

Geyer (1995) considers an arbitrary Markov chain as above and replaces $f(x)$ by its conditional expectation with respect to π given some function $h(x)$. He gives necessary and sufficient conditions for such a Rao–Blackwellization to reduce variance for all f simultaneously, and points out that this condition is unlikely to be satisfied in practice. His result does not cover the Rao–Blackwellization introduced here because, as mentioned earlier, we condition on the previous value of the chain and not on some function of the present value. Indeed, one could consider the chain of pairs (X^i, X^{i+1}) and condition on the function $h(X^i, X^{i+1}) = X^i$, but then one would be dealing with a specific class of functions $f(X^i, X^{i+1})$ that only depend on the second component X^{i+1}, and Geyer's result does not apply.

We now examine two specific examples in greater detail.

3.1 Data augmentation

Our result applies to data augmentation, which was introduced by Tanner and Wong (1987) and is also known as substitution sampler (Gelfand and Smith, 1990). Auxiliary variable methods used in statistical physics, including the algorithm of Swendsen and Wang (1987) for the Potts model, are of this type; cf. Besag and Green (1993).

Data augmentation is a two-step Gibbs sampler with deterministic sweep. For the sweep that updates $x = (x_1, x_2)$ in the order x_2, x_1, the transition distribution is

$$Q(x, dy) = p_2(x_1, dy_2)p_1(y_2, dy_1),$$

where p_1 and p_2 are the conditional distributions of each component of π given the other. For a function $f(x_1)$ of just the first component, the empirical estimator is $E_n f = \frac{1}{n}\sum_{i=1}^n f(X_1^i)$. Note that X_1^i, $i = 1, 2, \ldots$, is a Markov chain with transition distribution

$$P(x_1, dy_1) = \int p_2(x_1, dy_2)p_1(y_2, dy_1).$$

This chain is reversible, although, as mentioned earlier, the Gibbs sampler chain is not. Hence our result applies, and the Rao–Blackwellized empirical estimator

$$E_n P f = \frac{1}{n}\sum_{i=1}^n P(X_1^i, f)$$

has smaller asymptotic variance than $E_n f$. The estimator $E_n Pf$ differs from Liu et al.'s (1994) Rao–Blackwellized estimator

$$E_n p_1 f = \frac{1}{n} \sum_{i=1}^{n} p_1(X_2^i, f).$$

They show that $E_n p_1 f$ also has smaller variance than $E_n f$.

Which type of Rao–Blackwellization is better in terms of variance reduction? Note that our estimator is obtained by applying Liu et al.'s Rao–Blackwellization twice: $E_n Pf = E_n p_2(p_1 f)$. Hence, by their result, our Rao–Blackwellization leads to greater variance reduction. Of course, in some cases $p_1 f$ may be tractable, while $p_2(p_1 f)$ is not.

3.2 Ising models

In this subsection we show that our approach can lead to substantial variance reduction when applied to Ising model samplers with suitable sweeps. Applications of Ising models arise in spatial statistics, image analysis and statistical physics; see Winkler (1995). Consider a $k \times k$ integer lattice S with k even. Attach to each site a state space $\{-1, +1\}$ representing two spin orientations. Under the Gibbs distribution π on the configuration space $\{-1, +1\}^S$, a configuration x has mass $\pi(x)$ proportional to $\exp(-H(x))$, where the energy function H is given by $H(x) = -\beta \sum_{\langle s,t \rangle_1} x_s x_t$. Here β is the inverse temperature and the sum is over unordered pairs $\langle s, t \rangle_1$ of nearest neighbors s, t.

Simulations of Ising models are often carried out using Glauber dynamics (reversible single-site updating); see Neves and Schonmann (1992) for many examples of such schemes. The updating at site s depends only on the spins at the four neighboring sites and possibly the spin at site s itself. It is specified by the spin flip probability, say $p_s(x, -x_s)$. In particular, the Gibbs sampler uses

$$p_s(x, -x_s) = \Big(1 + \exp\Big(2\beta x_s \sum_{t:\langle t,s \rangle_1} x_t\Big)\Big)^{-1},$$

and the Metropolis sampler uses

$$p_s(x, -x_s) = \min\Big\{1, \exp\Big(-2\beta x_s \sum_{t:\langle t,s \rangle_1} x_t\Big)\Big\}.$$

To apply our result, we need a sweep that makes the transition distribution of the sampler reversible. The lattice can be partitioned into two sublattices consisting of sites with even or odd parity, as in a checkerboard. Partition a configuration $x = (x_e, x_o)$ correspondingly. We update the even subconfiguration by applying p_s independently at all even sites, resulting in the transition probability

$$Q_e(x, y) = \Big(\prod_{s \text{ even}} p_s(x, y_s) \Big) \delta_{x_o}(y_o),$$

which is in detailed balance with π. The transition probability Q_o for the odd subconfiguration is defined similarly and is in detailed balance with π. Usually one would alternate deterministically between these two subconfiguration updates (we call this the 'deterministic sweep'). Then the transition distribution of the sampler would be $Q_e Q_o$, which, as we have already mentioned, is not reversible. A natural modification of the sweep is to randomly choose between updating the even and odd sublattices at each step (we call this the 'random sweep'). The transition distribution of this sampler is $Q = \frac{1}{2}(Q_e + Q_o)$, which is reversible. We can apply our approach to the the resulting Markov chain $X^0, X^1, \ldots$.

