Elect. Comm. in Probab. **12** (2007), 454–464

ELECTRONIC
COMMUNICATIONS
in PROBABILITY

ON VARIANCE CONDITIONS FOR MARKOV CHAIN CLTS

OLLE HÄGGSTRÖM[1]
Mathematical Sciences, Chalmers University of Technology, SE-412 96 Göteborg, Sweden.
email: olleh@math.chalmers.se

JEFFREY S. ROSENTHAL[2]
Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3.
email: jeff@math.toronto.edu

Submitted 25 September 2007, *accepted in final form* 27 November 2007

AMS 2000 Subject classification: 60J10, 60F05.
Keywords: Markov chains, central limit theorems, asymptotic variance, spectral measure, reversibility

Abstract
Central limit theorems for Markov chains are considered, and in particular the relationships between various expressions for asymptotic variance known from the literature. These turn out to be equal under fairly general conditions, although not always. We also investigate the existence of CLTs, and pose some open problems.

1 Introduction

The existence of central limit theorems (CLTs) for Markov chains is well studied, and is is particularly important for Markov chain Monte Carlo (MCMC) algorithms, see e.g. [13], [23], [8], [5], [6], [10], [12], and [9]. In particular, the asymptotic variance σ^2 is very important in applications, and various alternate expressions for it are available in terms of limits, autocovariances, and spectral theory.

This paper considers three such expressions, denoted A, B, and C, which are known to "usually" equal σ^2. These expressions arise in different applications in different ways. For example, it is proved by Kipnis and Varadhan [13] that if $C < \infty$, then a $\sqrt{n}$-CLT exists for h, with $\sigma^2 = C$. In a different direction, it is proved by Roberts [17] that Metropolis algorithms satisfying a certain condition must have $A = \infty$. Such disparate results indicate the importance of sorting out the relationships between A, B, C, σ^2, and the existence of Markov chain CLTs.

In Sections 3 and 6 below, we consider the relationships between the quantities A, B, and C. In Section 4, we consider conditions under which the existence of a $\sqrt{n}$-CLT does or does not

[1]SUPPORTED BY THE SWEDISH RESEARCH COUNCIL AND BY THE GÖRAN GUSTAFSSON FOUNDATION FOR RESEARCH IN THE NATURAL SCIENCES AND MEDICINE.
[2]SUPPORTED IN PART BY THE NSERC OF CANADA.

imply the finiteness of these quantities. And, in Section 5, we present a number of questions that appear to be open.

2 Notation and Background

Let $\{X_n\}$ be a stationary, time homogeneous Markov chain on the measurable space $(\mathcal{X}, \mathcal{F})$, with transition kernel P, reversible with respect to the probability measure $\pi(\cdot)$, so $\mathbf{P}[X_n \in S] = \pi(S)$ for all $n \in \mathbf{N}$ and $S \in \mathcal{F}$. Let $P^n(x, S) = \mathbf{P}[X_n \in S \,|\, X_0 = x]$ be the n-step transitions. Say that P is *ergodic* if it is ϕ-irreducible and aperiodic, from which it follows (cf. [23], [21], [19]) that $\lim_{n\to\infty} \sup_{S\in\mathcal{F}} |P^n(x, S) - \pi(S)| = 0$ for π-a.e. $x \in \mathcal{X}$. Write $\pi(g) = \int_{\mathcal{X}} g(x)\, \pi(dx)$, and $(Pg)(x) = \int_{\mathcal{X}} g(y)\, P(x, dy)$, and $\langle f, g\rangle = \int_{\mathcal{X}} f(x)\, g(x)\, \pi(dx)$. By reversibility, $\langle f, Pg\rangle = \langle Pf, g\rangle$.
Let $h : \mathcal{X} \to \mathbf{R}$ be a fixed, measurable function, with $\pi(h) = 0$. Let $\gamma_k = \mathbf{E}[h(X_0)\, h(X_k)] = \langle h,\, P^k h\rangle$ be the corresponding lag-k autocovariance. Say that a $\sqrt{n}$-CLT exists for h if $n^{-1/2} \sum_{i=1}^n h(X_i)$ converges weakly to Normal$(0, \sigma^2)$ for some $\sigma^2 < \infty$:

$$n^{-1/2} \sum_{i=1}^{n} h(X_i) \;\Rightarrow\; \text{Normal}(0, \sigma^2)\,, \qquad \sigma^2 < \infty\,, \tag{1}$$

where we allow for the degenerate case $\sigma^2 = 0$ corresponding to a point mass at 0.

Remark 1. Below we shall generally assume that the Markov chain is ergodic. However, the convergence (1) does not necessarily require ergodicity; see e.g. Proposition 29 of [19].

Remark 2. The assumption of stationarity is not crucial. For example, it follows from [14] that for Harris recurrent chains, if a CLT holds when started in stationarity, then it holds from all initial distributions.

Remark 3. We note that (1), and the results below, are all specific to the $n^{-1/2}$ normalisation and the Normal limiting distribution. Other normalisations and limiting distributions may sometimes hold, but we do not consider them here.

