JUDGING MCMC ESTIMATORS BY THEIR ASYMPTOTIC VARIANCE

WOLFGANG WEFELMEYER

Fachbereich 6 Mathematik, Universität-GH Siegen, Hölderlinstr. 3, 57068 Siegen, Germany
e-mail: wefelmeyer@mathematik.uni-siegen.d400.de

Abstract

The expectation of a function can be estimated by the empirical estimator based on the output of a Markov chain Monte Carlo method. We review results on the asymptotic variance of the empirical estimator, and on improving the estimator by exploiting knowledge of the underlying distribution or of the transition distribution of the Markov chain.
AMS classification: 62G20, 62M05, 62M40, 65C05.
Key words: Markov chain Monte Carlo, empirical estimator, Rao–Blackwellization, Gibbs sampler, information bound, random field, local interaction.

1 Introduction

A Markov chain Monte Carlo (MCMC) method generates a Markov chain $X^0, \ldots, X^n$ with given invariant distribution π. It can be used to estimate the expectation πf of a function f under π by the empirical estimator $\frac{1}{n}\sum_{i=1}^n f(X^i)$. A given MCMC method may be judged by the speed at which the chain converges to stationarity, and by the variance of the empirical estimator.

Here we consider asymptotic variances of empirical estimators, and ways of improving them by exploiting knowledge of π or the MCMC method used. Other results on asymptotic variances of empirical estimators are found in the following references. Peskun (1973) introduces an inequality between transition distributions which entails a corresponding inequality between between asymptotic variances of empirical estimators; see also Liu (1996). Frigessi, Hwang and Younes (1992) identify the transition distributions which minimize the maximum of the asymptotic variance over functions f which have unit variance under π. Green and Han study asymptotic variances for the samplers introduced by Barone and Frigessi (1989). Fishman (1996) considers variances of sample means for Gibbs samplers with different sweeps.

Our results require that the empirical estimator is asymptotically normal, with variance

$$\pi(f-\pi f)^2 + 2\sum_{r=1}^{\infty} \pi((f-\pi f)Q^r(f-\pi f)),$$

where Q^r is the r-step transition distribution of the underlying Markov chain. This central limit theorem holds for $f^2 \le V$ if the chain is positive Harris recurrent and V-uniformly ergodic (Meyn and Tweedie, 1993, Theorem 17.0.1), and for π-square integrable f if the chain is also reversible (Roberts and Rosenthal, 1997, Corollary 2.1).

Section 2 describes a result of McKeague and Wefelmeyer (1996): If the chain $X^0, \ldots, X^n$ is reversible with transition distribution $Q(x, dy)$, then the estimator $\frac{1}{n}\sum_{i=1}^n Q(X^i, f)$ has smaller variance than the empirical estimator. To apply the result, we must be able to compute the conditional expectation.

Sections 3 and 4 contain some results of Greenwood, McKeague and Wefelmeyer (1997). Here π is a distribution on a product space. In Section 3 we calculate the asymptotic variances of empirical estimators for Gibbs samplers with random and deterministic sweep. At least if the components of π are not too strongly dependent, the asymptotic variance is about twice as large for random sweep. In Section 4 we present minimal asymptotic variances for regular estimators of πf for both types of sweep. They are equal if π is continuous in an appropriate sense. The empirical estimator is close to efficient for deterministic sweep, at least if the components of π are not too strongly dependent.

Section 5 presents a special case of the result of Greenwood, McKeague and Wefelmeyer (1996b). Let π be a random field on a product space E^S, where S is a finite square lattice with, say, K sites. Suppose that the field has local interactions, and that we have configurations $(X_1^0, \ldots, X_K^0), \ldots, (X_1^n, \ldots, X_K^n)$ from a Gibbs sampler with deterministic sweep. We may then combine certain components with different time indices into new configurations $(X_1^{i_1}, \ldots, X_K^{i_K})$ and obtain a new estimator for πf by averaging terms of the form $f(X_1^{i_1}, \ldots, X_K^{i_K})$. The method is reminiscent of generalized von Mises statistics for i.i.d. $(X_1^0, \dots, X_K^0), \dots, (X_1^n, \dots, X_K^n)$, which average over *all* such terms.

