2-9-2005

Fixed-Width Output Analysis for Markov Chain Monte Carlo

Galin L. Jones
School of Statistics, University of Minnesota, galin@stat.umn.edu

Murali Haran
Department of Statistics, Pennsylvania State University, mharan@stat.psu.edu

Brian S. Caffo
The Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, bcaffo@jhsph.edu

Ronald Neath
School of Statistics, University of Minnesota, rneath@stat.umn.edu

Fixed-Width Output Analysis for Markov Chain Monte Carlo

Galin L. Jones
School of Statistics
University of Minnesota
galin@stat.umn.edu

Murali Haran
Department of Statistics
Pennsylvania State University
mharan@stat.psu.edu

Brian S. Caffo
Department of Biostatistics
Johns Hopkins University
bcaffo@jhsph.edu

Ronald Neath
School of Statistics
University of Minnesota
rneath@stat.umn.edu

February 9, 2005

Abstract

Markov chain Monte Carlo is a method of producing a correlated sample in order to estimate features of a complicated target distribution via simple ergodic averages. A fundamental question in MCMC applications is when should the sampling stop? That is, when are the ergodic averages good estimates of the desired quantities? We consider a method that stops the MCMC sampling the first time the width of a confidence interval based on the ergodic averages is less than a user-specified value. Hence calculating Monte Carlo standard errors is a critical step in assessing the output of the simulation. In particular, we consider the regenerative simulation and batch means methods of estimating the variance of the asymptotic normal distribution. We describe sufficient conditions for the strong consistency and asymptotic normality of both methods and investigate their finite sample properties in a variety of examples.

1 Introduction

Suppose our goal is to calculate $E_\pi g := \int_{\mathsf{X}} g(x)\pi(dx)$ with π a probability distribution having support X and g a real-valued, π-integrable function. Also, suppose π is such that Markov chain Monte Carlo (MCMC) is the only viable method for estimating $E_\pi g$.

Let $X = \{X_0, X_1, X_2, \dots\}$ be a discrete-time, time-homogeneous, aperiodic, π-irreducible, positive Harris recurrent Markov chain with state space $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ and invariant distribution π. (See Meyn and Tweedie (1993) for definitions.) In this case, we say that X is Harris ergodic and the

Ergodic Theorem implies that, with probability 1,

$$\bar{g}_n := \frac{1}{n}\sum_{i=0}^{n-1} g(X_i) \to E_\pi g \quad \text{as } n \to \infty. \tag{1}$$

Given an MCMC algorithm that simulates X it is conceptually easy to generate large amounts of data and use $\bar{g}_n$ to obtain an arbitrarily precise estimate of $E_\pi g$.

There are several methods for deciding when n is sufficiently large; i.e., when to terminate the simulation. The simplest is to terminate the computation whenever patience runs out. This is clearly unsatisfactory since the user would not have any idea about the accuracy of $\bar{g}_n$. Alternatively, with several preliminary (and necessarily short) runs the user might be able to make an informed guess about the variability in $\bar{g}_n$ and hence make an a priori choice of n. Another alternative would be to monitor the sequence of $\bar{g}_n$ until it appears to have converged. Both of these alternatives are unsatisfactory in the sense that they are not automated and hence are inefficient uses of both user time and Monte Carlo resources. Moreover, without additional work they provide only a point estimate of $E_\pi g$.

An alternative approach is to calculate a Monte Carlo standard error and use it to terminate the simulation when the width of a confidence interval falls below some prespecified value. Under regularity conditions (that will be described in Section 2) the Markov chain X and function g will admit a central limit theorem (CLT); that is,

$$\sqrt{n}(\bar{g}_n - \mathrm{E}_\pi g) \xrightarrow{d} \mathrm{N}(0, \sigma_g^2) \tag{2}$$

as $n \to \infty$ where $\sigma_g^2 := \mathrm{var}_\pi\{g(X_0)\} + 2\sum_{i=1}^{\infty} \mathrm{cov}_\pi\{g(X_0), g(X_i)\} < \infty$. Given an estimate of σ_g^2, $\hat{\sigma}_n^2$ say, it is easy to form a confidence interval for $\mathrm{E}_\pi g$. If this interval is too large then the value of n is increased and simulation continues until the interval is sufficiently small and the simulation is terminated. Notice that this means the final Monte Carlo sample size is random. In this paper we study *fixed-width* methods which are a formalization of this approach. In particular, the simulation terminates the first time

$$t_* \frac{\hat{\sigma}_n}{\sqrt{n}} + p(n) \le \epsilon \tag{3}$$

where t_* is an appropriate quantile, $p(n) \ge 0$ on $\mathbb{Z}_+$ and $\epsilon > 0$ is the desired half-width. The role of p is to ensure that the simulation is not terminated prematurely due to a poor estimate of σ_g^2 so one possibility is to take $p(n) = I(n \le n^*)$ for some $n^* > 0$ and where I is the usual indicator function. Procedures based on (3) have been studied by Glynn and Whitt (1992) who established that these procedures are *asymptotically valid* in that if our goal is to have a $100(1-\delta)\%$ confidence interval with width 2ϵ then

$$\Pr(\mathrm{E}_\pi g \in \mathrm{Int}[T(\epsilon)]) \to 1 - \delta \quad \text{as } \epsilon \to 0 \tag{4}$$

where $T(\epsilon)$ is the first time that (3) is satisfied and $\mathrm{Int}[T(\epsilon)]$ is the interval at this time. Glynn and Whitt's conditions for asymptotic validity are substantial: (i) A functional central limit theorem

(FCLT) holds; (ii) $\hat{\sigma}_n^2 \to \sigma_g^2$ with probability 1 as $n \to \infty$; and (iii) $p(n) = o(n^{-1/2})$. There has been a large body of work done on the FCLT that indicates the Markov chains encountered in MCMC settings frequently enjoy an FCLT; see Billingsley (1968), Doukhan et al. (1994) and Kipnis and Varadhan (1986) among many others. However, in the context of MCMC, little work has been done on establishing conditions for (ii) to hold. Thus one of our goals is to give conditions under which some common methods provide strongly consistent estimators of σ_g^2. Specifically, our conditions require the sampler to be either uniformly or geometrically ergodic. The MCMC community has expended considerable effort in establishing such mixing conditions for a variety of samplers. See Jones and Hobert (2001) and Roberts and Rosenthal (1998) for some references and discussion about the implications of these mixing conditions.

We consider two methods for estimating the variance of the asymptotic normal distribution, regenerative simulation (RS) and non-overlapping batch means (BM). Both have strengths and weaknesses; essentially, BM may be easier to implement but RS is on a stronger theoretical footing. For example, it is well known that when used with fixed batch sizes BM *cannot* be even weakly consistent for σ_g^2 but is extremely easy to implement. We give conditions for the consistency of RS and show that BM can provide a consistent estimation procedure by allowing the batch sizes to increase (in a specific way) as n increases. In this case it is denoted CBM to distinguish it from the standard fixed-batch size version of BM. This has been previously addressed by Damerdji (1994) but, while the approach is similar, our regularity conditions on X are weaker. However, the regularity conditions required to obtain strong consistency of the batch means estimator are stronger than those required by RS.

The justification of fixed-width methods is entirely asymptotic and it is not at all clear how the finite sample properties of BM, CBM, and RS compare in typical MCMC settings. For this reason, we conduct a substantial simulation study in the context of two benchmark examples and two realistic examples, one of which is a complicated frequentist problem of estimating a pvalue and one which involves a high-dimensional posterior. Roughly speaking, we find that BM (with a fixed number of batches) performs poorly while RS performs slightly better than CBM.

The rest of this article is organized as follows. Section 2 fixes some notation and contains a brief discussion of some relevant Markov chain theory. In Section 3 we introduce RS and CBM and consider some of their asymptotic properties, specifically, consistency and asymptotic normality. Then in Section 4 we implement BM, CBM and RS in several examples. Some concluding remarks are given in Section 5.

2 Basic Markov Chain Theory

For $n \in \mathbb{N} := \{1, 2, 3, \ldots\}$ let $P^n(x, dy)$ be the n-step Markov transition kernel; that is, for $x \in \mathsf{X}$ and a measurable set A, $P^n(x, A) = \Pr(X_n \in A | X_0 = x)$. A Harris ergodic Markov chain X enjoys a strong form of convergence. Specifically, if $\lambda(\cdot)$ is a probability measure on $\mathcal{B}(\mathsf{X})$ then

$$\|P^n(\lambda, \cdot) - \pi(\cdot)\| \downarrow 0 \quad \text{as } n \to \infty, \tag{5}$$

where $P^n(\lambda, A) := \int_{\mathsf{X}} P^n(x, A)\lambda(dx)$ and $\|\cdot\|$ is the total variation norm. Suppose there exists an extended real-valued function $M(x)$ and a nonnegative decreasing function $\kappa(n)$ on $\mathbb{Z}_+$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\kappa(n) \ . \tag{6}$$

When $\kappa(n) = t^n$ for some $t < 1$ say X is *geometrically ergodic* if M is unbounded and *uniformly ergodic* if M is bounded. *Polynomial ergodicity of order m* where $m \geq 0$ means M may be unbounded and $\kappa(n) = n^{-m}$. These rates of convergence lead to conditions for the existence of a CLT.

