Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: http://www.tandfonline.com/loi/uasa20

Fixed-Width Output Analysis for Markov Chain Monte Carlo

Galin L Jones, Murali Haran, Brian S Caffo & Ronald Neath

To cite this article: Galin L Jones, Murali Haran, Brian S Caffo & Ronald Neath (2006) Fixed-Width Output Analysis for Markov Chain Monte Carlo, Journal of the American Statistical Association, 101:476, 1537-1547, DOI: 10.1198/016214506000000492

To link to this article: http://dx.doi.org/10.1198/016214506000000492

Published online: 01 Jan 2012.

Fixed-Width Output Analysis for Markov Chain Monte Carlo

Galin L. JONES, Murali HARAN, Brian S. CAFFO, and Ronald NEATH

Markov chain Monte Carlo is a method of producing a correlated sample to estimate features of a target distribution through ergodic averages. A fundamental question is when sampling should stop; that is, at what point the ergodic averages are good estimates of the desired quantities. We consider a method that stops the simulation when the width of a confidence interval based on an ergodic average is less than a user-specified value. Hence calculating a Monte Carlo standard error is a critical step in assessing the simulation output. We consider the regenerative simulation and batch means methods of estimating the variance of the asymptotic normal distribution. We give sufficient conditions for the strong consistency of both methods and investigate their finite-sample properties in various examples.

KEY WORDS: Batch means; Geometric ergodicity; Markov chain; Monte Carlo central limit theorem; Regeneration.

1. INTRODUCTION

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

Let $X = \{X_0, X_1, X_2, \ldots\}$ be a 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,

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

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 $\mathrm{E}_\pi g$.

Several methods can be used to determine when n is sufficiently large (i.e., when to terminate the simulation). The simplest approach is to terminate the computation whenever patience runs out. This approach is unsatisfactory, because 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 method would be to monitor the sequence of $\bar{g}_n$ until it appears to have stabilized. None of these methods is automated, however; and hence they all are inefficient uses of user time and Monte Carlo resources. Moreover, they provide only a point estimate of $\mathrm{E}_\pi g$ without additional work.

Convergence diagnostics are also sometimes used to terminate the simulation (Cowles and Carlin 1996). Some convergence diagnostics are available in software (e.g., the R package `boa`) and hence may be considered automated. However, none of the diagnostics of which we are aware explicitly addresses how well $\bar{g}_n$ estimates $\mathrm{E}_\pi g$; this is discussed again in Section 4.1.1.

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 a specified value. Under regularity conditions (see Sec. 2), the Markov chain X and function g will admit a central limit theorem (CLT),

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

as $n \to \infty$, where $\sigma_g^2 := \mathrm{var}_\pi\{g(X_0)\} + 2\sum_{i=1}^{\infty} \mathrm{cov}_\pi\{g(X_0), g(X_i)\}$. Given an estimate of σ_g^2 (say $\hat{\sigma}_n^2$), we can 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; this is a common way of choosing n (see, e.g., Fishman 1996; Geyer 1992; Jones and Hobert 2001). Note that the final Monte Carlo sample size is random. Here we study sequential fixed-width methods that formalize this approach. In particular, the simulation terminates the first time that

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

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. One possibility is to fix $n^* > 0$ and take $p(n) = \epsilon \times I(n \le n^*)$, where I is the usual indicator function.

Sequential statistical procedures have a long history (see Lai 2001 for an overview and commentary). Moreover, classical approaches to sequential fixed-width confidence intervals, such as those given by Chow and Robbins (1965), Liu (1997), and Nadas (1969), are known to work well. However, the classical procedures are not relevant to the current work, because they assume that the observations are random samples.

In a simulation context, procedures based on (3) were studied most notably 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

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

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: (a) A functional central

Galin L. Jones is Assistant Professor, School of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: *galin@stat.umn.edu*). Murali Haran is Assistant Professor, Department of Statistics, Pennsylvania State University, University Park, PA 16802 (E-mail: *mharan@stat.psu.edu*). Brian S. Caffo is Assistant Professor, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205 (E-mail: *bcaffo@jhsph.edu*). Ronald Neath is Graduate Student, School of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: *rneath@stat.umn.edu*). The authors thank Ansu Chatterjee, Jeff Rosenthal, and Bill Sudderth for helpful discussions.

Journal of the American Statistical Association
December 2006, Vol. 101, No. 476, Theory and Methods
DOI 10.1198/016214506000000492

limit theorem (FCLT) holds, (b) $\hat{\sigma}_n^2 \to \sigma_g^2$ with probability 1 as $n \to \infty$, and (c) $p(n) = o(n^{-1/2})$. Markov chains frequently enjoy an FCLT under the same conditions that ensure a CLT. However, in the context of MCMC, little work has been done on establishing conditions for (b) 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 that the sampler 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; Roberts and Rosenthal 1998, 2004 for some references and discussion).

We consider two methods for estimating the variance of the asymptotic normal distribution: regenerative simulation (RS) and nonoverlapping batch means (BM). Both have strengths and weaknesses; essentially, BM is easier to implement, but RS is on a stronger theoretical footing. For example, when used with a fixed number of batches, BM cannot be even weakly consistent for σ_g^2. 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, we denoted it as CBM to distinguish it from the standard fixed-batch size version, which we denote as BM. This was addressed by Damerdji (1994); but, although the approach is similar, our regularity conditions on X are weaker. Also, the regularity conditions required to obtain strong consistency of the batch means estimator are slightly stronger than those required by RS. Finally, it is important to note that RS and CBM do not require that X be stationary; hence burn-in is not required.

The justification of fixed-width methods is entirely asymptotic so it is not clear how the finite-sample properties of BM, CBM, and RS compare in typical MCMC settings. For this reason, we conduct a simulation study in the context of two benchmark examples and two realistic examples, one of which is a complicated frequentist problem and one which involves a high-dimensional posterior. Roughly speaking, we find that BM performs poorly, whereas RS and CBM are comparable.

The rest of the article is organized as follows. Section 2 fixes notation and contains a brief discussion of some relevant Markov chain theory. Section 3 considers RS and CBM, and Section 4 presents examples.

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 $A \in \mathcal{B}(\mathsf{X})$, $P^n(x, A) = \Pr(X_n \in A | X_0 = x)$. 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 that there exists an extended real-valued function $M(x)$ and a nonnegative decreasing function $\kappa(n)$ on $\mathbb{Z}_+$ such that

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

When $\kappa(n) = t^n$ for some $t < 1$, we say that X is *geometrically ergodic* if M is unbounded and *uniformly ergodic* if M is bounded. *Polynomial ergodicity of order m*, where $m \ge 0$, means that M may be unbounded and $\kappa(n) = n^{-m}$.

Also, P satisfies *detailed balance* with respect to π if

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx) \quad \text{for all } x, y \in \mathsf{X}. \tag{7}$$

Note that Metropolis–Hastings samplers satisfy (7) by construction, but many Gibbs samplers do not. We are now in position to give conditions for the existence of a CLT.

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

1. X is polynomially ergodic of order $m > 1$, $\mathrm{E}_\pi M < \infty$, and there exists $B < \infty$ such that $|g(x)| < B$ almost surely
2. X is polynomially ergodic of order m, $\mathrm{E}_\pi M < \infty$, and $\mathrm{E}_\pi |g(x)|^{2+\delta} < \infty$ for some $\delta > 0$, where $m\delta > 2 + \delta$
3. X is geometrically ergodic and $\mathrm{E}_\pi[g^2(x)(\log^+ |g(x)|)] < \infty$
4. X is geometrically ergodic and satisfies (7), and $\mathrm{E}_\pi g^2(x) < \infty$
5. X is uniformly ergodic and $\mathrm{E}_\pi g^2(x) < \infty$.

