J. Appl. Prob. **43,** 486–499 (2006)
Printed in Israel

A MONOTONICITY IN REVERSIBLE MARKOV CHAINS

ROBERT LUND,* ** *Clemson University*
YING ZHAO,*** *The University of Georgia*
PETER C. KIESSLER,* *Clemson University*

Abstract

In this paper we identify a monotonicity in all countable-state-space reversible Markov chains and examine several consequences of this structure. In particular, we show that the return times to *every* state in a reversible chain have a decreasing hazard rate on the subsequence of even times. This monotonicity is used to develop geometric convergence rate bounds for time-reversible Markov chains. Results relating the radius of convergence of the probability generating function of first return times to the chain's rate of convergence are presented. An effort is made to keep the exposition rudimentary.

Keywords: Convergence rate; decreasing hazard rate; Markov chain; monotonicity; renewal sequence

2000 Mathematics Subject Classification: Primary 60J10
Secondary 60K05

1. Introduction

In this paper we identify a decreasing hazard rate (DHR) monotonicity inherent in all reversible Markov chains on countable state spaces and study some ramifications of this structure. Time-reversible Markov chains frequently arise in practice (Ross (1996), Stroock (2005), Chen (2005)) and include many Markov chain Monte Carlo-generated chains. Here we show that the first return time to every state in a countable-state-space reversible Markov chain has the DHR property along the even time indices. This structure is first proven for finite-state Markov chains and then extended to countable state spaces via truncation. The DHR property identified is then used to derive a 'clean' geometric convergence rate bound for reversible Markov chains. The bound is even optimal in some cases.

The DHR result imparts a geometry to reversible chain analyses. For example, a DHR first return time distribution implies that a fixed state is less likely to be visited in the immediate future if it has not been visited in the immediate past (the notion is made precise in the next section). The DHR property identified here appears weaker than the classical stochastic monotonicity structure discussed in Stoyan (1983, pp. 64–67). In particular, the DHR property identified holds for all states in any reversible Markov chain (the class of reversible chains is vast); however, there are many Markov chains that are not stochastically monotone. Stochastic monotonicity was exploited by Lund and Tweedie (1996) to extract sharp chain convergence rates. Lindvall (1992) and Kijima (1997) are other prominent references in which stochastic orderings of various types are used to assess stability and other chain properties. Keilson and Kester (1978),

Received 5 January 2006; revision received 3 March 2006.
* Postal address: Department of Mathematical Sciences, Clemson University, O-106 Martin Hall Box 340975, Clemson, SC 29634-0975, USA.
** Email address: lund@ces.clemson.edu
*** Postal address: Department of Statistics, The University of Georgia, Athens, GA 30602-1952, USA.

Brown (1980), Shaked and Shanthikumar (1987), (1994), Liggett (1989), Hansen and Frenk (1991), and Berenhaut and Lund (2001), (2002) are other references linking renewal theory, Markov chains, and stochastic orderings.

That the DHR monotonicity is identified along the even integers only is also noteworthy. This implies that the two-step-ahead chain might be more amenable to analysis in reversible settings than the traditional one-step-ahead chain. For Markov chain Monte Carlo simulators, this merely involves iterating the chain twice at each step instead of once, a straightforward task.

The remainder of the paper is organized as follows. In Section 2 we clarify our notation and review results from renewal theory and distribution classes of discrete random variables. In Section 3 we present the DHR monotonicity. In Section 4 we apply the DHR structure to obtain convergence rate bounds for reversible chains. In Section 5 we present several examples, and in Section 6 prove the main result.