We shall also require spectral measures. Let $\mathcal{E}$ be the spectral decomposition measure (e.g. [22], Theorem 12.23) associated with P, so that

$$f(P) \;=\; \int_{-1}^{1} f(\lambda)\; \mathcal{E}(d\lambda)$$

for all bounded analytic functions $f : [-1, 1] \to \mathbf{R}$, and $\mathcal{E}(\mathbf{R}) = \mathcal{E}([-1, 1]) = I$ is the identity operator. (Of course, here $f(P)$ is defined in terms of power series, so that e.g. $\sin(P) = \sum_{j=0}^{\infty} (-1)^j\, P^{2j+1}$.) Let $\mathcal{E}_h$ be the induced spectral measure for h (cf. [5], p. 1753), viz.

$$\mathcal{E}_h(S) \;=\; \langle h,\; \mathcal{E}(S) h\rangle\,, \qquad S \subseteq [-1, 1] \;\; \text{Borel}$$

with $\mathcal{E}_h(\mathbf{R}) = \langle h,\, \mathcal{E}(\mathbf{R}) h\rangle = \langle h,\, h\rangle = \pi(h^2) < \infty$.
There are a number of possible formulae in the literature (e.g. [13], [8], [5]) for the limiting variance σ^2 in (1), including:

$$A \;=\; \lim_{n\to\infty} n^{-1} \mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right)\;;$$

$$B = \gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k\,;$$
$$C = \int_{-1}^{1}\frac{1+\lambda}{1-\lambda}\,\mathcal{E}_h(d\lambda)\,.$$

We consider A, B, and C below. Of course, if $\pi(h^2) < \infty$, then expanding the square gives $A = \gamma_0 + 2\lim_{n\to\infty}\sum_{k=1}^{n-1}\left(\frac{n-k}{n}\right)\gamma_k$. We shall also have occasion to consider versions of A and B where the limit is taken over odd integers only:

$$A' = \lim_{j\to\infty}(2j+1)^{-1}\mathbf{Var}\left(\sum_{i=1}^{2j+1}h(X_i)\right)\,;$$
$$B' = \gamma_0 + 2\lim_{j\to\infty}\sum_{k=1}^{2j+1}\gamma_k\,.$$

Obviously, $A' = A$ and $B' = B$ provided the limits in A and B exist. But it may be possible that, say, A' is well-defined even though A is not.

3 Relationships Between Variance Expressions

The following result is implicit in some earlier works (e.g. [13], [8], [5]), though it may not have previously been written down precisely.

Theorem 4. *If P is reversible and ergodic, and $\pi(h^2) < \infty$, then $A = B = C$ (though they may all be infinite).*

Theorem 4 is proved in Section 6. We first note that if ergodicity is not assumed, then we may have $A \neq B$:

Example 5. Let $\mathcal{X} = \{-1, 1\}$, with $\pi\{-1\} = \pi\{1\} = 1/2$, and $P(1, \{-1\}) = P(-1, \{1\}) = 1$, so P is reversible with respect to $\pi(\cdot)$. Let h be the identity function. Then $\left|\sum_{i=1}^{n}h(X_i)\right| \leq 1$, so $A = 0$. On the other hand, $\gamma_k = (-1)^k$, so $\gamma_0 + 2\sum_{k=1}^{2j+1}\gamma_k = 1 + 2(-1) = -1$, so $B' = 0$. However, B is an oscillating sum and thus undefined. So $A \neq B$, but Theorem 4 is not violated since the chain is periodic and hence not ergodic. And, a (degenerate) $\sqrt{n}$-CLT does hold, with $\sigma^2 = A = 0$. ■

Now, Kipnis and Varadhan [13] proved for reversible chains that if $C < \infty$, then a $\sqrt{n}$-CLT exists for h, with $\sigma^2 = C$. Combining this with Theorem 4, we have:

Corollary 6. *If P is reversible and ergodic, and $\pi(h^2) < \infty$, and any one of A, B, and C is finite, then a $\sqrt{n}$-CLT exists for h, with $\sigma^2 = A = B = C < \infty$.*

(Furthermore, it is easily seen [20] that $C < \infty$ whenever $\pi(h^2) < \infty$ and the spectrum of P is bounded away from 1.)

In a different direction, Roberts [17] considered the quantity $r(x) = \mathbf{P}[X_1 = x \,|\, X_0 = x]$, the probability of remaining at x, which is usually positive for Metropolis-Hastings algorithms. He proved that if $\lim_{n\to\infty} n\,\mathbf{E}[h^2(X_0)\,r(X_0)^n] = \infty$, then $A = \infty$ (and used this to prove that $A = \infty$ for some specific independence sampler examples). Combining his result with Theorem 4, we have:

Corollary 7. *If P is reversible and ergodic, and* $\lim_{n\to\infty} n\,\mathbf{E}[h^2(X_0)\,r(X_0)^n] = \infty$ *where* $\pi(h^2) < \infty$*, then* $A = B = C = \infty$*.*

Remark 8. If the Markov chain is not reversible, then the spectral measure required to define C becomes much more complicated, and we do not pursue that here. However, it is still possible to compare A and B. It follows immediately from the definitions and the dominated convergence theorem (cf. [3], p. 172; [5]) that if $\sum_k |\gamma_k| < \infty$, then $A = B < \infty$ (though this might not imply a $\sqrt{n}$-CLT for h). The condition $\sum_k |\gamma_k| < \infty$ is known to hold for uniformly ergodic chains (see [3]), and for reversible geometrically ergodic chains (since that implies [18] that $|\gamma_k| \le \rho^k\,\pi(h^2)$ for some $\rho < 1$), but it does not hold in general. For more about geometric ergodicity and CLTs see e.g. [10], [19], [12], and [9].

