Author(s): KG Latuszynski and GO Roberts

Article Title: CLTs and asymptotic variance of time-sampled Markov chains

Year of publication: 2011
Link to published article:
http://www2.warwick.ac.uk/fac/sci/statistics/crism/research/2011/paper 11-04

Publisher statement: None

CLTs and asymptotic variance of time-sampled Markov chains*

Krzysztof Łatuszyński and Gareth O. Roberts

K. Łatuszyński
Department of Statistics
University of Warwick
CV4 7AL, Coventry, UK
e-mail: latuch@gmail.com

G. O. Roberts
Department of Statistics
University of Warwick
CV4 7AL, Coventry, UK
e-mail: Gareth.O.Roberts@warwick.ac.uk

February 10, 2011

Abstract: For a Markov transition kernel P and a probability distribution μ on nonnegative 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.

AMS 2000 subject classifications: Primary 60J05, 65C05.
Keywords and phrases: time-sampled Markov chains, Barker's algorithm, Metropolis algorithm, Central Limit Theorem, asymptotic variance, variance bounding Markov chains, MCMC estimation.

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.

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

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

The following theorem from [KV86] is a fundamental result on conditions that guarantee (1) for reversible Markov chains.

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

Theorem 1.1 ([KV86]). *For a reversible and ergodic Markov chain, and a function $f \in L^2(\pi)$, if*

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

then (1) *holds with*

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

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

We refer to (2) as the Kipnis-Varadhan condition. Assuming that (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

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

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 [MT93]), and also in the context of computational algorithms [Ros03a, Ros03b].

Next we proceed to analyze efficiency of Barker's algorithm [Bar65]. 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

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

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.

2. Time-sampled Markov chains

In this section we work under assumptions of Theorem 1.1 which imply that the asymptotic variance $\sigma^2_{f,P}$ equals $Var(f,P)$ defined in (2) and satisfies (3). For other Markov chain CLT conditions we refer to [Jon04, RR04, MT93, BŁL08, RR08].

Theorem 2.1. *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 (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

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

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

Remark 2.2. 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. [KV86, RR97]). 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 $n \geq 0$. Thus in particular

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

and by [KV86] (c.f. also Theorem 4 of [HR07]) one obtains

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

Since $P^n_\mu \;=\; \sum_k \mu(k) P^k$, by the spectral mapping theorem [Con90], we have

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

and consequently, applying the same argument as [KV86, HR07], we obtain

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

Now (9) gives the claimed formula but we need to prove (9) is finite: by [KV86] finiteness of the integral in (9) implies a CLT for f and P_μ. Observe that

$$\begin{aligned} |G(x)| &\le 1 && \text{for all} \quad x \in [-1,1], \\ G(x) &\le \mu(0) + x(1-\mu(0)) && \text{for} \quad x \ge 0. \end{aligned}$$

Moreover, if (i) holds, then

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

hence we can write

$$\begin{aligned} \clubsuit &= \int_{[-1,0)} \frac{1+G_\mu(x)}{1-G_\mu(x)} E_{f,P}(dx) + \int_{[0,1]} \frac{1+G_\mu(x)}{1-G_\mu(x)} E_{f,P}(dx) \\ &\le \frac{1}{\mu_{\text{odd}}} \int_{[-1,0)} 2E_{f,P}(dx) + \frac{1}{1-\mu(0)} \int_{[0,1]} \frac{2}{1-x} E_{f,P}(dx). \qquad (10) \end{aligned}$$

The first integral in (10) is finite by (7) and the second by (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 [RR97] 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

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

Consequently

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

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

The most important special case of Theorem 2.1 is underlined and computed explicitly in the next corollary.

Corollary 2.3. *Let P be a reversible and ergodic transition kernel with stationary measure π, and assume that for f and P the CLT (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*

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

Proof. We use Theorem 2.1 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 2.1.

Corollary 2.4. *Let P and f be as in Theorem 2.1. 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

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

The conclusion follows from (6). □

In another direction of studying CLTs, the *variance bounding* property of Markov chains has been introduced in [RR08] 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 (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 2.5. *Assume P is reversible and variance bounding. Then P_μ is variance bounding 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.*

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 2.1 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 [RR08]. □

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

Example 2.6. 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 [BŁL08] that provides if and only if conditions for Markov chains CLTs in terms of regenerations. It will be apparent that the condition $\mu_{\mathrm{odd}} > 0$ in Theorem 2.1 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 [Num78, AN78] 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 (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 [BŁL08, RR04] for central limit theorems in this context. Denote

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

By Theorem 4.1 of [BŁL08], the CLT for P and f holds if and only if the following expression for the asymptotic variance is finite.