2. Background

We start by clarifying our notation. Let $\{\tau_i\}_{i=1}^{\infty}$ be a sequence of independent and identically distributed random variables (lifetimes) supported on $\{1, 2, \dots\}$. Let τ denote a lifetime with distribution equivalent to that of τ_n for any $n \geq 1$. Define $f_k = \mathrm{P}[\tau = k]$ for $k \geq 1$ and let $\bar{F}_k = \mathrm{P}[\tau > k]$ for $k \geq 0$. Let $S_n = \tau_1 + \tau_2 + \cdots + \tau_n$, $n \geq 1$, be the 'lifelength' of the first n items, with $S_0 = 0$, and define the nondelayed probability of a recurrent event (an item replacement) at time n by

$$u_n = \sum_{k=0}^{\infty} \mathrm{P}[S_k = n], \qquad n \geq 1$$

(we take $u_0 = 1$). The well-known elementary recurrent event relation is

$$u_n = \sum_{k=1}^{n} f_k u_{n-k}, \qquad n \geq 1, \tag{2.1}$$

and its limit is

$$\lim_{n\to\infty} u_n = \frac{1}{\sum_{k=1}^{\infty} k f_k} =: u_\infty$$

(Feller (1968, p. 313)) when τ is aperiodic (take $u_\infty = 0$ when $\mathrm{E}[\tau] = \infty$). Equation (2.1) is easily manipulated into the tail form

$$\bar{F}_n = \sum_{k=1}^{n} (u_{k-1} - u_k)\bar{F}_{n-k}, \qquad n \geq 1, \tag{2.2}$$

which will be of use later.

The hazard rate, h_i, of τ at index i is defined as

$$h_i = \mathrm{P}[\tau = i \mid \tau \geq i]$$

whenever $\mathrm{P}[\tau \geq i] > 0$. We say that τ has a decreasing hazard rate (or is DHR) if h_i is nonincreasing in i. It is easy to show that τ is DHR if and only if the log-convexity relation

$$\bar{F}_{n+1}^2 \leq \bar{F}_n \bar{F}_{n+2}, \qquad n \geq 0, \tag{2.3}$$

holds. Mimicking (2.3), we call the renewal sequence $\{u_n\}_{n=0}^{\infty}$ DHR or log-convex if

$$u_{n+1}^2 \le u_n u_{n+2}, \qquad n \ge 0.$$

Renewal theory and lifetime orderings have been previously studied in Brown (1980), Liggett (1989), Hansen and Frenk (1991), Sengupta *et al.* (1995), Kijima (1997), and Berenhaut and Lund (2001), (2002) (among others).

Now suppose that $\{X_n\}_{n=0}^{\infty}$ is an ergodic (irreducible, aperiodic, positive recurrent) Markov chain on the state space $\{0, 1, \dots\}$ with time-homogeneous transition probability matrix $\boldsymbol{P} = (p_{i,j})_{i,j=0}^{\infty}$. Here, $p_{i,j} = \mathrm{P}[X_{n+1} = j \mid X_n = i]$ for each $n \ge 0$. Such chains have a unique limiting (stationary) measure, denoted by π, that does not depend on the initial state, X_0. Of course, $\pi_i := \pi(\{i\})$ is the long-run frequency at which the chain resides in state i:

$$\pi_j = \lim_{n\to\infty} \mathrm{P}[X_n = j \mid X_0 = i].$$

Renewal sequences can be constructed from Markov chains; in fact, the times of visit to any fixed state k in the chain can be regarded as the renewal epochs. For each pair of states j and k, we use the notation

$$\tau_{j,k} = \inf\{n \ge 1 \colon X_n = k \mid X_0 = j\}$$

for the time of first passage into state k (a first return when $j = k$). Later, the return times restricted to the even integers will become important; we denote these by

$$\eta_{j,k} = \inf\{n \ge 1 \colon X_{2n} = k \mid X_0 = j\}. \tag{2.4}$$

A relationship that we will use repeatedly in proofs is the tail formulation

$$\mathrm{P}[\tau_{k,k} > n] = \prod_{i=1}^{n} (1 - h_i(k)), \tag{2.5}$$

where $h_i(k) = \mathrm{P}[\tau_{k,k} = i \mid \tau_{k,k} \ge i]$.

Now suppose that the chain is reversible in that, for each pair of states j and k,

$$\pi_j p_{j,k} = \pi_k p_{k,j}.$$

Reversible chains include birth-and-death chains, random walks, two-state chains, many Markov chain Monte Carlo-generated chains, and urn models.

