

**On the applicability of regenerative simulation
in Markov chain Monte Carlo**

BY JAMES P. HOBERT

*Department of Statistics, University of Florida,
Gainesville, Florida 32611, U.S.A.*

`jhobert@stat.ufl.edu`

GALIN L. JONES

*School of Statistics, University of Minnesota,
Minneapolis, Minnesota 55455, U.S.A.*

`galin@stat.umn.edu`

BRETT PRESNELL

*Department of Statistics, University of Florida,
Gainesville, Florida 32611, U.S.A.*

`presnell@stat.ufl.edu`

JEFFREY S. ROSENTHAL

*Department of Statistics, University of Toronto,
Toronto, Ontario, M5S 3G3, Canada*

`jeff@math.toronto.edu`

SUMMARY

We consider the central limit theorem and the calculation of asymptotic standard errors for the ergodic averages constructed in Markov chain Monte Carlo. Chan & Geyer (1994) established a central limit theorem for ergodic averages by assuming that the underlying Markov chain is geometrically ergodic and that a simple moment condition is satisfied. While it is relatively straightforward to check Chan and Geyer's conditions, their theorem does not lead to a consistent and easily computed estimate of the variance of the asymptotic normal distribution. Conversely, Mykland et al. (1995) discuss the use of regeneration to establish an alternative central limit theorem with the advantage that a simple, consistent estimator of the asymptotic variance is readily available. However, their result assumes a pair of unwieldy moment conditions whose verification is difficult in practice. In this paper, we show that the conditions of Chan and Geyer's theorem are sufficient to establish Mykland et al.'s central limit theorem. This result, in conjunction with other recent developments, should pave the way for more widespread use of the regenerative method in Markov chain Monte Carlo. Our results are illustrated in the context of the slice sampler.

Some key words: Asymptotic standard error; Burn-in; Central limit theorem; Geometric ergodicity; Minorisation condition; Slice sampler.

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 draws from π is prohibitively difficult. In such a case, we might resort to Markov chain Monte Carlo methods 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 & Tweedie (1993, Part I). The Ergodic Theorem implies that, with probability 1,

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

The Markov chain Monte Carlo method entails constructing a Markov chain Φ satisfying the regularity conditions described above and then simulating Φ for a finite number of steps, n say, and using \bar{g}_n as an estimate of $E_\pi g$. The popularity of the Markov chain Monte Carlo method is due to the ease with which such a Φ can be constructed and simulated (Robert & Casella, 1999, Ch. 6-7).

An important question that has received too little attention in the Markov chain Monte Carlo literature is that of how we construct a legitimate asymptotic standard error for \bar{g}_n . If the X_i 's comprising (1) were independent and identically distributed from π and $E_\pi g^2 < \infty$ then, by the central limit theorem,

$$\sqrt{n} (\bar{g}_n - E_\pi g) \rightarrow N \{0, E_\pi g^2 - (E_\pi g)^2\},$$

in distribution, and the obvious moment estimator of the variance of the asymptotic distribution is consistent. Unfortunately, when the X_i 's comprising (1) follow a Markov chain, the condition $E_\pi g^2 < \infty$ is no longer sufficient for a central limit theorem to hold. Indeed, the Markov chain must mix quickly in order to allow central limit theorems. Making this precise

requires a couple of definitions. For $n \in \mathbb{N} := \{1, 2, 3, \dots\}$ let $P^n(x, dy)$ be the n -step Markov transition kernel; i.e., for $x \in \mathcal{X}$ and a measurable set A , $P^n(x, A) = \text{pr}(X_n \in A | X_0 = x)$. The assumptions we have made so far about Φ guarantee that

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

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 $x \in \mathcal{X}$,

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

for all $n \in \mathbb{N}$. Chan & Geyer (1994) have shown that geometric ergodicity along with a moment condition on the function g guarantee a central limit theorem. 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*

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

in distribution, where

$$\gamma_g^2 = \text{var}_\pi\{g(X_0)\} + 2 \sum_{i=1}^{\infty} \text{cov}_\pi\{g(X_0), g(X_i)\}.$$

Remark 1. Roberts & Rosenthal (1997) have shown that, if Φ is reversible, the same result holds without the ε ; that is, a finite second moment is sufficient. See also Kipnis & Varadhan (1986). Furthermore, the necessity of the ε in the non-reversible case is an open question; that is, no one has given an example of a geometrically ergodic chain such that the central limit theorem fails to hold for a function g with $E_\pi g^2 < \infty$.