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

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

Remark 1. The theorem was proved by Ibragimov and Linnik (1971) (condition 5), Roberts and Rosenthal (1997) (condition 4), and Doukhan, Massart, and Rio (1994) (condition 3). See Jones (2004) for details on conditions 1 and 2.

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

Remark 3. The mixing conditions on the Markov chain X stated in the Theorem are not necessary for the CLT (see, e.g., Chen 1999; Meyn and Tweedie 1993; Nummelin 2002). However, checking the weaker conditions is often prohibitively difficult in situations where MCMC is appropriate.

Remark 4. There are constructive techniques for verifying the existence of an appropriate M and κ from (6) (Meyn and Tweedie 1993, chap. 15). For example, one method of establishing geometric ergodicity requires finding a function $V : \mathsf{X} \to [1, \infty)$ and a small set $C \in \mathcal{B}(\mathsf{X})$ such that

$$PV(x) \le \lambda V(x) + bI(x \in C) \qquad \forall x \in \mathsf{X}, \tag{8}$$

where $PV(x) := \int V(y)P(x, dy)$, $0 < \lambda < 1$, and $b < \infty$. Substantial effort has been devoted to establishing convergence rates for MCMC algorithms through (8) or related techniques. For example, Hobert and Geyer (1998), Hobert, Jones, Presnell, and Rosenthal (2002), 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, whereas Christensen, Moller, and Waagepetersen (2001), Douc, Fort, Moulines, 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) analyzed Metropolis–Hastings algorithms.

2.1 The Split Chain

An object important to the study of both RS and CBM is the *split chain*, $X' := \{(X_0, \delta_0), (X_1, \delta_1), (X_2, \delta_2), \ldots\}$, which has state space $\mathsf{X} \times \{0, 1\}$. The construction of X' requires a *minorization condition*, that is, a function $s : \mathsf{X} \to [0, 1]$ for which $\mathrm{E}_\pi s > 0$ and a probability measure Q such that

$$P(x, A) \geq s(x)Q(A) \quad \text{for all } x \in \mathsf{X} \text{ and } A \in \mathcal{B}(\mathsf{X}). \tag{9}$$

When X is countable, it is easy to see that (9) holds by fixing $x_* \in \mathsf{X}$ and setting $s(x) = I(x = x_*)$ and $Q(\cdot) = P(x_*, \cdot)$. Mykland, Tierney, and Yu (1995) and Rosenthal (1995) have given prescriptions that are often useful for establishing (9) in general spaces. Note that (9) 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 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 $X_{i+1} \sim Q(\cdot)$; otherwise, draw $X_{i+1} \sim R(x, \cdot)$.

The two chains X and X' are closely related because X' will inherit properties such as aperiodicity and positive Harris recurrence, and the sequence $\{X_i : i = 0, 1, \ldots\}$ obtained from X' has the same transition probabilities as X. In addition, X and X' converge to their respective stationary distributions at exactly the same rate.

If $\delta_i = 1$, then time $i + 1$ is a *regeneration time* at which X' probabilistically restarts itself. Specifically, suppose that we start X' with $X_0 \sim Q$. Then each time that $\delta_i = 1$, $X_{i+1} \sim Q$. Let $0 = \tau_0 < \tau_1 < \cdots$ be the regeneration times; that is, set $\tau_{r+1} = \min\{i > \tau_r : \delta_{i-1} = 1\}$. Also assume that X' is run for R tours; that is, the simulation is stopped the Rth time that a $\delta_i = 1$. Let τ_R denote the total length of the simulation and let N_r be the length of the rth tour, that is, $N_r = \tau_r - \tau_{r-1}$. Define

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

for $r = 1, \ldots, R$. The (N_r, S_r) pairs are iid because each is based on a different tour. In the sequel we will make repeated use of 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 (9) holds and that X is geometrically ergodic. Let $p \geq 1$ be an integer. Then the following results hold:

a. If $\mathrm{E}_\pi |g|^{2^{(p-1)}+\delta} < \infty$ for some $\delta > 0$, then $\mathrm{E}_Q N_1^p < \infty$ and $\mathrm{E}_Q S_1^p < \infty$.

b. If $\mathrm{E}_\pi |g|^{2^p+\delta} < \infty$ for some $\delta > 0$, then $\mathrm{E}_Q N_1^p < \infty$ and $\mathrm{E}_Q S_1^{p+\delta} < \infty$.

For the proof see Appendix A.

3. OUTPUT ANALYSIS

3.1 Regenerative Simulation

Regenerative simulation is based on directly simulating the split chain. However, using the mixture approach described earlier is problematic because simulation from $R(x, dy)$ is challenging. Mykland et al. (1995) suggested a method to avoid this issue. Suppose that (9) holds and that the measures $P(x, \cdot)$ and $Q(\cdot)$ admit densities $k(\cdot|x)$ and $q(\cdot)$, respectively. Then the following recipe allows us to simulate X'. Assume that $X_0 \sim q(\cdot)$, which is typically quite easy to do (see Mykland et al. 1995 for some examples). Note that this means that burn-in is irrelevant. Draw $X_{i+1} \sim k(\cdot|x)$; that is, draw from the sampler at hand, and obtain δ_i by simulating from the distribution of $\delta_i | X_i, X_{i+1}$ with

$$\Pr(\delta_i = 1 | X_i, X_{i+1}) = \frac{s(X_i)q(X_{i+1})}{k(X_{i+1}|X_i)}. \tag{10}$$

Example 1. In a slight abuse of notation, let π also denote the density of the target distribution. Consider an independence Metropolis–Hastings sampler with proposal density ν. This chain works as follows: Let the current state be $X_i = x$. Draw $y \sim \nu$ and independently draw $u \sim \text{uniform}(0, 1)$. If

$$u < \frac{\pi(y)\nu(x)}{\pi(x)\nu(y)},$$

then set $X_{i+1} = y$; otherwise, set $X_{i+1} = x$. Mykland et al. (1995) derived (10) for this case. Let $c > 0$ be a user-specified constant. Then, conditional on an acceptance, (i.e., $X_i = x$ and $X_{i+1} = y$),

$$\Pr(\delta_i = 1 | X_i = x, X_{i+1} = y) = \begin{cases} c \max\left\{\dfrac{\nu(x)}{\pi(x)}, \dfrac{\nu(y)}{\pi(y)}\right\} & \text{if } \min\left\{\dfrac{\pi(x)}{\nu(x)}, \dfrac{\pi(y)}{\nu(y)}\right\} > c \\ \dfrac{1}{c} \max\left\{\dfrac{\pi(x)}{\nu(x)}, \dfrac{\pi(y)}{\nu(y)}\right\} & \text{if } \max\left\{\dfrac{\pi(x)}{\nu(x)}, \dfrac{\pi(y)}{\nu(y)}\right\} < c \\ 1 & \text{otherwise.} \end{cases} \tag{11}$$

Note that we do not need to know the normalizing constants of π or ν to calculate (11).

In discrete state spaces, regenerations can be easy to identify. In particular, regeneration occurs 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 most useful when the state space is not too large but is potentially complicated; see Section 4.3. It is not useful when the state space is extremely large, because returns to the fixed state are too infrequent. Further practical advice on implementing and automating RS has been given by Brockwell and Kadane (2005), Gilks, Roberts, and Sahu (1998), Geyer and Thompson (1995), Hobert et al. (2002), Hobert, Jones, and Robert (2005), and Jones and Hobert (2001).