It is frequently desirable to assess convergence speeds. For renewal sequences, a geometric convergence rate of u_n to u_∞ is a bound of the form

$$|u_n - u_\infty| \le \kappa r^{-n}, \tag{2.6}$$

for some finite κ and geometric rate $r > 1$. The constants κ and r should be explicit (computable); moreover, one would clearly like κ to be as small and r to be as large as possible. It is important to note that (2.6) must hold for every $n \ge 0$; this differs fundamentally from asymptotic approximations. Much of our discourse pertains to obtaining actual values of κ and r.

A geometric renewal convergence rate exists if and only if $\mathrm{E}[s^\tau] < \infty$ for some $s > 1$ (Kendall (1959)). In particular, if $\mathrm{E}[s^\tau] = \infty$ for each $s > 1$, then no $r > 1$ satisfying (2.6) exists. Unfortunately, the finiteness of $\mathrm{E}[s^\tau]$ for a fixed $s > 1$ does not imply that (2.6) holds with $r = s$. Indeed, identifying an explicit rate and a first constant from the finiteness of $\mathrm{E}[s^\tau]$ for some fixed $s > 1$ is very difficult in general.

To assess the convergence speed of a chain $\{X_n\}_{n=0}^\infty$ to its limiting measure π, we seek a geometric rate $r > 1$ and a first constant $\kappa(j)$ such that

$$\sup_A |\,\mathrm{P}[X_n \in A \mid X_0 = j] - \pi(A)| \le \kappa(j) r^{-n},$$

where the supremum is taken over all measurable subsets of the state space. Here, $\kappa(j)$ is allowed to depend on the initial state j, but r is not. The fact that one can find some uniform rate $r > 1$ in an ergodic chain that holds for all states is generally attributed to Kendall (1959); see Baxendale (2005) for the latest developments. In addition to aiding conceptual understanding, quantitative geometric convergence rates are also useful for assessing sampled Markov chain convergence and for proving large-sample statistical results.

3. The DHR monotonicity

Our first result identifies the DHR monotonicity for reversible chains on a finite state space.

Theorem 3.1. *Suppose that $\{X_n\}_{n=0}^\infty$ is a reversible ergodic Markov chain on the finite state space $\{0, \dots, N\}$.*

1. *The lifetime $\eta_{k,k}$ (see (2.4)) has a DHR distribution for each k.*
2. $\mathrm{P}[X_{2n} = k \mid X_0 = k]$ *is nondecreasing and log-convex in n for every state k.*

The proof of Theorem 3.1 relies on the spectral decomposition of $\boldsymbol{P}$ and is presented in Section 6. To avoid the technicalities of spectral theory for unbounded operators, we use a truncation argument to extend Theorem 3.1 to countably infinite state spaces. In particular, suppose that $\{X_n\}_{n=0}^\infty$ is a reversible chain on the states $\{0, 1, \dots\}$ and let M be an arbitrary positive integer. We truncate $\{X_n\}_{n=0}^\infty$ to the states $\{0, \dots, M\}$ by disallowing transitions to states numbered higher than M (i.e. stay in the current state should a transition out of $\{0, \dots, M\}$ be suggested). In particular, the truncated chain $\{X_n^{(M)}\}_{n=0}^\infty$ has transition probability matrix $\boldsymbol{P}^{(M)}$ with $p_{i,j}^{(M)} = p_{i,j}$ if $0 \le i \ne j \le M$ and

$$p_{i,i}^{(M)} = p_{i,i} + \sum_{k=M+1}^{\infty} p_{i,k}, \qquad 0 \le i \le M.$$

We can assume that $\{X_n^{(M)}\}_{n=0}^\infty$ is defined on the same probability space for each $M \ge 1$ by enlarging the space supporting $\{X_n\}_{n=0}^\infty$ if necessary.

Lemma 3.1. *The chain $\{X_n^{(M)}\}_{n=0}^\infty$ is time reversible with limiting measure*

$$\pi_i^{(M)} := \lim_{n\to\infty} \mathrm{P}[X_n^{(M)} = i] = \frac{\pi_i}{\sum_{i=0}^M \pi_i}.$$

Proof. See Problem 4.44 of Ross (1996, p. 228).

