On the Applicability of Regenerative Simulation in Markov Chain Monte Carlo

James P. Hobert[1]*, Galin L. Jones[2], Brett Presnell[1], and Jeffrey S. Rosenthal[3]

[1]Department of Statistics
University of Florida

and

[2]School of Statistics
University of Minnesota

and

[3]Department of Statistics
University of Toronto

September 2001

*Hobert's research supported by NSF Grant DMS-00-72827. Rosenthal's research supported by ...

1 Introduction

Suppose we want to know the value of $E_\pi g := \int_{\mathcal{X}} g(x)\, \pi(dx)$, where π is a probability distribution with support $\mathcal{X}$ and g is a real-valued, π-integrable function on $\mathcal{X}$. Further suppose that this integral cannot be evaluated analytically nor by standard quadrature methods, and that classical Monte Carlo methods are not an option as obtaining independent and identically distributed (iid) draws from π is prohibitively difficult. In such a case, we might resort to Markov chain Monte Carlo methods (MCMC) which we now explain. Suppose that $\Phi = \{X_0, X_1, X_2, \dots\}$ is an aperiodic, irreducible, positive Harris recurrent Markov chain with state space $\mathcal{X}$ and invariant distribution π [for definitions see Meyn and Tweedie, 1993]. The Ergodic Theorem implies that, with probability 1,

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

The MCMC method entails constructing a Markov chain Φ satisfying the regularity conditions described above and then simulating Φ for a finite number of steps, say n, and using $\bar{g}_n$ as an estimate of $E_\pi g$. The popularity of the MCMC method is due to the ease with which such a Φ can be constructed and simulated [Robert and Casella, 1999].

An obvious and important question that has received far too little attention in the MCMC literature is "How do we construct a legitimate asymptotic standard error for $\bar{g}_n$?" If the X_i's comprising (1) were iid and $E_\pi g^2 < \infty$ then by the central limit theorem (CLT),

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

and the obvious moment estimator of the variance of the asymptotic distribution is consistent. Unfortunately, when the X_i's comprising (1) are a Markov chain, $E_\pi g^2 < \infty$ is no longer sufficient for a CLT to hold. Indeed, the Markov chain must *mix* quickly in order to have CLTs. More specifically, for $n \in \mathbb{N} := \{1, 2, 3, \dots\}$ let $P^n(x, dy)$ be the n-step Markov transition kernel; that is, for $x \in \mathcal{X}$ and a measurable set A, $P^n(x, A) = \Pr(X_n \in A | X_0 = x)$. The assumptions we have thus far made about Φ guarantee that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \downarrow 0 \quad \text{as } n \to \infty, \tag{2}$$

where the left-hand side is the *total variation* distance between $P^n(x, \cdot)$ and $\pi(\cdot)$; that is, the supremum over measurable A of $|P^n(x, A) - \pi(A)|$. We say that Φ is *geometrically ergodic* if this convergence occurs at a geometric rate; that is, if there exists a constant $0 < t < 1$ and a function $M : \mathcal{X} \mapsto \mathbb{R}^+$ such that for any $x \in \mathcal{X}$,

$$\|P^n(x,\cdot) - \pi(\cdot)\| \leq M(x)\, t^n \tag{3}$$

for $n \in \mathbb{N}$. Chan and Geyer [1994] have shown that geometric ergodicity along with a moment condition on the function g guarantee a CLT. Here is their theorem.

Theorem 1. *Suppose that $\Phi = \{X_0, X_1, X_2, \dots\}$ is an aperiodic, irreducible, positive Harris recurrent Markov chain with invariant distribution π. If Φ is geometrically ergodic and $E_\pi|g|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$, then*

$$\sqrt{n}\,(\bar{g}_n - E_\pi g) \xrightarrow{d} N(0, \gamma_g^2) \tag{4}$$

where

$$\gamma_g^2 = \mathrm{Var}_\pi(g(X_0)) + 2\sum_{i=1}^{\infty} \mathrm{Cov}_\pi(g(X_0), g(X_i)).$$

A couple of remarks are in order:

Remark 1. Roberts and Rosenthal [1997] have shown that, if Φ is reversible, the same result holds without the ε; that is, a finite second moment is sufficient. Moreover, Kipnis and Varadhan [1986] provide an even less restrictive CLT for reversible chains.

Remark 2. It is well-known that geometric convergence is not necessary for CLTs (see e.g. Nummelin [1984, Corollary 7.3]). On the other hand, the CLTs that involve weaker assumptions on the convergence rate of Φ do not hold for all functions with a $2+\varepsilon$ moment. For example, Corollary 7.3 of Nummelin [1984] holds for bounded functions and the CLT given in Chan [1993] holds for a single function. This attitude is summed up concisely by Roberts and Rosenthal [1998a] who state: "While not the weakest condition to imply central limit theorems, geometric ergodicity is one of the easiest to check and leads to clean statements."