4 Converse: What Does a CLT Imply about Variance?

The result from [13] raises the question of the *converse*. Suppose $\{X_n\}$ is a stationary Markov chain, and $n^{-1}\sum_{i=1}^n h(X_i)$ converges weakly to Normal$(0,\sigma^2)$ for some $\sigma^2 < \infty$. Does it necessarily follow that any of A, B, and C are finite? An affirmative answer to this question would, for example, allow a strengthening of Corollary 7 to conclude that no $\sqrt{n}$-CLT holds for such h, and in particular a $\sqrt{n}$-CLT does not hold for the independence sampler examples considered by Roberts [17].

Even in the i.i.d. case (where $P(x,S) = \pi(S)$ for all $x \in \mathcal{X}$ and $S \in \mathcal{F}$), this question is non-trivial. However, classical results (cf. Sections IX.8 and XVII.5 of Feller [7]; for related results see e.g. [4], [2]) provide an affirmative answer in this case:

Theorem 9. *If $\{X_i\}$ are i.i.d., and $n^{-1/2}\sum_{i=1}^n h(X_i)$ converges weakly to Normal$(0,\sigma^2)$, where $0 < \sigma < \infty$ and $\pi(h) = 0$, then A, B, and C are all finite, and $\sigma^2 = A = B = C$.*

Proof. Let $Y_i = h(X_i)$, and let $U(z) = \mathbf{E}[Y_1^2\,\mathbf{1}_{|Y_1|\le z}]$. Then since the $\{Y_i\}$ are i.i.d. with mean 0, Theorem 1a on p. 313 of [7] says that there are positive sequences $\{a_n\}$ with $a_n^{-1}(Y_1 + \ldots + Y_n) \Rightarrow$ Normal$(0,1)$ if and only if $\lim_{z\to\infty}\big[U(sz)/U(z)\big] = 1$ for all $s > 0$. Furthermore, equation (8.12) on p. 314 of [7] (see also equation (5.23) on p. 579 of [7]) says that in this case,

$$\lim_{n\to\infty} n\,a_n^{-2}\,U(a_n) \;=\; 1\,. \tag{2}$$

Now, the hypotheses imply that $a_n^{-1}(Y_1 + \ldots + Y_n) \Rightarrow$ Normal$(0,1)$ where $a_n = c\,n^{1/2}$ with $c = \sigma$. Thus, from (2), we have $\lim_{n\to\infty} c^{-2}\,U(cn^{1/2}) = 1$. It follows that $\lim_{z\to\infty} U(z) = c^2 = \sigma^2 < \infty$, i.e. $\mathbf{E}(Y_1^2) < \infty$. We then compute that $\gamma_k = 0$ for $k \ge 1$, so $B = \gamma_0 = \mathbf{E}(Y_1^2) = \sigma^2 < \infty$. Hence, by Corollary 6, $\sigma^2 = A = B = C = \mathbf{E}(Y_1^2) < \infty$. ∎

Remark 10. In the above proof, if $\mathbf{E}(Y_1^2) = \sigma^2 < \infty$, then of course $U(z) \to \sigma^2$, implying that $U(sz)/U(z) \to \sigma^2/\sigma^2 = 1$, and the (classical) CLT applies. On the other hand, there are many distributions for the $\{Y_i\}$ which have infinite variance, but for which the corresponding U is still slowly varying in this sense. Examples include the density function $|y|^{-3}\mathbf{1}_{|y|\ge 1}$, and the cumulative distribution function $(1-(1+y)^{-2})\mathbf{1}_{y\ge 0}$. The results from [7] say that we cannot have $a_n = c\,n^{1/2}$ in such cases. (In the $y^{-3}\mathbf{1}_{|y|\ge 1}$ example, we instead have $a_n = c\,(n\log n)^{1/2}$.)

If $\{X_n\}$ is *not* assumed to be i.i.d., then the question becomes more complicated. Of course, if $\{n^{-1}\sum_{i=1}^n h(X_i)^2\}_{n=1}^\infty$ is *uniformly integrable*, then whenever a $\sqrt{n}$-CLT exists we must have

$A = \sigma^2$, which implies by Theorem 4 (assuming reversibility) that $\sigma^2 = A = B = C < \infty$. However, it is not clear when this uniform integrability condition will be satisfied, and we now turn to some counter-examples.

If $\{X_n\}$ is not even *reversible*, then it is possible for a $\sqrt{n}$-CLT to hold even though A is not finite:

Example 11. Let the state space $\mathcal{X}$ be the integers, and let h be the identity function. Consider the Markov chain on $\mathcal{X}$ with transition probabilities given (for $j \geq 1$) by $P(0,0) = 1/2$, $P(j,-j) = P(-j,0) = 1$, and $P(0,j) = c/j^3$ where $c = \frac{1}{2}\,\zeta(3)^{-1}$ with $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ the Riemann zeta function. (That is, whenever the chain leaves 0, it cycles to some positive integer j, then to $-j$, and then back to 0.)