Corollary 3.1. *The statements in Theorem 3.1 also hold for reversible ergodic chains with countably infinite state spaces.*

https://doi.org/10.1239/jap/1152413736 Published online by Cambridge University Press

Proof. Given $\{X_n\}_{n=0}^{\infty}$, let $\{X_n^{(M)}\}_{n=0}^{\infty}$ be the truncated chain defined above. Since $\{X_n^{(M)}\}_{n=0}^{\infty}$ is also reversible, Theorem 3.1 implies that

$$(\bar{F}_{n+1}^{(M)})^2 \leq \bar{F}_n^{(M)} \bar{F}_{n+2}^{(M)}, \qquad n \geq 0,$$

for each fixed M, where

$$\bar{F}_n^{(M)} = \mathrm{P}[\eta_{k,k}^{(M)} > n] \quad \text{and} \quad \eta_{k,k}^{(M)} = \inf\{n \geq 1 \colon X_{2n}^{(M)} = k \mid X_0^{(M)} = k\}.$$

To obtain Theorem 3.1 for $\{X_n\}_{n=0}^{\infty}$, it is sufficient to establish that $\lim_{M\uparrow\infty} \bar{F}_n^{(M)} = \bar{F}_n$ for each fixed $n \geq 0$.

Since X_j is a proper random variable for each fixed j with $1 \leq j \leq n$,

$$\lim_{M\uparrow\infty} \mathrm{P}[X_j \leq M,\ 1 \leq j \leq n \mid X_0 = k] = 1.$$

Now, on the set $\{X_j \leq M,\ 1 \leq j \leq n\}$, $\eta_{k,k}^{(M)} \leq n$ if and only if $\eta_{k,k} \leq n$. Hence,

$$\begin{aligned}\lim_{M\uparrow\infty} \mathrm{P}[\eta_{k,k}^{(M)} \leq n \mid X_0 = k] &= \lim_{M\uparrow\infty} \mathrm{P}[\{\eta_{k,k}^{(M)} \leq n\} \cap \{X_j \leq M, 1 \leq j \leq n\} \mid X_0 = k] \\ &= \lim_{M\uparrow\infty} \mathrm{P}[\{\eta_{k,k} \leq n\} \cap \{X_j \leq M,\ 1 \leq j \leq n\} \mid X_0 = k] \\ &= \mathrm{P}[\eta_{k,k} \leq n \mid X_0 = k],\end{aligned}$$

and the result follows.

4. Convergence rates of reversible chains

Convergence rates of reversible Markov chains can be extracted from the DHR structure of the last section. Reversible chain convergence rates have received considerable recent attention; Diaconis and Stroock (1991), Rosenthal (1995), Stroock (2005), and Chen (2005) are excellent references. Some of the rates obtained here are good (tight), and others are not. The advance here lies primarily in connecting convergence rates of reversible chains to hazard rates and finiteness of the probability generating function of state return times; in particular, no eigenvalues of $\boldsymbol{P}$ are needed.

We begin by linking geometric Markov chain convergence rates to the hazard rates of the first return times of the chain. This topic has been studied further in Berenhaut and Lund (2001). The following power series bound will prove useful in the ensuing analysis.

Lemma 4.1. *Suppose that $\{a_k\}_{k=0}^{\infty}$ is a sequence of real numbers, and set*

$$S_n(t) = \sum_{k=n+1}^{\infty} (a_{k-1} - a_k) t^k \tag{4.1}$$

for each fixed $n \geq 0$ and $t \in (0, 1)$. If $a_k \in [0, C]$ for all $k \geq 0$ and some $C > 0$, then $|S_n(t)| \leq C t^{n+1}$ for all $n \geq 0$.