Theorem 1. *Let X be a Harris ergodic Markov chain on X with invariant distribution π and suppose $g : \mathsf{X} \to \mathbb{R}$ is a Borel function. Assume one of the following conditions:*

1. *X is polynomially ergodic of order $m > 1$, $E_\pi M < \infty$ and there exists $B < \infty$ such that $|g(x)| < B$ almost surely;*
2. *X is polynomially ergodic of order m, $E_\pi M < \infty$ and $E_\pi|g(x)|^{2+\delta} < \infty$ where $m\delta > 2 + \delta$;*
3. *X is geometrically ergodic and $E_\pi|g(x)|^{2+\delta} < \infty$ for some $\delta > 0$;*
4. *X is geometrically ergodic and $E_\pi[g^2(x)(\log^+ |g(x)|)] < \infty$;*
5. *X is geometrically ergodic, reversible and $E_\pi g^2(x) < \infty$; or*
6. *X is uniformly ergodic and $E_\pi g^2(x) < \infty$.*

Then for any initial distribution, as $n \to \infty$

$$\sqrt{n}(\bar{g}_n - E_\pi g) \xrightarrow{d} N(0, \sigma_g^2) \ .$$

Remark 1. The theorem was proven by Ibragimov and Linnik (1971) (condition 6), Roberts and Rosenthal (1997) (condition 5), Doukhan et al. (1994) (condition 4) and Chan and Geyer (1994) (condition 3). See Jones (2004) for a description of conditions 1 and 2.

Remark 2. Conditions 3, 4, 5, and 6 of the theorem are also sufficient to guarantee the existence of an FCLT; see Roberts and Rosenthal (1997), Doukhan et al. (1994) and Billingsley (1968), respectively.

Remark 3. Frequently, the rate of convergence (6) is established via drift and minorization. In this case, an ordinary CLT and an FCLT can hold immediately for certain functions without verifying a moment condition (see eg. Meyn and Tweedie, 1993, Theorems 17.0.1 and 17.4.4). Jones (2004) compares this approach with Theorem 1.

Remark 4. It is well known that the mixing conditions on the Markov chain X stated in Theorem 1 are not necessary for the CLT. For minimal conditions see, for example, Chen (1999), Meyn and Tweedie (1993) and Nummelin (1984, 2002). However, the weaker conditions are often prohibitively difficult to check in situations where MCMC is appropriate. Moreover, substantial effort has been devoted to establishing convergence rates for MCMC algorithms. For example, Hobert and Geyer (1998), Jones and Hobert (2004), Marchev and Hobert (2004), Mira and Tierney (2002), Robert (1995), Roberts and Polson (1994), Roberts and Rosenthal (1999) Rosenthal (1995, 1996) and Tierney (1994) examined Gibbs samplers while Christensen et al. (2001), Douc and Soulier (2004), Fort and Moulines (2000, 2003), Geyer (1999), Jarner and Hansen (2000), Jarner and Roberts (2002), Meyn and Tweedie (1994), and Mengersen and Tweedie (1996) considered Metropolis-Hastings algorithms.

2.1 The Split Chain

An object that is important to our study of both RS and CBM is the *split chain* (Athreya and Ney, 1978; Nummelin, 1978, 1984)

$$X' := \{(X_0, \delta_0), (X_1, \delta_1), (X_2, \delta_2), \dots\}$$

which has state space $\mathsf{X} \times \{0, 1\}$. The construction of X' requires a *minorization condition*; that is, we must find a function $s : \mathsf{X} \mapsto [0, 1]$ for which $E_\pi s > 0$ and a probability measure Q such that for all $x \in \mathsf{X}$ and all measurable sets A

$$P(x, A) \geq s(x)\, Q(A). \tag{7}$$

Nummelin (1984) calls s a *small function* and Q a *small measure*. Note that (7) allows us to write $P(x, dy)$ as a mixture of two distributions,

$$P(x, dy) = s(x)\, Q(dy) + [1 - s(x)]\, R(x, dy),$$

where $R(x, dy) := [1 - s(x)]^{-1} [P(x, dy) - s(x)\, Q(dy)]$ is called the *residual* distribution (define $R(x, dy)$ as 0 if $s(x) = 1$). This mixture gives us a recipe for simulating X': given $X_i = x$, generate $\delta_i \sim \text{Bernoulli}(s(x))$. If $\delta_i = 1$, then draw $X_{i+1} \sim Q(\cdot)$, else draw $X_{i+1} \sim R(x, \cdot)$.

The two chains, X and X' are closely related since X' will inherit properties such as aperiodicity and positive Harris recurrence and, marginally, the sequence $\{X_i : i = 0, 1, \dots\}$ obtained from X' has the same transition probabilities as the original chain, X. Moreover, X and X' are co-de-initializing (Roberts and Rosenthal, 2001) and hence converge to their respective stationary distributions at exactly the same rate.

If $\delta_i = 1$, then time $i+1$ is a *regeneration time* when X' probabilistically restarts itself. Specifically, suppose we start X' with $X_0 \sim Q$; this is often easy to do, see Mykland et al. (1995) for some examples. Then each time that $\delta_i = 1$, $X_{i+1} \sim Q$. Also assume that X' is run for R tours; that is, the simulation is stopped the Rth time that a $\delta_i = 1$. Thus, the total length of the simulation, τ_R, is random. Let N_r be the length of the rth tour; that is, $N_r = \tau_r - \tau_{r-1}$ and define

$$S_r = \sum_{i=\tau_{r-1}}^{\tau_r - 1} g(X_i)$$

for $r = 1, \dots, R$. The (N_r, S_r) pairs are iid since each is based on a different tour and the Ergodic Theorem implies that

$$\bar{g}_{\tau_R} = \frac{1}{\tau_R} \sum_{j=0}^{\tau_R - 1} g(X_j) \to \mathrm{E}_\pi g$$

with probability 1 as $R \to \infty$.

3 Monte Carlo Standard Errors

3.1 Regenerative Simulation

Regenerative simulation is based on directly simulating the split chain. Let E_Q denote the expectation for the split chain started with $X_0 \sim Q(\cdot)$. Also, let $\bar{N}$ be the average tour length; that is, $\bar{N} = R^{-1} \sum_{r=1}^{R} N_r$. Since the (N_r, S_r) pairs are iid the strong law implies with probability 1 $\bar{N} \to E_Q N_1$ which is finite by positive recurrence. Also, if $E_Q N_1^2 < \infty$ and $E_Q S_1^2 < \infty$ it follows that a CLT holds as $R \to \infty$

$$\sqrt{R}(\bar{g}_{\tau_R} - \mathrm{E}_\pi g) \xrightarrow{d} \mathrm{N}\left(0, \xi_g^2\right) \tag{8}$$

where

$$\xi_g^2 = \frac{E_Q(S_1 - N_1 E_\pi g)^2}{(E_Q N_1)^2} \; . \tag{9}$$

Also, note $\xi_g^2 = \sigma_g^2 E_\pi s$ (Hobert et al., 2002). Define

$$\hat{\xi}_{RS}^2 := \frac{1}{\bar{N}^2} \frac{1}{R} \sum_{r=1}^{R} (S_r - \bar{g}_{\tau_R} N_r)^2 \tag{10}$$

and

$$\xi_*^2 := \frac{1}{\bar{N}^2} \frac{1}{R} \sum_{r=1}^{R} (S_r - N_r E_\pi g)^2 \; .$$

By comparison with ξ_*^2 it is easy to show that $\hat{\xi}_{RS}^2 \to \xi_g^2$ w.p. 1 as $R \to \infty$; see also Hobert et al. (2002). Now assume that $E_Q(S_1 - N_1 E_\pi g)^4 < \infty$ and define $v^2 := Var_Q(S_1 - N_1 E_\pi g)^2$. An application of Slutsky's theorem shows that as $R \to \infty$

$$\sqrt{R}\, \xi_*^2 \xrightarrow{d} \mathrm{N}\left(\xi_g^2 \, , \; v^2/(E_Q N_1)^4\right) \; .$$

Now consider

$$\sqrt{R}(\hat{\xi}_{RS}^2 - \xi_*^2) = \frac{1}{\bar{N}^2} \left\{ \left[(\bar{g}_{\tau_R})^2 - (E_\pi g)^2 \right] \left[\frac{1}{\sqrt{R}} \sum_{r=1}^{R} N_r^2 \right] + 2 \left[\bar{g}_{\tau_R} - E_\pi g \right] \left[\frac{1}{\sqrt{R}} \sum_{r=1}^{R} N_r S_r \right] \right\}$$

Thus Slutsky's theorem implies that as $R \to \infty$, $\sqrt{R}(\hat{\xi}^2_{RS} - \xi^2_*) \xrightarrow{d} 0$ and hence also in probability. Putting this together we have that as $R \to \infty$

$$\sqrt{R}\,\hat{\xi}^2_{RS} \xrightarrow{d} \mathrm{N}\left(\xi^2_g\,,\ v^2/(E_Q N_1)^4\right)\ . \tag{11}$$

The moment conditions assumed appear prohibitively difficult to check directly in any given application. In an attempt to alleviate this difficulty we prove the following lemma which generalizes Theorem 2 of Hobert et al. (2002).

