Methodol Comput Appl Probab (2013) 15:237–247
DOI 10.1007/s11009-011-9237-8

CLTs and Asymptotic Variance of Time-Sampled Markov Chains

Krzysztof Łatuszyński · Gareth O. Roberts

Received: 8 March 2011 / Revised: 1 June 2011 /
Accepted: 6 June 2011 / Published online: 29 June 2011

Abstract For a Markov transition kernel P and a probability distribution μ on non-negative integers, a time-sampled Markov chain evolves according to the transition kernel $P_\mu = \sum_k \mu(k) P^k$. In this note we obtain CLT conditions for time-sampled Markov chains and derive a spectral formula for the asymptotic variance. Using these results we compare efficiency of Barker's and Metropolis algorithms in terms of asymptotic variance.

Keywords Time-sampled Markov chains · Barker's algorithm · Metropolis algorithm · Central Limit Theorem · Asymptotic variance · Variance bounding Markov chains · MCMC estimation

AMS 2000 Subject Classifications 60J05 · 60F05

1 Introduction

Let P be an ergodic transition kernel of a Markov chain $(X_n)_{n\geq 0}$ with limiting distribution π on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and let $f : \mathcal{X} \to \mathbb{R}$ be in $L^2(\pi)$. A typical MCMC procedure for estimating $I = \pi f := \int_{\mathcal{X}} f(x)\pi(dx)$ would use $\hat{I}_n := \frac{1}{n}\sum_{i=0}^{n-1} f(X_i)$. Under appropriate assumptions on P and f a CLT holds for $\hat{I}_n$, i.e.

$$\sqrt{n}(\hat{I}_n - I) \to \mathcal{N}(0, \sigma^2_{f,P}), \tag{1}$$

where the constant $\sigma^2_{f,P} < \infty$ is called asymptotic variance and depends only on f and P.

Supported by EPSRC grants EP/G026521/1 and EP/D002060/1 and by CRiSM.

K. Łatuszyński (✉) · G. O. Roberts
Department of Statistics, University of Warwick, CV4 7AL, Coventry, UK
e-mail: latuch@gmail.com

The following theorem from Kipnis and Varadhan (1986) is a fundamental result on conditions that guarantee Eq. 1 for reversible Markov chains.

Theorem 1 (Kipnis and Varadhan 1986) *For a reversible and ergodic Markov chain, and a function $f \in L^2(\pi)$, if*

$$Var(f, P) := \lim_{n\to\infty} n\mathrm{Var}_\pi(\hat{I}_n) \quad < \quad \infty, \tag{2}$$

then Eq. 1 holds with

$$\sigma^2_{f,P} = Var(f, P) \quad = \quad \int_{[-1,1]} \frac{1+x}{1-x} E_{f,P}(dx), \tag{3}$$

where $E_{f,P}$ is the spectral measure associated with f and P.

We refer to Eq. 2 as the Kipnis–Varadhan condition. Assuming that Eq. 2 holds and P is reversible, in Section 2 we obtain conditions for the CLT and derive a spectral formula for the asymptotic variance σ^2_{f,P_μ} of a time-sampled Markov chain of the form

$$P_\mu := \sum_{k=0}^{\infty} \mu(k) P^k, \tag{4}$$

where μ is a probability distribution on the nonnegative integers. Time-sampled Markov chains are of theoretical interest in the context of petite sets (cf. Chapter 5 of Meyn and Tweedie 1993), and also in the context of computational algorithms (Rosenthal 2003a, b).

Next we proceed to analyze efficiency of Barker's algorithm (Barker 1965). Barker's algorithm, similarly as Metropolis, uses an irreducible transition kernel Q to draw proposals. A move form $X_n = x$ to a proposal $Y_{n+1} = y$ is then accepted with probability

$$\alpha^{(\mathrm{B})}(x, y) = \frac{\pi(y)q(y, x)}{\pi(y)q(y, x) + \pi(x)q(x, y)}, \tag{5}$$

where $q(x, \cdot)$ is the transition density of $Q(x, \cdot)$. It is well known that with the same proposal kernel Q, the Metropolis acceptance ratio results in a smaller asymptotic variance then Barker's. In Section 3 we show that the asymptotic variance of Barker's algorithm is not bigger then, roughly speaking, two times that of Metropolis. We also motivate our considerations by recent advances in exact MCMC for diffusion models. The theoretical results are illustrated by a simulation study in Section 4.