Making practical use of Chan and Geyer's result requires (i) showing that $E_\pi|g|^{2+\varepsilon} < \infty$, (ii) establishing that Φ converges at a geometric rate, and (iii) finding an easily computed yet consistent estimate of γ_g^2. Since one must establish a moment condition for g even in the iid case, (i) is not unduly restrictive.

Regarding (ii), during the last ten or so years, many standard Markov chains used in MCMC have been shown to be geometrically ergodic. See, for example, Meyn and Tweedie [1994], Mengersen and Tweedie [1996], Roberts and Tweedie [1996], Hobert and Geyer [1998], Roberts and

Rosenthal [1998b], Roberts and Rosenthal [1999], and Jarner and Hansen [2000]. However, there are still many chains used in MCMC to which these results do not apply. The most straightforward method of establishing that a Markov chain is geometrically ergodic is through the development of drift and minorization conditions [Meyn and Tweedie, 1993, Chapter 15]. (See Jones and Hobert [2001] for an introduction to these ideas.) In our opinion, (ii) is becoming less and less of a problem all the time.

On the other hand, (iii) is often a problem. Specifically, finding a consistent estimate of γ^2 that is also easy to compute is often a challenge. There have been many estimators of γ^2 suggested in the Operations Research and Time Series literature. Two of the most commonly used methods are batch means [Bratley, Fox, and Schrage, 1987] and window estimators [Geyer, 1992]. It is well known that the method of batch means will not produce a consistent estimate of γ^2 as long as the batch sizes are fixed. However, batch means is extremely easy to implement and hence popular. Generally speaking, it is possible to impose enough regularity conditions to ensure consistency of window estimators [Geyer, 1992, Priestly, 1981]. However, the optimal choice of lag window is often unclear and, in general, window estimators can be computationally intensive. Many standard simulation texts claim that the methods of batch means and spectral analysis tend to be more effective when the chain is stationary [Bratley et al., 1987, Ripley, 1987]. Thus, they suggest that one must be careful about the burn-in period in order to obtain a good estimate of γ^2 when using these methods.

Fortunately, there is a method of analyzing the simulation output that alleviates all of these concerns. By identifying (random) times at which Φ probabilistically restarts itself, we can represent $\bar{g}_n$ as the ratio of two empirical averages each involving iid terms. This allows us to write the CLT for $\bar{g}_n$ in a slightly different way such that there is an obvious consistent estimator of the variance of the asymptotic normal distribution. This method is known as *regenerative simulation* (RS). Also, we note that RS does not require that the Markov chain be stationary and, in fact, the initial value is drawn from a prescribed distribution. Thus, it is unsurprising that RS, when available, is considered the preferred method for variance estimation [Bratley et al., 1987]. Moreover, in our experience RS has been nearly trivial to implement for many standard MCMC samplers.

In the next section, we discuss RS and the moment assumptions that are necessary to make legitimate use of RS in the MCMC context. Our main result, Theorem 2, is a checkable sufficient condition which guarantees that the appropriate moments are finite. In section 3 we provide a proof of this theorem, and in section 4 we apply these results to the slice sampler.

2 Minorization, Regeneration, and the Central Limit Theorem

In order to use RS in MCMC, we need a *minorization condition* on Φ; that is, we need a function $s : \mathcal{X} \mapsto \mathbb{R}^+$ for which $E_\pi \, s > 0$ and a probability measure Q such that for all $x \in \mathcal{X}$ and all measurable sets A

$$P(x, A) \geq s(x) \, Q(A). \tag{5}$$

Nummelin [1984] calls s a *small function* and Q a *small measure*. When $\mathcal{X}$ is finite it is trivial to establish (5) by fixing a point $x^* \in \mathcal{X}$ and taking $s(x) = I(x = x^*)$ and $Q(\cdot) = P(x^*, \cdot)$. However, when $\mathcal{X}$ is general, our assumptions about Φ do not guarantee the existence of an s and a Q satisfying (5). On the other hand, our assumptions are enough to guarantee that there exists a $k \geq 1$ such that a minorization condition holds for the k-step transition kernel, P^k. This could be difficult to exploit in practice. Fortunately, Mykland, Tierney, and Yu [1995] and Rosenthal [1995] have given recipes for establishing (5) for many of the Gibbs samplers and Metropolis–Hastings algorithms that arise in MCMC. Jones and Hobert [2001] use simple examples to demonstrate techniques for constructing s and Q.