Implementation of RS is simple once we can effectively simulate the split chain. For example, 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 $R \to \infty$, and hence estimating $\mathrm{E}_\pi g$ is routine.

We now turn our attention to calculating a Monte Carlo standard error for $\bar{g}_{\tau_R}$. Let E_Q denote the expectation for the split chain starting with $X_0 \sim Q(\cdot)$, and let $\bar{N}$ be the average tour length, that is, $\bar{N} = R^{-1}\sum_{r=1}^{R} N_r$. Because the (N_r, S_r) pairs are iid, the strong law implies, with probability 1, that $\bar{N} \to \mathrm{E}_Q N_1$ which is finite by positive recurrence. If $\mathrm{E}_Q N_1^2 < \infty$ and $\mathrm{E}_Q S_1^2 < \infty$, then it follows that a CLT holds; that is, as $R \to \infty$,

$$\sqrt{R}\big(\bar{g}_{\tau_R} - \mathrm{E}_\pi g\big) \xrightarrow{d} \mathrm{N}(0, \xi_g^2), \tag{12}$$

where, as shown by Hobert et al. (2002), $\xi_g^2 = \mathrm{E}_Q(S_1 - N_1\mathrm{E}_\pi g)^2/(\mathrm{E}_Q N_1)^2$. An obvious estimator of ξ_g^2 is

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

Now consider

$$\begin{aligned}\hat{\xi}_{RS}^2 - \xi_g^2 &= \frac{1}{\bar{N}^2}\frac{1}{R}\sum_{r=1}^{R}\big(S_r - \bar{g}_{\tau_R}N_r\big)^2 \\ &\quad - \frac{\mathrm{E}_Q(S_1 - N_1\mathrm{E}_\pi g)^2}{(\mathrm{E}_Q N_1)^2} \pm \frac{\mathrm{E}_Q(S_1 - N_1\mathrm{E}_\pi g)^2}{\bar{N}^2} \\ &= \frac{1}{\bar{N}^2}\frac{1}{R}\sum_{r=1}^{R}\big[\big(S_r - \bar{g}_{\tau_R}N_r\big)^2 \\ &\quad - \mathrm{E}_Q(S_1 - N_1\mathrm{E}_\pi g)^2 \pm (S_r - N_r\mathrm{E}_\pi g)^2\big] \\ &\quad + \left[\mathrm{E}_Q(S_1 - N_1\mathrm{E}_\pi g)^2\left(\frac{1}{\bar{N}^2} - \frac{1}{\mathrm{E}_Q N_1^2}\right)\right].\end{aligned}$$

Using this representation and repeated application of the strong law gives that $\hat{\xi}_{RS}^2 - \xi_g^2 \to 0$ with probability 1 as $R \to \infty$ (see also Hobert et al. 2002). It is typically difficult to check that $\mathrm{E}_Q N_1^2 < \infty$ and $\mathrm{E}_Q S_1^2 < \infty$; however, using Lemma 1 yields the following result.

Proposition 1. Let X be a Harris ergodic Markov chain with invariant distribution π. Assume that $\mathrm{E}_\pi |g|^{2+\delta} < \infty$ for some $\delta > 0$, that (9) holds, and that X is geometrically ergodic. Then (12) holds and $\hat{\xi}_{RS}^2 \to \xi_g^2$ with probability 1 as $R \to \infty$.

Fix $\epsilon > 0$ and let z denote an appropriate standard normal quantile. An asymptotically valid fixed-width procedure results by terminating the simulation the first time that

$$z\frac{\hat{\xi}_{RS}}{\sqrt{R}} + p(R) \le \epsilon. \tag{13}$$

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; compare the method of overlapping batch means.) Suppose that the algorithm is run for a total of $n = ab$ iterations (and hence $a = a_n$ and $b = b_n$ are implicit functions of n) and define

$$\bar{Y}_j := \frac{1}{b}\sum_{i=(j-1)b}^{jb-1} g(X_i) \quad \text{for } j = 1, \ldots, a.$$

The batch means estimate of σ_g^2 is

$$\hat{\sigma}_{BM}^2 = \frac{b}{a-1}\sum_{j=1}^{a}(\bar{Y}_j - \bar{g}_n)^2. \tag{14}$$

With a fixed number of batches, (14) 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 increases, then it may be possible to obtain consistency. The first result in this direction is due to Damerdji (1994). Damerdji's major assumption is the existence of a strong invariance principle. Let $B = \{B(t), t \ge 0\}$ 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

$$\left|\sum_{i=1}^{n} g(X_i) - n\mathrm{E}_\pi g - \sigma_g B(n)\right| = O(\gamma(n)) \tag{15}$$

with probability 1 as $n \to \infty$, where "with probability 1" means for almost all sample paths. In particular, Damerdji assumed that (15) held with $\gamma(n) = n^{1/2-\alpha}$, where $0 < \alpha \le 1/2$. However, directly checking this condition in any given application would seem a daunting task. In an attempt to alleviate this difficulty somewhat, 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. Then the following results hold:

a. If X is uniformly ergodic and $\mathrm{E}_\pi |g|^{2+\delta} < \infty$ for some $\delta > 0$, then (15) holds with $\gamma(n) = n^{1/2-\alpha}$, where $\alpha < \delta/(24 + 12\delta)$.

b. If X is geometrically ergodic, (9) holds, and $\mathrm{E}_\pi |g|^{4+\delta} < \infty$ for some $\delta > 0$, then (15) holds with $\gamma(n) = n^\alpha \log n$, where $\alpha = 1/(2 + \delta)$.

Proof. The first part of the lemma 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. The second part follows from our Lemma 1 and theorem 2.1 of Csáki and Csörgő (1995).

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

Proposition 2 (Damerdji 1994). Assume that $g : \mathsf{X} \to \mathbb{R}$ such that $\mathrm{E}_\pi |g|^{2+\delta} < \infty$ for some $\delta > 0$, and let X be a Harris ergodic Markov chain with invariant distribution π. Further, suppose that X is uniformly ergodic. If (a) $a_n \to \infty$ as $n \to \infty$; (b) $b_n \to \infty$ and $b_n/n \to 0$ as $n \to \infty$; (c) $b_n^{-1} n^{1-2\alpha} \log n \to 0$ as $n \to \infty$, where $\alpha \in (0, \delta/(24 + 12\delta))$; and (d) there exists a constant $c \ge 1$ such that $\sum_n (b_n/n)^c < \infty$, then, as $n \to \infty$, $\hat{\sigma}_{BM}^2 \to \sigma_g^2$ with probability 1.

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

Proposition 3. Assume that $g : \mathsf{X} \to \mathbb{R}$ such that $\mathrm{E}_\pi |g|^{4+\delta} < \infty$ for some $\delta > 0$, and let X be a Harris ergodic Markov chain with invariant distribution π. Further, suppose that X is geometrically ergodic. If (a) $a_n \to \infty$ as $n \to \infty$; (b) $b_n \to \infty$ and

$b_n/n \to 0$ as $n \to \infty$; (c) $b_n^{-1} n^{2\alpha} [\log n]^3 \to 0$ as $n \to \infty$, where $\alpha = 1/(2+\delta)$; and (d) there exists a constant $c \geq 1$ such that $\sum_n (b_n/n)^c < \infty$, then, as $n \to \infty$, $\hat{\sigma}^2_{BM} \to \sigma^2_g$ with probability 1.

Remark 5. There is no assumption of stationarity in Propositions 2 or 3. Hence burn-in is not required to implement CBM.