2 Time-sampled Markov Chains

In this section we work under assumptions of Theorem 1 which imply that the asymptotic variance $\sigma^2_{f,P}$ equals $Var(f, P)$ defined in Eq. 2 and satisfies Eq. 3. For other Markov chain CLT conditions we refer to Jones (2004), Roberts and Rosenthal (2004, 2008), Meyn and Tweedie (1993) and Bednorz et al. (2008).

Theorem 2 *Let P be a reversible and ergodic transition kernel with stationary measure π, and let $f \in L^2(\pi)$. Assume that the Kipnis–Varadhan condition (2) holds for f and P. For a probability distribution μ on nonnegative integers, let the time-sampled kernel P_μ be defined by Eq. 4. Then, if any of the following conditions hold*

(i) $\mu_{odd} := \mu(\{1, 3, 5, \dots\}) > 0$,
(ii) $\mu(0) < 1$ *and P is geometrically ergodic,*

the CLT holds for f and P_μ, moreover

$$\sigma^2_{f,P_\mu} = \int_{[-1,1]} \frac{1 + G_\mu(x)}{1 - G_\mu(x)} E_{f,P}(dx) \quad < \quad \infty, \tag{6}$$

where G_μ is the probability generating function of μ, i.e. $G_\mu(z) := \mathbb{E}_\mu z^K$, $|z| \le 1$, $K \sim \mu$, and $E_{f,P}$ is the spectral measure associated with f and P.

Remark 1 The condition $\mu_{\text{odd}} > 0$ in the above result is necessary, which we show below by means of a counterexample.

Proof The proof is based on the functional analytic approach (see e.g. Kipnis and Varadhan 1986; Roberts and Rosenthal 1997). Without loss of generality assume that $\pi f = 0$. A reversible transition kernel P with invariant distribution π is a self-adjoint operator on $L^2_0(\pi) := \{f \in L^2(\pi) : \pi f = 0\}$ with spectral radius bounded by 1. By the spectral decomposition theorem for self adjoint operators, for each $f \in L^2_0(\pi)$ there exists a finite positive measure $E_{f,P}$ on $[-1, 1]$, such that

$$\langle f, P^n f \rangle = \int_{[-1,1]} x^n E_{f,P}(dx),$$

for all integers $n \ge 0$. Thus in particular

$$\sigma^2_f = \pi f^2 = \int_{[-1,1]} 1 E_{f,P}(dx) \quad < \quad \infty, \tag{7}$$

and by Kipnis and Varadhan (1986) (c.f. also Theorem 4 of Häggström and Rosenthal 2007) one obtains

$$\sigma^2_{f,P} = \int_{[-1,1]} \frac{1 + x}{1 - x} E_{f,P}(dx) \quad < \quad \infty. \tag{8}$$

Since $P^n_\mu = \left(\sum_k \mu(k) P^k\right)^n$, by the spectral mapping theorem (Conway 1990), we have

$$\begin{aligned} \left\langle f, P^n_\mu f \right\rangle = \int_{[-1,1]} x^n E_{f,P_\mu}(dx) \quad &= \quad \int_{[-1,1]} \Big(\sum_k \mu(k) x^k\Big)^n E_{f,P}(dx) \\ &= \int_{[-1,1]} \Big(G_\mu(x)\Big)^n E_{f,P}(dx), \end{aligned}$$

and consequently, applying the same argument as Kipnis and Varadhan (1986) and Häggström and Rosenthal (2007), we obtain

$$\sigma^2_{f,P_\mu} = \int_{[-1,1]} \frac{1+x}{1-x} E_{f,P_\mu}(dx)$$
$$= \int_{[-1,1]} \frac{1+G_\mu(x)}{1-G_\mu(x)} E_{f,P}(dx) \quad =: \quad \clubsuit. \tag{9}$$