Remark 2. It is well known that geometric convergence is not necessary for central limit theorems. On the other hand, the central limit theorems that involve weaker assumptions on the convergence rate of Φ do not hold for all functions with a $2+\varepsilon$ moment. For example, the central limit theorem in Corollary 7.3 of Nummelin (1984) holds for bounded functions and the central limit theorem given in Chan (1993) holds for a single function. This attitude is summed up concisely by Roberts & Rosenthal (1998a) who state that ‘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 independent and identically distributed case, (i) is not unduly restrictive.

Regarding (ii), during the last ten or so years, many standard Markov chains used in Markov chain Monte Carlo have been shown to be geometrically ergodic; see for example Meyn & Tweedie (1994), Mengersen & Tweedie (1996), Roberts & Tweedie (1996), Hobert & Geyer (1998), Roberts & Rosenthal (1998b), Roberts & Rosenthal (1999) and Jarner & Hansen (2000). However, there are still many chains used in Markov chain Monte Carlo 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 minorisation conditions (Meyn & Tweedie, 1993, Ch. 15); see Jones & Hobert (2001) for an introduction to these ideas. In our opinion, (ii) is becoming less and less of a problem as time passes.

Remark 3. If a drift condition is used to establish that Φ is geometrically ergodic, then it may not be necessary to verify that $E_\pi |g|^{2+\varepsilon} < \infty$. To be specific, one method of establishing geometric ergodicity entails finding a function $V : \mathcal{X} \mapsto [1, \infty)$ and a petite set C such that for all $x \in \mathcal{X}$,

$$E \{V(X_{n+1}) | X_n = x\} \leq \lambda V(x) + bI_C(x) ,$$

where $\lambda \in (0, 1)$, $b < \infty$ and I is the usual indicator function; see Meyn & Tweedie (1993, Part III). This drift condition actually implies that a central limit theorem holds for every

function g such that $g^2(x) \leq V(x)$ for all $x \in \mathcal{X}$.

Unfortunately, (iii) is often a problem; finding an easily computed, consistent estimate of γ_g^2 can be quite challenging. Many estimators of γ_g^2 have been suggested in the operational research and time series literatures. Two of the most commonly used methods are batch means (Bratley et al., 1987, Ch. 3) and window estimators (Geyer, 1992). The method of batch means is popular because it is easily implemented, but it is well known that this method will not produce a consistent estimator of γ_g^2 as long as the batch sizes are fixed. Generally speaking, it is possible to impose enough regularity conditions to ensure consistency of window estimators (Geyer, 1992; Priestley, 1981, Ch. 6). However, the optimal choice of lag window is often unclear and, in general, window estimators are quite computationally intensive. Another disadvantage of batch means and window estimators is that they tend to be more effective when the chain is stationary. Thus it may be necessary to ascertain an appropriate burn-in period in order to use these methods effectively (see e.g. Bratley et al., 1987, Ch. 3; Ripley, 1987, Ch. 6).

Fortunately, there is an alternative central limit theorem that allows one to construct legitimate asymptotic standard errors without having to estimate γ_g^2 directly and without having to worry about burn-in. The basic idea is as follows. 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 independent and identically distributed terms. This allows us to write the central limit theorem for \bar{g}_n in a slightly different way such that there is an obvious, easily computed, consistent estimator of the variance of the asymptotic normal distribution. This method is known as regenerative simulation and does not require that the Markov chain be stationary. In fact, the initial value is drawn from a prescribed distribution. Given these advantages, it is not surprising that regenerative simulation, when available, is considered the preferred method for variance estimation (Bratley et al., 1987, Ch. 3).

