The Annals of Applied Probability
2002, Vol. 12, No. 1, 224–247

POLYNOMIAL CONVERGENCE RATES OF MARKOV CHAINS[1]

BY SØREN F. JARNER AND GARETH O. ROBERTS

Lancaster University

In this paper we consider Foster–Liapounov-type drift conditions for Markov chains which imply polynomial rate convergence to stationarity in appropriate V-norms. We also show how these results can be used to prove central limit theorems for functions of the Markov chain. We consider two examples concerning random walks on the half line and the independence sampler.

1. Introduction. This paper considers the use of Foster–Liapounov-type drift conditions to establish polynomial rates of convergence of the f-norm for a general state space Markov chain in discrete time. Results of this type involving geometric convergence rates are now well established; see, for example, Meyn and Tweedie (1992) and Chapters 15, 16 of Meyn and Tweedie (1993). However, the more subtle polynomial case is not nearly as well understood. The foundational work of Tuominen and Tweedie (1994) studies general subgeometric rates using a sequence of drift conditions. Our results build on this work, but because we restrict ourselves to polynomial rates we are able to take a different and more direct approach which ultimately leads to the derivation of the single drift condition (1). This condition is the natural analogue of the drift condition for geometric ergodicity and, as will be illustrated by examples, it is simple to apply in practice.

Let $\mathbf{X} = (X_0, X_1, \ldots)$ be a discrete-time Markov chain on a general state space with transition kernel P. Assume $\mathbf{X}$ is ψ-irreducible, aperiodic and positive recurrent. A main result of the paper is Theorem 3.6 which states that if there exists a test function $V \geq 1$, positive constants c and b, a petite set C and $0 \leq \alpha < 1$ such that

$$PV \leq V - cV^{\alpha} + b\mathbb{1}_C, \tag{1}$$

then the chain is positive recurrent and there is polynomial convergence of the n-step transition kernel P^n to the invariant distribution π in the sense that the following statement holds

$$n^{\beta-1}\|P^n(x,\cdot) - \pi\|_{V_\beta} \to 0, \qquad n \to \infty, \tag{2}$$

Received May 2000; revised May 2001.

[1]Supported in part by the Danish Research Training Office and the EU TMR network ERB-FMRX-CT96-0095 on “Computational and Statistical Methods for the Analysis of Spatial Data.”

AMS 2000 subject classifications. 60J05, 60J10.

Key words and phrases. Markov chains, Foster–Liapounov drift conditions, polynomial convergence, central limit theorems, independence sampler.

for all $1 \le \beta \le 1/(1-\alpha)$, where $V_\beta = V^{1-\beta(1-\alpha)}$ and the f-norm is defined for a signed measure μ as $\|\mu\|_f = \sup_{g\,:\,|g|\le f} |\mu(g)|$. This gives a precise characterization of the rate with which the different powers of V converges in terms of the exponent α: one order of convergence is gained for each $1-\alpha$ power lost. In particular, there is convergence of order $\alpha/(1-\alpha)$ in the total variation norm.

This result relies on a more general result, Theorem 3.2, in which we establish conditions implying polynomial convergence when a set of drift conditions is satisfied. This is related to the work of Fort and Moulines (2000). The connection to the existence of moments of the hitting times of the Markov chain to the set C is established in Theorem 3.4. In particular, it follows from that theorem that if there exists a petite set C such that the mth moments of the return times to this set exist and are bounded on C then there is convergence of order $m-1$ in total variation norm. Theorem 3.4 also shows that the convergence rate given by Theorem 3.2 is optimal when applied to the hitting time solution.

Section 2 gives the necessary definitions and results from Tuominen and Tweedie (1994). The main results are derived in Section 3, and Section 4 investigates implications for our work in terms of central limit theorems for Markov chains. Examples are given in Section 5. We begin with an analysis of the random walk on the half line, an example also considered in Tuominen and Tweedie (1994). Our second example considers an important type of Markov chain Monte Carlo algorithm known as the *independence sampler*.

2. Definitions and the Tuominen–Tweedie result. Let $\mathbf{X} = (X_0, X_1, \ldots)$ be a discrete-time Markov chain on a general state space $(E, \mathcal{E})$, where $\mathcal{E}$ is a countably generated σ-algebra. The Markov transition kernel for $\mathbf{X}$ is denoted by P. Let P^n, $n \in \mathbb{N}_0$, denote the n-step transition kernel for the Markov chain, that is,

$$P^n(x, A) = \mathsf{P}_x(X_n \in A), \qquad x \in E,\ A \in \mathcal{E}, \tag{3}$$

where P_x is the conditional distribution of the chain given $X_0 \equiv x$. The corresponding expectation operator will be denoted E_x. For any function f we write $Pf(x)$ for the function $\int f(y)P(x, dy)$, and for any signed measure μ we write $\mu(f)$ for $\int f(y)\mu(dy)$.

We assume that P is ψ-irreducible and aperiodic [cf. Meyn and Tweedie (1993)], where ψ is a maximal irreducibility measure, we let $\mathcal{E}^+ = \{A \in \mathcal{E} \mid \psi(A) > 0\}$. A set A is called full if $\psi(A^c) = 0$ and absorbing if $P(x, A) = 1$ for all $x \in A$.

We will need the notion of petite and regular sets, and the associated concepts of sampling distributions and hitting times. For a distribution $a = (a_n)_{n\in\mathbb{N}}$ let K_a be the transition kernel given by

$$K_a(x, A) = \sum_{n=1}^{\infty} a_n P^n(x, A), \qquad x \in E,\ A \in \mathcal{E}.$$

This is the Markov transition kernel for the Markov chain $\mathbf{X}$ observed at time points with intervals sampled according to a. A set $C \in \mathcal{E}$ is called petite if there exists a sampling distribution a and a nontrivial measure ν such that

$$K_a(x, \cdot) \geq \nu(\cdot), \qquad x \in C. \tag{4}$$

Under our assumptions of ψ-irreducibility and aperiodicity, this is the same as the set C being small [cf. Meyn and Tweedie (1993)]. A set C is called an atom if the measures $P(x, \cdot)$ are identical for all x in C.

For any set $B \in \mathcal{E}$ let $\tau_B = \min\{n \geq 1 \mid X_n \in B\}$ denote the first time the Markov chain returns to the set B. This is a stopping time with respect to the filtration $(\mathcal{F}_n)$, where $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. A set $A \in \mathcal{E}$ is said to be (f, r)-regular where $f\colon E \to [1, \infty)$ and $r\colon \mathbb{N}_0 \to [1, \infty)$ if, for every $B \in \mathcal{E}^+$,

$$\sup_{x \in A} \mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} r(k) f(X_k)\right] < \infty. \tag{5}$$

A point $x \in E$ is called (f, r)-regular if the singleton $A = \{x\}$ is (f, r)-regular.

In this paper we are interested in polynomial rate functions, that is, rate functions of the form $r(n) = (n+1)^\beta$ for some $\beta \geq 0$. These are contained in a more general class Λ of subgeometric rate functions considered in Tuominen and Tweedie (1994). For this class Tuominen and Tweedie (1994) have the following main result.

THEOREM 2.1. *Suppose P is ψ-irreducible and aperiodic and let $f\colon E \to [1, \infty)$ and $r \in \Lambda$ be given. The following conditions are equivalent*:

(i) *There exists a petite set C such that*

$$\sup_{x \in C} \mathsf{E}_x\left[\sum_{k=0}^{\tau_C - 1} r(k) f(X_k)\right] < \infty. \tag{6}$$

(ii) *There exist a sequence $(V_n)_{n \in \mathbb{N}_0}$ of functions $V_n\colon E \to [0, \infty]$, a petite set C and constant b such that V_0 is bounded on C, $\{V_1 < \infty\} \subset \{V_0 < \infty\}$ and*

$$PV_{n+1} \leq V_n - r(n) f + b r(n) \mathbb{1}_C, \qquad n \in \mathbb{N}_0. \tag{7}$$

(iii) *There exists an (f, r)-regular set $A \in \mathcal{E}^+$.*

Any of these conditions imply that there exists a unique invariant distribution π and that, for all (f, r)-regular points x,

$$r(n) \|P^n(x, \cdot) - \pi\|_f \to 0, \qquad n \to \infty; \tag{8}$$

the set of all (f, r)-regular points is full, absorbing and contains the set $\{V_0 < \infty\}$.

Note that P is called (f, r)-regular if the conditions of Theorem 2.1 are satisfied and every point is (f, r)-regular. In the remainder of this paper we consider only polynomial rate functions and for these we derive simpler and easier verifiable drift conditions than the system of drift conditions (7) needed for general rate functions. Fort and Moulines (2000) also considered the case of polynomial rate functions and they constructed a solution to (7) from a system of drift conditions similar to the system we consider in Theorem 3.2. However, the system of drift conditions we consider is more general and we use it to show the existence of (f, r)-regular sets instead. This is a simpler and more direct approach of showing (f, r)-regularity of the chain. The connection between solutions to the system of drift conditions and moments of hitting times established in Theorem 3.4 and the reduction to a single drift condition in Theorem 3.6 are also new.