This minorization condition can be used to divide the Markov chain into iid blocks. Specifically, note that (5) allows us to write $P(x, dy)$ as a mixture of two distributions,

$$P(x, dy) = s(x) \, Q(dy) + [1 - s(x)] \, R(x, dy), \tag{6}$$

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 can be used to generate X_{i+1} sequentially as follows. 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 $X_{i+1} \sim R(x, \cdot)$. This is actually a recipe for simulating the so-called *split chain* [Athreya and Ney, 1978, Nummelin, 1984, 1978]

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

which has state space $\mathcal{X} \times \{0, 1\}$ and Markov transition kernel

$$P'\left((x, \delta), dy \times \rho\right) = \begin{cases} Q(dy) s(y)^\rho \left(1 - s(y)\right)^{1-\rho} \, , & \text{if } \delta = 1, \\ R(x, dy) s(y)^\rho \left(1 - s(y)\right)^{1-\rho} \, , & \text{if } \delta = 0, \end{cases} \tag{7}$$

where $\delta, \rho \in \{0,1\}$ [Nummelin, 1984, Section 4.4]. Note that the split chain, Φ', retains the key properties (aperiodicity, irreducibility, and positive Harris recurrence) of the original chain, Φ [Nummelin, 1984, Section 4.4].

The times at which $\delta_i = 1$ are *regeneration times* when Φ' probabilistically restarts itself. Consider the starting value for Φ', call it (X_0, δ_0). The split chain is defined in such a way that, given X_i, the distribution of δ_i is Bernoulli($s(X_i)$). Thus, whenever we discuss starting Φ', we only specify a distribution for X_0 and we use E_Q and E_π to denote expectation for both the split chain Φ' and the marginal chain Φ started with $X_0 \sim Q(\cdot)$ and $X_0 \sim \pi(\cdot)$, respectively.

Remark 3. Sampling directly from the residual distribution can be problematic. Fortunately, there is a simple and clever way of avoiding R altogether. If we write the transition as $X_i \to \delta_i \to X_{i+1}$, we need to generate from $(\delta_i, X_{i+1})|X_i$. Above, we suggested doing this by first drawing from $\delta_i|X_i$ and then drawing from $X_{i+1}|\delta_i, X_i$, which, if $\delta_i = 0$, entails simulation from $R(X_i, dy)$. Mykland et al. [1995] note that simulating from the residual density can be avoided by first drawing from $X_{i+1}|X_i$ (in the usual way) and then drawing from $\delta_i|X_i, X_{i+1}$. Nummelin [1984, p.62] notes that

$$\Pr(\delta_i = 1|X_i, X_{i+1}) = \frac{s(X_i)\, q(X_{i+1})}{k(X_{i+1}|X_i)} \,, \tag{8}$$

where $q(\cdot)$ and $k(\cdot|x)$ are densities corresponding to $Q(\cdot)$ and $P(x, \cdot)$.

Assume Φ' is started with $X_0 \sim Q(\cdot)$. Let $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ be the (random) regeneration times; i.e., $\tau_{t+1} = \min\{i > \tau_t : \delta_{i-1} = 1\}$. Also assume that Φ is run for a fixed number, R, of *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_t be the length of the tth tour; that is, $N_t = \tau_t - \tau_{t-1}$ and define

$$S_t = \sum_{j=\tau_{t-1}}^{\tau_t - 1} g(X_j)$$

for $t = 1, \ldots, R$. The (N_t, S_t) pairs are iid since each is based on a different tour. Let $\bar{N}$ be the average tour length; that is, $\bar{N} = R^{-1} \sum_{t=1}^{R} N_t$ and, analogously, let $\bar{S} = R^{-1} \sum_{t=1}^{R} S_t$. Note that $\tau_R \to \infty$ w.p. 1 as $R \to \infty$. This combined with the Ergodic Theorem yields

$$\bar{g}_R = \frac{\sum_{t=1}^{R} S_t}{\sum_{t=1}^{R} N_t} = \frac{\bar{S}}{\bar{N}} = \frac{1}{\tau_R} \sum_{j=0}^{\tau_R - 1} g(X_j) \to E_\pi g \tag{9}$$

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

Now since $E_Q(N_1) = 1/(E_\pi s) < \infty$, it follows from the Strong Law of Large Numbers that $\bar{N} \to E_Q(N_1)$ w.p. 1 as $R \to \infty$. Hence, it now follows from (9) that $\bar{S} \to E_Q(N_1)\, E_\pi g$ w.p. 1. as