Section 6 describes a special case of a result in Greenwood, McKeague and Wefelmeyer (1996a). Let π be a random field as in Section 5, now with *periodic* boundary. This means that the lattice is wrapped around a torus. If π is invariant under a translation T of the lattice, we obtain an additional empirical estimator for πf. Its asymptotic variance is, in general, different from that of the original empirical estimator because the distribution of the Markov chain does not inherit an appropriate invariance property from π. Section 6 shows how best to combine "empirical" estimators from certain Gibbs samplers if the field has nearest neighbor interactions.

Much of the recent literature on MCMC methods can be obtained from http://www.stats.bris.ac.uk/MCMC/. Preprints of our papers can also be obtained from http://www.math.uni-siegen.de/statistik/wefelmeyer.html.

2 Rao–Blackwellization for reversible chains

Let $X^0, \dots, X^n$ be realizations of a Markov chain with transition distribution $Q(x, dy)$ and invariant distribution $\pi(dx)$. The *empirical estimator* for the expectation πf of a function f under π is

$$\frac{1}{n}\sum_{i=1}^{n} f(X^i).$$

Assume that the central limit theorem described in the Introduction holds for the empirical estimator.

In the MCMC literature, one considers *Rao–Blackwellized* empirical estimators of the form

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(f(X^i)|h(X^i)),$$

where h is a suitable function. See Schmeiser and Chen (1991, 1996) for the hit-and-run algorithm proposed by Belisle, Romeijn and Smith (1993); Pearl (1987), Gelfand and Smith (1990, 1991) and Liu, Wong and Kong (1994) for the data augmentation scheme, or substitution sampler; and Casella and Robert (1996) for the Metropolis–Hastings algorithm. Geyer (1995) shows that such a Rao–Blackwellization usually does not decrease the variance simultaneously for all functions f. McKeague and Wefelmeyer (1996) suggest a different form of Rao–Blackwellization,

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(f(X^{i+1})|X^i).$$

Repeating this Rao–Blackwellization k times, we obtain

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(f(X^{i+k})|X^i) = \frac{1}{n}\sum_{i=1}^{n} Q^k(X^i, f).$$

Assume that the chain is reversible,

$$\pi(dx)Q(x, dy) = \pi(dy)Q(y, dx).$$

McKeague and Wefelmeyer (1996) prove the following:

The asymptotic variance of the k times Rao–Blackwellized empirical estimator is

$$\pi((f-\pi f)Q^{2k}(f-\pi f)) + 2\sum_{r=2k+1}^{\infty} \pi((f-\pi f)Q^r(f-\pi f)).$$

The asymptotic variance tends to zero as k goes to infinity. The asymptotic variance reduction over the empirical estimator is

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

where $I(x, dy) = \varepsilon_x(dy)$.

In particular, our Rao–Blackwellization reduces the asymptotic variance of the empirical estimator for *all* f.

The primary application will be to MCMC methods, where the transition distribution Q is, at least in principle, known, and hence the conditional expectations can be calculated. Of course, the cost of calculating conditional expectations should be small compared to the cost of running the sampler.

3 Empirical estimators for Gibbs samplers

In this and the next section we assume that the distribution π lives on a product space $E_1 \times \cdots \times E_k$ with product σ-field. Writing $x \in E$ as $x = (x_{<j}, x_j, x_{>j}) = (x_j, x_{-j})$, we can factor π into the marginal distribution of x_{-j} and the conditional distribution of x_j given x_{-j},

$$\pi(dx) = m_{-j}(dx_{-j})p_j(x_{-j}, dx_j).$$

Gibbs samplers successively use the transition distributions

$$Q_j(x, dy) = p_j(x_{-j}, dy_j)\varepsilon_{x_{-j}}(dy_{-j})$$

which change only the j-th component of x. The conditional expectation $Q_j f$ can be written

$$(p_j f)(x) = \int p_j(x_{-j}, dx_j)f(x).$$

We will also use the notation

$$(m_j f)(x) = \int m_j(dx_j) f(x).$$

Both functions depend only on x_{-j}.

The Gibbs sampler with *random* sweep (with *equal* probabilities picks an index j according to the uniform distribution on $1, \ldots, k$. It has transition distribution $\frac{1}{k}\sum_{j=1}^k Q_j(x, dy)$.