Lemma 1. *Let X be a Harris ergodic Markov chain with invariant distribution π. Assume that (7) holds and that X is geometrically ergodic. Let $p \geq 1$ be an integer.*

1. *If $E_\pi |g|^{2(p-1)+\delta} < \infty$ for some $0 < \delta < 1$ then $E_Q N_1^p < \infty$ and $E_Q S_1^p < \infty$.*
2. *If $E_\pi |g|^{2p+\delta} < \infty$ for some $0 < \delta < 1$ then $E_Q N_1^p < \infty$ and $E_Q S_1^{p+\delta} < \infty$.*

Proof. See Appendix A. □

An application of Lemma 1 to our above work yields the following result.

Proposition 1. *Let X be a Harris ergodic Markov chain with invariant distribution π. Assume that (7) holds and that X is geometrically ergodic. Then*

1. *if $E_\pi |g|^{2+\delta} < \infty$ for some $\delta > 0$ then $\hat{\xi}^2_{RS} \to \xi^2_g$ w. p. 1 as $R \to \infty$ and*
2. *if $E_\pi |g|^{8+\delta} < \infty$ for some $\delta > 0$ then $\sqrt{R}\,\hat{\xi}^2_{RS} \xrightarrow{d} N(\xi^2_g, v^2/(E_Q N_1)^4)$ as $R \to \infty$.*

Based on these results, an asymptotically valid fixed-width procedure for estimating $E_\pi g$ results if we terminate the simulation the first time

$$z\,\frac{\hat{\xi}_{RS}}{\sqrt{R}} + p(R) \leq \epsilon \tag{12}$$

where z denotes the appropriate standard normal quantile.

Simulating the split chain in the fashion described above can be problematic since simulation from $R(x, dy)$ is challenging. Mykland et al. (1995) suggest an ingenious method for avoiding this by first drawing from the distribution of $X_{n+1}|X_n$ and then drawing from the distribution of $\delta_n|X_{n+1}, X_n$; see equation 3 on p. 235 of Mykland et al. (1995). This is the approach we use in our simulations. Further practical advice on simulating the split chain is given in Geyer and Thompson (1995), Hobert et al. (2002), Hobert et al. (2003) and Jones and Hobert (2001, 2004).

3.2 Batch Means

In standard batch means the output of the sampler is broken into batches of equal size that are assumed to be approximately independent. (This is not strictly necessary; c.f., the method of overlapping batch means.) Suppose the algorithm is run for a total of $n = ab$ iterations (hence $a = a_n$ and $b = b_n$ are implicit functions of n) and define

$$\bar{Y}_j := \frac{1}{b} \sum_{i=(j-1)b}^{jb-1} g(X_i) \; .$$

The batch means estimate of σ_g^2 is

$$\hat{\sigma}_{BM}^2 = \frac{b}{a-1} \sum_{j=1}^{a} (\bar{Y}_j - \bar{g}_n)^2 \; . \tag{13}$$

It is well known that for fixed batch sizes (13) is not a consistent estimator of σ_g^2 (Glynn and Iglehart, 1990; Glynn and Whitt, 1991). On the other hand, if the batch size and the number of batches are allowed to increase as the overall length of the simulation does it may be possible to obtain consistency. The first result in this direction is due to Damerdji (1994) whose result we now describe. The major assumption made by Damerdji (1994) is the existence of a strong invariance principle. Let $B = \{B(t), t \geq 0\}$ denote a standard Brownian motion. A strong invariance principle holds if there exists a nonnegative increasing function $\gamma(n)$ on the positive integers, a constant $0 < \sigma_g < \infty$ and a sufficiently rich probability space such that

$$\left| \sum_{i=1}^{n} g(X_i) - n\mathrm{E}_\pi g - \sigma_g B(n) \right| = O(\gamma(n)) \quad \text{w.p. 1 as } \; n \to \infty \tag{14}$$

where the w.p. 1 in (14) means for almost all sample paths. In particular, Damerdji (1994) assumed (14) held with $\gamma(n) = n^{1/2-\alpha}$ where $0 < \alpha \leq 1/2$. However, it would seem a daunting task to directly check this condition in any given application. In an attempt to somewhat alleviate this difficulty we state the following lemma.

Lemma 2. *Let $g : \mathsf{X} \to \mathbb{R}$ be a Borel function and let X be a Harris ergodic Markov chain with invariant distribution π.*

1. *If X is geometrically ergodic,* (7) *holds and $E_\pi|g|^{4+\delta} < \infty$ for some $\delta > 0$ then* (14) *holds with $\gamma(n) = n^\alpha \log n$ where $\alpha = 1/(2+\delta)$.*
2. *If X is uniformly ergodic and $E_\pi|g|^{2+\delta} < \infty$ for some $\delta > 0$ then* (14) *holds with $\gamma(n) = n^{1/2-\alpha}$ where $\alpha < \delta/(24+12\delta)$.*

Proof. The first part of the lemma follows from our Lemma 1 and Theorem 2.1 in Csáki and Csörgö (1995) whereas the second part is an immediate consequence of Theorem 4.1 of Philipp and Stout (1975) and the fact that uniformly ergodic Markov chains enjoy exponentially fast uniform mixing. □

Using part 2 of Lemma 2 we can state Damerdji's result as follows.

Proposition 2. *(Damerdji, 1994) Assume $g : \mathsf{X} \to \mathbb{R}$ such that $E_\pi|g|^{2+\delta} < \infty$ for some $0 < \delta < 1$ and let X be a Harris ergodic Markov chain with invariant distribution π. Further, suppose X is uniformly ergodic. If*

1. *$a_n \to \infty$ as $n \to \infty$,*
2. *$b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$,*
3. *$b_n^{-1} n^{1-2\lambda} \log n \to 0$ as $n \to \infty$ where $\lambda \in (0, \delta/(24 + 12\delta))$ and*
4. *there exists a constant $c \geq 1$ such that $\sum_n (b_n/n)^c < \infty$*

then as $n \to \infty$, $\hat{\sigma}^2_{BM} \to \sigma^2_g$ w. p. 1.

In Appendix B we use part 1 of Lemma 2 to extend Proposition 2 to geometrically ergodic Markov chains.

Proposition 3. *Assume $g : \mathsf{X} \to \mathbb{R}$ such that $E_\pi|g|^{4+\delta} < \infty$ for some $0 < \delta < 1$ and let X be a Harris ergodic Markov chain with invariant distribution π. Further, suppose X is geometrically ergodic. If*

1. *$a_n \to \infty$ as $n \to \infty$,*
2. *$b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$,*
3. *$b_n^{-1} n^{2\alpha} [\log n]^3 \to 0$ as $n \to \infty$ where $\alpha = 1/(2 + \delta)$ and*
4. *there exists a constant $c \geq 1$ such that $\sum_n (b_n/n)^c < \infty$*

then as $n \to \infty$, $\hat{\sigma}^2_{BM} \to \sigma^2_g$ w. p. 1.

Remark 5. There is no assumption of stationarity in Propositions 2 or 3.

Remark 6. Consider using $b_n = \lfloor n^\theta \rfloor$. Damerdji (1994) shows that in Proposition 2 it is acceptable to use $\theta > 1 - 2\lambda$ but in Proposition 3 we require $(1 + \delta/2)^{-1} < \theta < 1$.

The conditions for the asymptotic normality of $\hat{\sigma}^2_{BM}$ are more demanding than those required for RS. For example, Sherman and Goldsman (2002) show that $\hat{\sigma}^2_{BM}$ is asymptotically normal when $b_n = Kn^\theta$ for a constant K and some $1/3 < \theta < 1$, X is uniformly ergodic and $\mathrm{E}_\pi|g|^{12} < \infty$. It appears to be an open question as to whether this can be extended to geometrically ergodic case.

Under the conditions of Propositions 2 or 3 an asymptotically valid fixed-width procedure for estimating $E_\pi g$ results if we terminate the simulation the first time

$$t_{a-1} \frac{\hat{\sigma}_{BM}}{\sqrt{n}} + p(n) \leq \epsilon$$

where t_{a-1} is the appropriate quantile from a student's t distribution with $a - 1$ degrees of freedom.

3.3 Alternatives to BM and RS

We chose to focus on BM and RS since they seem to be the most commonly considered methods in MCMC. However, there are many other available methods for estimating the variance of the asymptotic distribution some of which may enjoy strong consistency; eg. see Damerdji (1991), Nummelin (2002) and Peligrad and Shao (1995). In particular, Damerdji (1991) uses a strong invariance principle to obtain strong consistency of certain spectral variance estimators under conditions similar to those required in Proposition 2. Apparently, this can be extended to geometrically ergodic chains via Lemma 2 to obtain a result with regularity conditions similar to Proposition 3. However, we do not pursue this further here.