$R \to \infty$. Thus, it must be true that $E_Q|S_1| < \infty$. Appealing to the Strong Law again, we know that $\bar{g}_R$ converges almost surely to $E_Q(S_1)/E_Q(N_1)$. Therefore, $E_Q(S_1) = E_Q(N_1) E_\pi g$ and hence the random variables $S_t - N_t\, E_\pi g$, $t = 1, \dots, R$, are iid with mean zero. Then if $E_Q N_1^2$ and $E_Q S_1^2$ are both finite, we can appeal to the CLT as follows

$$\sqrt{R}\,(\bar{g}_R - E_\pi g) = \frac{1}{\bar{N}}\, R^{-\frac{1}{2}} \sum_{t=1}^{R} (S_t - N_t\, E_\pi g) \xrightarrow{\text{d}} \frac{1}{E_Q(N_1)}\, \mathrm{N}\big(0, E_Q\left[(S_1 - N_1\, E_\pi g)^2\right]\big)\ .$$

Thus,

$$\sqrt{R}\,(\bar{g}_R - E_\pi g) \xrightarrow{\text{d}} \mathrm{N}\big(0, \sigma_g^2\big) \tag{10}$$

where

$$\sigma_g^2 = \frac{E_Q\left[(S_1 - N_1 E_\pi g)^2\right]}{\left[E_Q(N_1)\right]^2}\ .$$

The advantage of (10) over (4) is that there is an obvious and easily computed consistent estimate of σ_g^2. Indeed, consider the estimator

$$\hat{\sigma}_g^2 = \frac{\sum_{t=1}^{R} (S_t - \bar{g}_R N_t)^2}{R \bar{N}^2}. \tag{11}$$

A straightforward calculation shows that the difference between $\hat{\sigma}_g^2$ and

$$\frac{1}{\bar{N}^2} \frac{1}{R} \sum_{t=1}^{R} (S_t - N_t E_\pi g)^2 \tag{12}$$

converges almost surely to 0 as $R \to \infty$. Thus, since (12) is consistent, so is $\hat{\sigma}_g^2$.

All that we need to implement RS is the minorization condition (5) and the ability to simulate from Q. Given the work of Mykland et al. [1995], it is relatively easy to do both of these things with many MCMC samplers. Thus it is often painless to implement RS in MCMC while the benefits are substantial. In particular, since we will draw $X_0 \sim Q(\cdot)$, burn-in is not an issue and we have a consistent estimate of σ_g^2 that is simple to compute. Some applications of RS are discussed in Geyer and Thompson [1995], Gilks, Roberts, and Sahu [1998], Jones and Hobert [2001], and Robert [1995]

Recall that the derivation of the CLT (10) requires the assumption that $E_Q N_1^2$ and $E_Q S_1^2$ are both finite. In practice, this needs to be verified before one can make legitimate use of the regenerative method. Given Chan and Geyer's [1994] result, one might hope that geometric ergodicity of Φ along with $E_\pi |g|^{2+\varepsilon} < \infty$ would imply that $E_Q N_1^2$ and $E_Q S_1^2$ are finite. Our main result shows that this is indeed the case.

Theorem 2. *Let $\Phi = \{X_0, X_1, X_2, \dots\}$ be an aperiodic, irreducible, positive Harris recurrent Markov chain with invariant distribution π. Assume that (5) holds. If Φ is geometrically ergodic and $E_\pi |g|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$, then $E_Q N_1^2$ and $E_Q S_1^2$ are both finite.*

This theorem shows that, in conjunction with the minorization condition (5), the conditions of Chan and Geyer's [1994] CLT are sufficient to assure asymptotic normality of $\bar{g}_R$ and the consistency of the variance estimator $\hat{\sigma}_g^2$ given in (11). Note also that the conclusions of theorem 2 are precisely the moment conditions required by the CLT given in Theorem 17.2.2 of Meyn and Tweedie [1993], thus providing an alternative proof of Chan and Geyer's [1994] CLT, though again with the additional assumption of the minorization condition (5). Since a minorization condition is generally required to verify geometric ergodicity, this additional requirement is not as stringent as it may first appear. Since it is also the key element that make RS and the variance estimator $\hat{\sigma}_g^2$ possible, the practical payoff is great when an appropriate minorization condition can be developed for a given problem.