Consider the r-th nearest neighbor correlation

$$f(x) = \frac{1}{N} \sum_{\langle s,t \rangle_r} x_s x_t.$$

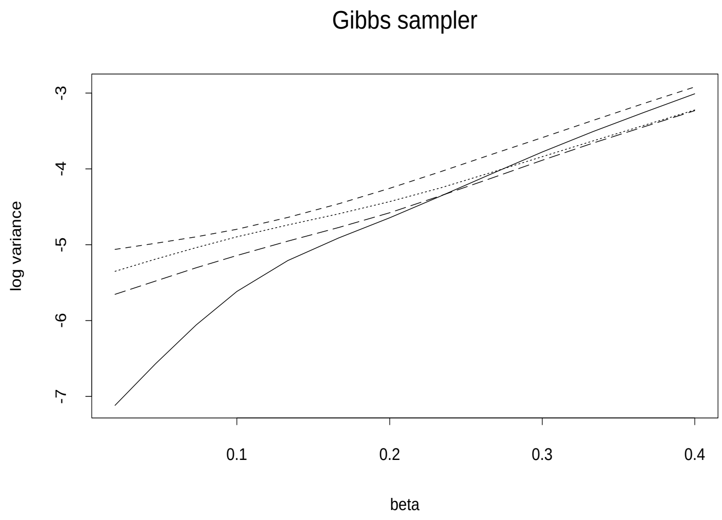
The sum is over all unordered pairs $\langle s, t\rangle_r$ of r-th nearest neighbors s, t, and N is the number of these pairs. We denote $\pi f = \rho_r$, the *expected* r-th nearest neighbor correlation. We easily obtain Qf: for r odd,

$$Q(x, f) = \frac{1}{N} \sum_{\langle s,t \rangle_r} x_s x_t (1 - p_s(x, -x_s) - p_t(x, -x_t)),$$

and for r even,

$$Q(x, f) = \frac{1}{2} \Big(f(x) + \frac{1}{N} \sum_{\langle s,t \rangle_r} x_s x_t (1 - 2p_s(x, -x_s))(1 - 2p_t(x, -x_t)) \Big).$$

Since the spin flip probabilities are already calculated by the sampler, there is no extra computational cost involved in the Rao–Blackwellized estimator $E_n Q f$.



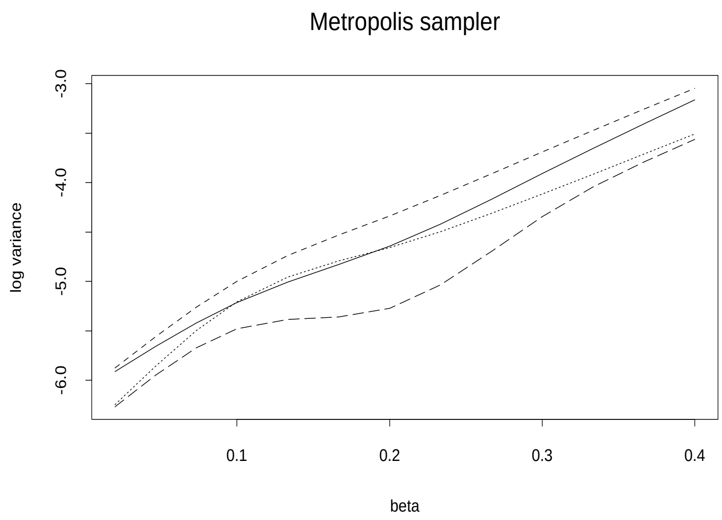


Figure 1: Variances (on a $\log_{10}$ scale) of the Rao–Blackwellized estimator $E_n Qf$ (solid line) and the empirical estimator $E_n f$ (short-dashed line) of ρ_3 as a function of the inverse temperature β. Variances of the von Mises type estimator $M_n f$ (long-dashed line) and the empirical estimator $G_n f$ (dotted line) under the deterministic sweep are plotted on the same scale.

For r even, E_nQf is of the form $\frac{1}{2}(E_nf + R_nf)$, which contains the original empirical estimator E_nf. We expect that R_nf has a smaller variance than E_nQf. Note that R_nf is an average of two empirical estimators that are Rao–Blackwellized in the sense of Liu et al. (1994): $R_nf = \frac{1}{2}(E_nQ_ef_e + E_nQ_of_o)$, where f_e and f_o are the r-th nearest neighbor correlations on the even and odd sublattices, respectively.