4 Examples

In this section we investigate the finite sample performance of fixed-width methodology using RS, CBM and BM with 30 batches (BM30) in four examples. In particular, we examine the coverage probabilities and half-widths of the resulting intervals as well as the required simulation effort. While each example concerns a different statistical model and MCMC sampler there are some commonalities. In each case we perform 2000 independent replications (or runs) of the given MCMC sampler. We used all three methods on the *same* output from each replication of the MCMC sampler. When the half-width of a 95% interval with $p(n) = I(n \geq n^*)$ (or $p(R) = I(R \geq R^*)$ for RS) is less than ϵ for a particular method, that procedure was stopped and the chain length recorded. Other procedures would continue until all of them were below the targeted half-width, at which time a single replication was complete. In order to estimate the coverage probabilities we need true values of the quantities of interest. These are not available in situations where MCMC is appropriate. Our solution is to obtain very precise estimates of the truth through independent methods which are different for each example. The details are described below. A summary of the results is reported in Table 1.

4.1 A Benchmark Example

Gaver and O'Muircheartaigh (1987) present a data set concerning the failure rates of 10 pumps at a nuclear power plant, each monitored for different amounts of time. The failure counts for pump i, having been monitored for time t_i, are assumed to follow a Poisson law with a pump-specific mean $t_i\lambda_i$ and observed count y_i. A multilevel model is assumed with $\lambda_i \sim \text{Gamma}(1.802, \beta)$ and $\beta \sim \text{Gamma}(.01, 1)$. (We say $W \sim \text{Gamma}(\alpha, \beta)$ if its density is proportional to $w^{\alpha-1}e^{-\beta w}I(w > 0)$.) Let $\pi(\beta, \lambda|y)$ be the resulting posterior and consider estimating the posterior between-pump mean $\text{E}[1.802/\beta \mid y]$.

A Harris ergodic Gibbs sampler having $\pi(\beta, \lambda|y)$ as its invariant density completes a one-step transition $(\beta', \lambda') \to (\beta, \lambda)$ by simulating $\beta \sim \text{Gamma}(18.03, \sum \lambda'_i + 1)$ then each $\lambda_i \sim \text{Gamma}(1.802 + y_i, t_i + \beta)$ independently. This Gibbs sampler has been analyzed by many authors including Tierney (1994) who established that it is uniformly ergodic. Also, Mykland et al. (1995) show that if $(\beta, \lambda_1, \ldots, \lambda_{10}) \in [d_1, d_2] \times \mathbb{R}^{10}$ then the conditional probability of a regeneration is

$$\Pr(\delta = 1 | \beta, \lambda, \beta', \lambda') = \exp\left[\left\{6.7 - \sum_i \lambda'_i\right\}\left\{d_1 I\left(\sum_i \lambda'_i < 6.7\right) + d_2 I\left(\sum_i \lambda'_i \geq 6.7\right) + \beta\right\}\right],$$

where I is the usual indicator function. Following Mykland et al. (1995) we use this with $[d_1, d_2] = [1.591, 3.109]$. The implementation of CBM is simpler to describe; we set $b_n = \lfloor \sqrt{n} \rfloor$.

To obtain a gold standard, we integrated $\pi(\beta, \lambda | y)$ to get the (non-standard) posterior distribution of β and used 10^9 importance sampling simulations with a shifted and scaled student's T candidate to obtain a precise estimate of $\text{E}[1.802/\beta \mid y]$ which we assumed to be the truth.

4.2 A Hierarchical Model

Efron and Morris (1975) present a famous data set that gives the raw batting averages (based on 45 official at-bats) and a transformation ($\sqrt{45} \arcsin(2x - 1)$) for 18 Major League Baseball players during the 1970 season. Rosenthal (1996) considers the following conditionally independent hierarchical model for the transformed data. Suppose for $i = 1, \ldots, K$ that

$$\begin{aligned} Y_i | \theta_i &\sim \text{N}(\theta_i, 1) \\ \theta_i | \mu, \lambda &\sim \text{N}(\mu, \lambda) \\ \lambda \sim \text{IG}(2, 2) & \qquad f(\mu) \propto 1 \, . \end{aligned} \tag{15}$$

(Note that if $X \sim \text{Gamma}(b, c)$ then $X^{-1} \sim \text{IG}(b, c)$.) Rosenthal (1996) introduces a Harris ergodic block Gibbs sampler that has the posterior, $\pi(\theta, \mu, \lambda | y)$, characterized by the hierarchy in (15) as its invariant distribution. This Gibbs sampler completes a one-step transition $(\lambda', \mu', \theta') \to (\lambda, \mu, \theta)$ by drawing from the distributions of $\lambda | \theta'$ then $\mu | \theta', \lambda$ and subsequently $\theta | \mu, \lambda$. The full conditionals needed to implement this sampler are given by

$$\lambda | \theta, y \sim \text{IG}\left(2 + \frac{K-1}{2}, 2 + \frac{\sum (\theta_i - \bar{\theta})^2}{2}\right), \quad \mu | \theta, \lambda, y \sim \text{N}\left(\bar{\theta}, \frac{\lambda}{K}\right),$$

$$\theta_i | \lambda, \mu, y \overset{\text{ind}}{\sim} \text{N}\left(\frac{\lambda y_i + \mu}{\lambda + 1}, \frac{\lambda}{\lambda + 1}\right).$$

Rosenthal proved that the corresponding Markov chain is geometrically ergodic. However, MCMC is not required to sample from the posterior; in Appendix C we develop an accept-reject sampler that produces an iid sample from the posterior. Also in Appendix C we derive an expression for the probability of regeneration. For CBM we set $b_n = \lfloor \sqrt{n} \rfloor$.

We focus on estimating the posterior mean of θ_9, the "true" long-run (transformed) batting average of the Chicago Cubs' Ron Santo. It is straightforward to check that the moment conditions for CBM and RS are met. Finally, we employed our accept-reject sampling algorithm to generate 9×10^7 independent draws from $\pi(\theta_9|y)$ which were then used to estimate the posterior mean of θ_9 which we assumed to be the truth.

4.3 Calculating Exact Conditional pvalues

Agresti (2002, p. 432) reports data that correspond to pairs of scorings of tumor ratings by two pathologists. A linear by linear association model specifies that the log of the Poisson mean in cell i, j satisfies

$$\log \mu_{ij} = \alpha + \beta_i + \gamma_j + \delta\, ij\ .$$

A parameter free null distribution for testing goodness-of-fit is obtained by conditioning on the sufficient statistics for the parameters, ie., the margins of the table and $\sum_{ij} n_{ij}\, ij$, where the n_{ij} are the observed cell counts. The resulting conditional distribution is a generalization of the hypergeometric distribution. An exact pvalue for goodness-of-fit versus a saturated alternative can be calculated by summing the conditional probabilities of all tables satisfying the margins and the additional constraint and having deviance statistics larger than the observed.

For the current data set there are over twelve billion tables that satisfy the margin constraints but an exhaustive search revealed that there are only roughly 34,000 tables that also satisfy the constraint induced by $\sum_{ij} n_{ij}\, ij$. We will denote this set of permissible tables by Γ. Now the desired pvalue is given by

$$\sum_{y \in \Gamma} I[d(y) \geq d(y_{obs})]\, \pi(y) \tag{16}$$

where $d(\cdot)$ is the deviance function and π denotes the generalized hypergeometric. Since we have enumerated Γ we find that the true exact pvalue is .044 whereas the chi-squared approximation yields a pvalue of .368. However, if we were given a different data set with different values of the sufficient statistics then we would have a different reference set which would need to be enumerated in order to find the exact pvalue. This would be too computationally burdensome to implement generally and hence it is common to resort to MCMC-based approximations (see eg. Caffo and Booth, 2001; Diaconis and Sturmfels, 1998; Forster et al., 1996).

To estimate (16) we will use the Metropolis-Hastings algorithm developed in Caffo and Booth (2001). This algorithm is also employed by the R package `exactLoglinTest`. It is easy to see that the associated Markov chain is Harris ergodic and its invariant distribution is the appropriate generalized hypergeometric distribution. Moreover, the chain is uniformly ergodic and since we are estimating the expectation of a bounded function the regularity conditions for both RS and CBM are easily met.

For CBM we set $b_n = \lfloor \sqrt{n} \rfloor$. Our implementation of RS requires more explanation. In finite state spaces regenerations occur whenever the chain returns to any fixed state; for example, when the Metropolis-Hastings chain accepts a move to the fixed state. This regeneration scheme is useful when the state space is small but potentially complicated. It will not be useful when the state space is extremely large because returns to the fixed state are too infrequent. In order to choose the fixed state we ran the algorithm for 1000 iterations and chose the state which had the highest probability with respect to the stationary distribution. The same fixed state was used in each of the 2000 replications.