The Gibbs sampler with *deterministic* (and *cyclic*) sweep applies Q_j cyclically according to the numbering of the coordinates of E. It has transition distributions $Q_1, \ldots, Q_k, Q_1, \ldots$.

Let $X^0, \ldots, X^n$ be realizations from a Gibbs sampler, and assume that the central limit theorem described in the Introduction is valid for the empirical estimator $\frac{1}{n}\sum_{i=1}^n f(X^i)$. The following results are proved in Greenwood et al. (1997).

For random sweep, the empirical estimator has asymptotic variance

$$\sigma^2 + 2\frac{k}{k-1}\sum_{s=1}^{\infty} \frac{1}{k(k-1)^{s-1}} \sum_{\substack{j_1,\ldots,j_s=1 \\ j_r \neq j_{r+1}}}^{k} \sigma^2_{j_1,\ldots,j_s}$$

with $\sigma^2 = \pi(f - \pi f)^2$ *and*

$$\sigma^2_{j_1,\ldots,j_s} = \pi((f-\pi f)p_{j_1}\cdots p_{j_s}(f - \pi f)).$$

For deterministic sweep, the empirical estimator has asymptotic variance

$$\sigma^2 + 2\sum_{s=1}^{\infty}\frac{1}{k}\sum_{j=1}^{k}\sigma^2_{j,\mathrm{cycl}\ s}$$

with

$$\sigma^2_{j,\mathrm{cycl}\ s} = \pi((f-\pi f)p_j^{\mathrm{cycl}\ s}(f-\pi f))$$

and $p_j^{\mathrm{cycl}\ s} = p_j p_{j+1}\cdots p_k p_1 p_2 \cdots$ *with* s *terms.*

The s-order terms in both infinite series are averages over s-order autocovariances. At first sight, it looks as if the asymptotic variance for random sweep is at most $\frac{k}{k-1}$ times that for deterministic sweep, a small factor for large k. Greenwood et al. (1997) show, however, that it is about *twice* as large, at least if the components of π are not too strongly dependent. The explanation is that a typical s-order autocovariance $\sigma^2_{j_1,\ldots,j_s}$ is larger than an s-order autocovariance $\sigma^2_{j,\mathrm{cycl}\ s}$. This follows, by a continuity argument, from the case where π has independent components $m_1, \ldots, m_k$. Then $\sigma^2_{j,\mathrm{cycl}\ s}$ is the variance under π of $m_j^{\mathrm{cycl}\ s}(f - \pi f)$, which vanishes for $s \geq k$, while $\sigma^2_{j_1,\ldots,j_s}$ vanishes only if all k components are present among $j_1 \ldots, j_s$. Also, if some of the j_r are equal, fewer than s components are integrated out, and $\sigma^2_{j_1,\ldots,j_s}$ tends to be larger than $m_j^{\mathrm{cycl}\ s}(f-\pi f)$.

By Gibbs sampler with deterministic sweep one often means the subchain of full sweeps, taking only every k-th step, with transition distribution $Q_1 \cdots Q_k$. For $n = pk$, the empirical estimator based on the subchain has asymptotic variance

$$k\sigma^2 + 2k\sum_{s=1}^{\infty}\sigma^2_{j,\mathrm{cycl}\ sk}.$$

Simulations in Greenwood et al. (1996a, 1997) show that the empirical estimator based on the subchain can be considerably worse than that based on the full chain. Subsampling of the subchain further increases the asymptotic variance, as observed by Geyer (1992, Theorem 3.3) and McEachern and Berliner (1994).

4 Information bounds for Gibbs samplers

As in the previous section we assume that the distribution π lives on a product space $E_1 \times \cdots \times E_k$, and consider Gibbs samplers for random and deterministic sweep. How much information about πf is contained in the sample $X^0, \ldots, X^n$? In particular: What fraction of the information is exploited by the empirical estimator? Is it worthwhile to construct improved estimators?

To answer these questions, we view π as an infinite-dimensional parameter of the transition distribution driving the sampler, and determine the minimal asymptotic variance, or *information bound*, of regular estimators of πf in the sense of an infinite-dimensional version of the convolution theorem of Hájek (1970). The following results are proved in Greenwood et al. (1997).