This Markov chain is irreducible and aperiodic, with stationary distribution given by $\pi(0) = 1/2$ and $\pi(j) = \pi(-j) = c'/j^3$ where $c' = \zeta(3)/4$. Furthermore, $\pi(h) = 0$.

Since $h(j) + h(-j) = 0$ and $h(0) = 0$, it is easy to see that for $n \geq 2$, we have $\sum_{i=1}^{n} h(X_i) = \mathbf{1}_{X_1<0}\, X_1 + \mathbf{1}_{X_n>0}\, X_n$. In particular, $\sum_{i=1}^{n} h(X_i) \leq |X_1| + |X_n|$, and since by stationarity $\mathbf{E}|X_1| = \mathbf{E}|X_n| = \sum_{x\neq 0} |x| c' |x|^{-3} < \infty$, it follows immediately that $n^{-1/2} \sum_{i=1}^{n} h(X_i)$ converges in distribution to 0, i.e. to $N(0,0)$. It also follows that for $n \geq 2$,

$$\mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right) \;=\; 2\,\mathbf{E}[X_1^2\,\mathbf{1}_{X_1>0}] \;=\; 2\sum_{j=1}^{\infty} j^2 (c'/j^3) \;=\; \infty\,.$$

Hence, $A = \lim_{n\to\infty} n^{-1} \mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right) = \infty$. ∎

Remark 12. If we wish, we can modify Example 11 to achieve convergence to $N(0,1)$ instead of $N(0,0)$, as follows. Replace the state space $\mathcal{X}$ by $\mathcal{X} \times \{-1,1\}$, let the first coordinate $\{X_n\}$ evolve as before, let the second coordinate $\{Y_n\}$ evolve independently of $\{X_n\}$ such that each Y_n is i.i.d. equal to -1 or 1 with probability 1/2 each, and redefine h as $h(x,y) = x + y$. Then $n^{-1/2} \sum_{i=1}^{n} h(X_i)$ will converge in distribution to $N(0,1)$.

Remark 13. If we wish, we can modify Example 11 to make the functional h bounded, as follows. Instead of jumping from 0 to a value j, the chain instead jumps from 0 to a deterministic path of length $2j+1$, where the first j states have $h = +1$, the next j states have $h = -1$, and then the chain jumps back to 0.

We now show that even if $\{X_n\}$ is reversible, the existence of a $\sqrt{n}$-CLT does *not* necessarily imply that A is finite:

Example 14. We again let the state space $\mathcal{X}$ be the integers, with h the identity function. Consider the Markov chain on $\mathcal{X}$ with transition probabilities given by $P(0,0) = 0$, $P(0,y) = c\,|y|^{-4}$ for $y \neq 0$ (where $c = 45/\pi^4$), and, for $x \neq 0$,

$$P(x,y) = \begin{cases} |x|^{-1}\,, & y = 0 \\ 1 - |x|^{-1}\,, & y = -x \\ 0\,, & \text{otherwise}\,. \end{cases}$$

That is, the chain jumps from 0 to a random site x, and then oscillates between $-x$ and x for a geometric amount of time with mean $|x|$, before returning to 0. This chain is irreducible and aperiodic, and is reversible with respect to the stationary distribution given by $\pi(x) = c'|x|^{-3}$ and $\pi(0) = c'/c$, where $c' = [c^{-1} + 2\,\zeta(3)]^{-1}$.

We prove the existence of a CLT by regeneration analysis (see e.g. [1]). We define each visit to the state 0 as a regeneration time, and write these regeneration times as $T_1, T_2, \ldots$. For convenience, we set $T_0 = 0$ (even though we usually will not have $X_0 = 0$, i.e. we do *not* impose a regeneration at time 0, unlike e.g. [15], [10]). These times break up the Markov chain into a collection of random paths ("tours"), of the form $\{(X_{T_j+1}, X_{T_j+2}, \ldots, X_{T_{j+1}})\}_{j=0}^{\infty}$. The tours from T_1 onwards each travel from 0 to 0, are all i.i.d. We note that for $j \geq 1$, the sum over a single tour, $\sum_{i=T_j+1}^{T_{j+1}} X_i$, is either X_{T_j+1}, $-X_{T_j+1}$, or 0. Furthermore, $\mathbf{P}(X_{T_j+1} = y) = P(0,y) = c|y|^{-4}$, so $\mathbf{E}(X_{T_j+1}^2) = \sum_{y \neq 0} y^2 c|y|^{-4} < \infty$. This implies that $\sum_{i=T_j+1}^{T_{j+1}} X_i$ has finite variance, say V. It then follows from the classical central limit theorem that as $J \to \infty$, $J^{-1/2} \sum_{i=T_1+1}^{T_{J+1}} X_i$ converges in distribution to $N(0,V)$. By the Law of Large Numbers, as $J \to \infty$, $T_J/J \to \tau$ where $\tau = \mathbf{E}[T_{j+1} - T_j]$. Hence, asymptotically for large n, if we find J with $T_J < n \leq T_{J+1}$, then $n^{-1/2} \sum_{i=1}^{n} X_i \approx (J\tau)^{-1/2} \sum_{i=T_1+1}^{T_{J+1}} X_i \approx \tau^{-1/2} N(0,V) = N(0, V/\tau)$. That is, a $\sqrt{n}$-CLT exists for h, with mean 0 and variance $\sigma^2 = V/\tau = \mathbf{Var}\Big[\sum_{i=T_j+1}^{T_{j+1}} X_i\Big] \Big/ \mathbf{E}[T_{j+1} - T_j]$.