Remark 4. Of course, the two CLTs (4) and (10) are equivalent. Note first that

$$\sqrt{R}\,\left(\bar{g}_R - E_\pi g\right) = \frac{1}{\sqrt{\tau_R / R}}\,\sqrt{\tau_R}\,\left(\bar{g}_{\tau_R} - E_\pi g\right)\ .$$

Thus, if τ_R were a deterministic sequence tending to ∞ as $r \to \infty$, it would follow from (4) that this sequence converges weakly to $\mathrm{N}\big(0, \gamma_g^2\, E_\pi s\big)$ and hence that $\sigma_g^2 = \gamma_g^2 E_\pi s$. In fact, the proof of Meyn and Tweedie's (1993) Theorem 17.2.2 shows that this remains true despite the fact that τ_R is actually a random sequence converging to ∞ w.p. 1.

3 Proof of the Main Result

Lemma 1. *Let $\Phi = \{X_0, X_1, X_2, \dots\}$ be an aperiodic, irreducible, positive Harris recurrent Markov chain with invariant distribution π. Assume that (5) holds. Then for any function $h : \mathcal{X}^\infty \to \mathbb{R}$ we have*

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

where $c = E_\pi\, s$.

Proof. For any measurable set A it follows from (5) that

$$\pi(A) = \int_{\mathcal{X}} \pi(dx)\, P(x, A) \geq Q(A) \int_{\mathcal{X}} \pi(dx)\, s(x) \tag{13}$$

and hence $\pi(\cdot) \geq c\, Q(\cdot)$. Next note that

$$E_\pi |h(X_0, X_1, \dots)| = E_\pi \left[E \left(|h(X_0, X_1, \dots)| \mid X_0 \right) \right]$$

The inner expectation is a function of X_0 not depending on the starting distribution. Thus, we can use (13) and the Markov property to obtain

$$E_\pi |h(X_0, X_1, \dots)| \geq c\, E_Q \left[E \left(|h(X_0, X_1, \dots)| \mid X_0 \right) \right] = c\, E_Q |h(X_0, X_1, \dots)|$$

□

In order to use Lemma 1 in conjunction with Φ', we need to establish that a minorization condition of the form (5) holds for Φ'. Fortunately, this is straightforward. From (7) we have

$$\begin{aligned} P'\left((x, \delta), dy \times \rho\right) &\geq Q(dy)\, s(y)^\rho \left(1 - s(y)\right)^{1-\rho} I(\delta = 1) \\ &= I(\delta = 1)\, Q'(dy \times \rho), \end{aligned}$$

where the probability measure Q' is defined in an obvious way. Thus, the split chain also satisfies a minorization condition [See also Meyn and Tweedie, 1993, Proposition 5.5.6].

Lemma 2. *Assume that $\Phi = \{X_0, X_1, X_2, \dots\}$ is an aperiodic, irreducible, positive Harris recurrent Markov chain with invariant distribution π. Assume further that* (5) *holds. If Φ is geometrically ergodic, then there exists a $\beta > 1$ such that $E_\pi \beta^{N_1} < \infty$.*

Proof. First $N_1 = \tau_1 = \min\{i > 0 : (X_{i-1}, \delta_{i-1}) \in \mathcal{X} \times \{1\}\}$; that is, N_1 is just the hitting time on the set $\mathcal{X} \times \{1\}$. Now note that Φ and Φ' are what Roberts and Rosenthal [2001] call *co-de-initializing* Markov chains. Consequently, the two chains converge to stationarity at exactly the same rate. In particular, since Φ is geometrically ergodic, so is Φ'. Let π' denote the invariant distribution of Φ'. Note that a random variable (X, δ) with distribution π' can be represented as follows. First, $X \sim \pi(\cdot)$, and conditional on X, $\delta | X \sim \text{Bernoulli}(s(X))$. Now since Φ' is geometrically ergodic and $\pi'(\mathcal{X} \times \{1\}) > 0$, Theorem 2.5 of Nummelin and Tuominen [1982] implies that there exists a $\beta > 1$ such that

$$E_\pi \beta^{N_1} < \infty.$$

□

Proof of Theorem 2. From lemmas 1 and 2, it follows that $E_Q \beta^{N_1} \leq E_\pi \beta^{N_1} < \infty$ for some $\beta > 1$. This of course implies that $E_Q N_1^p < \infty$ for any $p > 0$ and in particular that $E_Q N_1^2 < \infty$.