Now Eq. 9 gives the claimed formula but we need to prove Eq. 9 is finite: by Kipnis and Varadhan (1986) finiteness of the integral in Eq. 9 implies a CLT for f and P_μ. Observe that

$$|G(x)| \leq 1 \qquad \text{for all} \quad x \in [-1,1],$$
$$G(x) \leq \mu(0) + x(1-\mu(0)) \qquad \text{for} \quad x \geq 0.$$

Moreover, if (i) holds, then

$$G(x) \leq \sum_{k \text{ even}} \mu(k)x^k \quad \leq \quad 1 - \mu_{\text{odd}} \qquad \text{for} \quad x \leq 0,$$

hence we can write

$$\clubsuit = \int_{[-1,0)} \frac{1+G_\mu(x)}{1-G_\mu(x)} E_{f,P}(dx) + \int_{[0,1]} \frac{1+G_\mu(x)}{1-G_\mu(x)} E_{f,P}(dx)$$
$$\leq \frac{1}{\mu_{\text{odd}}} \int_{[-1,0)} 2E_{f,P}(dx) + \frac{1}{1-\mu(0)} \int_{[0,1]} \frac{2}{1-x} E_{f,P}(dx). \tag{10}$$

The first integral in Eq. 10 is finite by Eq. 7 and the second by Eq. 8 and we are done with (i).

Next assume that (ii) holds. By $S(P)$ denote the spectrum of P and let $s_P := \sup\{|\lambda| : \lambda \in S(P)\}$ be the spectral radius. From Roberts and Rosenthal (1997) we know that since P is reversible and geometrically ergodic, it has a spectral gap, i.e. $s_P < 1$. Hence for $x \in [-s_P, 0]$, we can write

$$G_\mu \leq \mu(0) + \sum_{k \text{ even}} \mu(k)x^k \quad \leq \quad \mu(0) + s_P(1-\mu(0)).$$

Consequently

$$\clubsuit = \int_{[-s_P,0)} \frac{1+G_\mu(x)}{1-G_\mu(x)} E_{f,P}(dx) + \int_{[0,s_P]} \frac{1+G_\mu(x)}{1-G_\mu(x)} E_{f,P}(dx)$$
$$\leq \frac{1}{1-\mu(0)} \int_{[-s_P,0)} \frac{2}{1-s_P} E_{f,P}(dx) + \frac{1}{1-\mu(0)} \int_{[0,s_P]} \frac{2}{1-x} E_{f,P}(dx). \tag{11}$$

The first integral in Eq. 11 is finite by Eq. 7 and the second by Eq. 8. □

The most important special case of Theorem 2 is underlined and computed explicitly in the next corollary.

Corollary 1 *Let P be a reversible and ergodic transition kernel with stationary measure π, and assume that for f and P the CLT (Eq. 1) holds. For $\varepsilon \in (0,1)$ let the* *lazy version of P be defined as* $P_\varepsilon := \varepsilon Id + (1-\varepsilon)P$. *Then the CLT holds for* f *and* P_ε *and*

$$\sigma^2_{f,P_\varepsilon} = \frac{1}{1-\varepsilon}\sigma^2_{f,P} + \frac{\varepsilon}{1-\varepsilon}\sigma^2_f. \tag{12}$$

Proof We use Theorem 2 with $\mu(0)=\varepsilon,\ \mu(1)=1-\varepsilon$. Hence $G_\mu = \varepsilon + (1-\varepsilon)x$, and consequently

$$\begin{aligned}
\sigma^2_{f,P_\varepsilon} &= \int_{[-1,1]} \frac{1+\varepsilon+(1-\varepsilon)x}{1-\varepsilon-(1-\varepsilon)x} E_{f,P}(dx) \\
&= \int_{[-1,1]} \frac{1}{1-\varepsilon}\left(\frac{1+x}{1-x}+\varepsilon\right) E_{f,P}(dx) \\
&= \frac{1}{1-\varepsilon}\int_{[-1,1]} \frac{1+x}{1-x} E_{f,P}(dx) + \frac{\varepsilon}{1-\varepsilon}\int_{[-1,1]} 1 E_{f,P}(dx) \\
&= \frac{1}{1-\varepsilon}\sigma^2_{f,P} + \frac{\varepsilon}{1-\varepsilon}\sigma^2_f.
\end{aligned}$$

□

Efficiency of time sampled Markov chains can be compared using the following corollary from Theorem 2.