The Gibbs and Metropolis samplers were used to estimate ρ_3 for the Ising model on a 6×6 lattice with free boundary. Figure 1 contains plots of the variances of the estimators E_nQf and E_nf for the random sweep samplers, as well as the variances of two competing estimators based on the deterministic sweep samplers: the empirical estimator G_nf and the von Mises estimator M_nf, as defined in Greenwood et al. (1995, 1996). The estimates were based on a run of 1000 sublattice sweeps, which amounts to 500 'full steps' of the deterministic sweep samplers. The variances were calculated over a grid of β values using 5000 sampler runs at each point on the grid. The smooth curves in Figure 1 were obtained by interpolation.

As temperature decreases (β increases), the variances increase at an exponential rate, and the two estimators for random sweep becomes consistently worse than the two estimators for deterministic sweep. Moreover, the Rao–Blackwell and von Mises type estimators show a declining improvement over the usual empirical estimators E_nf and G_nf, respectively, at lower temperature. This is probably due to the increasing strength of long range temporal and spatial dependence in the chains at low temperatures, and the fact that E_nQf and M_nf only exploit information about short range dependence.

There are marked differences between the performance of the estimators under Gibbs and Metropolis dynamics. At high temperatures, the Gibbs chain is only weakly correlated, so, as expected, Rao–Blackwellization is extremely effective; it gives an order of magnitude improvement over competing estimators when $\beta \leq .1$. Rao–Blackwellization is ineffective for the Metropolis sampler at high temperatures, however, since the Metropolis chain is highly correlated when β is small: the spin flip proposal is almost always accepted.

References

J. Besag and P. J. Green, Spatial statistics and Bayesian inference, *J. Roy. Statist. Soc. Ser. B* 55 (1993) 25–37.

G. Casella and C. P. Robert, Recycling rejected values in accept-reject methods, *C. R. Acad. Sci. Paris* 320, Série I (1995) 1621–1626.

G. Casella and C. P. Robert, Rao–Blackwellization of sampling schemes, *Biometrika* 83 (1996a) 81–94.

G. Casella and C. P. Robert, Rao–Blackwell theorem for sampling schemes with rejection, *C. R. Acad. Sci. Paris* 322, Série I (1996b) 571–576.

K. S. Chan and C. J. Geyer, Contribution to the discussion of the paper by L. Tierney, *Ann. Statist.* 22 (1994) 1747–1758.

A. Frigessi, C.-R. Hwang and L. Younes, Optimal spectral structure of reversible stochastic matrices, Monte Carlo methods and the simulation of Markov random fields, *Ann. Appl. Probab.* 2 (1992) 610–628.

A. E. Gelfand and A. F. M. Smith, Sampling-based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.* 85 (1990) 398–409.

A. E. Gelfand and A. F. M. Smith, Gibbs sampling for marginal posterior expectations, *Comm. Statist. Theory Methods* 20 (1991) 1747–1766.

C. J. Geyer, Conditioning in Markov chain Monte Carlo, *J. Comput. Graph. Statist.* 4 (1995) 148–154.

M. I. Gordin and B. A. Lifšic, The central limit theorem for stationary Markov processes, *Soviet Math. Dokl.* 19 (1978) 392–394.

P. J. Green and X.-L. Han, Metropolis methods, Gaussian proposals and antithetic variables, in: *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*

(P. Barone, A. Frigessi and M. Piccioni, eds.) (1992) 142–164, Lecture Notes in Statistics 74, Springer-Verlag, Berlin.

P. E. Greenwood, I. W. McKeague and W. Wefelmeyer, Outperforming the Gibbs sampler empirical estimator for nearest neighbor random fields, *Ann. Statist.* 24 (1996), in press.

P. E. Greenwood, I. W. McKeague and W. Wefelmeyer, Von Mises type statistics for single site updated local interaction random fields, (1995), Preprint.

J. S. Liu, W. H. Wong and A. Kong, Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes, *Biometrika* 81 (1994) 27–40.

S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, (1993), Springer-Verlag, London.

E. J. Neves and R. H. Schonmann, Behavior of droplets for a class of Glauber dynamics at very low temperature, *Probab. Theory Related Fields* 91 (1992) 331–354.

P. K. Pathak, Introduction to Rao (1945) "Information and the accuracy attainable in the estimation of statistical parameters," in: *Breakthroughs in Statistics Volume I* (S. Kotz and N. L. Johnson, eds.) (1992) 227–234, Springer-Verlag, New York.

G. O. Roberts and S. K. Sahu, Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler, (1996), preprint.

R. H. Swendsen and J.-S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, *Phys. Rev. Lett.* 58 (1987) 86–88.

M. A. Tanner and W. H. Wong, The calculation of posterior distributions by data augmentation, *J. Amer. Statist. Assoc.* 82 (1987) 528–540.

L. Tierney, Markov chains for exploring posterior distributions (with discussion), *Ann. Statist.* 22 (1994) 1701–1762.

G. Winkler, *Image Analysis, Random Fields and Dynamic Monte Carlo Methods,* (1995), Springer-Verlag, Berlin.