3. Drift conditions and regularity.

The most general result of this paper is Theorem 3.2 which identifies regular sets from a set of drift conditions. It uses the following lemma which is a straightforward generalization of Lemma 11.3.10 of Meyn and Tweedie (1993).

LEMMA 3.1. *For any stopping time τ, any sampling distribution a, any positive function w: $\mathbb{N}_0 \to [0, \infty)$ and any positive function f: $E \to [0, \infty)$,*

$$\mathsf{E}_x\left[\sum_{k=0}^{\tau-1} w(k) K_a(X_k, f)\right] = \sum_{i=1}^{\infty} a_i \mathsf{E}_x\left[\sum_{k=0}^{\tau-1} w(k) f(X_{k+i})\right]. \tag{9}$$

PROOF. Using the Markov property, Fubini's theorem and the fact that $(k < \tau) \in \mathcal{F}_k$, we get

$$\begin{aligned}
\mathsf{E}_x\left[\sum_{k=0}^{\tau-1} w(k) K_a(X_k, f)\right] &= \sum_{i=1}^{\infty} a_i \mathsf{E}_x\left[\sum_{k=0}^{\infty} w(k) P^i(X_k, f) \mathbb{1}_{(k<\tau)}\right] \\
&= \sum_{i=1}^{\infty}\sum_{k=0}^{\infty} a_i \mathsf{E}_x\big[w(k)\mathsf{E}[f(X_{k+i})|\mathcal{F}_k]\mathbb{1}_{(k<\tau)}\big] \\
&= \sum_{i=1}^{\infty}\sum_{k=0}^{\infty} a_i \mathsf{E}_x\big[\mathsf{E}[w(k) f(X_{k+i})\mathbb{1}_{(k<\tau)}|\mathcal{F}_k]\big] \\
&= \sum_{i=1}^{\infty}\sum_{k=0}^{\infty} a_i \mathsf{E}_x[w(k) f(X_{k+i})\mathbb{1}_{(k<\tau)}] \\
&= \sum_{i=1}^{\infty} a_i \mathsf{E}_x\left[\sum_{k=0}^{\tau-1} w(k) f(X_{k+i})\right]. \qquad \square
\end{aligned}$$

Following the terminology of Meyn and Tweedie (1993), we will write that an event A occurs P_*-a.s. if $\mathsf{P}_x(A^c) = 0$ for all $x \in E$.

THEOREM 3.2. *Suppose P is ψ-irreducible and aperiodic. Assume there exists a nonnegative function V_0: $E \to [0,\infty]$ finite for at least one $x_0 \in E$. Further, assume there exists an integer m such that, for each $i = 1, \dots, m$, there exist functions V_i: $E \to [1,\infty)$, constants $0 < c_i, b_i < \infty$ and petite sets C_i such that*

$$PV_{i-1} \le V_{i-1} - c_i V_i + b_i \mathbb{1}_{C_i}, \qquad i = 1, \dots, m. \tag{10}$$

Then, for each $i = 1, \dots, m$ and each $B \in \mathcal{E}^+$, there exists $c_i(B) < \infty$ such that

$$\mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} (k+1)^{i-1} V_i(X_k)\right] \le c_i(B)\big(V_0(x) + 1\big). \tag{11}$$

Hence, any set A on which V_0 is bounded is $(V_i, (n+1)^{i-1})$-regular, $i = 1, \dots, m$.

Further, the set $\{V_0 < \infty\}$ is absorbing and full, and

$$(n+1)^{i-1} \| P^n(x,\cdot) - \pi \|_{V_i} \to 0, \qquad n \to \infty, \tag{12}$$

for all $x \in \{V_0 < \infty\}$ and all $i = 1, \dots, m$.

PROOF. We show (11) by induction in i. For $i = 1$ we have

$$PV_0 \le V_0 - c_1 V_1 + b_1 \mathbb{1}_{C_1}, \tag{13}$$

and in this case (11) follows from (i) of Theorem 14.2.3 of Meyn and Tweedie (1993).

Now assume that (11) holds for i ($<m$) and show it for $i+1$. The induction hypothesis implies that, for any $B \in \mathcal{E}^+$,

$$\mathsf{E}_x[\tau_B^i] \le i \mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} (k+1)^{i-1} V_i(X_k)\right] \le i c_i(B)\big(V_0(x) + 1\big), \tag{14}$$

where we have used that $V_i \ge 1$ and the bound

$$\sum_{k=0}^{\tau_B - 1} (k+1)^{i-1} \ge \int_0^{\tau_B} x^{i-1}\, dx = i^{-1} \tau_B^i.$$

From Propositions 5.5.5 and 5.5.6 of Meyn and Tweedie (1993), we can assume without loss of generality that C_{i+1} is ν-petite, where ν is equivalent to ψ, and that the sampling distribution a has finite mean, $m_a = \sum_{j=1}^\infty j a_j < \infty$. From the definition of petiteness (4), we get the bound

$$\mathbb{1}_{C_{i+1}}(x) \le \nu(B)^{-1} K_a(x, B), \qquad x \in E,\ B \in \mathcal{E}^+. \tag{15}$$

Define $Z_k = k^i V_i(X_k)$. By using (10) for $i+1$, we get

$$\begin{aligned}
\mathsf{E}[Z_{k+1}|\mathcal{F}_k] &= (k+1)^i \mathsf{E}\big[V_i(X_{k+1})|\mathcal{F}_k\big] \\
&\le (k+1)^i \big(V_i(X_k) - c_{i+1}V_{i+1}(X_k) + b_{i+1}\mathbb{1}_{C_{i+1}}(X_k)\big) \\
&= Z_k + \sum_{j=0}^{i-1} \binom{i}{j} k^j V_i(X_k) - (k+1)^i c_{i+1} V_{i+1}(X_k) \\
&\quad + (k+1)^i b_{i+1}\mathbb{1}_{C_{i+1}}(X_k) \\
&\le Z_k - (k+1)^i c_{i+1} V_{i+1}(X_k) \\
&\quad + c\big(k^{i-1}V_i(X_k) + (k+1)^i \mathbb{1}_{C_{i+1}}(X_k)\big), \quad \mathsf{P}_*\text{-a.s.}
\end{aligned}$$

for some constant c independent of k. Proposition 11.3.2 of Meyn and Tweedie (1993) then gives that

$$\begin{aligned}
&\mathsf{E}_x\left[\sum_{k=0}^{\tau_B-1}(k+1)^i V_{i+1}(X_k)\right] \\
&\qquad \le \tilde{c}\mathsf{E}_x\left[\sum_{k=0}^{\tau_B-1} k^{i-1}V_i(X_k)\right] + \tilde{c}\mathsf{E}_x\left[\sum_{k=0}^{\tau_B-1}(k+1)^i \mathbb{1}_{C_{i+1}}(X_k)\right]
\end{aligned}$$

for any set $B \in \mathcal{E}^+$ and any state x, where $\tilde{c} = c/c_{i+1}$. The first expectation on the right-hand side is bounded by $c_i(B)(V_0(x)+1)$ by the induction hypothesis. For the second expectation we get from (15) and Lemma 3.1, with $w(k) = (k+1)^i$,

$$\begin{aligned}
\mathsf{E}_x\left[\sum_{k=0}^{\tau_B-1}(k+1)^i \mathbb{1}_{C_{i+1}}(X_k)\right] &\le \frac{1}{\nu(B)}\mathsf{E}_x\left[\sum_{k=0}^{\tau_B-1}(k+1)^i K_a(X_k, B)\right] \\
&= \frac{1}{\nu(B)}\sum_{j=1}^{\infty} a_j \mathsf{E}_x\left[\sum_{k=0}^{\tau_B-1}(k+1)^i \mathbb{1}_B(X_{k+j})\right] \\
&= \frac{1}{\nu(B)}\sum_{j=1}^{\infty} a_j \mathsf{E}_x\left[\sum_{k=j}^{\tau_B-1+j}(k-j+1)^i \mathbb{1}_B(X_k)\right] \\
&= \frac{1}{\nu(B)}\sum_{j=1}^{\infty} a_j \mathsf{E}_x\left[\sum_{k=\tau_B\vee j}^{\tau_B-1+j}(k-j+1)^i \mathbb{1}_B(X_k)\right] \\
&\le \frac{1}{\nu(B)}\sum_{j=1}^{\infty} a_j j \mathsf{E}_x[\tau_B^i] = \frac{m_a}{\nu(B)}\mathsf{E}_x[\tau_B^i].
\end{aligned}$$

And (11) now follows by applying the bound (14) to this expression.