Remark 6. Consider using $b_n = \lfloor n^\theta \rfloor$ and $a_n = \lfloor n/b_n \rfloor$. Proposition 2 requires that $1 > \theta > 1 - 2\alpha > 1 - \delta/(12 + 6\delta) > 5/6$, but Proposition 3 requires only that $1 > \theta > (1 + \delta/2)^{-1} > 0$.

Under the conditions of Proposition 2 or 3, an asymptotically valid fixed-width procedure for estimating $\mathrm{E}_\pi g$ results if we terminate the simulation the first time

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

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

3.3 Practical Implementation Issues

Making practical use of the preceding theory requires (a) having a moment condition, (b) establishing geometric ergodicity of the sampler at hand, (c) choosing $p(n)$, (d) using (9) or at least (10) to use RS, and (e) choosing a_n and b_n to use CBM. Because a moment condition is required even in the iid case, we do not view (a) as restrictive. Consider (b). It is easy to construct examples in which the convergence rate is so slow that a Markov chain CLT does not hold (Roberts 1999), so the importance of establishing the rate of convergence in (6) should not be underestimated. On the other hand, the MCMC community has expended considerable effort trying to understand when certain Markov chains are geometrically ergodic; see the references cited in Remark 4. In our view, this is not the obstacle that it once was.

Regarding (c), we know of no work on choosing an optimal $p(n)$. Recall that the theory requires that $p(n) = o(n^{-1/2})$. In our examples we use $p(n) = \epsilon I(n \leq n^*)$, where $n^* > 0$ is fixed. Because n^* is typically chosen based on empirical experience with the sampler at hand, we might want a penalty for sample sizes greater than n^*, so another reasonable choice might be $p(n) = \epsilon I(n \leq n^*) + Cn^{-k}$ for some $k > 1/2$ and $C > 0$.

The issue in (d) [i.e., calculating (9) or (10)] is commonly viewed as overly burdensome; however, in our experience, this calculation need not be troublesome. For example, Mykland et al. (1995) gave recipes for constructing (9) and (10) for Metropolis–Hastings independence and random-walk samplers; recall (11). There also has been some work on establishing these conditions for very general models (see Hobert et al. 2005). Finally, Brockwell and Kadane (2005) and Geyer and Thompson (1995) have shown that regenerations can be made to occur naturally through simulated tempering.

Consider (e). As noted in Remark 6, it is common to choose the batch sizes according to $b_n = \lfloor n^\theta \rfloor$ for some θ. Song and Schmeiser (1995) and Chien (1988) have addressed the issue of which value of θ to use from different theoretical standpoints. In particular, Chien (1988) showed that under regularity conditions, using $\theta = 1/2$ results in the batch means estimator approaching asymptotic normality at the fastest rate. Song and Schmeiser (1995) showed that under different regularity conditions, using $\theta = 1/3$ minimizes the asymptotic mean squared error (MSE) of $\hat{\sigma}^2_{BM}$. Note that Remark 6 shows that $\theta = 1/3$ requires a stronger moment condition than $\theta = 1/2$. We address this issue in more detail in Section 4.

3.4 Alternatives to BM and RS

We chose to focus on BM and RS because in MCMC settings these seem to be the most common methods for estimating the variance of the asymptotic normal distribution. However, other methods may enjoy strong consistency (see, e.g., Damerdji 1991; Geyer 1992; Nummelin 2002; Peligrad and Shao 1995). In particular, Damerdji (1991) used 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 through Lemma 2 to obtain a result with regularity conditions similar to those of Proposition 3. However, we do not pursue this further here.

4. EXAMPLES

In this section we investigate the finite-sample performance of RS, BM with 30 batches, and CBM with $b_n = \lfloor n^{1/3} \rfloor$ and $b_n = \lfloor n^{1/2} \rfloor$ in four examples. In particular, we examine the coverage probabilities and half-widths of the resulting intervals, as well as the required simulation effort. Although each example concerns a different statistical model and MCMC sampler, there are some commonalities. In each case we performed many independent replications of the given MCMC sampler. The number of replications ranged from 2,000 to 9,000, depending on the complexity of the example. We used all methods on the same output from each replication of the MCMC sampler. When the half-width of a 95% interval with $p(n) = \epsilon I(n \geq n^*)$ [or $p(R) = \epsilon I(R \geq R^*)$ for RS] was less than ϵ for a particular method, that procedure was stopped and the chain length recorded. Our choice of n^* was different for each example and was chosen based on our empirical experience with the given Markov chain. Other procedures would continue until all of them were below the targeted half-width, at which time a single replication was complete. To estimate the coverage probabilities, we need true values of the quantities of interest. These are not analytically available in three of our examples. Our solution is to obtain precise estimates of the truth through independent methods that are different for each example. The details are described herein; the results are reported in Table 1.

4.1 Toy Example

Consider estimating the mean of a Pareto(α, β) distribution, [i.e., $\alpha\beta/(\beta - 1)$, $\beta > 1$], using a Metropolis–Hastings independence sampler with a Pareto(α, λ) candidate. Let π be the target density and let ν be the proposal density. Assume that $\beta \geq \lambda$. Then, for $x \geq \alpha$,

$$\frac{\pi(x)}{\nu(x)} = \frac{\beta}{\lambda} \alpha^{\beta - \lambda} x^{\lambda - \beta} \leq \frac{\beta}{\lambda}.$$

By theorem 2.1 of Mengersen and Tweedie (1996), this sampler is uniformly ergodic and

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq \left(1 - \frac{\lambda}{\beta}\right)^n.$$

Table 1. Summary Statistics for BM, CBM, and RS

Example section	Method	b_n	n^*/R^*	Average half-width	Average chain length	Coverage probability
4.1	CBM	$\lfloor n^{1/2} \rfloor$	45	$.0048_{(1.9\times 10^{-6})}$	$2{,}428_{(5)}$	$.923_{(.003)}$
9,000 reps	CBM	$\lfloor n^{1/3} \rfloor$	45	$.0049_{(8.0\times 10^{-7})}$	$2{,}615_{(3)}$	$.943_{(.002)}$
$\epsilon = .005$	BM	$\lfloor n/30 \rfloor$	45	$.0047_{(2.4\times 10^{-6})}$	$2{,}342_{(6)}$	$.908_{(.003)}$
	RS		30	$.0049_{(4.0\times 10^{-7})}$	$2{,}653_{(2)}$	$.948_{(.002)}$
4.2	CBM	$\lfloor n^{1/2} \rfloor$	2,000	$.0194_{(7.2\times 10^{-6})}$	$5{,}549_{(13)}$	$.930_{(.004)}$
5,000 reps	CBM	$\lfloor n^{1/3} \rfloor$	2,000	$.0198_{(3.3\times 10^{-6})}$	$5{,}778_{(6)}$	$.947_{(.003)}$
$\epsilon = .02$	BM	$\lfloor n/30 \rfloor$	2,000	$.0191_{(1.1\times 10^{-5})}$	$5{,}279_{(18)}$	$.915_{(.004)}$
	RS		50	$.0198_{(2.3\times 10^{-6})}$	$5{,}818_{(12)}$	$.945_{(.003)}$
4.3	CBM	$\lfloor n^{1/2} \rfloor$	4,000	$.0049_{(1.6\times 10^{-6})}$	$56{,}258_{(405)}$	$.920_{(.006)}$
2,000 reps	CBM	$\lfloor n^{1/3} \rfloor$	4,000	$.0049_{(1.8\times 10^{-6})}$	$46{,}011_{(499)}$	$.869_{(.008)}$
$\epsilon = .005$	BM	$\lfloor n/30 \rfloor$	4,000	$.0049_{(1.7\times 10^{-6})}$	$45{,}768_{(478)}$	$.874_{(.007)}$
	RS		20	$.0049_{(4.3\times 10^{-6})}$	$58{,}265_{(642)}$	$.894_{(.007)}$
4.4	CBM	$\lfloor n^{1/2} \rfloor$	10,000	$.00396_{(8.0\times 10^{-7})}$	$168{,}197_{(270)}$	$.934_{(.005)}$
2,000 reps	CBM	$\lfloor n^{1/3} \rfloor$	10,000	$.00398_{(4.0\times 10\_7)}$	$137{,}119_{(125)}$	$.900_{(.006)}$
$\epsilon = .004$	BM	$\lfloor n/30 \rfloor$	10,000	$.00394_{(1.2\times 10^{-6})}$	$132{,}099_{(809)}$	$.880_{(.007)}$
	RS		25	$.00398_{(2.0\times 10^{-7})}$	$179{,}338_{(407)}$	$.942_{(.005)}$

NOTE: Standard errors of estimates are in parentheses.