A major stumbling block that has prevented more widespread use of regenerative simulation in the Markov chain Monte Carlo context is the pair of complicated moment conditions that must be verified in order to ensure that the alternative central limit theorem, upon

which regenerative simulation is based, holds. Mykland et al. (1995) introduced the statistical community to regenerative simulation as a way of calculating standard errors of estimators based on Markov chain Monte Carlo, but these authors did not address the problem of verifying these moment conditions. In this paper, we provide a solution to this problem.

In the next section, we discuss regenerative simulation and the moment assumptions that are necessary to make legitimate use of regenerative simulation in the Markov chain Monte Carlo context. This section also contains a statement of our main result, Theorem 2, which provides a checkable sufficient condition that guarantees that the moment conditions are satisfied. The proof of Theorem 2 is given in § 3, and our results are illustrated using the slice sampler in § 4.

2. MINORISATION, REGENERATION AND THE CENTRAL LIMIT THEOREM

In order to use regenerative simulation in Markov chain Monte Carlo, we need a minorisation condition on Φ ; that is, we need a function $s : \mathcal{X} \mapsto [0, 1]$ 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 countable, it is trivial to establish (5) by fixing a point $x^* \in \mathcal{X}$ and taking $s(x) = I(x = x^*)$ and $Q(\cdot) = P(x^*, \cdot)$, but our assumptions about Φ do not guarantee the existence of an s and a Q satisfying (5) for general \mathcal{X} and P ; they do guarantee the existence of a $k \geq 1$ such that a minorisation condition holds for the k -step transition kernel, P^k , but it may be difficult to exploit this in practice. Generally speaking, it can be very difficult, if not impossible, to establish a viable minorisation condition for a complex Markov chain Monte Carlo algorithm. Fortunately, Mykland et al. (1995) and Rosenthal (1995) have given recipes for constructing s and Q satisfying (5) for many of the Gibbs samplers and Metropolis–Hastings algorithms that arise in Markov chain Monte Carlo. Jones & Hobert (2001) use simple examples to

demonstrate techniques for constructing s and Q .

This minorisation condition can be used to divide the Markov chain into independent and identically distributed blocks. To be specific, note that (5) allows us to write $P(x, dy)$ as a mixture of two distributions,

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

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 $s(x) = 1$. In practice, this mixture can be used to generate X_{i+1} sequentially as follows. Given $X_i = x$, generate $\delta_i \sim \text{Ber}\{s(x)\}$. If $\delta_i = 1$, then draw $X_{i+1} \sim Q(\cdot)$; otherwise draw $X_{i+1} \sim R(x, \cdot)$. This is actually a recipe for simulating the so-called ‘split chain’ (Athreya & Ney, 1978; Nummelin, 1978, 1984)

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

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

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

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

Let (X_0, δ_0) be the starting value for Φ' . The split chain is defined in such a way that, given X_i , the distribution of δ_i is $\text{Ber}\{s(X_i)\}$. 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.

If $\delta_i = 1$, then time $i + 1$ is a ‘regeneration time’ when Φ' probabilistically restarts itself. To be specific, suppose we start Φ' with $X_0 \sim Q$. Then, each time that $\delta_i = 1$, $X_{i+1} \sim Q$ and the chain is, in effect, probabilistically starting over again.