Equation (13) shows that the set $\{V_0 < \infty\}$ is absorbing, and since $V_0(x_0) < \infty$ the set is nonempty, and hence it is full by Proposition 4.2.3 of Meyn and Tweedie (1993). Therefore, there exists a constant c such that $\{V_0 \le c\} \in \mathcal{E}^+$, and this set is $(V_i, (n+1)^{i-1})$-regular, $i = 1, \dots, m$, from (11). Condition (iii) of Theorem 2.1 is then satisfied and (12) follows from (8) of Theorem 2.1, since every state $x \in \{V_0 < \infty\}$ is $(V_i, (n+1)^{i-1})$-regular, $i = 1, \dots, m$. $\square$

Note that since we allow an arbitrary constant in front of V_i it is enough to assume that the test functions are all bounded away from 0. Also note that, opposed to Theorem 2.1, V_0 need not be bounded on C in Theorem 3.2. On the other hand, we do not gain any generality by allowing the constants b_i and the petite sets C_i to depend on i; the same set of drift equations will be satisfied with the constant $b = \max\{b_1, \dots, b_m\}$ and the set $C = \bigcup_{i=1}^m C_i$, which is petite by Proposition 5.5.5 of Meyn and Tweedie (1993). However, the theorem is more convenient to use in the following in the version stated here. It also stresses the fact that the different constants and petite sets are unrelated; the drift equations are only tied together via the test functions.

The drift conditions in (10) also have the following function space interpretation which was provided by a helpful referee. Assume $1 \le V_0 < \infty$ and $\pi(V_0) < \infty$. Let $L_i = L_\infty^{V_i}$ be the Banach space of functions on E which are bounded by a constant times V_i, equipped with the norm $\|g\|_i = \sup |g(x)|/V_i(x)$. For a function g let $\bar{g} = g - \pi(g)$. The *fundamental kernel* $Z = (I - P + \Pi)^{-1}$, where Π is the kernel given by $\Pi(x, \cdot) \equiv \pi(\cdot)$, is a bounded linear operator from L_i to L_{i-1}, and for each $g \in L_i$ the function $\hat{g} = Z\bar{g}$ solves the Poisson equation, $\bar{g} = \hat{g} - P\hat{g}$; see Glynn and Meyn (1996). From Meyn and Tweedie [(1993), page 433], it follows that Z has the representation $Z = U_1$, where

$$U_\alpha = \sum_{n=0}^{\infty} \alpha^n (P - \Pi)^n = \big[I - \alpha(P - \Pi)\big]^{-1}.$$

Differentiating this expression with respect to α and evaluating at $\alpha = 1$ give $U_1' = \sum_{n=1}^{\infty} n(P - \Pi)^n = Z(P - \Pi)Z$, which is a bounded linear operator from L_i to L_{i-2}. In particular, we have for $g \in L_2$ that $U_1' g \in L_0$, or

$$\left| \sum_{n=1}^{\infty} n\big(P^n g - \pi(g)\big) \right| \le \text{const } V_0. \tag{16}$$

By continued differentiation it follows that any function g which is bounded by a constant times V_i converges in the sense of (16) at a polynomial rate of order $i - 1$. This is a conclusion similar to the $(V_i, (n+1)^{i-1})$-regularity established in Theorem 3.2 but based on a different interpretation of the drift equations.

3.1. *Hitting time solutions to the drift conditions.* The existence of solutions to the set of drift equations (10) is closely related to the existence of moments of the hitting times τ_B. From (11) with $i = m$ it follows that $\mathsf{E}_x[\tau_B^m]$ is finite when $V_0(x)$ is finite. Theorem 3.4 gives the reverse implication. It states that if there exists a (petite) set C such that $\mathsf{E}_x[\tau_C^m]$ is bounded on C then there is a solution to (10) with $V_i(x)$ equivalent to $\mathsf{E}_x[\tau_C^{m-i}]$.

The following lemma which is interesting in its own right demonstrates how the rate can be interchanged with the moment of τ_C.

LEMMA 3.3. *For any set C and any nonnegative exponents m and n, there exists a positive constant d such that, for all x,*

$$d\mathsf{E}_x[\tau_C^{m+n+1}] \le \mathsf{E}_x\left[\sum_{k=0}^{\tau_C-1}(k+1)^m \mathsf{E}_{X_k}[\tau_C^n]\right] \le \mathsf{E}_x[\tau_C^{m+n+1}]. \tag{17}$$

PROOF. By Fubini's theorem and the Markov property,

$$\begin{aligned}
&\mathsf{E}_x\left[\sum_{k=0}^{\tau_C-1}(k+1)^m \mathsf{E}_{X_k}[\tau_C^n]\right] \\
&\quad = \sum_{k=0}^{\infty} \mathsf{E}_x\left[\mathbb{1}_{(k<\tau_C)}(k+1)^m \sum_{h=1}^{\infty} h^n \mathsf{P}_{X_k}(\tau_C = h)\right] \\
&\quad = \sum_{k=0}^{\infty}\sum_{h=1}^{\infty} \mathsf{E}_x\big[\mathbb{1}_{(k<\tau_C)}(k+1)^m h^n \mathsf{E}[\mathbb{1}_{(X_{k+1}\notin C,\dots,X_{k+h-1}\notin C,\, X_{k+h}\in C)}|\mathcal{F}_k]\big] \\
&\quad = \sum_{k=0}^{\infty}\sum_{h=1}^{\infty} \mathsf{E}_x\big[(k+1)^m h^n \mathsf{E}[\mathbb{1}_{(\tau_C=k+h)}|\mathcal{F}_k]\big] \\
&\quad = \mathsf{E}_x\left[\sum_{k=0}^{\infty}\sum_{h=1}^{\infty}(k+1)^m h^n \mathbb{1}_{(\tau_C=k+h)}\right] = \mathsf{E}_x\left[\sum_{k=0}^{\tau_C-1}(k+1)^m(\tau_C-k)^n\right].
\end{aligned}$$

Applying Lemma A.1 to the integrand in the last expectation gives (17). □

For any two (substochastic) kernels R and S on $(E, \mathcal{E})$, we define the composite kernel RS by

$$(RS)(x, A) = \int R(x, dy)S(y, A), \qquad x \in E,\ A \in \mathcal{E}, \tag{18}$$

and for any set $B \in \mathcal{E}$ we define the n-step taboo probability by

$${}_BP^n(x, A) = \mathsf{P}_x(X_n \in A,\ \tau_B \ge n), \qquad x \in E,\ A \in \mathcal{E}. \tag{19}$$

If we let I_B denote the kernel given by $I_B(x, A) = \mathbb{1}_{A\cap B}(x)$, then the taboo probability can be written as

$$ {}_B P^n(x, A) = [(P I_{B^c})^{n-1} P](x, A). \tag{20} $$

Instead of I_E we write I for the kernel given by $I(x, A) = \mathbb{1}_A(x)$. For any set $B \in \mathcal{E}$ we define the random variable $\sigma_B = \min\{n \geq 0 \mid X_n \in B\}$. This stopping time is 0 if $X_0 \in B$ and equal to τ_B otherwise.

THEOREM 3.4. *Suppose P is ψ-irreducible and aperiodic. Assume there exist a petite set C and an integer $m \geq 1$ such that*

$$ \sup_{x\in C} \mathsf{E}_x[\tau_C^m] < \infty. \tag{21} $$

Then (10) *is satisfied on the absorbing set $\{x : \mathsf{E}_x[\tau_C^m] < \infty\}$ with*

$$ V_i(x) = \mathsf{E}_x\left[\sum_{k=0}^{\sigma_C} \mathsf{E}_{X_k}[\tau_C^{m-1-i}]\right] \qquad \textit{for } i = 0, \ldots, m-1 \textit{ and } V_m(x) \equiv 1. \tag{22} $$

Further, there exist positive constants d_i and D_i such that

$$ d_i \mathsf{E}_x[\tau_C^{m-i}] \leq V_i(x) \leq D_i \mathsf{E}_x[\tau_C^{m-i}], \qquad i = 0, \ldots, m. \tag{23} $$

PROOF. Define the two kernels U_C and G_C by

$$ U_C = \sum_{k=0}^{\infty} (P I_{C^c})^k P, \qquad G_C = I + I_{C^c} U_C. $$

Then, for any starting state x and any nonnegative function f,

$$ U_C(x, f) = \sum_{k=1}^{\infty} {}_C P^k(x, f) = \sum_{k=1}^{\infty} \mathsf{E}_x[\mathbb{1}_{(k\leq\tau_C)} f(X_k)] = \mathsf{E}_x\left[\sum_{k=1}^{\tau_C} f(X_k)\right], $$

$$ G_C(x, f) = [I + I_{C^c} U_C](x, f) = \mathsf{E}_x\left[\sum_{k=0}^{\sigma_C} f(X_k)\right]. $$