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

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*

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

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 (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 (17) as

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

3. Barker's algorithm

When assessing efficiency of Markov chain Monte Carlo algorithms, the asymptotic variance criterion is one of natural choices. Peskun ordering [Pes73] (see also [Tie98, MG99]) 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 $\pi-$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 [Pes73, Tie98] that for a given proposal kernel Q the standard Metropolis-Hastings [MRR+53, Has70] acceptance rule

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

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 [Bar65] that uses acceptance rule

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

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 [Tie98] for a more general statement and discussion.

Exact Algorithms introduced in [BPRF06, BR05, BPR06, BPR08] 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 (19) is feasible, as well as the 'lazy' version of the Metropolis-Hastings kernel

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

We refer to [GRŁ11, ŁPR11, ŁKPR11] 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 3.1. *Let $P^{(B)}$ denote the transition kernel of the Barker's algorithm and let $P^{(MH)}$ and $P_{\varepsilon}^{(MH)}$ be as defined in (20). If the CLT (1) holds for f and $P^{(MH)}$, then it holds also for*

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

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

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

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

Proof. The first claim (i) is a restatement of Corollary 2.3 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

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

Now since

$$\min\big\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\big\} \quad \geq \quad \frac{\pi(y)q(y,x)}{\pi(y)q(y,x)+\pi(x)q(x,y)} \;\geq\; \min\big\{\frac{1}{2}, \frac{\pi(y)q(y,x)}{2\pi(x)q(x,y)}\big\},$$

the result follows from Peskun ordering and Corollary 2.3. □

4. Acknowledgements

We thank Jeffrey S. Rosenthal for a helpful discussion.

References

[AN78] K.B. Athreya and P. Ney. A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, 245(Nov):493–501, 1978.

[Bar65] A.A. Barker. Monte Carlo calculations of the radial distribution functions for a proton-electron plasma. *Australian Journal of Physics*, 18:119, 1965.

[BŁL08] W. Bednorz, K. Łatuszyński, and R. Latała. A regeneration proof of the central limit theorem for uniformly ergodic markov chains. *Electronic Communications in Probability*, 13:85–98, 2008.

[BPR06] A. Beskos, O. Papaspiliopoulos, and G.O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077, 2006.

[BPR08] A. Beskos, O. Papaspiliopoulos, and G.O. Roberts. A factorisation of diffusion measure and finite sample path constructions. *Methodology and Computing in Applied Probability*, 10(1):85–104, 2008.

[BPRF06] A. Beskos, O. Papaspiliopoulos, G.O. Roberts, and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382, 2006.

[BR05] A. Beskos and G.O. Roberts. Exact simulation of diffusions. *Annals of Applied Probability*, 15(4):2422–2444, 2005.

[Con90] J.B. Conway. *A course in functional analysis*. Springer, 1990.

[GRŁ11] F.B. Gonçalves, G.O. Roberts, and K. Łatuszyński. Exact mcmc inference for jump diffusion models with stochastic jump rate. 2011.

[Has70] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97, 1970.

[HR07] O. Häggström and J.S. Rosenthal. On variance conditions for Markov chain CLTs. *Elect. Comm. in Probab*, 12:454–464, 2007.

[Jon04] G.L. Jones. On the Markov chain central limit theorem. *Probability surveys*, 1:299–320, 2004.

[KV86] C. Kipnis and SRS Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1):1–19, 1986.

[ŁKPR11] K. Łatuszyński, I. Kosmidis, O. Papaspiliopoulos, and G.O. Roberts. Simulating events of unknown probabilities via reverse time martingales. *Random Structures & Algorithms*, 2011.

[ŁPR11] K. Łatuszyński, J. Palczewski, and G.O. Roberts. Exact inference for a markov switching diffusion model with discretely observed data. 2011.

[MG99] A. Mira and C.J. Geyer. Ordering Monte Carlo Markov chains. In *School of Statistics, University of Minnesota.* technical report, 1999.

[MRR+53] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computational machine. *Journal of Chemical Physics*, 21(6):1087–1091, 1953.

[MT93] S.P. Meyn and R.L Tweedie. *Markov chains and stochastic stability.* Springer London et al., 1993.

[Num78] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Probability Theory and Related Fields*, 43(4):309–318, 1978.

[Pes73] PH Peskun. Optimum monte-carlo sampling using markov chains. *Biometrika*, 60(3):607, 1973.

[Ros03a] J.S. Rosenthal. Asymptotic variance and convergence rates of nearly-periodic Markov chain Monte Carlo algorithms. *Journal of the American Statistical Association*, 98(461):169–177, 2003.

[Ros03b] J.S. Rosenthal. Geometric convergence rates for time-sampled markov chains. *Journal of Theoretical Probability*, 16(3):671–688, 2003.

[RR97] G.O. Roberts and J.S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab*, 2(2):13–25, 1997.

[RR04] G.O. Roberts and J.S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71, 2004.

[RR08] G.O. Roberts and J.S. Rosenthal. Variance bounding Markov chains. *Annals of applied probability*, 18(3):1201, 2008.

[Tie98] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Annals of Applied Probability*, 8(1):1–9, 1998.