Next note that

$$S_1^2 = \left(\sum_{j=0}^{\tau_1 - 1} g(X_j) \right)^2 \leq \left(\sum_{j=0}^{\tau_1 - 1} |g(X_j)| \right)^2 = \left(\sum_{j=0}^{\infty} I(0 \leq j \leq \tau_1 - 1) |g(X_j)| \right)^2$$
$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} I(0 \leq i \leq \tau_1 - 1)\, I(0 \leq j \leq \tau_1 - 1)\, |g(X_i)|\, |g(X_j)|.$$

Thus,

$$E_\pi S_1^2 = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E_\pi \left[I(0 \leq i \leq \tau_1 - 1)\, I(0 \leq j \leq \tau_1 - 1)\, |g(X_i)|\, |g(X_j)| \right],$$

and by the Cauchy-Schwartz inequality,

$$E_\pi S_1^2 \leq \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sqrt{E_\pi \left[(I(0 \leq i \leq \tau_1 - 1)\, |g(X_i)|)^2 \right] E_\pi \left[(I(0 \leq j \leq \tau_1 - 1)\, |g(X_j)|)^2 \right]}$$
$$= \left(\sum_{i=0}^{\infty} \sqrt{E_\pi \left[(I(0 \leq i \leq \tau_1 - 1)\, |g(X_i)|)^2 \right]} \right)^2$$
$$= \left(\sum_{i=0}^{\infty} \sqrt{E_\pi \left[I(0 \leq i \leq \tau_1 - 1)\, |g(X_i)|^2 \right]} \right)^2.$$

Now set $q = 1 + \varepsilon/2$ and $p = 1 + 2/\varepsilon$. By Hölder's inequality,

$$E_\pi \left[I(0 \leq i \leq \tau_1 - 1)\, |g(X_i)|^2 \right] \leq \left[E_\pi I(0 \leq i \leq \tau_1 - 1) \right]^{\frac{1}{p}} \left[E_\pi |g(X_i)|^{2q} \right]^{\frac{1}{q}},$$

and since $\left[E_\pi |g(X_i)|^{2q} \right]^{\frac{1}{q}} = \left[E_\pi |g(X_0)|^{2+\varepsilon} \right]^{\frac{1}{q}} = c' < \infty$, it follows that

$$E_\pi S_1^2 \leq c' \left(\sum_{i=0}^{\infty} \left[E_\pi I(0 \leq i \leq \tau_1 - 1) \right]^{\frac{1}{2p}} \right)^2 = c' \left(\sum_{i=0}^{\infty} \left[\Pr_\pi(\tau_1 \geq i + 1) \right]^{\frac{1}{2p}} \right)^2.$$

We know from Lemma 2 that there exists a $\beta > 1$ such that $E_\pi \beta^{N_1} < \infty$, and an simple calculation shows that for any $i = 0, 1, 2, \ldots$,

$$E_\pi \beta^{N_1} \geq \beta^{(i+1)} \Pr_\pi(\tau_1 \geq i + 1).$$

Thus,

$$\sum_{i=0}^{\infty} \left[\Pr_{\pi'}(\tau_1 \geq i + 1) \right]^{\frac{1}{2p}} \leq \left(E_\pi \beta^{N_1} \right)^{\frac{1}{2p}} \sum_{i=0}^{\infty} \beta^{\frac{-(i+1)}{2p}} < \infty.$$

Therefore, $E_\pi S_1^2$ is finite and an application of Lemma 1 again yields the result. □

4 Regeneration and the Slice Sampler

4.1 Background

Let $\pi : \mathbb{R}^d \to [0, \infty)$ be a d-dimensional probability density function. Suppose that π can be factored as $\pi(x) = q(x)\, l(x)$ where q is nonnegative and l is strictly positive. Consider a univariate auxiliary variable ω such that the joint density of x and ω is given by

$$\pi(x, \omega) = q(x)\, I[0 < \omega < l(x)].$$

Note that $\int \pi(x, \omega)\, d\omega = \pi(x)$. The *simple slice sampler* [Neal, 2000] is just the Gibbs sampler applied to the joint density $\pi(x, \omega)$. So our Markov chain takes the form $\Phi = \{(\omega_0, x_0), (\omega_1, x_1), \dots\}$ and the Markov transition density is simply

$$k(x, \omega | x', \omega') = \pi(\omega | x')\, \pi(x | \omega),$$

where $\omega | x \sim \text{uniform}(0, l(x))$ and $\pi(x|\omega) \propto q(x) I[l(x) > \omega]$. The Markov chain Φ is aperiodic, π-irreducible, and Harris recurrent [Mira and Tierney, 2001, Roberts and Rosenthal, 1999].

In order to use regenerative simulation in conjunction with Φ, we must show that Φ is geometrically ergodic and we must establish a minorization condition of the form (5). As we pointed out in Section 1, a great deal of work has been done over the last few years establishing conditions under which some popular MCMC algorithms are geometrically ergodic. The following result is due to Roberts and Rosenthal [1999].