For deterministic sweep, the information bound is

$$\sigma^2 + \frac{k}{k-1}\sum_{s=1}^{\infty}\frac{1}{k(k-1)^{s-1}}\sum_{\substack{j_1,\ldots,j_s=1 \\ j_r\neq j_{r+1}}}^{k}\sigma^2_{j_1,\ldots,j_s}.$$

The information bound for random sweep is complicated if the conditional distributions $p_j(x_{-j}, dx_j)$ have atoms; see Greenwood et al. (1997, Theorem 2). However:

The information bound for random sweep is equal to that for deterministic sweep if the conditional distributions $p_j(x_{-j}, dx_j)$ *have no atoms.*

One sees that the information bound for deterministic sweep is about half the asymptotic variance of the empirical estimator for random sweep if the leading term σ^2 is small compared to the infinite series, which is usually the case. On the other hand, as mentioned in the previous section, the asymptotic variance of the empirical estimator for deterministic sweep is also about half that for random sweep, at least if the components of π are not too strongly dependent. This means that the empirical estimator for deterministic sweep is close to efficient.

The information bound for deterministic sweep does not depend on the order of the sweep, so changing the sweep has a limited effect on the performance of the empirical estimator.

If π has only two components, then the empirical estimator is fully efficient under deterministic sweep.

The problem considered in this section belongs to a class of problems on which some progress has been made recently: Given a semiparametric Markov chain model, how well can one estimate the expectation πf of a function under the invariant law? Greenwood and Wefelmeyer (1998) consider models described by restrictions on the invariant law of the chain, in particular reversible chains. Schick and Wefelmeyer (1998) study models given by restrictions on the transition distribution of the chain and outline how to construct efficient estimators for quasi-likelihood models and nonlinear heteroscedastic autoregression models. Kessler, Schick and Wefelmeyer (1998) treat the situation in which we have a parametric model for π but do not know the transition distribution. The answer depends very much on the type of model.

5 Improved estimators for local interaction random fields

In this and the next section we assume that the distribution π is a random field on a finite square lattice, i.e., a distribution on a product space E^S with $S = \{0, \dots, k-1\}^2$ and product σ-field. It will be convenient to take k even. Similarly as in Section 3, for each site s we can factor π into the marginal distribution of x_{-s} and the conditional distribution of x_s given x_{-s},

$$\pi(dx) = m_{-s}(dx_{-s}) p_s(x_{-s}, dx_s).$$

In this section we assume that the field has nearest neighbor interactions with free boundary,

$$p_s(x_{-s}, dx_s) = p_s(x_{\partial s}, dx_s),$$

with $\partial s = \{(s_1 \pm 1, s_2), (s_1, s_2 \pm 1)\}$.

Consider the Gibbs sampler with deterministic sweep. Write $X^0, \dots, X^n$ for the output, and $Z^0 = X^0, Z^1 = X^{k^2}, \dots, Z^p = X^n$, with $n = pk^2$, for the subchain of full sweeps considered at the end of Section 3.

We choose a *checkerboard* sweep, updating first the sites s with, say, even parity $s_1 + s_2$ and then those with odd parity. Note that the sites in ∂s have opposite parity to s. Therefore, all even, or all odd, sites can be updated simultaneously, and we can write the k^2-step Gibbs sampler as a two-step Gibbs sampler. More explicitly, $E = E_e \times E_o$ and $x = (y_e, y_o)$, where y_e and y_o are the subconfigurations on the even and odd sites, respectively. The conditional distributions of y_e and y_o are

$$\begin{aligned} p_e(y_e, dy_o) &= \prod_{s \text{ even}} (x_{\partial s}, dx_s), \\ p_o(y_o, dy_e) &= \prod_{s \text{ odd}} (x_{\partial s}, dx_s). \end{aligned}$$

Let $X^0 = (Y^0, Y^1)$ be the initial configuration. The conditional distributions $p_e, p_o, p_e, \dots$ generate subconfigurations $Y^2, Y^3, Y^4, \dots$, and the subchain of full sweeps is $Z^q = (Y^{2q}, Y^{2q+1})$, $q = 0, \dots, p$.