Remark 4. Sampling directly from the residual distribution can be problematic. Fortunately, there is a simple and clever way around this. If we write the transition as $X_i \rightarrow \delta_i \rightarrow X_{i+1}$,

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

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

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

We now explain how this regenerative structure can be exploited to construct an alternative central limit theorem that leads to a simple method for computing asymptotic standard errors; see Geyer & Thompson (1995) and Ripley (1987, Ch. 6) for similar developments. Assume that Φ' is started with $X_0 \sim Q(\cdot)$. Let $0 = \tau_0 < \tau_1 < \tau_2 < \dots$ 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 on the R th occasion that a $\delta_i = 1$. Thus, the total length of the simulation, τ_R , is random. Let N_t be the length of the t th 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 $t = 1, \dots, R$. The (N_t, S_t) pairs are independent and identically distributed 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 $\tau_R \rightarrow \infty$ with probability 1 as $R \rightarrow \infty$. This combined with the Ergodic Theorem yields

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

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

By Kac's theorem, $E_Q(N_1) = 1/(E_{\pi}s) < \infty$. It follows from the Strong Law of Large Numbers that $\bar{N} \rightarrow E_Q(N_1)$ with probability 1 as $R \rightarrow \infty$, which together with (9) implies

that $\bar{S} \rightarrow E_Q(N_1) E_\pi g$ with probability 1 as $R \rightarrow \infty$. Thus, it must be true that $E_Q|S_1| < \infty$. Appealing to the Strong Law again, we know that \bar{g}_{τ_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 independent and identically distributed with mean zero. Then, if $E_Q N_1^2$ and $E_Q S_1^2$ are both finite, we can appeal to the central limit theorem to show that

$$\sqrt{R}(\bar{g}_{\tau_R} - E_\pi g) = \frac{1}{\bar{N}} R^{-\frac{1}{2}} \sum_{t=1}^R (S_t - N_t E_\pi g) \rightarrow \frac{1}{E_Q(N_1)} \text{N}[0, E_Q \{(S_1 - N_1 E_\pi g)^2\}] ,$$

in distribution. Thus, also in distribution,

$$\sqrt{R}(\bar{g}_{\tau_R} - E_\pi g) \rightarrow \text{N}(0, \sigma_g^2) , \quad (10)$$

where

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

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

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

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

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

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

There is another important advantage of regenerative simulation over other methods. The regenerative simulation method does not require the Markov chain to be stationary. In fact, the algorithm dictates that $X_0 \sim Q(\cdot)$. Therefore, burn-in is not an issue.

We reiterate that it may be difficult, if not impossible, to establish the minorisation condition necessary for the implementation of the regenerative simulation procedure. However, in our experience, once such a minorisation condition is established, implementing the procedure is nearly trivial. Some applications of regenerative simulation are discussed in Geyer & Thompson (1995), Gilks et al. (1998), Jones & Hobert (2001) and Robert (1995).

Recall that the derivation of the central limit theorem (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 & 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 minorisation condition (5), the conditions of Chan & Geyer's (1994) central limit theorem are sufficient to assure asymptotic normality of \bar{g}_{τ_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 central limit theorem given in Theorem 17.2.2 of Meyn & Tweedie (1993), thus providing an alternative proof of Chan & Geyer's (1994) central limit theorem, though again with the additional assumption of the minorisation condition (5). Since a minorisation condition is generally established en route to verifying geometric ergodicity, this additional requirement is not as stringent as it may first appear. Since it is also the key element that make regenerative simulation and the variance estimator $\hat{\sigma}_g^2$ possible, the practical payoff is great when an appropriate minorisation condition can be developed for a given problem.

Remark 5. Of course, the two central limit theorems (4) and (10) are closely connected. Note that

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

Thus, if τ_R were a deterministic sequence satisfying $\tau_R/R \rightarrow 1/E_\pi(s)$ as $R \rightarrow \infty$, it would follow from (4) that $\sqrt{R} (\bar{g}_{\tau_R} - E_\pi g) \rightarrow N(0, \gamma_g^2 E_\pi s)$, in distribution, and hence that $\sigma_g^2 = \gamma_g^2 E_\pi s$ in (10). In fact, the proof of Meyn & 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 ∞ with probability 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 $h : \mathcal{X}^\infty \rightarrow \mathbb{R}$ we have that*

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

where $c = E_\pi s$.