To ensure the moment conditions required for Proposition 3, we set $\beta = 10$ and $\lambda = 9$, in which case the right side is 10^{-n}. Hence this sampler converges rapidly. Implementation of RS was accomplished using (11) with $c = 1.5$.

4.1.1 Comparing Convergence Diagnostics With CBM. As noted by a referee, one method for terminating the simulation is through convergence diagnostics. Consider the method of Geweke (1992), which is a diagnostic that seems close in spirit to the current work. Geweke's diagnostic (GD) is based on a Markov chain CLT and hence does not apply much more generally than CBM; the same can be said for many other diagnostics. GD uses a hypothesis test to ascertain when $\bar{g}_n$ has stabilized.

In the remainder of this section we compare GD and CBM in terms of MSE and chain length. Toward this end, we ran 9,000 independent replications of the independence sampler with $\alpha = 1$, $\beta = 10$, and $\lambda = 9$. We used CBM and GD on the output in the following manner. For each replication, we set $n^* = 45$, but the R package `boa` required a minimum of 120 iterations to calculate GD. After the minimum was achieved and the cutoff for a particular method was attained, we noted the chain length and the current estimate of $\mathrm{E}_\pi g$. The cutoff for CBM was to set the desired half-width to $\epsilon = .005$. The result of using GD is a p value. We chose four values (.05, .10, .2, and .4) for the threshold in an attempt to tune the computation; the results are reported in Table 2. As we noted previously, this sampler mixes extremely well. Thus it is not surprising that using GD results in a small estimated MSE; however, using CBM results in much smaller MSE than GD. The average chain lengths make it clear that GD stops the simulation much too soon. Moreover, changing the p value threshold for GD does not result in any substantial improvement in estimation accuracy.

Table 2. Summary Statistics for CBM versus GD for Example 4.1

Method	Cutoff	Estimated MSE	Average chain length
CBM ($b_n = \lfloor n^{1/3} \rfloor$)	$\epsilon = .005$	$6.65\times 10^{-6}{}_{(9.9\times 10^{-8})}$	$2{,}428_{(5)}$
CBM ($b_n = \lfloor n^{1/2} \rfloor$)	$\epsilon = .005$	$7.34\times 10^{-6}{}_{(1.2\times 10^{-8})}$	$2{,}615_{(3)}$
Geweke	p value $= .4$	$1.17\times 10^{-4}{}_{(2\times 10^{-6})}$	$202.6_{(3.4)}$
Geweke	p value $= .2$	$1.30\times 10^{-4}{}_{(2\times 10^{-6})}$	$148.9_{(1.6)}$
Geweke	p value $= .1$	$1.34\times 10^{-4}{}_{(2\times 10^{-6})}$	$133.4_{(.9)}$
Geweke	p value $= .05$	$1.37\times 10^{-4}{}_{(2\times 10^{-6})}$	$127.4_{(.5)}$

NOTE: Standard errors of estimates are in parentheses.

4.2 A Hierarchical Model

Efron and Morris (1975) presented a dataset 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) considered the following conditionally independent hierarchical model for the transformed data. Suppose that for $i = 1, \ldots, K$,

$$\begin{aligned} Y_i|\theta_i &\sim \mathrm{N}(\theta_i, 1), \qquad \theta_i|\mu, \lambda \sim \mathrm{N}(\mu, \lambda), \\ \lambda &\sim \mathrm{IG}(2, 2), \qquad f(\mu) \propto 1. \end{aligned} \tag{16}$$

[Note that we say that $W \sim \text{gamma}(\alpha, \beta)$ if its density is proportional to $w^{\alpha-1}e^{-\beta w}I(w > 0)$, and if $X \sim \text{gamma}(b, c)$, then $X^{-1} \sim \mathrm{IG}(b, c)$.] Rosenthal (1996) introduced a Harris ergodic block Gibbs sampler with posterior, $\pi(\theta, \mu, \lambda|y)$, characterized by the hierarchy in (16) 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

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

$$\mu|\theta, \lambda, y \sim \mathrm{N}\left(\bar{\theta}, \frac{\lambda}{K}\right),$$

and

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

Rosenthal proved geometric ergodicity of the associated Markov chain. 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. In Appendix C we also derive an expression for the probability of regeneration (10).

We focused 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 used 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 *p* Values

Agresti (2002, p. 432) reported data corresponding 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, that is, the margins of the table and $\sum_{ij} n_{ij}\, ij$, where the n_{ij}'s are the observed cell counts. The resulting conditional distribution is a generalization of the hypergeometric distribution. An exact p value 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 dataset, there are more than 12 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 denote this set of permissible tables by Γ. Now the desired p value is given by

$$\sum_{y \in \Gamma} I[d(y) \geq d(y_{\text{obs}})]\pi(y), \tag{17}$$

where $d(\cdot)$ is the deviance function and π denotes the generalized hypergeometric. Because we have enumerated Γ, we find that the true exact p value is .044, whereas the chi-squared approximation yields a p value of .368. However, a different dataset with different values of the sufficient statistics will have a different reference set that must be enumerated to find the exact p value. In general, implementing this would be too computationally burdensome, and hence it is common to resort to MCMC-based approximations (see, e.g., Caffo and Booth 2001; Diaconis and Sturmfels 1998).

To estimate (17), we use the Metropolis–Hastings algorithm developed by Caffo and Booth (2001). This algorithm is also used by the R package `exactLoglinTest`. The associated Markov chain is Harris ergodic, and its invariant distribution is the appropriate generalized hypergeometric distribution. Moreover, the chain is uniformly ergodic, and because we are estimating the expectation of a bounded function, the regularity conditions for both RS and CBM are easily met.

Implementation of RS is straightforward. As mentioned earlier, in finite state spaces regenerations occur whenever the chain returns to any fixed state. To choose the fixed state, we ran the algorithm for 1,000 iterations and chose the state that had the highest probability with respect to the stationary distribution. We used the same fixed state in each replication.

4.4 A Model-Based Spatial Statistics Application

Consider the Scottish lip cancer dataset (Clayton and Kaldor 1987) consisting of the number of cases of lip cancer registered in each of the 56 (prereorganization) counties of Scotland, together with the expected number of cases given the age–sex structure of the population. We assume a Poisson likelihood for areal (spatially aggregated) data. Specifically, for $i = 1, \ldots, N$, we assume that given μ_i, the disease counts Y_i are conditionally independent and

$$Y_i|\mu_i \sim \text{Poisson}(E_i e^{\mu_i}), \tag{18}$$

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. Set $\phi = (\phi_1, \ldots, \phi_N)^T$. Each μ_i is modeled as $\mu_i = \theta_i + \phi_i$, where

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

and

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

where n_i is the number of neighbors for the ith region. Each θ_i captures the ith region's extra-Poisson variability due to areawide heterogeneity, whereas each ϕ_i captures the ith region's excess variability attributable to regional clustering. The priors on the precision parameters are $\tau_h \sim \text{gamma}(1, .01)$ and $\tau_c \sim \text{gamma}(1, .02)$. This is a challenging model to consider, because 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).