Theorem 3. *Let Φ be the simple slice sampler described above. Define $Q(\omega) = \int q(x)\, I[l(x) > \omega]\, dx$ and*

$$G(\omega) = \omega^{\frac{1}{\alpha}+1} \frac{\partial}{\partial \omega} Q(\omega).$$

If π is bounded and there exists an $\alpha > 1$ such that $G(\omega)$ is non-increasing on an open set containing 0, then Φ is geometrically ergodic.

We now use a technique described by Mykland et al. [1995] to construct a minorization condition for Φ. Fix a "distinguished point" $\tilde{x} \in \mathbb{R}^d$. Now

$$\begin{aligned}
k(x, \omega | x', \omega') &= \pi(x|\omega) \, \frac{I[0 < \omega < l(x')]}{l(x')} \\
&\geq \pi(x|\omega) \, \frac{I[0 < \omega < l(x')]}{l(x')} \, I[0 < \omega < l(\tilde{x})] \, I[l(x') \geq l(\tilde{x})] \\
&= \pi(x|\omega) \, \frac{I[0 < \omega < l(\tilde{x})]}{l(\tilde{x})} \, \left[\frac{l(\tilde{x})}{l(x')} \, I[l(x') \geq l(\tilde{x})] \right] \\
&= q(\omega, x) s(x')
\end{aligned}$$

where $q(\omega, x) = \pi(\omega|\tilde{x}) \, \pi(x|\omega)$ is simply a special case of the MTD of Φ. In this particular case, it is easy to sample from the residual density, which is given by

$$r(x, \omega | x', \omega') = \left\{ \frac{I[l(\tilde{x}) < \omega < l(x')]}{l(x') - l(\tilde{x})} I\big[l(x') \geq l(\tilde{x})\big] + \pi(\omega|x') \, I\big[l(x') < l(\tilde{x})\big] \right\} \pi(x|\omega).$$

Here's an overview of simulating the split chain. Suppose the current value is (ω_i, x_i). If $l(x_i) < l(\tilde{x})$, then $\delta_i = 0$ w.p. 1 and we draw (ω_{i+1}, x_{i+1}) as usual from $\pi(\omega|x_i) \, \pi(x|\omega)$. Now suppose that $l(x_i) > l(\tilde{x})$. First, draw $\delta_i \sim \text{Bernoulli}\,(l(\tilde{x})/l(x_i))$. If $\delta_i = 1$, draw (ω_{i+1}, x_{i+1}) from $\pi(\omega|\tilde{x}) \, \pi(x|\omega)$. If, on the other hand, $\delta_i = 0$, draw ω_{i+1} uniformly from the interval $(l(\tilde{x}), l(x_i))$ and then, conditional on ω_{i+1}, draw $x_{i+1} \sim \pi(x|\omega_{i+1})$.

Now, suppose we know (x_i, ω_i) and (x_{i+1}, ω_{i+1}) and consider trying to infer the value of δ_i. If $l(x_i) < l(\tilde{x})$, then we know $\delta_i = 0$. Now suppose that $l(x_i) > l(\tilde{x})$. If $\omega_{i+1} < l(\tilde{x})$ then δ_i must have been 1. Conversely, if $\omega_{i+1} \in (l(\tilde{x}), l(x_i))$, then δ_i must have been 0. Thus, it's easy to see (without using (8)) that

$$\Pr\left[\delta_i = 1 | (x_i, \omega_i), (x_{i+1}, \omega_{i+1})\right] = I[0 < \omega_{i+1} < l(\tilde{x}) < l(x_i)].$$

A specific example is examined in the following subsection.

4.2 An Example

The following example was introduced in Damien, Wakefield, and Walker [1999]. Fix $\tau \in \mathbb{R}$ and consider the univariate density

$$\pi(x; \tau) \propto \exp\left\{ -e^x - \frac{1}{2}(x - \tau)^2 \right\}.$$

Suppose we want to know $E_\pi \, g$ where $g(x) = x$; that is, we want to calculate

$$\int_{\mathbb{R}} x \, \pi(x; \tau) \, dx = \frac{\int_{\mathbb{R}} x \, \exp\left\{-e^x - \frac{1}{2}(x - \tau)^2\right\} \, dx}{\int_{\mathbb{R}} \exp\left\{-e^x - \frac{1}{2}(x - \tau)^2\right\} \, dx}.$$

While these integrals have no closed form solution, $\pi(x;\tau)$ is univariate and hence it is quite straightforward to approximate μ using numerical integration or rejection sampling. We will use this simple example to illustrate the application of regeneration in the slice sampler. The result will be checked against an essentially exact answer based on rejection sampling.