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

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

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

$$E_\pi |h(X_0, X_1, \dots)| = E_\pi [E\{|h(X_0, X_1, \dots)| | X_0\}].$$

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

$$E_\pi |h(X_0, X_1, \dots)| \geq c E_Q [E\{|h(X_0, X_1, \dots)| | X_0\}] = c E_Q |h(X_0, X_1, \dots)|. \quad \square$$

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

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

where the probability measure Q' is defined in the obvious way. Thus, the split chain also satisfies a minorisation condition; see also Meyn & 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 note that $N_1 = \tau_1 = \min \{i > 0 : (X_{i-1}, \delta_{i-1}) \in \mathcal{X} \times \{1\}\}$; 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 & 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' and note that a random vector (X, δ) with distribution π' satisfies $X \sim \pi(\cdot)$, and, conditional on X , $\delta|X \sim \text{Ber}\{s(X)\}$. Thus $\pi'(\mathcal{X} \times \{1\}) = E_\pi(s) > 0$, and, since Φ' is geometrically ergodic, Theorem 2.5 of Nummelin & Tuominen (1982) then implies that there exists a $\beta > 1$ such that

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

Proof of Theorem 2. From Lemmas 1 and 2, it follows that $E_Q \beta^{N_1} \leq c^{-1} E_\pi \beta^{N_1} < \infty$ 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

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

Thus,

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

and, by the Cauchy-Schwartz inequality,

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

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

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

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

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

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

$$E_\pi \beta^{N_1} \geq \beta^{(i+1)} \text{pr}_\pi(\tau_1 \geq i + 1). \quad (14)$$

Thus,

$$\sum_{i=0}^{\infty} \{\text{pr}_\pi(\tau_1 \geq i + 1)\}^{\frac{1}{2p}} \leq (E_\pi \beta^{N_1})^{\frac{1}{2p}} \sum_{i=0}^{\infty} \beta^{-\frac{(i+1)}{2p}} < \infty.$$

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

Remark 6. One of the hypotheses of Theorem 2 is that Φ is geometrically ergodic. Note that, in proving Theorem 2, the only place we use this assumption is in Lemma 2 to conclude that there exists a $\beta > 1$ such that $E_\pi \beta^{N_1} < \infty$. In fact, the arguments after Lemma 2 still go through, with minor changes, if we replace the assumption that $E_\pi \beta^{N_1} < \infty$ for some $\beta > 1$ with the weaker assumption that $E_\pi N_1^m < \infty$ for all $m \in \{0, 1, \dots\}$. The key difference is that (14) is replaced by $E_\pi N_1^m \geq (i + 1)^m \text{pr}_\pi(\tau_1 \geq i + 1)$. Therefore, we could replace the assumption that Φ is geometrically ergodic with the weaker assumption that Φ is polynomially ergodic of all orders (Jarner & Roberts, 2002; Nummelin & Tuominen, 1983; Tuominen & Tweedie, 1994), a consequence of which is that, for all $x \in \mathcal{X}$ and all $m \in \{0, 1, \dots\}$,

$$n^m \|P^n(x, \cdot) - \pi(\cdot)\| \downarrow 0 \quad \text{as } n \rightarrow \infty.$$

The difference between geometric and polynomial convergence can be important from a practical standpoint because some commonly used Markov chain Monte Carlo algorithms are not geometrically ergodic, but are polynomially ergodic of all orders; see Fort & Moulines (2000) for an example involving the Metropolis–Hastings algorithm.

4. REGENERATION AND THE SLICE SAMPLER

4.1. Background

Let $\pi : \mathbb{R}^d \rightarrow [0, \infty)$ 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

$$\pi(x, \omega) = q(x) I\{0 < \omega \leq l(x)\}.$$

Note that $\int \pi(x, \omega) d\omega = \pi(x)$. The simple slice sampler is just the Gibbs sampler applied to the joint density $\pi(x, \omega)$. Our Markov chain therefore 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 $\omega | x \sim \text{Un}\{0, l(x)\}$ and $\pi(x | \omega) \propto q(x) I\{l(x) \geq \omega\}$. The Markov chain Φ is aperiodic, π -irreducible and Harris recurrent (Mira & Tierney, 2001; Roberts & Rosenthal, 1999); see Neal (2002) for a general treatment of slice sampling.