Proof. Fix $n \geq 0$ and regroup terms in (4.1) to obtain

$$S_n(t) = a_n t^{n+1} + \sum_{k=n+1}^{\infty} a_k (t^{k+1} - t^k). \tag{4.2}$$

Since $t \in (0, 1)$, $t^{k+1} - t^k \le 0$ for all $k \ge 0$ and a sign analysis of (4.2) shows that $S_n(t)$ is maximized for each fixed t (over all $\{a_k\}_{k=0}^{\infty}$ with $a_k \in [0, C]$) by taking $a_n = C$ and $a_k = 0$ for $k \ge n+1$. Similarly, $S_n(t)$ is minimized over all such $\{a_k\}_{k=0}^{\infty}$ by taking $a_n = 0$ and $a_k = C$ for $k \ge n+1$. Both choices of $\{a_k\}_{k=0}^{\infty}$ give the same value of $|S_n(t)|$. The claimed bound for $|S_n(t)|$ follows by inserting the maximizing/minimizing values of the a_k into (4.1).

The following result cleanly bounds chain convergence rates. The result applies to any ergodic chain on a countable state space.

Theorem 4.1. *Suppose that* $\{X_n\}_{n=0}^{\infty}$ *is an ergodic Markov chain on the state space* $\{0, 1, \dots\}$. *Then*

$$|\,\mathrm{P}[X_n = k \mid X_0 = k] - \pi_k| \le (1 - h_{\min}(k))^{n+1}, \tag{4.3}$$

where $h_{\min}(k) = \inf\{h_i(k) \colon i \ge 1\}$ *and, recall,* $h_i(k) = \mathrm{P}[\tau_{k,k} = i \mid \tau_{k,k} \ge i]$.

Proof. Inequality (4.3) clearly holds when $h_{\min}(k) = 0$. Hence, suppose that $h_{\min}(k) > 0$ and choose a ρ such that $1 < \rho < (1 - h_{\min}(k))^{-1}$. Let $\tau^*_{k,k}$ be a random variable with distribution

$$\mathrm{P}[\tau^*_{k,k} = n] = \mathrm{P}[\tau_{k,k} > n-1]\rho^{n-1} - \mathrm{P}[\tau_{k,k} > n]\rho^n, \qquad n \ge 1, \tag{4.4}$$

whence

$$\mathrm{P}[\tau^*_{k,k} > n] = \mathrm{P}[\tau_{k,k} > n]\rho^n \tag{4.5}$$

for all $n \ge 0$. We show below that (4.4) defines a legitimate discrete lifetime distribution over $\{1, 2, \dots\}$.

The governing recurrent event recursion, (2.1), and its tail form, (2.2), also apply to $\tau^*_{k,k}$ (in our notation we will use u^*_n as the probability of a renewal at time n in a renewal process with independent lifetimes each having the same distribution as $\tau^*_{k,k}$).

Multiplying both sides of (2.2) by ρ^n and applying (4.5) yields

$$\mathrm{P}[\tau^*_{k,k} > n] = \sum_{\ell=1}^{n} \rho^{\ell}(u_{\ell-1} - u_{\ell})\,\mathrm{P}[\tau^*_{k,k} > n - \ell], \qquad n \ge 1. \tag{4.6}$$

Comparing (4.6) with the version of (2.2) that applies for $\tau^*_{k,k}$ and inductively equating coefficients gives

$$(u_{\ell-1} - u_{\ell})\rho^{\ell} = u^*_{\ell-1} - u^*_{\ell}, \qquad \ell \ge 1.$$

From this, we find that

$$|u_n - u_\infty| = \left|\sum_{i=n+1}^{\infty} (u_{i-1} - u_i)\right| = \left|\sum_{i=n+1}^{\infty} (u^*_{i-1} - u^*_i)\rho^{-i}\right|.$$

Since u^*_k is a probability for each k, we have $0 \le u^*_k \le 1$, and Lemma 4.1 with $C = 1$ and $t = \rho^{-1}$ gives $|u_n - u_\infty| \le \rho^{-(n+1)}$. Letting $\rho \uparrow (1 - h_{\min}(k))^{-1}$ gives the result of the theorem.