Haran and Tierney (2004) established uniform ergodicity of a Harris-ergodic Metropolis–Hastings independence sampler with invariant distribution $\pi(\theta, \phi, \tau_h, \tau_c|y)$, where $\theta = (\theta_1, \ldots, \theta_N)^T$, and a heavy-tailed proposal. In our implementation of RS, we used the formula for the probability of a regeneration given by (11) with $\log c = -342.72$. Using the empirical supremum of the ratio of the invariant density to the proposal density (based on several draws from the proposal) guided the choice of c.

We focused on estimating the posterior expectation of ϕ_7, the log-relative risk of disease for County 7 attributable to spatial clustering. Finally, we used an independent run of length 10^7 to obtain an estimate that we treated as the "true value."

4.5 Summary

Table 1 reveals that the estimates of the coverage probabilities are all below the desired .95. But examining the standard errors shows that only BM is significantly lower in all of the examples and the estimated coverage probability for RS is not significantly different than .95 in three of four examples. The story for CBM is more complicated, in that the coverage depends on the choice of b_n. Using $b_n = \lfloor n^{1/3} \rfloor$ gives the best coverage for the examples given in Sections 4.1 and 4.2, whereas $b_n = \lfloor n^{1/2} \rfloor$ is superior for the examples in Sections 4.3 and 4.4. This is because the Markov chains in Sections 4.1 and 4.2 mix exceptionally well, and hence smaller batch sizes can be tolerated. But the examples in Sections 4.3 and 4.4 are realistic problems, and hence the chains do not mix as well, so that larger batch sizes are required. Thus we would generally recommend using $b_n = \lfloor n^{1/2} \rfloor$.

The example in Section 4.3 deserves to be singled out because of 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.

Whereas RS and CBM appear comparable in terms of coverage probability, RS tends to result in slightly longer runs than CBM, which in turn results in longer runs than BM. Moreover, RS and CBM are comparable in their ability to produce intervals that meet the target half-width more closely than BM. Also, the intervals for RS are apparently more stable than those of CBM and BM. Finally, BM underestimates the Monte Carlo standard error and thus suggests stopping the chain too early.

Although RS has a slight theoretical advantage over CBM, the two methods' finite-sample properties appear comparable. Also, like RS, CBM avoids the burn-in issue, which has been a long-standing obstacle to MCMC practitioners. In addition, CBM is slightly easier to implement. Thus CBM clearly has a place in the tool kit of MCMC practitioners.

APPENDIX A: PROOF OF LEMMA 1

A.1 Preliminary Results

Recall the split chain X' and that $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ denotes the regeneration times, that is, $\tau_{r+1} = \min\{i > \tau_r : \delta_{i-1} = 1\}$.

Lemma A.1 (Hobert et al. 2002, lemma 1). Let X be a Harris ergodic Markov chain and assume that (9) holds. Then, for any function $h : \mathsf{X}^\infty \to \mathbb{R}$,

$$\mathrm{E}_\pi |h(X_0, X_1, \ldots)| \geq c\mathrm{E}_Q |h(X_0, X_1, \ldots)|,$$

where $c = \mathrm{E}_\pi s$.

Lemma A.2 (Hobert et al. 2002, lemma 2). Let X be a Harris ergodic Markov chain and assume that (9) holds. If X is geometrically ergodic, then there exists a $\beta > 1$ such that $\mathrm{E}_\pi \beta^{\tau_1} < \infty$.

Corollary A.1. Assume the conditions of Lemma A.2. For any $a > 0$,

$$\sum_{i=0}^{\infty} [\mathrm{Pr}_\pi(\tau_1 \geq i+1)]^a \leq (\mathrm{E}_\pi \beta^{\tau_1})^a \sum_{i=0}^{\infty} \beta^{-a(i+1)} < \infty.$$

A.2 Proof of Lemma 1

We prove only part b of the lemma; the proof of part a is similar. Without loss of generality, we assume that $0 < \delta < 1$. By Lemma A.1, it is sufficient to verify that $\mathrm{E}_\pi \tau_1^p < \infty$ and $\mathrm{E}_\pi S_1^{p+\delta} < \infty$. Lemma A.2 shows that $\mathrm{E}_\pi \tau_1^p < \infty$ for any $p > 0$. Note that

$$\begin{aligned}
&\left(\sum_{i=0}^{\tau_1 - 1} g(X_i) \right)^{p+\delta} \\
&\quad \leq \left(\sum_{i=0}^{\tau_1 - 1} |g(X_i)| \right)^{p+\delta} \\
&\quad = \left(\sum_{i=0}^{\infty} I(0 \leq i \leq \tau_1 - 1) |g(X_i)| \right)^{p+\delta} \\
&\quad \leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} \sum_{i_{p+1}=0}^{\infty} \left[\prod_{j=1}^{p} I(0 \leq i_j \leq \tau_1 - 1) \left| g\left(X_{i_j}\right) \right| \right] \\
&\qquad \times I(0 \leq i_{p+1} \leq \tau_1 - 1) \left| g\left(X_{i_{p+1}}\right) \right|^{\delta},
\end{aligned}$$

and hence

$$\begin{aligned}
\mathrm{E}_\pi S_1^{p+\delta} &\leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} \sum_{i_{p+1}=0}^{\infty} \mathrm{E}_\pi \left(\left[\prod_{j=1}^{p+1} I(0 \leq i_j \leq \tau_1 - 1) \right] \right. \\
&\qquad \left. \times \left[\prod_{j=1}^{p} \left| g\left(X_{i_j}\right) \right| \right] \left| g\left(X_{i_{p+1}}\right) \right|^{\delta} \right) \\
&\leq \sum_{i_1=0}^{\infty} \cdots \sum_{i_p=0}^{\infty} \sum_{i_{p+1}=0}^{\infty} \left[\mathrm{E}_\pi I(0 \leq i_1 \leq \tau_1 - 1) \left| g\left(X_{i_1}\right) \right|^2 \right]^{1/2} \\
&\qquad \times \cdots \times \left[\mathrm{E}_\pi I(0 \leq i_p \leq \tau_1 - 1) \left| g\left(X_{i_p}\right) \right|^{2^p} \right]^{1/2^p} \\
&\qquad \times \left[\mathrm{E}_\pi I(0 \leq i_{p+1} \leq \tau_1 - 1) \left| g\left(X_{i_{p+1}}\right) \right|^{2^p \delta} \right]^{1/2^p},
\end{aligned}$$

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

$$\begin{aligned}
&\mathrm{E}_\pi I(0 \leq i_j \leq \tau_1 - 1) \left| g\left(X_{i_j}\right) \right|^{2^j} \\
&\qquad \leq [\mathrm{E}_\pi I(0 \leq i_j \leq \tau_1 - 1)]^{1/a_j} \left[\mathrm{E}_\pi \left| g\left(X_{i_j}\right) \right|^{2^j + \delta} \right]^{1/b_j}.
\end{aligned}$$