We now claim that $\mathbf{Var}[\sum_{i=0}^{n} X_i]$ is infinite for n even. Indeed, in the special case $n = 0$,

$$\mathbf{Var}\Big[\sum_{i=0}^{n} X_i\Big] \;=\; \mathbf{Var}[X_0] \;=\; \sum_{x \in \mathcal{X}} \mathbf{P}[X_0 = x]\, x^2 \;=\; \sum_{x \in \mathcal{X}} c'|x|^{-3}\, x^2 \;=\; \sum_{x \in \mathcal{X}} c'|x|^{-1} \;=\; \infty\,.$$

Assume now that $n \geq 2$. Let $S_n = \sum_{i=0}^{n} X_n$, and let D_n be the event that $X_i = 0$ for some $0 \leq i \leq n$ (so that $0 < \mathbf{P}(D_n) < 1$). Since n is even, we have that $S_n = X_0$ on the event D_n^C (because of cancellation), and so

$$\mathbf{E}[S_n^2\, \mathbf{1}_{D_n^C}] \;=\; \mathbf{E}[X_0^2\, \mathbf{1}_{D_n^C}] \;=\; \sum_{x \in \mathcal{X}} \mathbf{P}[X_0 = x\,,\, D_n^C]\; x^2 \;=\; \sum_{x \in \mathcal{X}} c'\,|x|^{-3}\,(1 - |x|^{-1})^n\, x^2$$

$$=\; \sum_{x \in \mathcal{X}} c'\,|x|^{-1}\,(1 - |x|^{-1})^n \;=\; \infty\,.$$

Hence, $\mathbf{Var}[S_n] = \mathbf{E}[S_n^2] = \infty$. (In fact, $\mathbf{Cov}[X_m, X_k] = +\infty$ for $k - m$ even, and $-\infty$ for $k - m$ odd, but we do not use that here.)

This proves the claim that $\mathbf{Var}[\sum_{i=0}^{n} X_i]$ is infinite for all even n. In particular, the limit in the definition of A is either infinite or undefined, so A is certainly not finite. ∎

Remark 15. We now show that in Example 14, in fact A is undefined rather than infinite. (In particular, it is *not* true that $A = C$, thus showing that the condition $\pi(h^2) < \infty$ cannot be dropped from Theorem 4 above.)

To prove this, it suffices to show that $\mathbf{Var}(n^{-1/2} S_n)$ remains bounded over all odd n (so that the limit defining A oscillates between bounded and infinite values). But for n odd, we have that $S_n = 0$ on D_n^C (again because of cancellation), so the claim will follow from showing that $n^{-1}\, \mathbf{Var}[S_n\, \mathbf{1}_{D_n}]$ remains bounded for all odd n.

Let $\{T_j\}$ be the regeneration times as above, so $X_{T_j} = 0$, and let $Y_j = X_{T_j+1}$. Then $v \equiv \mathbf{Var}(Y_j) = \sum_{y \neq 0} c y^{-4} y^2 < \infty$. Now, on D_n, the sequence $(X_{T_1}, \ldots, X_n)$ breaks up into some number $1 \leq m \leq \frac{n-1}{2}$ of complete tours [say, $(X_{T_1}, \ldots, X_{T_2 - 1})$, $(X_{T_2}, \ldots, X_{T_3 - 1})$, …, $(X_{T_m}, \ldots, X_{T_{m+1}-1})$], plus one possibly-incomplete final tour [say, $(X_{T_{m+1}}, \ldots, X_n)$]. Now, for $j = 1, \ldots, m+1$, we have that $X_{T_j} + \ldots + X_{\min\{T_{j+1}-1, n\}}$ equals either Y_j or 0, and it follows

that $Var[X_{T_j}+\ldots+X_{\min\{T_{j+1}-1,n\}}] \leq v$. The contributions of different such blocks are clearly uncorrelated, so $\mathbf{Var}[\mathbf{1}_{D_n}\sum_{i=T_1}^{n} X_i] \leq (\frac{n-1}{2}+1)v \leq nv$.
Now let $T_n^- = \max\{i < n : X_i = 0\}$ be the time of the last visit to 0 before time n. Then arguing as before,

$$\mathbf{Var}[\mathbf{1}_{D_n}\sum_{i=T_1}^{T_n^-} X_i] \;\leq\; \mathbf{Var}[\mathbf{1}_{D_n}\sum_{i=T_1}^{n} X_i] \;\leq\; nv\,.$$

But by time reversibility, the distribution of $\mathbf{1}_{D_n}\sum_{i=0}^{T_n^-} X_i$ is identical to that of $\mathbf{1}_{D_n}\sum_{i=T_1}^{n} X_i$. Hence, $\mathbf{Var}[\mathbf{1}_{D_n}\sum_{i=0}^{T_n^-} X_i] = \mathbf{Var}[\mathbf{1}_{D_n}\sum_{i=T_1}^{n} X_i] \leq nv$.
Finally, we note that

$$\mathbf{Var}[\mathbf{1}_{D_n}S_n] \;=\; \mathbf{Var}\Big[\mathbf{1}_{D_n}\sum_{i=0}^{T_n^-} X_i + \mathbf{1}_{D_n}\sum_{i=T_1}^{n} X_i - \mathbf{1}_{D_n}\sum_{i=T_1}^{T_n^-} X_i\Big]\,.$$