Corollary 2 *Let* P *and* f *be as in Theorem 2. If* P *is positive as an operator on* $L^2(\pi)$ *and* μ_1 *dominates stochastically* μ_2 *(i.e.* $\mu_1 \geq_{st} \mu_2$*), then* P_{μ_1} *dominates* P_{μ_2} *in the efficiency ordering, i.e.* $\sigma^2_{f,P_{\mu_1}} \leq \sigma^2_{f,P_{\mu_2}}$.

Proof If P is positive self-adjoint then $\text{supp} E_{f,P} \subseteq [0,1]$. Moreover

$$\mu_1 \geq_{st} \mu_2 \Rightarrow G_{\mu_1}(x) \leq G_{\mu_2}(x) \qquad \text{for} \quad x \in [-1,1].$$

The conclusion follows from Eq. 6. □

In another direction of studying CLTs, the *variance bounding* property of Markov chains has been introduced in Roberts and Rosenthal (2008) and is defined as follows. P is variance bounding if there exists $K < \infty$ such that $Var(f,P) \leq K\text{Var}_\pi(f)$ for all f. Here $Var(f,P)$ is defined in Eq. 2 and $\text{Var}_\pi(f) = \pi f^2 - (\pi f)^2$. We prove that for time-sampled Markov chains the variance bounding property propagates the same way the CLT does.

Theorem 3 *Assume* P *is reversible and variance bounding. Then* P_μ *is variance bounding if any of the following conditions hold*

(i) $\mu_{\text{odd}} := \mu(\{1,3,5,\dots\}) > 0$,
(ii) $\mu(0) < 1$ *and* P *is geometrically ergodic.*

Proof For any f such that $\text{Var}_\pi f < \infty$, the Kipnis–Varadhan condition holds due to variance bounding property of P and thus the assumptions of Theorem 2 are met. Hence for every $f \in L^2(\pi)$ there is a CLT for f and P_μ. Therefore P_μ is variance bounding by Theorem 7 of Roberts and Rosenthal (2008). □

The next example shows that in case of Markov chains that are not geometrically ergodic, the condition $\mu_{\text{odd}} > 0$ is necessary.

Example 1 We set $f(x) = x$ and give an example of an ergodic and reversible transition kernel P on $\mathcal{X} = [-1, 1]$, and such that there is a CLT for P and f but *not* for P^2 and f. We shall rely on Theorem 4.1 of Bednorz et al. (2008) that provides if and only if conditions for Markov chains CLTs in terms of regenerations. It will be apparent that the condition $\mu_{\text{odd}} > 0$ in Theorem 2 is necessary.

Set $s(x) := \sqrt{1 - |x|}$, let $U(\cdot)$ be the uniform distribution on $[-1, 1]$, and let the kernel P be of the form

$$P(x, \cdot) = (1 - s(x))\delta_{-x}(\cdot) + s(x)U(\cdot), \qquad \text{hence} \tag{13}$$

$$P^2(x, \cdot) = (1 - s(x))^2\delta_x(\cdot) + (2s(x) - s(x)^2)U(\cdot). \tag{14}$$

To find the stationary distribution of P (and also P^2), we verify reversibility with $\pi(x) \propto 1/s(x)$.

$$\begin{aligned}\pi(\mathrm{d}x)P(x, \mathrm{d}y) &\propto \frac{1}{s(x)}\delta_{-x}(y) + \delta_{-x}(y) + U(\mathrm{d}y) \\ &= \frac{1}{s(y)}\delta_{-y}(x) + \delta_{-y}(x) + U(\mathrm{d}x) \quad \propto \quad \pi(\mathrm{d}y)P(y, \mathrm{d}x).\end{aligned}$$

Hence $\pi(x)$ is a reflected Beta$(1, \frac{1}{2})$. Clearly $\pi(f^2) < \infty$.