4.4 A Model-Based Spatial Statistics Application

Consider the well-known Scottish lip cancer data set (Clayton and Kaldor, 1987) which consists of the number of cases of lip cancer registered in each of the 56 (pre-reorganization) counties of Scotland, together with the expected number of cases given the age-sex structure of the population. Following the work of Besag et al. (1991) we assume a Poisson likelihood for areal (spatially aggregated) data. Specifically, for $i = 1, ..., N$ we assume that given μ_i the disease counts Y_i are conditionally independent and

$$Y_i | \mu_i \sim \text{Poi}(E_i e^{\mu_i}) \tag{17}$$

where E_i is the known 'expected' number of disease events in the ith region assuming constant risk and μ_i is the log-relative risk of disease for the ith region. Each μ_i is modeled linearly as $\mu_i = \theta_i + \phi_i$ where

$$\theta_i | \tau_h \sim \text{ N}(0, 1/\tau_h), \qquad \phi | \tau_c \sim \text{ CAR}(\tau_c) \propto \tau_c^{M/2} \exp\left(-\frac{\tau_c}{2}\phi^T Q \phi\right),$$

where $\phi = (\phi_1, \ldots, \phi_N)^T$ and

$$Q_{ij} = \begin{cases} n_i & \text{if } i = j \\ 0 & \text{if } i \text{ is not adjacent to } j \\ -1 & \text{if } i \text{ is adjacent to } j \end{cases}$$

with n_i is the number of neighbors for the ith region. Each θ_i captures the ith region's extra-Poisson variability due to area-wide heterogeneity, while each ϕ_i captures the ith region's excess variability attributable to regional clustering. The priors on the precision parameters are

$$\tau_h \sim \text{Gamma}(1, .01), \qquad \tau_c \sim \text{Gamma}(1, .02)\,.$$

This is a challenging model to consider since the random effects parameters (θ_i, ϕ_i) are not identified in the likelihood, and the spatial prior used is improper. Also, no closed form expressions are available for the marginal distributions of the parameters, and the posterior distribution has $2N + 2$ dimensions (114 for the lip cancer data) making drawing random samples from the posterior difficult, at best.

Haran and Tierney (2004) describe a Harris ergodic independence Metropolis-Hastings sampler with invariant distribution $\pi(\theta, \phi, \tau_h, \tau_c | y)$ and joint proposal distribution $R(\theta, \phi, \tau_h, \tau_c)$ where $\theta =$

$(\theta_1, \ldots, \theta_N)^T$. Haran and Tierney (2004) establish that R dominates π by showing there exists $B > 0$, such that

$$\frac{\pi(\theta, \phi, \tau_h, \tau_c)}{R(\theta, \phi, \tau_h, \tau_c)} \leq B, \quad \text{for } \theta \in R^N, \phi \in R^N, \tau_h, \tau_c > 0$$

and hence this sampler is uniformly ergodic (Mengersen and Tweedie, 1996). In our implementation of RS we used the formula for the probability of a regeneration for independence samplers given in Mykland et al. (1995) while for for CBM we used $b_n = \lfloor \sqrt{n} \rfloor$.

We focus on estimating the posterior expectation of ϕ_7, the log-relative risk of disease for County 7 attributable to spatial clustering. It is straightforward to check that the moment conditions for CBM and RS are met. Finally, we used an independent run of length 10^7 to obtain an estimate which we treated as the 'true value'.

4.5 Summary

The results presented in Table 1 reveal that the estimated coverage probabilities for all of the procedures is less than the desired .95. However, only BM30 is significantly less in all of the examples. While CBM has higher estimated coverage than BM30 it is significantly lower than the nominal level in 3 out of the 4 examples. On the other hand, the coverage probability for RS is *not* significantly different from .95 in 3 out of 4. The example in subsection 4.3 deserves to be singled out due to the low estimated coverage probabilities. The goal in this example was to estimate a fairly small probability, a situation in which the Wald interval is known to have poor coverage even in iid settings. We suspect that the trouble in subsection 4.3 was due to the use of the Wald interval rather than the use of CBM, BM30 or RS.

While RS appears superior in terms of coverage probability it tends to result in slightly longer runs than CBM which in turn results in longer runs than BM30. Moreover, RS appears to result in intervals that meet the target half-width more closely than CBM which in turn appears to do a better job at this than BM30. Also, the intervals for RS are apparently more stable than those of CBM and BM30.

Based on our experience, it would be hard to recommend BM30 since it appears to underestimate the Monte Carlo standard error and therefore suggests stopping the chain too early. Also, the finite sample properties of RS seem to be slightly better than those of CBM.

5 Concluding Remarks

While we would generally recommend RS as the preferred procedure due to its (slight) theoretical and (slight) empirical advantages, CBM clearly has a place in the tool kit of MCMC users. We believe the more important distinction is between consistent estimation methods such as CBM and RS and inconsistent methods such as BM30. In part this is because none of these techniques will improve the situation if a poorly mixing sampler is used: think "garbage in, garbage out."

Finally, we come to an issue which is usually addressed only informally in most MCMC-based investigations. Using a stopping rule based on just a single parameter of interest may not be appropriate for a multidimensional distribution. Designing multidimensional stopping rules would be a useful area of future research since most settings where MCMC is useful are multidimensional. However, consistently estimating an asymptotic covariance matrix appears difficult. In particular, it poses practical challenges as monitoring all parameters can be extremely inefficient and may not even be the optimal use of resources. We believe that given computational constraints and the lack of theoretical work in this area, the methodology we describe here is useful and represents one more positive step towards automating the decision of stopping chains for MCMC-based inference.

A Proof of Lemma 1

A.1 Preliminary Results

We first recall a few results that will be useful in proving the claim. Recall the split chain and that $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ denote the regeneration times; i.e., $\tau_{r+1} = \min\{i > \tau_r : \delta_{i-1} = 1\}$.

Lemma 3. *(Hobert et al., 2002, Lemma 1) Let X be a Harris ergodic Markov chain and assume that (7) holds. Then for any function $h : \mathsf{X}^\infty \to \mathbb{R}$*

$$E_\pi |h(X_0, X_1, \ldots)| \geq c E_Q |h(X_0, X_1, \ldots)|$$

where $c = E_\pi s$.

Lemma 4. *(Hobert et al., 2002, Lemma 2) Let X be a Harris ergodic Markov chain and assume that (7) holds. If X is geometrically ergodic, then there exists a $\beta > 1$ such that $E_\pi \beta^{\tau_1} < \infty$.*

It is easy to see that Lemma 4 implies the following result:

Corollary 1. *Assume the conditions of Lemma 4. For any $a > 0$*

$$\sum_{i=0}^{\infty} [Pr_\pi(\tau_1 \geq i+1)]^a \leq (E_\pi \beta^{\tau_1})^a \sum_{i=0}^{\infty} \beta^{-a(i+1)} < \infty \, .$$

A.2 Proof of Lemma 1

We will prove only part 2 of the lemma as part 1 is similar. By Lemma 3, it is enough to verify that $\mathrm{E}_\pi \tau_1^p < \infty$ and $\mathrm{E}_\pi S_1^{p+\delta} < \infty$. Lemma 4 shows that $\mathrm{E}_\pi \tau_1^p < \infty$ for any $p > 0$. Note that

$$\left(\sum_{i=0}^{\tau_1 - 1} |f(X_i)|\right)^{p+\delta} = \left(\sum_{i=0}^{\infty} I(0 \leq i \leq \tau_1 - 1)|f(X_i)|\right)^{p+\delta}$$
$$\leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} \sum_{i_{p+1}=0}^{\infty} \left[\prod_{j=1}^{p} I(0 \leq i_j \leq \tau_1 - 1)|f(X_{i_j})|\right] I(0 \leq i_{p+1} \leq \tau_1 - 1)|f(X_{i_{p+1}})|^\delta$$

and hence

$$
\begin{aligned}
&\mathrm{E}_\pi\left(\sum_{i=0}^{\tau_1-1}|f(X_i)|\right)^{p+\delta}\\
&\le \sum_{i_1=0}^{\infty}\cdots\sum_{i_p=0}^{\infty}\sum_{i_{p+1}=0}^{\infty}\mathrm{E}_\pi\left(\left[\prod_{j=1}^{p+1}I(0\le i_j\le\tau_1-1)\right]\left[\prod_{j=1}^{p}|f(X_{i_j})|\right]|f(X_{i_{p+1}})|^{\delta}\right)\\
&\le \sum_{i_1=0}^{\infty}\cdots\sum_{i_p=0}^{\infty}\sum_{i_{p+1}=0}^{\infty}\left[\mathrm{E}_\pi I(0\le i_1\le\tau_1-1)|f(X_{i_1})|^2\right]^{1/2}\times\cdots\times\left[\mathrm{E}_\pi I(0\le i_p\le\tau_1-1)|f(X_{i_p})|^{2^p}\right]^{1/2^p}\times\\
&\qquad\qquad\times\left[\mathrm{E}_\pi I(0\le i_{p+1}\le\tau_1-1)|f(X_{i_{p+1}})|^{2^p\delta}\right]^{1/2^p}
\end{aligned}
$$

where the second inequality follows with repeated application of Cauchy-Schwartz. Set $a_j = 1+2^j/\delta$ and $b_j = 1+\delta/2^j$ for $j = 1, 2, \ldots, p$ and apply Hölder's inequality to obtain

$$\mathrm{E}_\pi I(0\le i_j\le\tau_1-1)|f(X_{i_j})|^{2^j}\le\left[\mathrm{E}_\pi I(0\le i_j\le\tau_1-1)\right]^{1/a_j}\left[\mathrm{E}_\pi|f(X_{i_j})|^{2^j+\delta}\right]^{1/b_j}.$$