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

Theorem 3. *Let Φ be the simple slice sampler described above. Define*

$$Q(\omega) = \int q(x) I\{l(x) \geq \omega\} dx,$$

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 nonincreasing on $(0, \epsilon)$ for some $\epsilon > 0$, then Φ is geometrically ergodic.

We now use a technique described by Mykland et al. (1995) to construct a minorisation condition for Φ . Fix a ‘distinguished point’ $\tilde{x} \in \mathbb{R}^d$. Now

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

where $q(x, \omega) = \pi(\omega|\tilde{x}) \pi(x|\omega)$ is simply a special case of the Markov transition density of Φ and $s(x', \omega') = s(x') = I\{l(x') > l(\tilde{x})\} l(\tilde{x})/l(x')$ is a function of x' only. With this choice of s and q the residual density is given by

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

which is easily derived after noting that $\{1 - s(x')\}^{-1} = I\{l(x') \leq l(\tilde{x})\} + I\{l(x') > l(\tilde{x})\} l(x')/\{l(x') - l(\tilde{x})\}$.

In this particular case, it is easy to sample from the residual density. Here is an overview of how to simulate the split chain. Suppose the current value is (ω_i, x_i) . If $l(x_i) \leq l(\tilde{x})$, then $\delta_i = 0$ with probability 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 $\delta_i \sim \text{Ber}\{l(\tilde{x})/l(x_i)\}$. 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 , as in Remark 4. If $l(x_i) \leq l(\tilde{x})$, then we know that $\delta_i = 0$. Now suppose that $l(x_i) > l(\tilde{x})$. If $\omega_{i+1} \leq l(\tilde{x})$ 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 is easy to see, without using (8), that

$$\text{pr}\{\delta_i = 1 | (x_i, \omega_i), (x_{i+1}, \omega_{i+1})\} = I\{0 < \omega_{i+1} \leq l(\tilde{x}) < l(x_i)\}.$$

4.2. An Example

The following example was introduced by Damien et al. (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 $E_\pi g$ 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; to be specific, $\pi(x|\omega) \propto \phi(x - \tau) I[x < \log \log(1/\omega)]$ 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,

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

Thus,

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

where q' denotes the derivative of q . A straightforward calculation shows that

$$q' \{ \log \log(1/\omega) \} - q \{ \log \log(1/\omega) \}$$

is negative as long as $\omega < \exp(-e^{\tau-1}) < 1$. Hence, for any $\alpha > 1$, $G(\omega)$ is nonincreasing for $\omega < \exp(-e^{\tau-1})$. Hence, by Theorem 3, this simple slice sampler is geometrically ergodic.

Note that the moment generating function associated with the density $\pi(x; \tau)$ exists, so that $E_\pi |X|^{2+\varepsilon} < \infty$ for any positive ε . We set $\tau = 0$ and $\tilde{x} = -1/2$ 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 $\bar{g}_{\tau_R} = \bar{S}/\bar{N} = -1.5383/2.2671 = -0.6785$ and $\hat{\sigma}_g^2 = 2.0795$. Thus, the asymptotic standard error is about 0.0014. As a check, we used a rejection sampler with a $N(-1/2, 1)$ candidate to get an independent and identically distributed 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 helpful conversations and to an anonymous referee for several insightful comments, one of which led to Remark 6. This research was partially supported by grants from the U.S. National Science Foundation and the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

- ATHREYA, K. B. & NEY, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Am. Math. Soc.* **245**, 493–501.
- BRATLEY, P., FOX, B. L. & SCHRAGE, L. E. (1987). *A Guide to Simulation*. New York: Springer-Verlag, 2nd ed.
- CHAN, K. S. (1993). On the central limit theorem for an ergodic Markov chain. *Stoch. Proces. Applic.* **47**, 113–7.
- CHAN, K. S. & GEYER, C. J. (1994). Comment on ‘Markov chains for exploring posterior distributions’ by L. Tierney. *Ann. Statist.* **22**, 1747–58.