Note that

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

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

$$\begin{aligned}
&\mathrm{E}_\pi I(0 \leq i_{p+1} \leq \tau_1 - 1) \left| g\left(X_{i_{p+1}}\right) \right|^{2^p \delta} \\
&\quad \leq [\mathrm{E}_\pi I(0 \leq i_{p+1} \leq \tau_1 - 1)]^{1/(a_{p+1})} \left[\mathrm{E}_\pi \left| g\left(X_{i_{p+1}}\right) \right|^{\delta(2^p + \delta)} \right]^{1/(b_{p+1})}.
\end{aligned}$$

Note that

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

and set $c = \max\{c_1, \ldots, c_{p+1}\}$. Then an appeal to Corollary A.1 yields

$$\begin{aligned}
\mathrm{E}_\pi S_1^{p+\delta} &\leq c \left[\prod_{j=1}^{p} \sum_{i_j=0}^{\infty} \{\mathrm{Pr}_\pi(\tau_1 \geq i_j + 1)\}^{1/(a_j 2^j)} \right] \\
&\qquad \times \left[\sum_{i_{p+1}=0}^{\infty} \{\mathrm{Pr}_\pi(\tau_1 \geq i_j + 1)\}^{1/(a_{p+1} 2^p)} \right] < \infty.
\end{aligned}$$

APPENDIX B: PROOF OF PROPOSITION 3

B.1 Preliminary Results

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

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

where $\bar{B}_j(b_n) = b_n^{-1}(B((j+1)b_n) - B(jb_n))$ and $\bar{B}(n) = n^{-1}B(n)$.

Lemma B.1 (Damerdji 1994, p. 508). For all $\epsilon > 0$ and for almost all sample paths, there exists $n_0(\epsilon)$ such that for all $n \geq n_0$,

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

Lemma B.2 (Csörgő and Révész 1981). For all $\epsilon > 0$ and for almost all sample paths, there exists $n_0(\epsilon)$ such that for all $n \geq n_0$,

$$|B(n)| < (1+\epsilon)[2n\log\log n]^{1/2}. \tag{B.3}$$

B.2 Proof of Proposition 3

Proposition 3 follows from Lemma 2 and the following two lemmas.

Lemma B.3 (Damerdji 1994, prop. 3.1). Assume that (a) $b_n \to \infty$ and $n/b_n \to \infty$ as $n \to \infty$ and (b) 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 B.4. Assume that (15) holds with $\gamma(n) = n^\alpha \log n$, where $\alpha = 1/(2+\delta)$. If (a) $a_n \to \infty$ as $n \to \infty$, (b) $b_n \to \infty$ and $n/b_n \to \infty$ as $n \to \infty$, and (c) $b_n^{-1} n^{2\alpha}[\log n]^3 \to 0$ 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. 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

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

where $\bar{Y}_j(b_n) = b_n^{-1} \sum_{i=1}^{b_n} Y_{jb_n + i}$ for $j = 0, \ldots, a_n - 1$ and $\bar{Y}(n) = n^{-1} \sum_{i=1}^{n} Y_i$. Because

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

we have

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

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

a. Our assumptions say that there exists a constant C such that for all large n,

$$\left|\sum_{i=1}^{n} g(X_i) - n\mathrm{E}_\pi g - \sigma_g B(n)\right| < Cn^\alpha \log n \quad \text{a.s.} \tag{B.4}$$

Note that

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

and hence, by (B.4),

$$\begin{aligned}
&|\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n)| \\
&\quad \leq \frac{1}{b_n}\left[\left|\sum_{i=1}^{(j+1)b_n} Y_i - \sigma_g B((j+1)b_n)\right| + \left|\sum_{i=1}^{jb_n} Y_i - \sigma_g B(jb_n)\right|\right] \\
&\quad < \frac{2}{b_n} Cn^\alpha \log n.
\end{aligned} \tag{B.5}$$

Then

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

as $n \to \infty$ by conditions (a) and (c).

b. Apply (B.4) to obtain

$$|\bar{Y}(n) - \sigma_g \bar{B}(n)| = n^{-1}\left|\sum_{i=1}^{n} Y_i - \sigma_g B(n)\right| < Cn^{\alpha - 1} \log n. \tag{B.6}$$

Then

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

as $n \to \infty$ by conditions (a) and (b) and because $1 - 2\alpha > 0$.

c. By (B.5) and (B.6),

$$\big|2\big(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n)\big)(\bar{Y}(n) - \sigma_g \bar{B}(n))\big| < 4C^2 b_n^{-1} n^{2\alpha - 1} (\log n)^2.$$

Thus

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

as $n \to \infty$ by condition (a) and because $1 - 2\alpha > 0$.

d. Because $b_n \geq 2$, (B.2) and (B.5) together imply that

$$\begin{aligned}
&\big|\big(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n)\big)\bar{B}_j(b_n)\big| \\
&\quad < 2^{3/2} C(1+\epsilon) b_n^{-1} \\
&\qquad \times \big[b_n^{-1} n^{2\alpha} (\log n)^2 \log(n/b_n) + b_n^{-1} n^{2\alpha} (\log n)^2 \log\log n\big]^{1/2}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\frac{b_n}{a_n - 1} \sum_{j=0}^{a_n - 1} \big|2\sigma_g\big(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n)\big)\bar{B}_j(b_n)\big| \\
&\quad \leq 8\sigma_g C(1+\epsilon) \frac{a_n}{a_n - 1} \\
&\qquad \times \big[b_n^{-1} n^{2\alpha} (\log n)^2 \log(n/b_n) + b_n^{-1} n^{2\alpha} (\log n)^2 \log\log n\big]^{1/2} \\
&\quad \to 0
\end{aligned}$$

as $n \to \infty$ by conditions (a) and (c).

e. By (B.3) and (B.5), $|(\bar{Y}_j(b_n) - \sigma_g \bar{B}_j(b_n))\bar{B}(n)| < 4C(1+\epsilon)b_n^{-1} n^{-1/2+\alpha}(\log n)(\log\log n)^{1/2}$, so that

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

as $n \to \infty$ by condition (a) and because $1/2 - \alpha > 0$.

f. Use (B.2) and (B.6) to get

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

Hence using conditions (a), (b), and (c) shows that as $n \to \infty$,

$$\begin{aligned}&\frac{b_n}{a_n-1}\sum_{j=0}^{a_n-1}\left|2\sigma_g(\bar{Y}(n) - \sigma_g\bar{B}(n))\bar{B}_j(b_n)\right| \\ &< 4\sigma_g C(1+\epsilon)\frac{a_n}{a_n-1}\frac{b_n}{n} \\ &\quad\times\left[b_n^{-1}n^{2\alpha}\left((\log n)^2\log(n/b_n) + (\log n)^2\log\log n\right)\right]^{1/2} \\ &\to 0.\end{aligned}$$

g. Now (B.3) and (B.6) imply that $|(\bar{Y}(n) - \sigma_g\bar{B}(n))\bar{B}(n)| < 2C \times (1+\epsilon)n^{-3/2+\alpha}(\log n)^{3/2}$. Hence

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

as $n \to \infty$ by conditions (a) and (b) and because $1/2 - \alpha > 0$.