We have shown that each of the three terms on the right-hand side has variance $\leq nv$. It follows by Cauchy-Schwarz that their covariances are also $\leq nv$. Hence, expanding the square, $\mathbf{Var}[\mathbf{1}_{D_n}S_n] \leq 9nv$, so $n^{-1}\,\mathbf{Var}[S_n\,\mathbf{1}_{D_n}] \leq 9v$ which is bounded independently of n. ■

5 Some Open Problems

The following questions are not answered by the above results, and appear (so far as we can tell) to be open problems. Let $\{X_n\}$ be a stationary Markov chain.

1. If the Markov chain is ergodic, but not necessarily reversible (cf. Theorem 4) nor geometrically ergodic nor uniformly ergodic (cf. Remark 8), does it necessarily follow that $A = B$ (allowing that they may both be infinite)? What if we assume that $P = P_1P_2$ where each P_i is reversible (which holds if, say, P is a two-variable systematic-scan Gibbs sampler, see e.g. [23])?

2. Is there a reversible, ergodic example where a $\sqrt{n}$-CLT exists, but $A = \infty$? (As discussed in Remark 15, in Example 14 the limit in the definition of A is actually an undefined, oscillating limit as opposed to equaling positive infinity.)

3. In particular, is there a reversible, ergodic example where a $\sqrt{n}$-CLT exists, but where Roberts' condition $\lim_{n\to\infty} n\,\mathbf{E}[h^2(X_0)\,r(X_0)^n] = \infty$ holds?

4. If the chain is reversible and ergodic and a $\sqrt{n}$-CLT exists, what further conditions (weaker than i.i.d. or uniformly ergodic) still imply that $A < \infty$? (There are various results in the stationary process literature, e.g. [11] and [16], that are somewhat related to those from [7] used in the proof of Theorem 9, but their applicability to generalising Theorem 9 is unclear.)

5. In particular, if the chain is reversible and a $\sqrt{n}$-CLT exists, and there is $\delta > 0$ such that $r(x) \geq \delta$ for all $x \in \mathcal{X}$, does this imply that $A < \infty$? (Note that the last condition fails for Example 14.)

6. In a different direction, if the chain is reversible and a $\sqrt{n}$-CLT exists, and $\pi(h^2) < \infty$, does this imply that $A < \infty$? (Of course, Example 14 has $\pi(h^2) = \infty$. Also, Example 11 is not reversible, although as discussed in Remark 13 it is possible to make h bounded and thus $\pi(h^2) < \infty$ in that case.)

7. Related to this, can the condition $\pi(h^2) < \infty$ be dropped from Corollaries 6 and 7? (It *cannot* be dropped from Theorem 4, on account of Remark 15.)

6 Proof of Theorem 4

In this section, we prove Theorem 4. We assume throughout that P is a reversible and ergodic Markov chain, and that $\pi(h^2) < \infty$. We begin with a lemma (somewhat similar to Theorem 3.1 of [8]).

Lemma 16. *$\gamma_{2i} \geq 0$, and $|\gamma_{2i+1}| \leq \gamma_{2i}$, and $|\gamma_{2i+2}| \leq \gamma_{2i}$.*

Proof. By reversibility, $\gamma_{2i} = \langle f, P^{2i} f\rangle = \langle P^i f, P^i f\rangle = \|P^i f\|^2 \geq 0$.
Also, $|\gamma_{2i+1}| = \langle f, P^{2i+1} f\rangle = |\langle P^i f, P(P^i f)\rangle| \leq \|P^i f\|^2 \|P\| \leq \|P^i f\|^2 = \gamma_{2i}$.
Similarly, $|\gamma_{2i+2}| = \langle f, P^{2i+2} f\rangle = |\langle P^i f, P^2(P^i f)\rangle| \leq \|P^i f\|^2 \|P^2\| \leq \|P^i f\|^2 = \gamma_{2i}$. ■

Lemma 17. $\lim_{k\to\infty} \gamma_k = 0$.

Proof. Since P is ergodic, it does not have an eigenvalue 1 or -1. Hence (cf. [22], Theorem 12.29(b)) its spectral measure $\mathcal{E}$ does not have an atom at 1 or -1, i.e. $\mathcal{E}(\{-1,1\}) = 0$, so also $\mathcal{E}_h(\{-1,1\}) = 0$ (cf. [8], Lemma 5). Hence, by the dominated convergence theorem (since $|\lambda^k| \leq 1$ for $-1 \leq \lambda \leq 1$, and $\int_{-1}^{1} 1\, \mathcal{E}_h(d\lambda) = \pi(h^2) < \infty$), we have:

$$\lim_{k\to\infty} \gamma_k \;=\; \lim_{k\to\infty} \langle h,\, P^k h\rangle \;=\; \lim_{k\to\infty} \int_{-1}^{1} \lambda^k\, \mathcal{E}_h(d\lambda)$$

$$=\; \int_{-1}^{1} \left(\lim_{k\to\infty} \lambda^k\right) \mathcal{E}_h(d\lambda) \;=\; \int_{-1}^{1} 0\, \mathcal{E}_h(d\lambda) \;=\; 0\,. \quad ■$$