Note that

$$\left[\left(\mathrm{E}_\pi|f(X_{i_j})|^{2^j+\delta}\right)^{1/b_j}\right]^{1/2^p} := c_j < \infty .$$

Also, if $a_{p+1} = 1+2^p$ and $b_{p+1} = 1+1/2^p$ then

$$\mathrm{E}_\pi I(0\le i_{p+1}\le\tau_1-1)|f(X_{i_{p+1}})|^{2^p\delta}\le\left[\mathrm{E}_\pi I(0\le i_{p+1}\le\tau_1-1)\right]^{1/a_{p+1}}\left[\mathrm{E}_\pi|f(X_{i_{p+1}})|^{\delta(2^p+\delta)}\right]^{1/b_{p+1}}.$$

Notice that

$$c_{p+1} := \left[\left(\mathrm{E}_\pi|f(X_{i_{p+1}})|^{\delta(2^p+\delta)}\right)^{1/b_j}\right]^{1/2^p} < \infty$$

and set $c = \max\{c_1, \ldots, c_{p+1}\}$. Then

$$\mathrm{E}_\pi\left(\sum_{i=0}^{\tau_1-1}|f(X_i)|\right)^{p+\delta}\le c\left[\prod_{j=1}^{p}\sum_{i_j=0}^{\infty}\{\Pr_\pi(\tau_1\ge i_j+1)\}^{1/(a_j2^j)}\right]\left[\sum_{i_{p+1}=0}^{\infty}\{\Pr_\pi(\tau_1\ge i_j+1)\}^{1/(a_{p+1}2^p)}\right]$$

Now an appeal to Corollary 1 yields the result.

B Proof of Proposition 3

B.1 Preliminary Results

Recall that $B = \{B(t), t \ge 0\}$ denotes a standard Brownian motion. Define

$$\tilde{\sigma}_*^2 = \frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}\left(\bar{B}_j(b_n)-\bar{B}(n)\right)^2 \tag{18}$$

where

$$\bar{B}_j(b_n) = \frac{1}{b_n}\left(B((j+1)b_n)-B(jb_n)\right) \quad \text{and} \quad \bar{B}(n) = \frac{1}{n}B(n) .$$

Lemma 5. *(Damerdji, 1994, p. 508) For all $\epsilon > 0$ and for almost all sample paths there exists $n_0(\epsilon)$ such that for all $n \geq n_0$*

$$|\bar{B}_j(b_n)| \leq \sqrt{2}(1+\epsilon)b_n^{-1/2}[\log(n/b_n) + \log\log n]^{1/2} \ . \tag{19}$$

Lemma 6. *(Csörgő and Révész, 1981) For all $\epsilon > 0$ and for almost all sample paths there exists $n_0(\epsilon)$ such that for all $n \geq n_0$*

$$|B(n)| < (1+\epsilon)[2n \log\log n]^{1/2} \ . \tag{20}$$

B.2 Proof of Proposition 3

Proposition 3 follows from Lemma 2, Lemma 1 and the following two lemmas:

Lemma 7. *(Damerdji, 1994, Proposition 3.1) Assume*

1. *$b_n \to \infty$ and $n/b_n \to \infty$ as $n \to \infty$ and*
2. *there exists a constant $c \geq 1$ such that $\sum_n (b_n/n)^c < \infty$*

then as $n \to \infty$, $\tilde{\sigma}_*^2 \to 1$ a.s.

Lemma 8. *Assume that* (14) *holds with $\gamma(n) = n^\alpha \log n$ where $\alpha = 1/(2+\delta)$. If*

1. *$a_n \to \infty$ as $n \to \infty$,*
2. *$b_n \to \infty$ and $n/b_n \to \infty$ as $n \to \infty$ and*
3. *$b_n^{-1} n^{2\alpha} [\log n]^3 \to 0$ as $n \to \infty$ where $\alpha = 1/(2+\delta)$*

then as $n \to \infty$, $\hat{\sigma}_{BM}^2 - \sigma_g^2 \tilde{\sigma}_*^2 \to 0$ a.s.

Proof. We begin with a preliminary matter. Define $h(x) = \frac{(\log x)^2}{x^\gamma}$ for $x > 0$ and $\gamma > 0$. Then

$$h'(x) = [2 - \gamma \log x] \frac{\log x}{x^{\gamma+1}} \ .$$

and hence $h' < 0$ if either $0 < x < 1$ or $x > e^{2/\gamma}$. For sufficiently large x, h is then a decreasing function and, in fact, $h(x) \to 0$ as $x \to \infty$.

Recall that $X = \{X_1, X_2, \ldots\}$ is a Harris ergodic Markov chain. Define the process Y by $Y_i = g(X_i) - \mathrm{E}_\pi g$ for $i = 1, 2, 3, \ldots$. Then

$$\hat{\sigma}_{BM}^2 = \frac{b_n}{a_n - 1} \sum_{j=0}^{a_n - 1} \left(\bar{Y}_j(b_n) - \bar{Y}(n)\right)^2$$

where

$$\bar{Y}_j(b_n) = \frac{1}{b_n}\sum_{i=1}^{b_n} Y_{jb_n+i} \quad \text{for} \quad j = 0, \ldots, a_n - 1$$

and

$$\bar{Y}(n) = \frac{1}{n}\sum_{i=1}^{n} Y_i \; .$$

Since

$$\bar{Y}_j(b_n) - \bar{Y}(n) = \bar{Y}_j(b_n) - \bar{Y}(n) \pm \sigma\bar{B}_j(b_n) \pm \sigma\bar{B}(n)$$

we have

$$\begin{aligned}\left|\hat{\sigma}^2_{BM} - \sigma^2\tilde{\sigma}^2_{BM}\right| \leq \frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1} \Big[&(\bar{Y}_j(b_n) - \sigma\bar{B}_j(b_n))^2 + (\bar{Y}(n) - \sigma\bar{B}(n))^2 \\ &+ |2(\bar{Y}_j(b_n) - \sigma\bar{B}_j(b_n))(\bar{Y}(n) - \sigma\bar{B}(n))| + |2\sigma(\bar{Y}_j(b_n) - \sigma\bar{B}_j(b_n))\bar{B}_j(b_n)| \\ &+ |2\sigma(\bar{Y}_j(b_n) - \sigma\bar{B}_j(b_n))\bar{B}(n)| + |2\sigma(\bar{Y}(n) - \sigma\bar{B}(n))\bar{B}_j(b_n)| \\ &+ |2\sigma(\bar{Y}(n) - \sigma\bar{B}(n))\bar{B}(n)|\Big] \; .\end{aligned}$$

Now we will consider each term in the sum and show that it tends to 0.

1. First, recall that (14) implies that there exists a constant C such that for all n

$$\left|\sum_{i=1}^{n} g(X_i) - n\mathrm{E}_\pi g - \sigma B(n)\right| < Cn^\alpha \log n \quad a.s. \tag{21}$$

   Note that

$$\bar{Y}_j(b_n) - \sigma\bar{B}_j(b_n) = \frac{1}{b_n}\left[\sum_{i=1}^{(j+1)b_n} Y_i - \sigma B((j+1)b_n)\right] - \frac{1}{b_n}\left[\sum_{i=1}^{jb_n} Y_i - \sigma B(jb_n)\right]$$

   and hence by (21)

$$|\bar{Y}_j(b_n) - \sigma\bar{B}_j(b_n)| \leq \frac{1}{b_n}\left[|\sum_{i=1}^{(j+1)b_n} Y_i - \sigma B((j+1)b_n)| + |\sum_{i=1}^{jb_n} Y_i - \sigma B(jb_n)|\right] < \frac{2}{b_n}Cn^\alpha \log n \tag{22}$$

   Then

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}(\bar{Y}_j(b_n) - \sigma\bar{B}_j(b_n))^2 < 4C^2\frac{a_n}{a_n-1}b_n^{-1}n^{2\alpha}(\log n)^2 \;\to 0$$

   as $n \to \infty$ by conditions 1 and 3.

2. Apply (21) to obtain

$$|\bar{Y}(n) - \sigma\bar{B}(n)| = \frac{1}{n}|\sum_{i=1}^{n} Y_i - \sigma B(n)| < Cn^{\alpha-1}\log n \; . \tag{23}$$

Then

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}(\bar{Y}(n)-\sigma\bar{B}(n))^2 < C^2\frac{a_n}{a_n-1}\frac{b_n}{n}\frac{(\log n)^2}{n^{1-2\alpha}} \to 0$$

as $n \to \infty$ by conditions 1 and 2 and since $1-2\alpha > 0$.

3. By (22) and (23)

$$|2(\bar{Y}_j(b_n)-\sigma\bar{B}_j(b_n))(\bar{Y}(n)-\sigma\bar{B}(n))| < 2C^2 b_n^{-1} n^{2\alpha-1}(\log n)^2 .$$

Thus

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}|2(\bar{Y}_j(b_n)-\sigma\bar{B}_j(b_n))(\bar{Y}(n)-\sigma\bar{B}(n))| < 4C^2\frac{a_n}{a_n-1}\frac{(\log n)^2}{n^{1-2\alpha}} \to 0$$

as $n \to \infty$ by condition 1 and since $1-2\alpha > 0$.