Recall now the split chain construction (Nummelin 1978; Athreya and Ney 1978) of the bivariate Markov chain $\{X_n, \Gamma_n\}$ on $\{0, 1\} \times \mathcal{X} = \{0, 1\} \times [0, 1]$. If $(X_n)_{n \geq 0}$ evolves according to P defined in Eq. 14, we have the following transition rule from $\{X_{n-1}, \Gamma_{n-1}\}$ to $\{X_n, \Gamma_n\}$ for the split chain.

$$\begin{aligned}\check{\mathbb{P}}(X_n \in \cdot | \Gamma_{n-1} = 1, X_{n-1} = x) &= U(\cdot), \\ \check{\mathbb{P}}(X_n \in \cdot | \Gamma_{n-1} = 0, X_{n-1} = x) &= \delta_{-x}(\cdot), \\ \check{\mathbb{P}}(\Gamma_n = 1 | \Gamma_{n-1}, X_n = x) &= s(x), \\ \check{\mathbb{P}}(\Gamma_n = 0 | \Gamma_{n-1}, X_n = x) &= 1 - s(x).\end{aligned}$$

The notation $\check{\mathbb{P}}$ above indicates that we consider the extended probability space for (X_n, Γ_n), not the original one of X_n. The appropriate modification of the above holds if the dynamics of X_n is P^2, namely

$$\begin{aligned}\check{\mathbb{P}}(X_n \in \cdot | \Gamma_{n-1} = 1, X_{n-1} = x) &= U(\cdot), \\ \check{\mathbb{P}}(X_n \in \cdot | \Gamma_{n-1} = 0, X_{n-1} = x) &= \delta_x(\cdot), \\ \check{\mathbb{P}}(\Gamma_n = 1 | \Gamma_{n-1}, X_n = x) &= 2s(x) - s^2(x), \\ \check{\mathbb{P}}(\Gamma_n = 0 | \Gamma_{n-1}, X_n = x) &= (1 - s(x))^2.\end{aligned}$$

We refer to to the original papers for more details on the split chain construction and to Bednorz et al. (2008) and Roberts and Rosenthal (2004) for central limit theorems in this context. Denote

$$\tau := \min\{k \geq 0 : \Gamma_k = 1\}. \tag{15}$$

By Theorem 4.1 of Bednorz et al. (2008), the CLT for P and f holds if and only if the following expression for the asymptotic variance is finite.

$$\sigma^2_{f,P} = \int_{[-1,1]} s(x)\pi(x)\mathrm{d}x \, \check{\mathbb{E}}_U \left(\sum_{k=0}^{\tau} f(X_n) \right)^2, \tag{16}$$

where (X_n, Γ_n) follow the dynamics of P. Respectively, the CLT for P^2 and f holds in our setting, *if and only if*

$$\sigma^2_{f,P^2} = \int_{[-1,1]} (2s(x) - s^2(x))\pi(x)\mathrm{d}x \, \check{\mathbb{E}}_U \left(\sum_{k=0}^{\tau} f(X_n) \right)^2 \tag{17}$$

is finite, where (X_n, Γ_n) follow the dynamics of P^2.

Now observe that if $(X_n)_{n\geq 0}$ evolves according to P, then $(\sum_{k=0}^{\tau} f(X_n))^2$ equals 0 if τ is odd, or $(\sum_{k=0}^{\tau} f(X_n))^2 = X_0^2$, if τ is even. Consequently Eq. 16 is finite. However, if $(X_n)_{n\geq 0}$ evolves according to P^2, then $(\sum_{k=0}^{\tau} f(X_n))^2 = (\tau+1)^2 X_0^2$ and the distribution of τ is geometric with parameter $2s(X_0) - s^2(X_0) = 1 - (1 - s(x))^2$. Therefore we compute σ^2_{f,P^2} in Eq. 17 as