We combine components of subconfigurations Y^i with different time indices into new configurations

$$X^I = (Y_s^{I(s)})_{s \in S},$$

where $I : S \to \{0, 1, \dots, 2q+1\}$ is a graph with $I(s)$ even for s even, and odd otherwise. Call a graph *admissible* if its values at any two neighboring sites differ by 1. Greenwood et al. (1996b) show that fields with nearest neighbor interactions have the following property:

If the graph I is admissible, then $X^I, X^{I+2}, \dots$ is distributed as a Gibbs sampler for π.

The empirical estimator is

$$\frac{1}{n} \sum_{i=1}^{n} f(X^i) = \frac{1}{k^2} \sum_{s=1}^{k^2} \frac{1}{p} \sum_{q=1}^{p} f(Z^q_{\leq s}, Z^{q-1}_{>s}).$$

The graph of the configuration $(Z^q_{\leq s}, Z^{q-1}_{>s})$ for even s is

$$I(t) = \begin{cases} 2q, & t \leq s,\ t \text{ even}, \\ 2q-2, & t > s,\ t \text{ even}, \\ 2q-1, & t \text{ odd}, \end{cases}$$

and for odd s,

$$I(t) = \begin{cases} 2q+1, & t \leq s,\ t \text{ odd}, \\ 2q-1, & t > s,\ t \text{ odd}, \\ 2q, & t \text{ even}. \end{cases}$$

We obtain new estimators by averaging $f(X^I)$ over other sets of admissible graphs. Simulations in Greenwood et al. (1996b) show that the variance reduction over the empirical estimator can be noticeable if the components of π are not too strongly dependent.

The approach works particularly well for deterministic sweeps following the checkerboard pattern, for nearest neighbor models and for free boundary. It can be adapted to other deterministic sweeps and also to random sweeps, and to more general local interactions, but in all these cases there are fewer admissible graphs and hence less variance reduction. The approach works also for other MCMC methods, as long as they update a single or only a few sites at a time and condition only on values at nearby sites. Examples are the Metropolis algorithm and many other Glauber dynamics for Ising models, see Neves and Schonmann (1992).

6 Symmetrized estimators for homogeneous random fields

As in the previous section we assume that π is a random field on E^S with $S = \{0, \ldots, k-1\}^2$ and k even. We also assume that the field has nearest neighbor interactions, now with *periodic* boundary. This means that addition on S, and in particular in $\partial s = \{(s_1 \pm 1, s_2), (s_1, s_2 \pm 1)\}$, is modulo k. We consider again the Gibbs sampler with (deterministic) checkerboard sweep.

Le k be a multiple of u and v, say $k = au = vb$. Let T be a horizontal translation by u sites and U a vertical translation by v sites,

$$(TX)_s = X_{(s_1-u,s_2)}, \quad (UX)_s = X_{(s_1,s_2-v)}.$$

Suppose that π is invariant under T and U. Then it is also invariant under combinations $T^\alpha U^\beta$ of these translations, defined by

$$(T^\alpha U^\beta X)_s = X_{(s_1-\alpha u,s_2-\beta v)}, \\ \alpha = 0, \ldots, a-1, \ \beta = 0, \ldots, b-1.$$

As in the previous section, let $X^0, \ldots, X^n$ be the output of the Gibbs sampler, and write $Z^0 = X^0, Z^1 = X^{k^2}, \ldots, Z^p = X^n$ for the subchain of full sweeps, and $Z^q = (Y^{2q}, Y^{2q+1})$ with Y^{2q} and Y^{2q+1} the subconfigurations on the even and odd sites, respectively. The empirical estimator based on the corresponding two-step Gibbs sampler is

$$G_n f = \frac{1}{2p} \sum_{q=1}^{p} (f(Y^{2q-1}, Y^{2q}) + f(Y^{2q}, Y^{2q+1})).$$

Greenwood et al. (1996a, Section 5) prove the following result:

The estimators $G_n f \circ T^\alpha U^\beta$, $\alpha = 0, \ldots, a-1$, $\beta = 0, \ldots, b-1$, have equal asymptotic variances, and the best linear combination is the average.