Consider an application of the simple slice sampler with $q(x) = \exp\left\{-\frac{1}{2}(x-\tau)^2\right\}$ and $l(x) = \exp\{-e^x\}$. Note that

$$\{x : l(x) > \omega\} = \{x : x < \log\log(1/\omega)\}\,.$$

Therefore, in this case, $\pi(x|\omega)$ is just a truncated normal density; specifically, $\pi(x|\omega) \propto \phi(x-\tau)\, I\left[x < \log\log(1/\omega)\right]$ where $\phi(\cdot)$ is the standard normal density. We now show that this simple slice sampler satisfies the conditions of Theorem 3 and is thus geometrically ergodic.

First, π is clearly bounded. Now,

$$G(\omega) = \frac{\omega^{\frac{1}{\alpha}}}{\log(\omega)}\, q\left(\log\log(1/\omega)\right).$$

Thus,

$$\frac{\partial}{\partial\omega}\, G(\omega) = \frac{\omega^{\frac{1}{\alpha}-1}}{\log(\omega)} \left\{ \frac{q\left(\log\log(1/\omega)\right)}{\alpha} + \frac{1}{\log(\omega)} \left[q'\left(\log\log(1/\omega)\right) - q\left(\log\log(1/\omega)\right)\right] \right\},$$

where q' denotes the derivative of q. A straightforward calculation shows that $q'\left(\log\log(1/\omega)\right) - q\left(\log\log(1/\omega)\right)$ is negative as long as $\omega < \exp\left\{-e^{\tau-1}\right\} < 1$. Hence, for any $\alpha > 1$, $G(\omega)$ is non-increasing for $\omega < \exp\left\{-e^{\tau-1}\right\}$. Hence, by Theorem 3, this simple slice sampler is geometrically ergodic.

Note that the moment generating function associated with $\pi(x;\tau)$ exists so $E_\pi|X|^{2+\varepsilon} < \infty$ for any positive ε. We set $\tau = 0$, $\tilde{x} = -1/2$ and ran the simple slice sampler for 1 million regenerations. This took about two minutes on a fast workstation. The resulting estimate of $E_\pi\, g$ was $\bar{g}_R = \bar{S}/\bar{N} = -1.5383/2.2671 = -0.6785$ and $\hat{\sigma}_g^2 = 2.0795$. Thus, the asymptotic standard error is about .0014.

As a check, we used a rejection sampler with a $\mathrm{N}(-1/2, 1)$ candidate to get an iid sample of size 10 million from $\pi(x;0)$. Based on this sample, an asymptotic 95% confidence interval for $E_\pi\, g$ is -0.6782 ± 0.0005.

Acknowledgment. The authors are grateful to Charlie Geyer for some useful conversations.

References

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

Paul Bratley, Bennet L. Fox, and Linus E. Schrage. *A Guide to Simulation*. Springer–Verlag, New York, 1987.

K. S. Chan. On the central limit theorem for an ergodic Markov chain. *Stochastic Processes and their Applications*, 47:113–117, 1993.

Kung Sik Chan and Charles J. Geyer. Comment on "Markov chains for exploring posterior distributions". *The Annals of Statistics*, 22:1747–1758, 1994.

Paul Damien, Jon Wakefield, and Stephen Walker. Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *Journal of the Royal Statistical Society,* Series B, 61:331–344, 1999.

Charles J. Geyer. Practical Markov chain Monte Carlo (with discussion). *Statistical Science*, 7: 473–511, 1992.

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

Walter R. Gilks, Gareth O. Roberts, and Sujit K. Sahu. Adaptive Markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association*, 93:1045–1054, 1998.

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

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

Galin L. Jones and James P. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science* (to appear), 2001.

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

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

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

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

Antoinetta Mira and Luke Tierney. On the use of auxiliary variables in Markov chain Monte Carlo sampling. *Scandinavian Journal of Statistics* (to appear), 2001.

Per Mykland, Luke Tierney, and Bin Yu. Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, 90:233–241, 1995.

Radford M. Neal. Slice sampling. Technical report, University of Toronto, 2000.

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

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

Esa Nummelin and Pekka Tuominen. Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stochastic Processes and their Applications*, 12:187–202, 1982.

M. B. Priestly. *Spectral Analysis and Time Series*. Academic, London, 1981.

Brian D. Ripley. *Stochastic Simulation*. John Wiley and Sons, New York, 1987.

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

Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer, New York, 1999.

G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83:95–110, 1996.

Gareth O. Roberts and Jeffrey S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2:13–25, 1997.

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

Gareth O. Roberts and Jeffrey S. Rosenthal. On convergence rates of Gibbs samplers for uniform distributions. *The Annals of Applied Probability*, 8:1291–1302, 1998b.

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

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

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