The two kernels satisfy the relationship

$$ P G_C = P + P I_{C^c} U_C = U_C = G_C - I + I_C U_C. \tag{24} $$

Let $f_i(x) = \mathsf{E}_x[\tau_C^{m-1-i}]$ for $i = 0, \ldots, m-1$. Then $V_i(x) = G_C(x, f_i)$ and we can use (24) to get, for $i = 1, \ldots, m$,

$$ \begin{aligned} P V_{i-1}(x) &= V_{i-1}(x) - f_{i-1}(x) + \mathbb{1}_C(x) U_C(x, f_{i-1}) \\ &= V_{i-1}(x) - \mathsf{E}_x[\tau_C^{m-i}] + \mathbb{1}_C(x) \mathsf{E}_x\left[\sum_{k=1}^{\tau_C} \mathsf{E}_{X_k}[\tau_C^{m-i}]\right]. \end{aligned} $$

By Lemma 3.3 and assumption (21),

$$\sup_{x\in C} \mathsf{E}_x\left[\sum_{k=1}^{\tau_C} \mathsf{E}_{X_k}[\tau_C^{m-i}]\right] \le \sup_{x\in C} \mathsf{E}_x\left[\sum_{k=1}^{\tau_C} \mathsf{E}_{X_k}[\tau_C^{m-1}]\right] \le 2 \sup_{x\in C} \mathsf{E}_x[\tau_C^m] < \infty.$$

Thus there exists a constant b such that

$$PV_{i-1}(x) \le V_{i-1}(x) - \mathsf{E}_x[\tau_C^{m-i}] + b\mathbb{1}_C(x), \qquad i = 1, \dots, m. \tag{25}$$

We next show the bound (23). For $i = m$ it is trivially satisfied. For $i < m$ we have

$$V_i(x) = G_C(x, f_i) = \begin{cases} \mathsf{E}_x[\tau_C^{m-1-i}], & \text{for } x \in C, \\ \mathsf{E}_x\left[\displaystyle\sum_{k=0}^{\tau_C} \mathsf{E}_{X_k}[\tau_C^{m-1-i}]\right], & \text{for } x \in C^c, \end{cases}$$

and (23) follows from Lemma 3.3, since $\mathsf{E}_x[\tau_C^{m-i}]$ is bounded on C for all i by assumption.

From (25) and the upper bound of (23), we get

$$PV_{i-1}(x) \le V_{i-1}(x) - c_i V_i + b\mathbb{1}_C(x), \qquad i = 1, \dots, m, \tag{26}$$

which shows that (10) is satisfied on the set $\{x : \mathsf{E}_x[\tau_C^m] < \infty\}$. Since $V_0(x)$ and $\mathsf{E}_x[\tau_C^m]$ are equivalent, this set is the same as $\{V_0 < \infty\}$, and from (26) with $i = 1$ it follows that this set is absorbing, and since it is nonempty it is also full. $\square$

Using the solution (22), we get from Theorem 3.2 that

$$\mathsf{E}_x\left[\sum_{k=0}^{\tau_C - 1} (k+1)^{i-1} \mathsf{E}_{X_k}[\tau_C^{m-i}]\right] \le R\mathsf{E}_x[\tau_C^m]. \tag{27}$$

From Lemma 3.3 it follows that (with another constant) the inequality also holds the other way around. Thus Theorem 3.2 gives the correct rate of convergence when the hitting time solution is used. This also shows the connection between the degree of regularity of the chain and the existence of moments of hitting times: when the hitting times have mth moments, there is convergence of order $m - 1$ in the total variation norm.

Also note that condition (21) via Lemma 3.3 is equivalent to

$$\sup_{x\in C} \mathsf{E}_x\left[\sum_{k=0}^{\tau_C - 1} (k+1)^{i-1} \mathsf{E}_{X_k}[\tau_C^{m-i}]\right] < \infty, \tag{28}$$

which is condition (i) of Theorem 2.1. This gives an alternative route to proving $(V_i, (n+1)^{i-1})$-regularity.

3.2. *Reduction to a single drift condition.* The set of drift equations in Theorem 3.2 can be generated from a single drift equation by use of the following lemma.

LEMMA 3.5. *Suppose a function* $V\colon E \to [1,\infty)$, *constants* $0 < c$, $b < \infty$, *a set* C *and* $\alpha < 1$ *jointly satisfy*

$$PV \le V - cV^{\alpha} + b\mathbb{1}_C. \tag{29}$$

Then, for every $0 < \eta \le 1$, *there exist constants* $0 < c_1, b_1 < \infty$ *such that*

$$PV^{\eta} \le V^{\eta} - c_1 V^{\alpha+\eta-1} + b_1\mathbb{1}_C. \tag{30}$$

PROOF. By Jensen's inequality we get from (29)

$$PV^{\eta}(x) \le \big(PV(x)\big)^{\eta} \le \big(V(x) - cV^{\alpha}(x) + b\mathbb{1}_C(x)\big)^{\eta}. \tag{31}$$

Let $f(x) = x^{\eta}$, $x \ge 0$. Then $f(0) = 0$ and $f'(x) = \eta x^{\eta-1}$, $x > 0$, which tends to 0 from above as x tends to ∞. Hence, for $y \ge 0$ and $z \ge 0$,

$$(y+z)^{\eta} = y^{\eta} + \int_y^{y+z} f'(x)\,dx \le y^{\eta} + \int_0^z f'(x)\,dx = y^{\eta} + z^{\eta}, \tag{32}$$

and, for $y > 0$ and $0 \le z \le y$,

$$(y-z)^{\eta} = y^{\eta} - \int_{y-z}^{y} f'(x)\,dx \le y^{\eta} - zf'(y) = y^{\eta} - \eta z y^{\eta-1}. \tag{33}$$

For x not in C it follows from (29) that $1 \le PV(x) \le V(x) - cV^{\alpha}(x)$. This implies that $0 < cV^{\alpha}(x) < V(x)$ and we can use (33) to bound the right-hand side of (31)

$$\big(V(x) - cV^{\alpha}(x)\big)^{\eta} \le V^{\eta}(x) - \eta c V^{\alpha+\eta-1}(x).$$

For x in C we consider two situations. Either $V(x) - cV^{\alpha}(x) \ge 0$ in which case

$$\begin{aligned}\big(V(x) - cV^{\alpha}(x) + b\big)^{\eta} &\le \big(V(x) - cV^{\alpha}(x)\big)^{\eta} + b^{\eta}\\ &\le V^{\eta}(x) - \eta c V^{\alpha+\eta-1}(x) + b^{\eta},\end{aligned}$$

where the first inequality is (32) and the second inequality is (33). Otherwise, $V(x) - cV^{\alpha}(x) < 0$ which implies $V(x) < c^{1/1-\alpha}$. Since also $V(x) \ge 1$ we have that in this case

$$\big(V(x) - cV^{\alpha}(x) + b\big)^{\eta} \le b^{\eta} \quad \text{and} \quad V^{\alpha+\eta-1}(x) \le 1 \vee c^{(\alpha+\eta-1)/(1-\alpha)},$$

and thus there exists a positive constant b_0 such that

$$\big(V(x) - cV^{\alpha}(x) + b\big)^{\eta} \le b^{\eta} + V^{\eta}(x) - \eta c V^{\alpha+\eta-1}(x) + b_0.$$

Combining the three bounds yields

$$PV^{\eta}(x) \le V^{\eta}(x) - c_1 V^{\alpha+\eta-1}(x) + b_1\mathbb{1}_C(x),$$

with $c_1 = \eta c$ and $b_1 = b^{\eta} + b_0$. $\square$

Lemma 3.5 can be used either inductively or directly with the same result, in the sense that if we apply Lemma 3.5 to

$$PV^{\eta_1} \le V^{\eta_1} - c_1(V^{\eta_1})^{(\alpha+\eta_1-1)/\eta_1} + b_1\mathbb{1}_C,$$

then we get

$$PV^{\eta_1\eta_2} \le V^{\eta_1\eta_2} - c_2V^{\alpha+\eta_1\eta_2-1} + b_2\mathbb{1}_C,$$

which is the same as applying Lemma 3.5 with $\eta = \eta_1\eta_2$.

THEOREM 3.6. *Suppose P is ψ-irreducible and aperiodic. Assume there exist a function $V\colon E \to [1,\infty)$, constants $0 < c$, $b < \infty$, a petite set C and $0 \le \alpha < 1$ such that*

$$PV \le V - cV^\alpha + b\mathbb{1}_C. \tag{34}$$

Then P is (V_β, r_β)-regular for each $1 \le \beta \le 1/(1-\alpha)$, where

$$V_\beta(x) = V^{1-\beta(1-\alpha)}(x), \qquad r_\beta(n) = (n+1)^{\beta-1}. \tag{35}$$

In particular, the following polynomial convergence statements hold for all x:

$$(n+1)^{\beta-1}\|P^n(x,\cdot) - \pi\|_{V_\beta} \to 0, \qquad n \to \infty. \tag{36}$$

PROOF. We will show that, for each $1 \le \beta \le 1/(1-\alpha)$ and each $B \in \mathcal{E}^+$, there exists $c_\beta(B) < \infty$ such that

$$\mathsf{E}_x\left[\sum_{k=0}^{\tau_B-1} r_\beta(k)V_\beta(X_k)\right] \le c_\beta(B)\bigl(V(x)+1\bigr). \tag{37}$$

From this it follows that any set on which V is bounded is (V_β, r_β)-regular, and since V is everywhere finite (iii) of Theorem 2.1 is satisfied with a sufficiently large level set $\{V \le \alpha\}$ and it follows that P is (V_β, r_β)-regular and that (36) holds for all x.