Proposition 18. *$A' = B'$. (We allow for the possibility that $A' = B' = \infty$.)*

Proof. We have that

$$(2j+1)^{-1}\, \mathbf{Var}\left(\sum_{i=1}^{2j+1} h(X_i)\right) \;=\; \gamma_0 + 2\gamma_1 + 2\sum_{i=1}^{j}\left(\frac{2j+1-2i}{2j+1}\,\gamma_{2i} + \frac{2j+1-2i-1}{2j+1}\,\gamma_{2i+1}\right)$$

$$=\; \gamma_0 + \frac{4j}{2j+1}\gamma_1 + 2\sum_{i=1}^{j} \frac{\gamma_{2i}}{2j+1} + 2\sum_{i=1}^{j} \frac{2j+1-2i-1}{2j+1}\,(\gamma_{2i} + \gamma_{2i+1})\,. \tag{3}$$

By Lemma 16, $\gamma_{2i} + \gamma_{2i+1} \geq 0$, so as $j \to \infty$, for fixed i,

$$\frac{2j+1-2i-1}{2j+1}\,(\gamma_{2i} + \gamma_{2i+1}) \;\nearrow\; \gamma_{2i} + \gamma_{2i+1}\,,$$

i.e. the convergence is *monotonic*. Hence, by the monotone convergence theorem,

$$\lim_{j\to\infty} 2\sum_{i=1}^{j} \frac{2j+1-2i-1}{2j+1}\,(\gamma_{2i}+\gamma_{2i+1}) \;=\; \lim_{j\to\infty} 2\sum_{i=1}^{j}(\gamma_{2i}+\gamma_{2i+1})\,.$$

By Lemma 17, $\gamma_{2i} \to 0$ as $i \to \infty$, so $\sum_{i=1}^{j} \frac{\gamma_{2i}}{2j+1} \to 0$ as $j \to \infty$. Finally, $\frac{4j}{2j+1} \to 2$. Putting this all together, we conclude from (3) that

$$\lim_{j\to\infty}\,(2j+1)^{-1}\,\mathbf{Var}\left(\sum_{i=1}^{2j+1} h(X_i)\right) \;=\; \gamma_0 \,+\, 2\gamma_1 \,+\, 2\,\lim_{j\to\infty}\sum_{i=1}^{j}(\gamma_{2i}+\gamma_{2i+1})\,,$$

i.e. $A' = B'$. ∎

Corollary 19. $A = B$. *(We allow for the possibility that $A = B = \infty$.)*

Proof. If P is ergodic, then by Lemma 17, $\gamma_k \to 0$, so $B = B'$. Also,

$$(n+1)^{-1}\mathbf{Var}\left(\sum_{i=1}^{n+1} h(X_i)\right) \,-\, n^{-1}\mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right) \tag{4}$$

$$=\; n^{-1}\left[\mathbf{Var}\left(\sum_{i=1}^{n+1} h(X_i)\right) \,-\, \mathbf{Var}\left(\sum_{i=1}^{n} h(X_i)\right)\right] \,-\, [n(n+1)]^{-1}\mathbf{Var}\left(\sum_{i=1}^{n+1} h(X_i)\right)$$

Now, the first term above is equal to $n^{-1}\sum_{i=1}^{n}\gamma_i$ (which goes to 0 since $\gamma_k \to 0$), plus $n^{-1}\mathbf{E}[h^2(X_{i+1})]$ (which goes to 0 since $\pi(h^2) < \infty$). The second term is equal to

$$\frac{\gamma_0}{n} \,+\, 2\sum_{k=1}^{n-1}\frac{n+1-k}{n(n+1)}\,\gamma_k$$

which also goes to 0. We conclude that the difference in (4) goes to 0 as $n \to \infty$, so that $A = A'$. Hence, by Proposition 18, $A = A' = B' = B$. ∎

Proposition 20. $B = C$. *(We allow for the possibility that $B = C = \infty$.)*

Proof. We compute that

$$\begin{aligned}
B \;&=\; \lim_{k\to\infty}\Big(\langle h,h\rangle + 2\,\langle h,\,Ph\rangle + 2\,\langle h,\,P^2h\rangle + \ldots + 2\,\langle h,\,P^kh\rangle\Big) \\
&=\; \lim_{k\to\infty}\Big\langle h,\;(I+2P+2P^2+\ldots+2P^k)h\Big\rangle \\
&=\; \lim_{k\to\infty}\int_{-1}^{1}(1+2\lambda+2\lambda^2+\ldots+2\lambda^k)\;\mathcal{E}_h(d\lambda) \\
&=\; \lim_{k\to\infty}\int_{-1}^{1}\Big(2\frac{1-\lambda^{k+1}}{1-\lambda}-1\Big)\;\mathcal{E}_h(d\lambda)
\end{aligned}$$

$$= \lim_{k\to\infty} \int_{-1}^{1} \left(\frac{1+\lambda-2\lambda^{k+1}}{1-\lambda} \right) \mathcal{E}_h(d\lambda)$$