We have taken the lattice two-dimensional, but this is not essential. However, the proof makes essential use of the fact that the sampler can be written as a two-step Gibbs sampler. In particular, it does not work for random fields with local interactions more extended than nearest neighbor. The result is a special case of the following situation.

Consider a distribution π on a product space $E_1 \times E_2$. Call a transformation T on $E_1 \times E_2$ *parallel* if it maps E_1 into E_1 and E_2 into E_2, say $T(x_1, x_2) = (T_1 x_1, T_2 x_2)$, and *transverse* if it maps E_1 into E_2 and E_2 into E_1, say $T(x_1, x_2) = (T_{21} x_2, T_{12} x_1)$. Let $X^0, \ldots, X^n$ be output of the Gibbs sampler with deterministic sweep.

If π is invariant under two commuting transformations T and U, each of which is either parallel or transverse, and if $T^a = T^0$ and $U^b = U^0$, then the estimators

$$\frac{1}{n} \sum_{i=1}^{n} f(T^\alpha U^\beta X^i), \\ \alpha = 0, \ldots, a-1, \ \beta = 0, \ldots, b-1,$$

have equal asymptotic variances, and the best linear combination is the average.

The proof is based on the observation that the parallel transformation T leaves the stationary law of the Gibbs sampler invariant, while the transverse transformation U reverses time. Then one checks that the asymptotic covariance matrix of the estimators is a circulant block matrix with circulant blocks.

7 Bibliography

Barone, P. and Frigessi, A. (1989). Improving stochastic relaxation for gaussian random fields. *Probab. Engrg. Inform. Sci.* 4, 369–389.

Belisle, C. J. P., Romeijn, H. E. and Smith, R. L. (1993). Hit-and-run algorithms for generating multivariate distributions. *Math. Oper. Res.* 18, 255-266.

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

Chen, M.-H. and Schmeiser, B. W. (1996). General hit-and-run Monte Carlo sampling for evaluating multidimensional integrals. *Oper. Res. Lett.*. 19, 161–169.

Fishman, G. (1996). Coordinate selection rules for Gibbs sampling. *Ann. Appl. Probab.* 6, 444–465.

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

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

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

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

Green, P. J. and Han, X.-l. (1992). 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.), 142–164, Lecture Notes in Statistics 74, Springer-Verlag, Berlin.

Greenwood, P. E., McKeague, I. W. and Wefelmeyer, W. (1996a). Outperforming the Gibbs sampler empirical estimator for nearest neighbor random fields. *Ann. Statist.* 24, 1433–1456.

Greenwood, P. E., McKeague, I. W. and Wefelmeyer, W. (1996b). Von Mises type statistics for single site updated local interaction random fields. Technical report, Department of Statistics, Florida State University.

Greenwood, P. E., McKeague, I. W. and Wefelmeyer, W. (1997). Information bounds for Gibbs samplers. Under revision for: *Ann. Statist.*

Greenwood, P. E. and Wefelmeyer, W. (1998). Reversible Markov chains and optimality of symmetrized empirical estimators. To appear in: *Bernoulli.*

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* 14, 323–330.

Kessler, M., Schick, A. and Wefelmeyer, W. (1998). The information in the marginal law of a Markov chain. In preparation.

Liu, J. S. (1996). Peskun's theorem and a modified discrete-state Gibbs sampler. *Biometrika* 83, 681–682.

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

McEachern, S. N. and Berliner, L. M. (1994). Subsampling the Gibbs sampler. *Amer. Statist.* 48, 188–190.

McKeague, I. W. and Wefelmeyer, W. (1996). Markov chain Monte Carlo and Rao–Blackwellization. Technical report, Department of Statistics, Florida State University.

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

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

Pearl, J. (1987). Evidential reasoning using stochastic simulation of causal models. *Artificial Intelligence* 32, 245–257.

Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika* 60, 607–612.

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.* 2, 13–25.

Schick, A. and Wefelmeyer, W. (1998). Efficient estimation of invariant distributions of semiparametric Markov chain models: an outline. Technical report.

Schmeiser, B. and Chen, M.H. (1991). On random-direction Monte Carlo sampling for evaluating multidimensional integrals. Technical report, Department of Statistics, Purdue University.