We will first show (37) for integer values of β. Let $\gamma = 1-\alpha$, $m = \lceil \gamma^{-1} \rceil$ and $V_i = V^{1-i\gamma}$ for $i = 0,\dots,m$. Then $V_0 = V$ and $V_i \ge 1$ for $i = 1,\dots,m-1$, since $1 - i\gamma > 0$ for $i < m$.

From the proof of Theorem 3.2 it can be seen that it is enough to assume that $V_i \ge 1$ for $i < m$ to reach the conclusion (11). The assumption $V_i \ge 1$ is only used to get the bound in (14) and this bound is only needed for $i < m$. Thus, if we can show

$$PV_{i-1} \le V_{i-1} - c_iV_i + b_i\mathbb{1}_C, \qquad i = 1,\dots,m, \tag{38}$$

for some positive constants c_i, b_i, then we can conclude from Theorem 3.2 that

$$\mathsf{E}_x\left[\sum_{k=0}^{\tau_B-1}(k+1)^{i-1}V_i(X_k)\right] \le c_i(B)\bigl(V(x)+1\bigr), \qquad i = 1,\dots,m. \tag{39}$$

But for $i = 1$ (38) is the same as (34), and for $i = 2, \dots, m$ (38) follows from Lemma 3.5 with $\eta = 1 - (i-1)\gamma > 0$.

For $i < \beta < i + 1$ we can use Lemma A.2 with $a_1 = i - 1$, $a_2 = i$, $b_1 = 1 - (i+1)\gamma$ and $b_2 = 1 - i\gamma$. Then $c = 1/\gamma$ and since $a = \beta - 1$ and $b = 1 - \beta\gamma$ satisfy

$$a + bc = \beta - 1 + \frac{1 - \beta\gamma}{\gamma} = -1 + \frac{1}{\gamma} = i - 1 + \frac{1 - i\gamma}{\gamma} = a_1 + b_2 c,$$

we can use (72) and (39) to get

$$\begin{aligned}
&\mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} r_\beta(k) V_\beta(X_k)\right] \\
&\quad = \mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} (k+1)^a \big(V(X_k)\big)^b\right] \\
&\quad \le \mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} \big((k+1)^{a_1} \big(V(X_k)\big)^{b_2} \vee (k+1)^{a_2} \big(V(X_k)\big)^{b_1}\big)\right] \\
&\quad \le \mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} (k+1)^{a_1} \big(V(X_k)\big)^{b_2}\right] + \mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} (k+1)^{a_2} \big(V(X_k)\big)^{b_1}\right] \\
&\quad = \mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} (k+1)^{i-1} V_i(X_k)\right] + \mathsf{E}_x\left[\sum_{k=0}^{\tau_B - 1} (k+1)^{i} V_{i+1}(X_k)\right] \\
&\quad \le \big(c_i(B) + c_{i+1}(B)\big)\big(V(x) + 1\big),
\end{aligned}$$

and we are finished. $\square$

Theorem 3.6 demonstrates directly the trade-off between moments and convergence rates in the polynomial case. The gap $1 - \alpha$ is the power lost for each order of convergence gained. If (34) holds for $\alpha = 1$ we get the drift condition for geometric ergodicity

$$PV \le \lambda V + b\mathbb{1}_C, \tag{40}$$

where $\lambda < 1$, $V \ge 1$, $b < \infty$ and C is a petite set, which implies a geometric rate of convergence in (36), that is,

$$\rho^n \| P^n(x, \cdot) - \pi \|_V \to 0, \qquad n \to \infty, \tag{41}$$

for some $\rho > 1$; see Meyn and Tweedie (1992, 1993). Thus $\alpha = 1$ represents a qualitative change in the ergodic behavior of the chain.

4. Central limit theorems. In this section we consider central limit theorems for the ergodic average

$$S_n(g) = \frac{1}{n}\sum_{i=1}^{n} g(X_i), \qquad n \ge 1, \tag{42}$$

of a function g on the state space. The central limit theorem (CLT) is said to hold if there exists $0 < \sigma_g^2 < \infty$ such that, for every starting state $X_0 = x$,

$$\sqrt{n}\big(S_n(g) - \pi(g)\big) \xrightarrow{w} N(0, \sigma_g^2), \qquad n \to \infty. \tag{43}$$

When (43) is satisfied for a function g and $\sigma_g^2 > 0$, we say that the CLT holds for g.

Chapter 17 of Meyn and Tweedie (1993) gives criteria for the CLT to hold in the presence of a drift condition of the form

$$PV \le V - f + b\mathbb{1}_C. \tag{44}$$

When C is a petite set, $f \ge 1$ and V is nonnegative and finite, (44) implies that P is positive Harris recurrent. For a function $g \in L_1(\pi)$ let $\bar{g} = g - \pi(g)$.

THEOREM 4.1. *Suppose P is ψ-irreducible and aperiodic. Assume there exist a function $V\colon\ E \to [0, \infty)$, a function $f\colon\ E \to [1, \infty)$, a petite set C and a constant b satisfying* (44). *Assume $\pi(V^2) < \infty$. Then, for any function g with $|g| \le f$, the constant*

$$\sigma_g^2 = \lim_{n\to\infty} n\mathsf{E}_\pi\big[\big(S_n(\bar{g})\big)^2\big] = \mathsf{E}_\pi[\bar{g}^2(X_0)] + 2\sum_{k=1}^{\infty}\mathsf{E}_\pi[\bar{g}(X_0)\bar{g}(X_k)] \tag{45}$$

is well defined, nonnegative and finite. If $\sigma_g^2 > 0$ the CLT holds for g.

PROOF. Lemma 17.5.2 and Theorem 17.5.3 of Meyn and Tweedie (1993). □

Under the same conditions the conclusion of Theorem 4.1 can be strengthened to give weak convergence of the partial sum process of $\bar{g}(X_k)$ to Brownian motion; see Theorem 17.4.4 of Meyn and Tweedie (1993) and Theorem 4.1 of Glynn and Meyn (1996). Theorem 4.1 directly translates to the situation where the drift condition (34) is satisfied.

THEOREM 4.2. *Suppose P is ψ-irreducible and aperiodic. Assume there exist a function $V\colon\ E \to [1, \infty)$, constants $0 < c$, $b < \infty$, a petite set C and $0 \le \alpha < 1$ such that*

$$PV \le V - cV^\alpha + b\mathbb{1}_C. \tag{46}$$

Assume there exists $1 - \alpha \le \eta \le 1$ with $\pi(V^{2\eta}) < \infty$. Then, for any function g with

$$|g| \le V^{\alpha+\eta-1}, \tag{47}$$

the constant σ_g^2 in (45) *is well defined and finite. If $\sigma_g^2 > 0$ the CLT holds for g.*

PROOF. Apply Theorem 4.1 with $f = V^{\alpha+\eta-1}$ to (30) of Lemma 3.5. $\square$

In particular, we get a CLT for all bounded functions if $\pi(V^{2(1-\alpha)}) < \infty$. Since $\pi(V^\alpha) < \infty$ by (46) this gives a CLT for all bounded functions when $\alpha \geq 2/3$.

However, as pointed out to us by Eric Moulines, it is possible to improve this since the assumption $\pi(V^2) < \infty$ of Theorem 4.1 is in fact too strong. For a positive Harris chain satisfying (44), it is sufficient that $\pi(Vf) < \infty$ for the CLT to hold for functions $|g| \leq f$. The following results also cover the case where $\alpha + \eta - 1 < 0$, that is, when $f = V^{\alpha+\eta-1}$ is not bounded away from 0. For ease of exposition we will only state the results in the case where an atom A exists. A similar result holds in the general case via the split chain; see Nummelin (1984) and Meyn and Tweedie (1993). We will write E_A for the common expectation E_x for $x \in A$.

THEOREM 4.3. *Suppose P is ψ-irreducible, aperiodic and positive Harris recurrent. Assume there exists an atom $A \in \mathcal{E}^+$. If, for a function g,*

$$\mathsf{E}_A\left[\left(\sum_{k=1}^{\tau_A} |\bar{g}(X_k)|\right)^2\right] < \infty, \tag{48}$$

then

$$\sigma_g^2 = \pi(A)\mathsf{E}_A\left[\left(\sum_{k=1}^{\tau_A} \bar{g}(X_k)\right)^2\right] \tag{49}$$

is well defined, nonnegative and finite, and if $\sigma_g^2 > 0$ then the CLT holds for g.

PROOF. This is Theorem 17.2.2 of Meyn and Tweedie (1993) with weakened conditions. They assumed that (48) holds for g, instead of $\bar{g}$, and also that $\mathsf{E}_A \tau_A^2 < \infty$. However, by inspection of the proof, it can be seen that only (48) and $\mathsf{E}_A \tau_A < \infty$ are needed, and the latter is automatically satisfied since $\mathsf{E}_A \tau_A = \pi(A)^{-1}$ by Kac's theorem.