4. Since $b_n \geq 2$, (19) and (22) together imply

$$|(\bar{Y}_j(b_n)-\sigma\bar{B}_j(b_n))\bar{B}_j(b_n)| < 2^{3/2}C(1+\epsilon)b_n^{-1}\left[b_n^{-1}n^{2\alpha}(\log n)^2\log(n/b_n)+b_n^{-1}n^{2\alpha}(\log n)^2\log\log n\right]^{1/2}$$

Hence

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}|2\sigma(\bar{Y}_j(b_n)-\sigma\bar{B}_j(b_n))\bar{B}_j(b_n)| \leq 4\sigma C(1+\epsilon)\frac{a_n}{a_n-1}\left[b_n^{-1}n^{2\alpha}(\log n)^2\log(n/b_n)\right.$$
$$\left.+\, b_n^{-1}n^{2\alpha}(\log n)^2\log\log n\right]^{1/2} \to 0$$

as $n \to \infty$ by conditions 1 and 3.

5. By (22) and (20)

$$|(\bar{Y}_j(b_n)-\sigma\bar{B}_j(b_n))\bar{B}(n)| < 4C(1+\epsilon)b_n^{-1}\frac{(\log n)(\log\log n)^{1/2}}{n^{1/2-\alpha}}$$

so that

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}|2\sigma(\bar{Y}_j(b_n)-\sigma\bar{B}_j(b_n))\bar{B}(n)| < 8\sigma C(1+\epsilon)\frac{a_n}{a_n-1}\frac{(\log n)(\log\log n)^{1/2}}{n^{1/2-\alpha}} \to 0$$

as $n \to \infty$ by condition 1 and since $1/2-\alpha > 0$.

6. Use (19), (23) and that $b_n \geq 2$ to get

$$|(\bar{Y}(n)-\sigma\bar{B}(n))\bar{B}_j(b_n)| < \sqrt{2}C(1+\epsilon)\frac{n^{\alpha-1}\log n}{\sqrt{b_n}}\left[\log(n/b_n)+\log\log n\right]^{1/2}$$

and hence

$$\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}|2\sigma(\bar{Y}(n)-\sigma\bar{B}(n))\bar{B}_j(b_n)| < 2^{3/2}\sigma C(1+\epsilon)\frac{a_n}{a_n-1}\frac{b_n}{n}\left[b_n^{-1}n^{2\alpha}(\log n)^2\log(n/b_n)+\right.$$
$$\left.+(\log n)^2\log\log n\right]^{1/2} \to 0$$

as $n \to \infty$ by conditions 1, 2 and 3.

7. Now (20) and (23) imply

$$|(\bar{Y}(n) - \sigma\bar{B}(n))\bar{B}(n)| < 2C(1+\epsilon)\frac{(\log n)^{3/2}}{n^{3/2-\alpha}} \ .$$

Hence

$$\frac{b_n}{a_n - 1}\sum_{j=0}^{a_n-1}|2\sigma(\bar{Y}(n) - \sigma\bar{B}(n))\bar{B}(n)| < 4C(1+\epsilon)\frac{a_n}{a_n-1}\frac{b_n}{n}\frac{(\log n)^{3/2}}{n^{1/2-\alpha}} \quad \to 0$$

as $n \to \infty$ by conditions 1 and 2 and since $1/2 - \alpha > 0$.

This completes the proof of the lemma. □

C Calculations for Example 4.2

We consider a slightly more general formulation of the model given in (15). Suppose for $i = 1, \dots, K$

$$\begin{aligned} Y_i|\theta_i &\sim \mathrm{N}(\theta_i, a) \\ \theta_i|\mu,\lambda &\sim \mathrm{N}(\mu,\lambda) \\ \lambda \sim \mathrm{IG}(b,c) & \qquad f(\mu) \propto 1 \ . \end{aligned} \tag{24}$$

where a, b, c are all known positive constants.

C.1 Sequential Sampling from $\pi(\theta, \mu, \lambda|y)$

Let $\pi(\theta, \mu, \lambda|y)$ be the posterior distribution corresponding to the hierarchy in (24). Note that θ is a vector containing all of the θ_i and that y is a vector containing all of the data. Consider the factorization

$$\pi(\theta,\mu,\lambda|y) = \pi(\theta|\mu,\lambda,y)\pi(\mu|\lambda,y)\pi(\lambda|y). \tag{25}$$

The factorization given in (25) suggests that if it is possible to sequentially simulate from each of the three densities on the right-hand side of (25) we can produce iid draws from the posterior. Now $\pi(\theta|\mu,\lambda,y)$ is the product of independent univariate normal densities, i.e. $\theta_i|\mu,\lambda,y \sim \mathrm{N}((\lambda y_i + a\mu)/(\lambda+a), a\lambda/(\lambda+a))$. Also, $\pi(\mu|\lambda,y)$ is a normal distribution, ie. $\mu|\lambda,y \sim \mathrm{N}(\bar{y}, (\lambda+a)/K)$. Next

$$\pi(\lambda|y) \propto \frac{1}{\lambda^{b+1}(\lambda+a)^{(K-1)/2}}e^{-c/\lambda - s^2/2(\lambda+a)}$$

where $\bar{y} = K^{-1}\sum_{i=1}^{K} y_i$ and $s^2 = \sum_{i=1}^{K}(y_i - \bar{y})^2$. An accept-reject sampler with an $\mathrm{IG}(b,c)$ candidate can be used to sample from $\pi(\lambda|y)$ since if we let $g(\lambda)$ be the kernel of an $\mathrm{IG}(b,c)$ density

$$\sup_{\lambda\geq 0}\frac{1}{g(\lambda)\lambda^{b+1}(\lambda+a)^{(K-1)/2}}e^{-c/\lambda - s^2/2(\lambda+a)} = \sup_{\lambda\geq 0}(\lambda+a)^{(1-K)/2}e^{-s^2/2(\lambda+a)} = M < \infty$$

It is easy to show that the only critical point is $\hat{\lambda} = s^2/(K-1) - a$ which is where the maximum occurs if $\hat{\lambda} > 0$. But if $\hat{\lambda} \leq 0$ then the maximum occurs at 0.

C.2 Implementing regenerative simulation

We begin by establishing the minorization condition (7) for Rosenthal's (1996) block Gibbs sampler. For the one-step transition $(\lambda', \mu', \theta') \to (\lambda, \mu, \theta)$ the Markov transition density, p, is given by

$$p(\lambda, \mu, \theta | \lambda', \mu', \theta') = f(\lambda, \mu | \theta') f(\theta | \lambda, \mu)$$

Note that $\mathsf{X} = \mathbb{R}^+ \times \mathbb{R}^1 \times \mathbb{R}^K$. Fix a point $(\tilde{\lambda}, \tilde{\mu}, \tilde{\theta}) \in \mathsf{X}$ and let $D \subseteq \mathsf{X}$. Then

$$\begin{aligned}
p(\lambda, \mu, \theta | \lambda', \mu', \theta') &= f(\lambda, \mu | \theta') f(\theta | \lambda, \mu) \\
&\geq f(\lambda, \mu | \theta') f(\theta | \lambda, \mu) I_{\{(\lambda,\mu,\theta) \in D\}} \\
&= \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} f(\lambda, \mu | \tilde{\theta}) f(\theta | \lambda, \mu) I_{\{(\lambda,\mu,\theta) \in D\}} \\
&\geq \left\{ \inf_{(\lambda,\mu,\theta) \in D} \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} \right\} f(\lambda, \mu | \tilde{\theta}) f(\theta | \lambda, \mu) I_{\{(\lambda,\mu,\theta) \in D\}}
\end{aligned}$$

and hence (7) will follow by setting

$$\varepsilon = \int_D f(\lambda, \mu | \tilde{\theta}) f(\theta | \lambda, \mu) \, d\lambda \, d\mu \, d\theta,$$

$$s(\lambda', \mu', \theta') = \varepsilon \inf_{(\lambda,\mu,\theta) \in D} \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} \quad \text{and} \quad q(\lambda, \mu, \theta) = \varepsilon^{-1} f(\lambda, \mu | \tilde{\theta}) f(\theta | \lambda, \mu) I_{\{(\lambda,\mu,\theta) \in D\}}.$$

Now using equation 3 on p.235 of Mykland et al. (1995) shows that when $(\lambda, \mu, \theta) \in D$ the probability of regeneration is given by

$$\Pr(\delta = 1 | \lambda', \mu', \theta', \lambda, \mu, \theta) = \left\{ \inf_{(\lambda,\mu,\theta) \in D} \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} \right\} \frac{f(\lambda, \mu | \tilde{\theta})}{f(\lambda, \mu | \theta')} \tag{26}$$