$$\begin{aligned}
\sigma^2_{f,P^2} &= \int_{[-1,1]} \left(2s(x) - s^2(x)\right) \pi(x)\,\mathrm{d}x \int_{[-1,1]} \frac{2 - \left(1 - \left(1 - s(x)\right)^2\right)}{2\left(1 - (1 - s(x))^2\right)^2} x^2\,\mathrm{d}x \\
&= C \int_{[-1,1]} \frac{\left(1 + (1 - s(x))^2\right) x^2}{2\left(1 - |x| - 2\sqrt{1 - |x|}\right)^2}\,\mathrm{d}x \\
&\geq C \int_{[-1,1]} \frac{x^2}{8(1 - |x|)}\,\mathrm{d}x = \infty.
\end{aligned}$$

3 Barker's Algorithm

When assessing efficiency of Markov chain Monte Carlo algorithms, the asymptotic variance criterion is one of natural choices. Peskun ordering (Peskun 1973) (see also Tierney 1998; Mira and Geyer 1999) provides a tool to compare two reversible transition kernels P_1, P_2 with the same limiting distribution π and is defined as follows. $P_1 \succ P_2 \iff$ for π−almost every $x \in \mathcal{X}$ and all $A \in \mathcal{B}(\mathcal{X})$ holds $P_1(x, A - \{x\}) \geq P_2(x, A - \{x\})$. If $P_1 \succ P_2$ then $\sigma^2_{f,P_1} \leq \sigma^2_{f,P_2}$ for every $f \in L^2(\pi)$.

Consider now a class of algorithms where the transition kernel P is defined by applying an irreducible proposal kernel Q and an acceptance rule α, i.e. given $X_n = x$, the value of X_{n+1} is a result of performing the following two steps.

1. Draw a proposal $y \sim Q(x, \cdot)$,
2. Set $X_{n+1} := y$ with probability $\alpha(x, y)$ and $X_{n+1} = x$ otherwise,

where $\alpha(x, y)$ is such that the resulting kernel P is reversible with stationary distribution π. It follows Peskun (1973) and Tierney (1998) that for a given proposal kernel Q the standard Metropolis–Hastings (Metropolis et al. 1953; Hastings 1970) acceptance rule

$$\alpha^{(\mathrm{MH})}(x, y) = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\} \tag{18}$$

yields a transition kernel $P^{(\mathrm{MH})}$ that is maximal with respect to Peskun ordering and thus minimal with respect to asymptotic variance. In particular, the Barker's algorithm (Barker 1965) that uses acceptance rule

$$\alpha^{(\mathrm{B})}(x, y) = \frac{\pi(y)q(y,x)}{\pi(y)q(y,x) + \pi(x)q(x,y)} \tag{19}$$

is inferior to Metropolis–Hastings when the asymptotic variance is considered. In the above notation we assume that all the involved distributions have common denominating measure and $q(x, \cdot)$ are transition densities of Q. See Tierney (1998) for a more general statement and discussion.

Exact Algorithms introduced in Beskos et al. (2006a, b, 2008) and Beskos and Roberts (2005) allow for inference in diffusion models without Euler discretization error. In recent advances in Exact MCMC inference for complex diffusion models a particular setting is reoccurring, where the Metropolis–Hastings acceptance step requires a specific Bernoulli Factory and is not possible to execute. However, in this diffusion context the Barker's algorithm (Eq. 19) is feasible, as well as the 'lazy' version of the Metropolis–Hastings kernel

$$P_\varepsilon^{(\mathrm{MH})} := \varepsilon Id + (1-\varepsilon)P^{(\mathrm{MH})}. \tag{20}$$

We refer to Gonçalves et al. (2011) and Łatuszyński et al. (2011a, b) for the background on exact MCMC inference for diffusions and the Bernoulli Factory problem. This motivates us to investigate performance of these alternatives in comparison to the standard Metropolis–Hastings.