When the state space is countable (and ψ is the counting measure), this is Theorem I.16.1 of Chung (1679). $\square$

Using this result, we can improve Theorem 4.2 in the case where an atom exists.

THEOREM 4.4. *Suppose P is ψ-irreducible and aperiodic. Assume there exist a function $V: E \to [1, \infty)$, constants $0 < c$, $b < \infty$, an atom A and $0 \leq \alpha < 1$ such that*

$$PV \leq V - cV^\alpha + b\mathbb{1}_A. \tag{50}$$

Assume there exists $0 < \eta \leq 1$ *with* $\pi(V^{\alpha+2\eta-1}) < \infty$ *and* $\pi(\mathbb{1}_A V^{\eta}) < \infty$. *Then, for any function* g *with*

$$|\bar{g}| \leq V^{\alpha+\eta-1} + \mathbb{1}_A, \tag{51}$$

the constant σ_g^2 *in* (49) *is well defined and finite. If* $\sigma_g^2 > 0$ *the CLT holds for* g.

PROOF. Condition (50) implies that P is positive Harris recurrent and that $A \in \mathcal{E}^+$. Applying Lemma 3.5 to (50) gives

$$PW \leq W - f + b_1 \mathbb{1}_A, \tag{52}$$

where $W = V^{\eta}$ and $f = c_1(V^{\alpha+\eta-1} + \mathbb{1}_A)$ for some positive constant c_1. By assumption $\pi(Wf) < \infty$ and we will show that this together with (52) implies that

$$\mathsf{E}_A\left[\left(\sum_{k=1}^{\tau_A} f(X_k)\right)^2\right] < \infty, \tag{53}$$

from which the conclusion follows via Theorem 4.3. To show (53), we define

$$\hat{f}(x) = \mathsf{E}_x\left[\sum_{k=0}^{\sigma_A} f(X_k)\right]. \tag{54}$$

[This is the function $G_A(x, f)$ in the notation used in the proof of Theorem 3.4.] From the Comparison Theorem 14.2.2 of Meyn and Tweedie (1993) and (52), we get

$$\hat{f}(x) = \mathsf{E}_x\left[\sum_{k=0}^{\sigma_A} f(X_k)\right] \leq W(x) + b_1 \mathsf{E}_x\left[\sum_{k=0}^{\sigma_A} \mathbb{1}_A(X_k)\right] = W(x) + b_1 \leq RW(x)$$

for some constant R since $W \geq 1$. Since $\pi(Wf) < \infty$ this implies that $\pi(\hat{f} f) < \infty$ and then also $\pi(f^2) < \infty$ since trivially $f \leq \hat{f}$. By the representation of π using the atom A, Theorem 10.0.1 of Meyn and Tweedie (1993), we then have

$$\mathsf{E}_A\left[\sum_{k=1}^{\tau_A} \bigl(2\hat{f}(X_k) f(X_k) - f^2(X_k)\bigr)\right] = \pi(A)^{-1} \pi(2\hat{f} f - f^2) < \infty. \tag{55}$$

Now, inserting the definition of $\hat{f}$ and using the Markov property give

$$\begin{aligned}
&\mathsf{E}_A\left[\sum_{k=1}^{\tau_A} \bigl(2\hat{f}(X_k) f(X_k) - f^2(X_k)\bigr)\right] \\
&\qquad = \mathsf{E}_A\left[\sum_{k=1}^{\tau_A} \left(2 f(X_k) \mathsf{E}_{X_k}\left[\sum_{j=0}^{\sigma_A} f(X_j)\right] - f^2(X_k)\right)\right]
\end{aligned}$$

$$= \mathsf{E}_A\left[\sum_{k=1}^{\tau_A}\left(2f(X_k)\mathsf{E}\left[\sum_{j=k}^{\tau_A} f(X_j)|\mathcal{F}_k\right] - f^2(X_k)\right)\right]$$

$$= \mathsf{E}_A\left[\sum_{k=1}^{\tau_A}\mathsf{E}\left[\sum_{j=k}^{\tau_A} 2f(X_k)f(X_j) - f^2(X_k)|\mathcal{F}_k\right]\right]$$

$$= \mathsf{E}_A\left[\sum_{k=1}^{\tau_A}\left(\sum_{j=k}^{\tau_A} 2f(X_k)f(X_j) - f^2(X_k)\right)\right] = \mathsf{E}_A\left[\left(\sum_{k=1}^{\tau_A} f(X_k)\right)^2\right].$$

Hence (53) holds and we are finished. $\square$

Since $\pi(V^\alpha) < \infty$ by (50) the first condition, $\pi(V^{\alpha+2\eta-1}) < \infty$, always holds for $\eta \le 1/2$. For the same reason the second condition, $\pi(\mathbb{1}_A V^\eta) < \infty$, always holds for $\eta \le \alpha$. [Of course, if V is bounded on A, then $\pi(\mathbb{1}_A V^\eta) < \infty$ for any η.] Hence, for $\alpha > 0$, it is always possible to use Theorem 4.4 with $\eta = \alpha \wedge 1/2$. If $\alpha \ge 1/2$, then $\alpha + \eta - 1 \ge 0$ with this choice of η and, in particular, we get a CLT for all bounded functions in this case. If $\alpha < 1/2$, then $\alpha + \eta - 1$ is negative with this choice of η and Theorem 4.4 only gives us a CLT for functions g which are bounded on A and with deviations from their mean bounded by $V^{\alpha+\eta-1}$ outside of A. This is a rather narrow class of functions, which does not include the bounded functions in general.

That $\alpha \ge 1/2$ implies the existence of CLTs for bounded functions also follows from Theorem 4.3 since in this case $\mathsf{E}_A[\tau_A^2] < \infty$ by (50). More generally, we can use Theorem 3.4 to derive CLTs in terms of finiteness of moments of return times.

THEOREM 4.5. *Suppose P is ψ-irreducible, aperiodic and positive Harris recurrent. Assume there exist an atom A and an integer $m \ge 1$ such that $\mathsf{E}_x[\tau_A^m] < \infty$ for all x and*

$$\mathsf{E}_A[\tau_A^{2m}] < \infty. \tag{56}$$

Then, for any function g with $|g| \le \mathsf{E}_x[\tau_A^{m-1}]$, the constant σ_g^2 in (49) is well defined and finite. If $\sigma_g^2 > 0$ the CLT holds for g.

PROOF. By Theorem 3.4 there exist functions W and f equivalent to $\mathsf{E}_x[\tau_A^m]$ and $\mathsf{E}_x[\tau_A^{m-1}]$, respectively, such that (52) holds, and the conclusion then follows as in the proof of Theorem 4.4 if we can show that $\pi(Wf) < \infty$.

By two applications of Jensen's inequality,

$$\mathsf{E}_x[\tau_A^{m-1}]\mathsf{E}_x[\tau_A^m] \le \mathsf{E}_x[\tau_A^m]^{(m-1)/m}\mathsf{E}_x[\tau_A^m] = \mathsf{E}_x[\tau_A^m]^{(2m-1)/m} \le \mathsf{E}_x[\tau_A^{2m-1}],$$

and it is thus enough to show that $\pi(\mathsf{E}_x[\tau_A^{2m-1}]) < \infty$. However, by assumption (56), Lemma 3.3 and Theorem 10.4.9 of Meyn and Tweedie (1993), we have

$$\infty > \mathsf{E}_A[\tau_A^{2m}] \ge \mathsf{E}_A\left[\sum_{k=0}^{\tau_A-1}\mathsf{E}_{X_k}[\tau_A^{2m-1}]\right] = \pi(A)^{-1}\pi\big(\mathsf{E}_x[\tau_A^{2m-1}]\big),$$

and we are finished. $\square$

5. Examples. In this section we illustrate the applicability of the drift condition (34) by two examples.

EXAMPLE 5.1. We first consider Example 5.1 of Tuominen and Tweedie (1994) to show that we get the same rate of convergence using the drift condition (34) as they do using the system of drift conditions (7).

Let P be the Markov transition kernel for the random walk on $[0, \infty)$ given by

$$X_{n+1} = (X_n + W_{n+1})^+, \qquad n \in \mathbb{N}_0, \tag{57}$$

where (W_n) is a sequence of i.i.d. real-valued random variables with common law Γ. If $\mathsf{E}[W] < 0$, then this chain is δ_0-irreducible, aperiodic and positive and all compact sets are petite.

PROPOSITION 5.1. *Assume that* $\mathsf{E}[W] < 0$ *and that there exists an integer* $m \geq 2$ *such that*

$$\mathsf{E}\big[(W^+)^m\big] < \infty. \tag{58}$$

Then there exist a finite interval $C = [0, z_0]$ *and constants* $0 < c, b < \infty$ *such that*

$$PV \leq V - cV^\alpha + b\mathbb{1}_C, \tag{59}$$

where $V(x) = (x+1)^m$ *and* $\alpha = (m-1)/m$.