Thus we need to calculate the infimum and plug into (26). To this end let $0 < d_1 < d_2 < \infty$, $-\infty < d_3 < d_4 < \infty$ and set $D = [d_1, d_2] \times [d_3, d_4] \times \mathbb{R}^K$. Define $V(\theta, \mu) = \sum_{i=1}^K (\theta_i - \mu)^2$ and note that

$$\inf_{(\lambda,\mu,\theta) \in D} \frac{f(\lambda, \mu | \theta')}{f(\lambda, \mu | \tilde{\theta})} = \inf_{\lambda \in [d_1, d_2], \ \mu \in [d_3, d_4]} \exp\left\{ \frac{V(\tilde{\theta}, \mu) - V(\theta', \mu)}{2\lambda} \right\} = \exp\left\{ \frac{V(\tilde{\theta}, \hat{\mu}) - V(\theta', \hat{\mu})}{2\hat{\lambda}} \right\}$$

where $\hat{\mu}$ and $\hat{\lambda}$ are given by

$$\hat{\mu} = \begin{cases} d_4 & \bar{\theta}' \leq \bar{\tilde{\theta}} \\ d_3 & \bar{\theta}' > \bar{\tilde{\theta}} \end{cases}$$

with $\bar{\theta} = \frac{1}{K} \sum_{i=1}^K \theta_i$, and

$$\hat{\lambda} = \begin{cases} d_2 & V(\theta'; \hat{\mu}) \leq V(\tilde{\theta}; \hat{\mu}) \\ d_1 & V(\theta'; \hat{\mu}) > V(\tilde{\theta}; \hat{\mu}) \end{cases}$$

We find the fixed point with a preliminary estimate of the mean of the stationary distribution, and D to be centered at that point. Let $(\tilde{\lambda}, \tilde{\mu}, \tilde{\theta})$ be the ergodic mean for a preliminary Gibbs sampler run, and let S_λ and S_μ denote the usual sample standard deviations of λ and μ respectively. After some trial and error we took

$$d_1 = \max\left\{.01, \tilde{\lambda} - .5S_\lambda\right\}, \quad d_2 = \tilde{\lambda} + .5S_\lambda, \quad d_3 = \tilde{\mu} - S_\mu \quad \text{and} \quad d_4 = \tilde{\mu} + S_\mu.$$

References

Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York.

Athreya, K. B. and Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, 245:493–501.

Besag, J., York, J., and Mollie, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43:1–20.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Brockwell, A. E. and Kadane, J. B. (2002). Identification of regeneration times in MCMC simulation, with application to adaptive schemes. Technical report, Carnegie Mellon, Department of Statistics.

Caffo, B. S. and Booth, J. G. (2001). A Markov chain Monte Carlo algorithm for approximating exact conditional probabilities. *Journal of Compuatational and Graphical Statistics*, 10:730–745.

Chan, K. S. and Geyer, C. J. (1994). Comment on "Markov chains for exploring posterior distributions". *The Annals of Statistics*, 22:1747–1758.

Chen, X. (1999). Limit theorems for functionals of ergodic Markov chains with general state space. *Memoirs of the American Mathematical Society*, 139.

Christensen, O. F., Moller, J., and Waagepetersen, R. P. (2001). Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalized linear mixed models. *Methodology and Computing in Applied Probability*, 3:309–327.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43:671–681.

Csáki, E. and Csörgö, M. (1995). On additive functionals of Markov chains. *Journal of Theoretical Probability*, 8:905–919.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press.

Damerdji, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37:1424–1440.

Damerdji, H. (1994). Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research*, 19:494–512.

Diaconis, P. and Sturmfels, B. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of Statistics*, 26:363–397.

Douc, R., F. G. M. E. and Soulier, P. (2004). Practical drift conditions for subgeometric rates of convergence. *The Annals of Applied Probability*, 14:1353–1377.

Doukhan, P., Massart, P., and Rio, E. (1994). The functional central limit theorem for strongly mixing processes. *Annales de l'Institut Henri Poincare, Section B, Calcul des Probabilities et Statistique*, 30:63–82.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70:311–319.

Forster, J. J., McDonald, J. W., and Smith, P. W. F. (1996). Monte Carlo exact conditional tests for log-linear and logistic models. *Journal of the Royal Statistical Society, Series B, Methodological*, 58:445–453.

Fort, G. and Moulines, E. (2000). V-subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statistics and Probability Letters*, 49:401–410.

Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Processes and their Applications*, 103:57–99.

Gaver, D. P. and O'Muircheartaigh, I. G. (1987). Robust empirical Bayes analyses of event rates. *Technometrics*, 29:1–15.

Geyer, C. J. (1999). Likelihood inference for spatial point processes. In Barndorff-Nielsen, O. E., Kendall, W. S., and van Lieshout, M. N. M., editors, *Stochastic Geometry: Likelihood and Computation*, pages 79–140. Chapman & Hall/CRC, Boca Raton.

Geyer, C. J. and Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90:909–920.

Glynn, P. W. and Iglehart, D. L. (1990). Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15:1–16.

Glynn, P. W. and Whitt, W. (1991). Estimating the asymptotic variance with batch means. *Operations Research Letters*, 10:431–435.

Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2:180–198.

Haran, M. and Tierney, L. (2004). Perfect sampling for a Bayesian spatial model. Technical report, Pennsylvania State University, Department of Statistics.

Hobert, J. P. and Geyer, C. J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis*, 67:414–430.

Hobert, J. P., Jones, G. L., Presnell, B., and Rosenthal, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, 89:731–743.

Hobert, J. P., Jones, G. L., and Robert, C. P. (2003). Using a Markov chain to construct a tractable approximation of an intractable probability distribution. Technical report, University of Minnesota, School of Statistics.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Walters-Noordhoff, The Netherlands.

Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Processes and Their Applications*, 85:341–361.

Jarner, S. F. and Roberts, G. O. (2002). Polynomial convergence rates of Markov chains. *Annals of Applied Probability*, 12:224–247.

Jones, G. L. (2004). On the Markov chain central limit theorem. *Probability Surveys*, 1:299–320.

Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, 16:312–334.

Jones, G. L. and Hobert, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32:784–817.

Kipnis, C. and Varadhan, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104:1–19.

Marchev, D. and Hobert, J. P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *Journal of the American Statistical Association*, 99:228–238.

Mengersen, K. and Tweedie, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24:101–121.

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

Meyn, S. P. and Tweedie, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability*, 4:981–1011.

Mira, A. and Tierney, L. (2002). Efficiency and convergence properties of slice samplers. *Scandinavian Journal of Statistics*, 29:1–12.

Mykland, P., Tierney, L., and Yu, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, 90:233–241.

Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 43:309–318.

Nummelin, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, London.

Nummelin, E. (2002). MC's for MCMC'ists. *International Statistical Review*, 70:215–240.

Peligrad, M. and Shao, Q.-M. (1995). Estimation of the variance of partial sums for ρ-mixing random variables. *Journal of Multivariate Analysis*, 52:140–157.

Philipp, W. and Stout, W. (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Memoirs of the American Mathematical Society*, 2:1–140.

Robert, C. P. (1995). Convergence control methods for Markov chain Monte Carlo algorithms. *Statistical Science*, 10:231–253.

Roberts, G. O. and Polson, N. G. (1994). On the geometric convergence of the Gibbs sampler. *Journal of the Royal Statistical Society,* Series B, 56:377–384.

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2:13–25.

Roberts, G. O. and Rosenthal, J. S. (1998). Markov chain Monte Carlo: Some practical implications of theoretical results (with discussion). *Canadian Journal of Statistics*, 26:5–31.

Roberts, G. O. and Rosenthal, J. S. (1999). Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society,* Series B, 61:643–660.

Roberts, G. O. and Rosenthal, J. S. (2001). Markov chains and de-initializing processes. *Scandinavian Journal of Statistics*, 28:489–504.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90:558–566.

Rosenthal, J. S. (1996). Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Statistics and Computing*, 6:269–275.

Sherman, M. and Goldsman, D. (2002). Large-sample normality of the batch-means variance estimator. *Operations Research Letters*, 30:319–326.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics*, 22:1701–1762.

Example Section	Target half-width	n^* / R^*	Method	Average half width	Average Chain Length	Coverage Probability
4.1	.02	4500	CBM	.0193 (1.2×10^{-5})	16732 (44)	.937 (.005)
		4500	BM30	.0188 (2.3×10^{-5})	15536 (79)	.922 (.006)
		1000	RS	.0197 (4.0×10^{-6})	16466 (17)	.941 (.005)
4.2	.02	5000	CBM	.0192 (1.4×10^{-5})	5832 (15)	.941 (.005)
		5000	BM30	.0188 (2.3×10^{-5})	5899 (21)	.929 (.006)
		100	RS	.0197 (5.4×10^{-6})	5893 (19)	.945 (.005)
4.3	.02	4000	CBM	.0197 (6.0×10^{-6})	56429 (425)	.882 (.007)
		4000	BM30	.0197 (7.0×10^{-6})	45975 (519)	.870 (.008)
		20	RS	.0197 (1.5×10^{-5})	58574 (659)	.890 (.007)
4.4	.002	10000	CBM	.00198 (4.0×10^{-7})	168197 (270)	.934 (.005)
		10000	BM30	.00197 (6.0×10^{-7})	132099 (809)	.880 (.007)
		25	RS	.00199 (1.0×10^{-7})	179338 (407)	.942 (.005)

Table 1: Summary statistics for BM30, CBM and RS. Standard errors of estimates are in parentheses.