APPENDIX C: CALCULATIONS FOR THE EXAMPLE IN SECTION 4.2

We consider a slightly more general formulation of the model given in (16). Suppose that for $i = 1, \ldots, K$,

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

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

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

Let $\pi(\theta,\mu,\lambda|y)$ be the posterior distribution corresponding to the hierarchy in (C.1). Note that θ is a vector containing all of the θ_i's and y is a vector containing all of the data. Consider the factorization

$$\pi(\theta,\mu,\lambda|y) = \pi(\theta|\mu,\lambda,y)\pi(\mu|\lambda,y)\pi(\lambda|y). \tag{C.2}$$

If it is possible to sequentially simulate from each of the densities on the right side of (C.2), then we can produce iid draws from the posterior. Now $\pi(\theta|\mu,\lambda,y)$ is the product of independent univariate normal densities, that is, $\theta_i|\mu,\lambda,y \sim \mathrm{N}((\lambda y_i + a\mu)/(\lambda+a), a\lambda/(\lambda+a))$. In addition, $\pi(\mu|\lambda,y)$ is a normal distribution, that is, $\mu|\lambda,y \sim \mathrm{N}(\bar{y}, (\lambda+a)/K)$. 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 $\pi(\lambda|y)$ if we let $g(\lambda)$ be the kernel of an $\mathrm{IG}(b,c)$ density, then

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

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 $\hat{\lambda} \le 0$, then the maximum occurs at 0.

C.2 Implementing Regenerative Simulation

We begin by establishing the minorization condition (9) 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

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

and hence (9) will follow by setting

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

$$s(\lambda',\mu',\theta') = \varepsilon\inf_{(\lambda,\mu,\theta)\in D}\frac{f(\lambda,\mu|\theta')}{f(\lambda,\mu|\tilde{\theta})},$$

and

$$q(\lambda,\mu,\theta) = \varepsilon^{-1}f(\lambda,\mu|\tilde{\theta})f(\theta|\lambda,\mu)I_{\{(\lambda,\mu,\theta)\in D\}}.$$

Now using (10) shows that when $(\lambda,\mu,\theta) \in D$, the probability of regeneration is given by

$$\Pr(\delta = 1|\lambda',\mu',\theta',\lambda,\mu,\theta) = \left\{\inf_{(\lambda,\mu,\theta)\in D}\frac{f(\lambda,\mu|\theta')}{f(\lambda,\mu|\tilde{\theta})}\right\}\frac{f(\lambda,\mu|\tilde{\theta})}{f(\lambda,\mu|\theta')}. \tag{C.3}$$

Thus we need to calculate the infimum and plug it into (C.3). Toward this end, let $0 < d_1 < d_2 < \infty$ and $-\infty < d_3 < d_4 < \infty$ 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

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

where $\hat{\mu} = d_4 I(\bar{\theta}' \le \bar{\tilde{\theta}}) + d_3 I(\bar{\theta}' > \bar{\tilde{\theta}})$ and $\hat{\lambda} = d_2 I(V(\theta',\hat{\mu}) \le V(\tilde{\theta},\hat{\mu})) + d_1 I(V(\theta',\hat{\mu}) > V(\tilde{\theta},\hat{\mu}))$. We determine the fixed point with a preliminary estimate of the mean of the stationary distribution, and take 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. After some trial and error, we took $d_1 = \max\{.01, \tilde{\lambda} - .5S_\lambda\}$, $d_2 = \tilde{\lambda} + .5S_\lambda$, $d_3 = \tilde{\mu} - S_\mu$, and $d_4 = \tilde{\mu} + S_\mu$.

[Received February 2005. Revised December 2005.]

REFERENCES

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

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

Brockwell, A. E., and Kadane, J. B. (2005), "Identification of Regeneration Times in MCMC Simulation, With Application to Adaptive Schemes," *Journal of Computational and Graphical Statistics*, 14, 436–458.

Caffo, B. S., and Booth, J. G. (2001), "A Markov Chain Monte Carlo Algorithm for Approximating Exact Conditional Probabilities," *Journal of Computational and Graphical Statistics*, 10, 730–745.

Chen, X. (1999), "Limit Theorems for Functionals of Ergodic Markov Chains With General State Space," *Memoirs of the American Mathematical Society*, 139.

Chien, C.-H. (1988), "Small-Sample Theory for Steady-State Confidence Intervals," in *Proceedings of the 1988 Winter Simulation Conference*, eds. M. Abrams, P. Haigh, and J. Comfort, pp. 408–413.

Chow, Y. S., and Robbins, H. (1965), "On the Asymptotic Theory of Fixed-Width Sequential Confidence Intervals for the Mean," *The Annals of Mathematical Statistics*, 36, 457–462.

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.

Cowles, M. K., and Carlin, B. P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, 91, 883–904.

Csáki, E., and Csörgő, 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*, New York: Academic Press.

Damerdji, H. (1991), "Strong Consistency and Other Properties of the Spectral Variance Estimator," *Management Science*, 37, 1424–1440.

——— (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., Fort, G., Moulines, 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.

Fishman, G. S. (1996), *Monte Carlo: Concepts, Algorithms, and Applications*, New York: Springer-Verlag.

Fort, G., and Moulines, E. (2000), "V-Subgeometric Ergodicity for a Hastings–Metropolis Algorithm," *Statistics and Probability Letters*, 49, 401–410.

——— (2003), "Polynomial Ergodicity of Markov Transition Kernels," *Stochastic Processes and Their Applications*, 103, 57–99.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments" (with discussion), in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, London: Clarendon Press, pp. 169–188.

Geyer, C. J. (1992), "Practical Markov Chain Monte Carlo" (with discussion), *Statistical Science*, 7, 473–511.

——— (1999), "Likelihood Inference for Spatial Point Processes," in *Stochastic Geometry: Likelihood and Computation*, eds. O. E. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, Boca Raton, FL: Chapman & Hall/CRC, pp. 79–140.

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.

Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1998), "Adaptive Markov Chain Monte Carlo Through Regeneration," *Journal of the American Statistical Association*, 93, 1045–1054.

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.

——— (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, Dept. 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. (2005), "Using a Markov Chain to Construct a Tractable Approximation of an Intractable Probability Distribution," *Scandinavian Journal of Statistics*, 33, 37–51.

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

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," *The 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.

——— (2004), "Sufficient Burn-in for Gibbs Samplers for a Hierarchical Random-Effects Model," *The Annals of Statistics*, 32, 784–817.

Lai, T. L. (2001), "Sequential Analysis: Some Classical Problems and New Challenges," *Statistica Sinica*, 11, 303–351.

Liu, W. (1997), "Improving the Fully Sequential Sampling Scheme of Anscombe–Chow–Robbins," *The Annals of Statistics*, 25, 2164–2171.

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*, London: Springer-Verlag.

——— (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.

Nadas, A. (1969), "An Extension of a Theorem of Chow and Robbins on Sequential Confidence Intervals for the Mean," *The Annals of Mathematical Statistics*, 40, 667–671.

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. (1999), "A Note on Acceptance Rate Criteria for CLTs for Metropolis–Hastings Algorithms," *Journal of Applied Probability*, 36, 1210–1217.

Roberts, G. O., and Polson, N. G. (1994), "On the Geometric Convergence of the Gibbs Sampler," *Journal of the Royal Statistical Society*, Ser. 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.

——— (1998), "Markov Chain Monte Carlo: Some Practical Implications of Theoretical Results" (with discussion), *Canadian Journal of Statistics*, 26, 5–31.

——— (1999), "Convergence of Slice Sampler Markov Chains," *Journal of the Royal Statistical Society*, Ser. B, 61, 643–660.

——— (2004), "General State Space Markov Chains and MCMC Algorithms," *Probability Surveys*, 1, 20–71.

Rosenthal, J. S. (1995), "Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo," *Journal of the American Statistical Association*, 90, 558–566.

——— (1996), "Analysis of the Gibbs Sampler for a Model Related to James–Stein Estimators," *Statistics and Computing*, 6, 269–275.

Song, W. T., and Schmeiser, B. W. (1995), "Optimal Mean-Squared-Error Batch Sizes," *Management Science*, 41, 110–123.

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics*, 22, 1701–1762.