Hence, P *is* (V_i, r_i)*-regular, where*

$$V_i(x) = (x+1)^{m-i}, \qquad r_i(n) = (n+1)^{i-1}, \qquad i = 1, \dots, m. \tag{60}$$

PROOF. Take $x_0 > 0$ so large that $\int_{[-x_0,\infty)} y\Gamma(dy) < 0$.

For $x > x_0$ we bound $PV(x)$ by considering jumps smaller than $-x_0$ and jumps larger than $-x_0$ separately,

$$PV(x) \leq V(x - x_0)\Gamma\big((-\infty, -x_0)\big) + \int_{[-x_0,\infty)} V(x+y)\Gamma(dy). \tag{61}$$

First, we bound $V(x - x_0)$ in terms of $V(x)$ and $(V(x))^\alpha$,

$$\begin{aligned} V(x) - V(x - x_0) &= \int_{x-x_0}^{x} m(y+1)^{m-1}\,dy \geq x_0 m(x - x_0 + 1)^{m-1} \\ &= x_0 m\left(\frac{x - x_0 + 1}{x+1}\right)^{m-1}(x+1)^{m-1} \geq c_1(x+1)^{m-1}, \end{aligned}$$

where $c_1 = x_0 m(1/(x_0+1))^{m-1}$.

To bound the second term in (61), we first note that, for $x \geq 0$ and $y \geq 0$,

$$(x+y+1)^{m-2} \leq (x+1)^{m-2}(y+1)^{m-2}, \tag{62}$$

since

$$\log(x+y+1) - \log(x+1) = \int_{x+1}^{x+1+y} \frac{1}{z}\,dz \leq \int_{1}^{1+y} \frac{1}{z}\,dz = \log(y+1).$$

For $y > 0$ we then get

$$\begin{aligned} V(x+y) &\leq V(x) + m(x+1)^{m-1}y + \tfrac{1}{2}m(m-1)(x+y+1)^{m-2}y^2 \\ &\leq V(x) + m(x+1)^{m-1}y + \tfrac{1}{2}m(m-1)(x+1)^{m-2}(y+1)^{m}, \end{aligned}$$

and for $-x_0 \leq y \leq 0$ we get

$$V(x+y) \leq V(x) + m(x+1)^{m-1}y + \tfrac{1}{2}m(m-1)(x+1)^{m-2}x_0^2.$$

Inserting these bounds on V into (61) and using the assumption $\int_{[0,\infty)}(y+1)^m \Gamma(dy) < \infty$, we find, for $x > x_0$,

$$PV(x) \leq V(x) - c_2(x+1)^{m-1} + c_3(x+1)^{m-2}$$

for some constants $0 < c_2, c_3 < \infty$. Hence, there exist a positive constant c and a real number $z_0 \geq x_0$ such that, for $x > z_0$,

$$PV(x) \leq V(x) - c(x+1)^{m-1}.$$

Finally, since $PV(x)$ and $(x+1)^{m-1}$ are bounded on $C = [0, z_0]$, there exists a constant b such that (59) holds. The last assertion follows from Theorem 3.6. □

For comparison, we note that when $\mathsf{E}[e^{sW}] < \infty$ for some $s > 0$ the chain is geometrically ergodic and there is a solution to the drift equation (40) with test function $V(x) = e^{tx}$ for $t < s$; see Nummelin and Tuominen (1982) and Tweedie (1983).

A more sophisticated queueing application is considered in Dai and Meyn (1995) where polynomial rates of convergence of moments for multiclass queueing networks are derived using the results of Tuominen and Tweedie (1994). These results could also be derived by simpler means using the drift condition (34).

EXAMPLE 5.2. The second example is concerned with the MCMC algorithm known as the independence sampler. This is an algorithm to obtain samples from a target distribution π known only up to an unknown normalization constant; see Tierney (1994).

The independence sampler is described as follows. Suppose that π is a probability measure on E and suppose Q is another probability measure on E

such that π is absolutely continuous with respect to Q with Radon–Nikodym derivative

$$\frac{d\pi}{dQ}(x) = \frac{1}{q(x)} \tag{63}$$

for $x \in E$. Suppose the chain is currently at x. A move is proposed to y drawn from Q and accepted with probability

$$\alpha(x, y) = \frac{q(x)}{q(y)} \wedge 1. \tag{64}$$

If the proposed move is not accepted, the chain remains at x.

Clearly, this algorithm generates a Markov chain $\mathbf{X}$ on E, and it can easily be verified that the chain is ψ-irreducible and has unique stationary measure π; see, for example, Roberts and Tweedie (2001). A complete spectral analysis of the n-step transition probabilities of $\mathbf{X}$ is available [see Smith and Tierney (1996)], but such an analysis is not illuminating when q is not bounded away from 0. This corresponds to the case where the algorithm fails to be either uniformly or geometrically ergodic. Here we examine this case using the results of previous sections.

It is now well known [cf. Mengersen and Tweedie (1996)] that if π and Q both have densities denoted π and q, respectively, w.r.t. some common reference measure and if there exists $\beta > 0$ such that

$$\frac{q(x)}{\pi(x)} \geq \beta \tag{65}$$

for all x then the independence sampler is uniformly ergodic, and if (65) does not hold π-almost surely then the independence sampler is not even geometrically ergodic. However, using the results of this paper, it is possible to obtain polynomial rate results when (65) is violated.

First consider the case where π is the uniform distribution on [0, 1] and Q has density q w.r.t. Lebesgue measure on [0, 1]. The acceptance probability for accepting a proposed move from state x to state y is given by (64).

For each x define the regions of acceptance and possible rejection by

$$A(x) = \{y : q(y) \leq q(x)\}, \quad R(x) = \{y : q(y) > q(x)\}.$$

PROPOSITION 5.2. *Assume $\pi = U[0, 1]$ and that Q has density q w.r.t. Lebesgue measure on* [0, 1] *of the form*

$$q(x) = (r + 1)x^r \tag{66}$$

for some $r > 0$. Then, for each $r < s < r + 1$, the independence sampler P satisfies

$$PV \leq V - cV^\alpha + b\mathbb{1}_C, \tag{67}$$

where $V(x) = 1/x^s$, $\alpha = 1 - r/s$ and C is a petite set.

PROOF. The acceptance and rejection regions are $A(x) = [0, x]$ and $R(x) = (x, 1]$, and all sets of the form $[y, 1]$ are petite sets.

$$\begin{aligned}
PV(x) &= \int_0^x V(y)q(y)\,dy + \int_x^1 V(y)\alpha(x,y)q(y)\,dy \\
&\quad + V(x)\int_x^1 \bigl(1-\alpha(x,y)\bigr)q(y)\,dy \\
&= \int_0^x (r+1)y^{r-s}\,dy + \int_x^1 \frac{1}{y^s}(r+1)x^r\,dy + \frac{1}{x^s}\int_x^1 (r+1)(y^r - x^r)\,dy \\
&= (r+1)\left[\frac{1}{r-s+1}y^{r-s+1}\right]_0^x + (r+1)x^r\left[\frac{1}{-s+1}y^{-s+1}\right]_x^1 \\
&\quad + \frac{1}{x^s}[y^{r+1}]_x^1 - (r+1)x^{r-s}(1-x) \\
&= \frac{r+1}{r-s+1}x^{r-s+1} + \frac{r+1}{-s+1}x^r - \frac{r+1}{-s+1}x^{r-s+1} + \frac{1}{x^s} - x^{r-s+1} \\
&\quad - (r+1)x^{r-s}(1-x) \\
&= V(x) - (r+1)V(x)^{1-r/s}(1-x) + c_1x^{r-s+1} + c_2x^r.
\end{aligned}$$

Now, since $r - s + 1$ and r are both positive, x^{r-s+1} and x^r tend to 0 as x tends to 0, while $V(x)^{1-r/s} = x^{r-s}$ tends to ∞ as x tends to 0. Thus (67) is satisfied with $C = [x_0, 1]$ for x_0 sufficiently small and some constants b and c. $\square$

The choice of s leading to the best rate of convergence is $r + 1 - \varepsilon$ which gives $\alpha \approx 1 - r/(r+1)$. Hence, the independence sampler converges in total variation at a polynomial rate of order $1/r$.

This result extends to general independence samplers. First, we note that, in the notation of Roberts and Rosenthal (2001), the chain $q(\mathbf{X})$ is itself Markov and is *partially de-initializing* for $\mathbf{X}$. This implies that we could apply Corollary 10 of Roberts and Rosenthal (2001) to study total variation convergence of $\mathbf{X}$ in terms of that of $q(\mathbf{X})$. Now $q(\mathbf{X})$ is itself an independence sampler on $\mathbb{R}^+$ with proposal $q(Q)$ and target distribution $q(\pi)$. In addition, a further monotone transformation converts this Markov chain to an independence sampler on $[0, 1]$ with stationary distribution $U[0, 1]$, and we could then apply Proposition 5.2. However, a more direct approach is possible. For simplicity, we assume that π and Q are equivalent which is no restriction.