Theorem 4 *Let $P^{(B)}$ denote the transition kernel of the Barker's algorithm and let $P^{(MH)}$ and $P_\varepsilon^{(MH)}$ be as defined in Eq. 20. If the CLT (Eq. 1) holds for f and $P^{(MH)}$, then it holds also for*

(i) *f and $P_\varepsilon^{(MH)}$ with*

$$\sigma^2_{f,P_\varepsilon^{(MH)}} = \frac{1}{1-\varepsilon}\sigma^2_{f,P^{(MH)}} + \frac{\varepsilon}{1-\varepsilon}\sigma^2_f. \tag{21}$$

(ii) *f and $P^{(B)}$ with*

$$\sigma^2_{f,P^{(MH)}} \le \sigma^2_{f,P^{(B)}} \le \sigma^2_{f,P_{1/2}^{(MH)}} = 2\sigma^2_{f,P^{(MH)}} + \sigma^2_f. \tag{22}$$

Proof The first claim (i) is a restatement of Corollary 1 for Metropolis–Hastings chains. To obtain the second claim (ii), note that $P_{1/2}^{(\mathrm{MH})}$ can be viewed as an algorithm that uses proposals from Q and acceptance rule

$$\alpha(x, y) = \min\left\{\frac{1}{2}, \frac{\pi(y)q(y,x)}{2\pi(x)q(x,y)}\right\}.$$

Now since

$$\min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\} \geq \frac{\pi(y)q(y,x)}{\pi(y)q(y,x)+\pi(x)q(x,y)} \geq \min\left\{\frac{1}{2}, \frac{\pi(y)q(y,x)}{2\pi(x)q(x,y)}\right\},$$

the result follows from Peskun ordering and Corollary 1. □

4 Numerical Examples

To illustrate the theoretical findings, we consider two numerical examples. The first focuses on time sampling, the second on efficiency of the Barker's algorithm.

4.1 Time Sampled Contracting Normals

Consider the contracting normals example, i.e. a Markov chain with transition probabilities

$$P(x,\cdot) = N\left(\theta x, 1-\theta^2\right) \tag{23}$$

for some $\theta \in (-1, 1)$. It is easy to check that the stationary distribution is $\pi(\cdot) = N(0, 1)$. Moreover the transition kernel is geometrically ergodic and reversible for all $\theta \in (-1, 1)$ and also positive for $\theta \in [0, 1)$ (Baxendale 2005; Łatuszyński and Niemiro 2011). For the target function we take $f(x) = x$ and estimate the asymptotic variance using the batch means estimator of Jones et al. (2006) based on a trajectories of length 10^7. We set θ to 0.9 and −0.9 in the following settings:

- CN: Contracting normals;
- LCN: Lazy contracting normals with $\varepsilon = 0.5$;
- TSCN1: Time sampled contracting normals for sampling distribution

$$\mu = 1 + Pois(1);$$

- TSCN2: Time sampled contracting normals for sampling distribution

$$\mu = 1 + Pois(5).$$

The first two columns of Table 1 report how laziness increases asymptotic variance and illustrate Corollary 1. Note that the stationary variance $\sigma_f^2 = 1$ is substantial compared to the asymptotic variance of contracting normals for $\theta = -0.9$ and thus the lazy version LCN becomes severely inefficient compared to CN. The stochastic ordering of the sampling distributions in the above scenarios is LCN $<_{st}$ CN $<_{st}$ TSCN1 $<_{st}$ TSCN2 therefore the simulation shows how the asymptotic variance decreases for stochastically bigger sampling distributions (Corollary 2) in case of positive operators ($\theta = 0.9$) and how this property fails if the operator is not positive, i.e for $\theta = -0.9$.

Table 1 Estimated asymptotic variance of the contracting normals Markov chain for different sampling scenarios

	CN	LCN	TSCN1	TSCN2
$\theta = 0.9$	19.1	38.5	9.28	3.43
$\theta = -0.9$	0.053	1.14	0.80	0.96

Table 2 Estimated asymptotic variance of the Metropolis, Barker's and lazy Metropolis algorithms

	Metropolis	Barker's	Lazy Metropolis
Asymptotic variance	3.69	5.67	8.32

4.2 Efficiency of the Barker's Algorithm

We compare the estimated asymptotic variance of the random walk Metropolis algorithm, the Barker's algorithm and lazy version of the random walk Metropolis with $\varepsilon = 0.5$ to illustrate the bounds of Theorem 4. For the stationary distribution we take $N(0, 1)$ and the increment proposal is $U([-2, 2])$. The results based on a simulation length 10^7 are reported in Table 2.