$$= \lim_{k\to\infty} \int_{-1}^{0} \left(\frac{1+\lambda-2\lambda^{k+1}}{1-\lambda} \right) \mathcal{E}_h(d\lambda) \; + \; \lim_{k\to\infty} \int_{0}^{1} \left(\frac{1+\lambda-2\lambda^{k+1}}{1-\lambda} \right) \mathcal{E}_h(d\lambda)$$

$$= \int_{-1}^{0} \left(\frac{1+\lambda}{1-\lambda} \right) \mathcal{E}_h(d\lambda) \; + \; \int_{0}^{1} \left(\frac{1+\lambda}{1-\lambda} \right) \mathcal{E}_h(d\lambda)$$

$$= C\,,$$

where the first limit follows from the dominated convergence theorem since $|\frac{1+\lambda-2\lambda^{k+1}}{1-\lambda}| \leq 3$ for $\lambda \leq 0$, and the second limit (which may be infinite) follows from the monotone convergence theorem since $\left\{1+\lambda-2\lambda^{k+1}1-\lambda\right\}_{k=1}^{\infty} \nearrow 1+\lambda 1-\lambda$ for $\lambda \geq 0$. (For definiteness, if $\mathcal{E}_h\{0\}$ is non-zero, then we include the point $\lambda = 0$ in the integrals from -1 to 0, and not in the integrals from 0 to 1.) ∎

Theorem 4 then follows immediately from Corollary 19 and Proposition 20.

Acknowledgement. We thank J.H. Blanchet, R. Douc, C.J. Geyer, J.P. Hobert, G.L. Jones, N. Madras, E. Mossel, E. Moulines, G.O. Roberts, V. Roy, and T. Salisbury for very helpful discussions of these issues, and thank the anonymous referee for a very careful reading that led to many improvements.

References

[1] S. Asmussen (1987), Applied Probability and Queues. John Wiley & Sons, New York. MR0889893

[2] S.K. Basu (1984), A local limit theorem for attraction to the standard normal law: The case of infinite variance. Metrika **31**, 245–252. Available at http://www.springerlink.com/content/v616742404102u13/ MR0754965

[3] P. Billingsley (1968), Convergence of Probability Measures. Wiley, New York. MR0233396

[4] M. Breth, J. S. Maritz, and E. J. Williams (1978), On Distribution-Free Lower Confidence Limits for the Mean of a Nonnegative Random Variable. Biometrika **65**, 529–534.

[5] K.S. Chan and C.J. Geyer (1994), Discussion paper. Ann. Stat. **22**, 1747–1758.

[6] X. Chen (1999). Limit theorems for functionals of ergodic Markov chains with general state space. Mem. Amer. Math. Soc. **139**. MR1491814

[7] W. Feller (1971), An introduction to Probability Theory and its applications, Vol. *II*, 2^{nd} ed. Wiley & Sons, New York. MR0270403

[8] C.J. Geyer (1992), Practical Markov chain Monte Carlo. Stat. Sci., Vol. **7**, No. **4**, 473-483.

[9] O. Häggström (2005), On the central limit theorem for geometrically ergodic Markov chains. Probab. Th. Relat. Fields **132**, 74–82. MR2136867

[10] J.P. Hobert, G.L. Jones, B. Presnell, and J.S. Rosenthal (2002), On the Applicability of Regenerative Simulation in Markov Chain Monte Carlo. *Biometrika* **89**, 731–743. MR1946508

[11] I.A. Ibragimov and Y.V. Linnik (1971), Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen (English translation). MR0322926

[12] G.L. Jones (2004), On the Markov chain central limit theorem. Prob. Surveys **1**, 299–320. MR2068475

[13] C. Kipnis and S.R.S. Varadhan (1986), Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. Comm. Math. Phys. **104**, 1-19. MR0834478

[14] S.P. Meyn and R.L. Tweedie (1993), Markov chains and stochastic stability. Springer-Verlag, London. Available at probability.ca/MT. MR1287609

[15] P.A. Mykland, L. Tierney, and B. Yu (1995), Regeneration in Markov chain samplers. J. Amer. Stat. Assoc. **90**, 233–241. MR1325131

[16] M. Peligrad (1986), Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey). In *Dependence in Probability and Statistics: A Survey of Recent Results*, E. Eberlein and M.S. Taqqu, eds., Birkhauser, Cambridge, Mass., pp. 193–223. MR0899991

[17] G.O. Roberts (1999), A note on acceptance rate criteria for CLTs for Metropolis-Hastings algorithms. J. Appl. Prob. **36**, 1210–1217. MR1742161

[18] G.O. Roberts and J.S. Rosenthal (1997), Geometric ergodicity and hybrid Markov chains. Electronic Comm. Prob. **2**, Paper no. 2, 13–25. MR1448322

[19] G.O. Roberts and J.S. Rosenthal (2004), General state space Markov chains and MCMC algorithms. Prob. Surveys **1**, 20–71. MR2095565

[20] G.O. Roberts and J.S. Rosenthal (2006), Variance Bounding Markov Chains. Preprint. MR2288716

[21] J.S. Rosenthal (2001), A review of asymptotic convergence for general state space Markov chains. Far East J. Theor. Stat. **5**, 37–50. MR1848443

[22] W. Rudin (1991), Functional Analysis, 2nd ed. McGraw-Hill, New York. MR1157815

[23] L. Tierney (1994), Markov chains for exploring posterior distributions (with discussion). Ann. Stat. **22**, 1701–1762. MR1329166