- DAMIEN, P., WAKEFIELD, J. & WALKER, S. (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *J. R. Statist. Soc. B* **61**, 331–44.
- FORT, G. & MOULINES, E. (2000). V -Subgeometric ergodicity for a Hastings-Metropolis algorithm. *Statist. Prob. Lett.* **49**, 401–10.
- GEYER, C. J. (1992). Practical Markov chain Monte Carlo (with Discussion). *Statist. Sci.* **7**, 473–511.
- GEYER, C. J. & THOMPSON, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *J. Am. Statist. Assoc.* **90**, 909–20.
- GILKS, W. R., ROBERTS, G. O. & SAHU, S. K. (1998). Adaptive Markov chain Monte Carlo through regeneration. *J. Am. Statist. Assoc.* **93**, 1045–54.
- HOBERT, J. P. & GEYER, C. J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *J. Mult. Anal.* **67**, 414–30.
- JARNER, S. F. & HANSEN, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stoch. Proces. Applic.* **85**, 341–61.
- JARNER, S. F. & ROBERTS, G. O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Prob.* **12**, 224–47.
- JONES, G. L. & HOBERT, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16**, 312–34.
- KIPNIS, C. & VARADHAN, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Commun. Math. Phys.* **104**, 1–19.
- MENGERSEN, K. & TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24**, 101–21.

- MEYN, S. P. & TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. London: Springer-Verlag.
- MEYN, S. P. & TWEEDIE, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Prob.* **4**, 981–1011.
- MIRA, A. & TIERNEY, L. (2001). Efficiency and convergence properties of slice samplers. *Scand. J. Statist.* **29**, 1035–53.
- MYKLAND, P., TIERNEY, L. & YU, B. (1995). Regeneration in Markov chain samplers. *J. Am. Statist. Assoc.* **90**, 233–41.
- NEAL, R. M. (2002). Slice sampling. *Ann. Statist.* (to appear) .
- NUMMELIN, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahr. Verw. Geb.* **43**, 309–18.
- NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. London: Cambridge University Press.
- NUMMELIN, E. & TUOMINEN, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proces. Applic.* **12**, 187–202.
- NUMMELIN, E. & TUOMINEN, P. (1983). The rate of convergence in Orey’s theorem for Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proces. Applic.* **15**, 295–311.
- PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic.
- RIPLEY, B. D. (1987). *Stochastic Simulation*. New York: John Wiley and Sons.
- ROBERT, C. P. (1995). Convergence control methods for Markov chain Monte Carlo algorithms. *Statist. Sci.* **10**, 231–53.

- ROBERT, C. P. & CASELLA, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.
- ROBERTS, G. O. & ROSENTHAL, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2**, 13–25.
- ROBERTS, G. O. & ROSENTHAL, J. S. (1998a). Markov chain Monte Carlo: Some practical implications of theoretical results (with Discussion). *Can. J. Statist.* **26**, 5–31.
- ROBERTS, G. O. & ROSENTHAL, J. S. (1998b). On convergence rates of Gibbs samplers for uniform distributions. *Ann. Appl. Probab.* **8**, 1291–302.
- ROBERTS, G. O. & ROSENTHAL, J. S. (1999). Convergence of slice sampler Markov chains. *J. R. Statist. Soc. B* **61**, 643–60.
- ROBERTS, G. O. & ROSENTHAL, J. S. (2001). Markov chains and de-initializing processes. *Scand. J. Statist.* **28**, 489–504.
- ROBERTS, G. O. & TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83**, 95–110.
- ROSENTHAL, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Am. Statist. Assoc.* **90**, 558–66.
- TUOMINEN, P. & TWEEDIE, R. L. (1994). Subgeometric rates of convergence of f -ergodic Markov chains. *Adv. Appl. Probab.* **26**, 775–798.