Acknowledgement We thank Jeffrey S. Rosenthal for a helpful discussion.

References

Athreya K, Ney P (1978) A new approach to the limit theory of recurrent Markov chains. Trans Am Math Soc 245:493–501

Barker A (1965) Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. Aust J Phys 18:119

Baxendale P (2005) Renewal theory and computable convergence rates for geometrically ergodic Markov chains. Ann Appl Probab 15(1B):700–738

Bednorz W, Łatuszyński K, Latała R (2008) A regeneration proof of the central limit theorem for uniformly ergodic markov chains. Electron Commun Probab 13:85–98

Beskos A, Roberts G (2005) Exact simulation of diffusions. Ann Appl Probab 15(4):2422–2444

Beskos A, Papaspiliopoulos O, Roberts G, Fearnhead P (2006a) Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). J R Stat Soc, Ser B Stat Methodol 68(3):333–382

Beskos A, Papaspiliopoulos O, Roberts G (2006b) Retrospective exact simulation of diffusion sample paths with applications. Bernoulli 12(6):1077

Beskos A, Papaspiliopoulos O, Roberts G (2008) A factorisation of diffusion measure and finite sample path constructions. Methodol Comput Appl Probab 10(1):85–104

Conway J (1990) A course in functional analysis. Springer, New York

Gonçalves F, Roberts G, Łatuszyński K (2011) Exact MCMC inference for jump diffusion models with stochastic jump rate (in preparation)

Häggström O, Rosenthal J (2007) On variance conditions for Markov chain CLTs. Electron Commun Probab 12:454–464

Hastings W (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57(1):97

Jones G (2004) On the Markov chain central limit theorem. Probab Surv 1:299–320

Jones G, Haran M, Caffo B, Neath R (2006) Fixed-width output analysis for Markov chain Monte Carlo. J Am Stat Assoc 101(476):1537–1547

Kipnis C, Varadhan S (1986) Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. Commun Math Phys 104(1):1–19

Łatuszyński K, Niemiro W (2011) Rigorous confidence bounds for MCMC under a geometric drift condition. J Complex 27(1):23–38

Łatuszyński K, Palczewski J, Roberts G (2011a) Exact inference for a markov switching diffusion model with discretely observed data (in preparation)

Łatuszyński K, Kosmidis I, Papaspiliopoulos O, Roberts G (2011b) Simulating events of unknown probabilities via reverse time martingales. Random Struct Algorithms 38(4):441–452

Metropolis N, Rosenbluth A, Rosenbluth M, Teller A, Teller E (1953) Equations of state calculations by fast computational machine. J Chem Phys 21(6):1087–1091

Meyn S, Tweedie R (1993) Markov chains and stochastic stability. Springer, London

Mira A, Geyer C (1999) Ordering Monte Carlo Markov chains. Technical report, School of Statistics, University of Minnesota
Nummelin E (1978) A splitting technique for Harris recurrent Markov chains. Probab Theory Relat Fields 43(4):309–318
Peskun P (1973) Optimum monte-carlo sampling using markov chains. Biometrika 60(3):607
Roberts G, Rosenthal J (1997) Geometric ergodicity and hybrid Markov chains. Electron Commun Probab 2(2):13–25
Roberts G, Rosenthal J (2004) General state space Markov chains and MCMC algorithms. Probab Surv 1:20–71
Roberts G, Rosenthal J (2008) Variance bounding Markov chains. Ann Appl Probab 18(3):1201
Rosenthal J (2003a) Asymptotic variance and convergence rates of nearly-periodic Markov chain Monte Carlo algorithms. J Am Stat Assoc 98(461):169–177
Rosenthal J (2003b) Geometric convergence rates for time-sampled markov chains. J Theor Probab 16(3):671–688
Tierney L (1998) A note on Metropolis–Hastings kernels for general state spaces. Ann Appl Probab 8(1):1–9