THEOREM 5.3. *Assume π and Q are equivalent measures on E with Radon–Nikodym derivative given by* (63). *If*

$$\pi(A_\varepsilon) = O(\varepsilon^{1/r}) \qquad \textit{for } \varepsilon \to 0 \tag{68}$$

for some $r > 0$, *where* $A_\varepsilon = \{x \in E : q(x) \le \varepsilon\}$, *then for each* $r < s < r + 1$ *the independence sampler* P *satisfies*

$$PV \le V - cV^\alpha + b\mathbb{1}_C, \tag{69}$$

where $V(x) = (1/q(x))^{s/r}$, $\alpha = 1 - r/s$ *and* C *is a petite set.*

PROOF. Note that $A(x) = A_{q(x)}$ and that all the sets A_ε^c are petite. Let F denote the CDF of the distribution of q under π.

$$\begin{aligned}
PV(x) &= \int_{A_{q(x)}} V(y)q(y)\pi(dy) + \int_{A_{q(x)}^c} V(y)\alpha(x, y)q(y)\pi(dy) \\
&\quad + V(x)\int_{A_{q(x)}^c} \big(1 - \alpha(x, y)\big)q(y)\pi(dy) \\
&= \int_{A_{q(x)}} q(y)^{1-s/r}\pi(dy) + \int_{A_{q(x)}^c} q(x)q(y)^{-s/r}\pi(dy) \\
&\quad + V(x)\int_{A_{q(x)}^c} \big(q(y) - q(x)\big)\pi(dy) \\
&\le \int_{A_{q(x)}} q(y)^{1-s/r}\pi(dy) + \int_{A_{q(x)}^c} q(x)q(y)^{-s/r}\pi(dy) \\
&\quad - V(x)^\alpha \pi(A_{q(x)}^c) + V(x) \\
&= \int_{[0,q(x)]} y^{1-s/r}F(dy) + \int_{(q(x),\infty)} q(x)y^{-s/r}F(dy) \\
&\quad - V(x)^\alpha \pi(A_{q(x)}^c) + V(x),
\end{aligned} \tag{70}$$

where the inequality uses $\int_{A_{q(x)}^c} q(y)\pi(dy) = Q(A_{q(x)}^c) \le 1$. From the assumption (68) there exist positive K and y_0 such that

$$F(y) \le Ky^{1/r} \quad \text{for } y \le y_0.$$

Since $Ky^{1/r}$ is the CDF for the measure with density $K_1 y^{1/r-1}$ w.r.t. Lebesgue measure and since $y^{1-s/r}$ and $y^{-s/r}$ are decreasing functions, we get, for all $q(x) \le y_0$,

$$\begin{aligned}
\int_{[0,q(x)]} y^{1-s/r}F(dy) &\le K_1 \int_{[0,q(x)]} y^{1-s/r}y^{1/r-1}\,dy, \\
\int_{(q(x),\infty)} q(x)y^{-s/r}F(dy) &\le K_1 \int_{(q(x),y_0]} q(x)y^{-s/r}y^{1/r-1}\,dy \\
&\quad + \int_{(y_0,\infty)} q(x)y^{-s/r}F(dy).
\end{aligned}$$

From this it follows that the two integrals in (70) both tend to 0 as $q(x)$ tends to 0. Since $V(x)^\alpha$ tends to ∞ and $\pi(A_{q(x)}^c)$ tends to 1 as $q(x)$ tends to 0, (69) is satisfied with $C = A_\varepsilon^c$ for ε sufficiently small. $\square$

APPENDIX

This appendix contains two algebraic lemmas used in the proofs of Lemma 3.3 and Theorem 3.6, respectively.

LEMMA A.1. *For any nonnegative exponents m and n, there exists a positive constant d such that, for any integer $T \geq 1$,*

$$dT^{m+n+1} \leq \sum_{k=0}^{T-1} (k+1)^m (T-k)^n \leq T^{m+n+1}. \tag{71}$$

PROOF. We first prove the lower bound. For T divisible by 3 we can write

$$\begin{aligned}\sum_{k=0}^{T-1} (k+1)^m (T-k)^n &\geq \sum_{k=T/3}^{2(T/3)-1} (k+1)^m (T-k)^n \\ &\geq \frac{T}{3}\left(\frac{T}{3}+1\right)^m \left(\frac{T}{3}+1\right)^n \geq dT^{m+n+1},\end{aligned}$$

with $d = 3^{-(m+n+1)}$. A minor modification of the argument holds for general T and the left inequality in (71) follows.

The upper bound follows by observing that $k+1 \leq T$ and $T-k \leq T$ such that

$$\sum_{k=0}^{T-1} (k+1)^m (T-k)^n \leq \sum_{k=0}^{T-1} T^m T^n = T^{m+n+1}. \qquad \square$$

LEMMA A.2. *Suppose $a_1 < a_2$ and $b_1 < b_2$ and let $c = (a_2 - a_1)/(b_2 - b_1)$. Then, for any $k > 0$ and $h > 0$,*

$$k^a h^b \leq k^{a_1} h^{b_2} \vee k^{a_2} h^{b_1} \tag{72}$$

for all pairs $a_1 \leq a \leq a_2$ and $b_1 \leq b \leq b_2$ satisfying $a + bc = a_1 + b_2 c = a_2 + b_1 c$.

PROOF. Consider two cases. Either $k \leq h^{1/c}$, in which case

$$\begin{aligned}\log(k^a h^b) - \log(k^{a_1} h^{b_2}) &= \log\bigl(k^a (h^{1/c})^{bc}\bigr) - \log\bigl(k^{a_1} (h^{1/c})^{b_2 c}\bigr) \\ &= (a - a_1)\log k + (bc - b_2 c)\log h^{1/c} \\ &\leq (a - a_1 + bc - b_2 c)\log h^{1/c} = 0,\end{aligned}$$

or $k > h^{1/c}$, in which case

$$\begin{aligned}\log(k^a h^b) - \log(k^{a_2} h^{b_1}) &= \log\bigl(k^a (h^{1/c})^{bc}\bigr) - \log\bigl(k^{a_2} (h^{1/c})^{b_1 c}\bigr) \\ &= (a - a_2)\log k + (bc - b_1 c)\log h^{1/c} \\ &\leq (a - a_2 + bc - b_1 c)\log k = 0. \qquad \square\end{aligned}$$

Acknowledgments. We are grateful to Eric Moulines, Richard Tweedie and Charlie Geyer for useful discussions. In particular, we thank Eric Moulines for the discussions about CLTs leading to Theorem 4.4. We also thank the referee for very useful comments, including the function space interpretation of the drift conditions in Theorem 3.2.

REFERENCES

CHUNG, K. L. (1967). *Markov Chains with Stationary Transition Probabilities*, 2nd ed. Springer, Berlin.

DAI, J. and MEYN, S. (1995). Stability and convergence of moments for multiclass queueing networks via fluid limit models. *IEEE Trans. Automat. Control* **40** 1889–1904.

FORT, G. and MOULINES, E. (2000). V-subgeometric ergodicity for a Hastings–Metropolis algorithm. *Statist. Probab. Lett.* **49** 401–410.

GLYNN, P. and MEYN, S. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931.

MENGERSEN, K. and TWEEDIE, R. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121.

MEYN, S. and TWEEDIE, R. (1992). Stability of Markovian processes I: Criteria for discrete-time chains. *Adv. in Appl. Probab.* **24** 542–574.

MEYN, S. and TWEEDIE, R. (1993). *Markov Chains and Stochastic Stability*. Springer, London.

NUMMELIN, E. (1984). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge Univ. Press.

NUMMELIN, E. and TUOMINEN, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process. Appl.* **12** 187–202.

ROBERTS, G. and ROSENTHAL, J. (2001). Markov chains and de-initialising processes. *Scand. J. Statist*. **28** 489–504.

ROBERTS, G. and TWEEDIE, R. (2001). *Understanding Markov Chain Monte Carlo*. Springer, New York. To appear.

SMITH, R. and TIERNEY, L. (1996). Exact transition probabilities for the independence Metropolis sampler. Technical report, Statistical Laboratory, Cambridge Univ. Press.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762.

TUOMINEN, P. and TWEEDIE, R. (1994). Subgeometric rates of convergence of f-ergodic Markov chains. *Adv. in Appl. Probab.* **26** 775–798.

TWEEDIE, R. (1983). Criteria for rates of convergence of Markov chains with application to queueing and storage theory. In *Probability, Statistics and Analysis* (J. F. C. Kingman and G. E. H. Reuter, eds.). Cambridge Univ. Press.

DEPARTMENT OF MATHEMATICS AND STATISTICS
LANCASTER UNIVERSITY
LANCASTER LA1 4YF
UNITED KINGDOM
E-MAIL: s.jarner@lancaster.ac.uk
g.o.roberts@lancaster.ac.uk