It remains to show that (4.4) defines a legitimate lifetime distribution on $\{1, 2, \dots\}$. To do so, we must show that $\mathrm{P}[\tau_{k,k} > n]\rho^n$ is nonincreasing in n and that $\lim_{n\to\infty} \mathrm{P}[\tau_{k,k} > n]\rho^n = 0$. By choice of ρ and use of (2.5), we have

$$\frac{\mathrm{P}[\tau_{k,k} > n]}{\mathrm{P}[\tau_{k,k} > n-1]} = 1 - h_n < \rho^{-1}$$

for all $n \geq 0$. It now follows that

$$\frac{\mathrm{P}[\tau_{k,k} > n]\rho^n}{\mathrm{P}[\tau_{k,k} > n-1]\rho^{n-1}} < 1$$

for all $n \geq 0$; thus, $\mathrm{P}[\tau_{k,k} > n]\rho^n$ is nonincreasing in n. To show that $\mathrm{P}[\tau_{k,k} > n]\rho^n \to 0$ as $n \to \infty$, we use (2.5) to obtain

$$\mathrm{P}[\tau_{k,k} > n] = \rho^n \prod_{j=1}^{n} (1 - h_j) \to 0$$

as $n \to \infty$, which holds because $\sup_{j \geq 1} \rho(1 - h_j) < 1$ by choice of ρ.

Note that the convergence rate in Theorem 4.1 depends on the initial state k. This can be advantageous if there is a state in the chain that lends itself to easy analysis. Theorem 4.1 only addresses convergence of the state-k probabilities. Frequently, more global convergence measures are needed in applications. For this, we consider the total variation separation of the chain at time n, defined by $\sup_A |\mathrm{P}[X_n \in A \mid X_0 = k] - \pi(A)|$. Our next result uses Theorem 4.1 to extract a total variation convergence rate.

Theorem 4.2. *For a reversible ergodic Markov chain* $\{X_n\}_{n=0}^{\infty}$ *on the state space* $\{0, 1, \dots\}$ *we have*

$$\sup_A |\mathrm{P}[X_n \in A \mid X_0 = k] - \pi(A)| \leq \sqrt{\frac{(1 - h_{\min}(k))^{2n+1}}{4\pi_k}}.$$

Proof. The argument of Proposition 3 of Diaconis and Stroock (1991) yields

$$\sup_A |\mathrm{P}[X_n \in A \mid X_0 = k] - \pi(A)| \leq \sqrt{\frac{\mathrm{P}[X_{2n} = k \mid X_0 = k] - \pi_k}{4\pi_k}} \tag{4.7}$$

using only reversibility and the Cauchy–Schwarz inequality; in particular, a finite state space is not needed. Substituting the bound for $\mathrm{P}[X_{2n} = k \mid X_0 = k] - \pi_k$ of Theorem 4.1 into (4.7) proves the result.

It is worth noting that the quantity under the radical in (4.7) must be positive; Theorem 3.1 shows this to be the case (the quantity is also nondecreasing and log-convex in n).

We close this section by linking convergence rates of reversible chains to the radii of convergence of state return times. In general, reversible chains may not converge at rates up to the radius of convergence of $\mathrm{E}[r^{\tau_{k,k}}]$. Such a convergence property holds for many stochastically monotone chains (Lund and Tweedie (1996)) and for reversible chains in continuous time (see Theorem 4.1 of Chen (2000)). Example 5.2, below, provides a reversible chain counterexample in discrete time. There is, however, a connection between chain convergence rates and the convergence radius of $\mathrm{E}[r^{\eta_{k,k}}]$, the latter of which we denote by $R_\eta > 1$.

The tail form for probability generating functions (see Meyn and Tweedie (1993, p. 527)) gives

$$\begin{aligned}\mathrm{E}[r^{\eta_{k,k}}] &= 1 + (r-1)\sum_{n=0}^{\infty} r^n \,\mathrm{P}[\eta_{k,k} > n] \\ &= 1 + (r-1)\sum_{n=0}^{\infty}\prod_{i=1}^{n} r(1 - h_i^{\eta}(k)),\end{aligned} \tag{4.8}$$

where $h_i^\eta(k)$ denotes the hazard rate of $\eta_{k,k}$ at index i and (2.5) has been applied. Theorem 3.1 shows that $h_i^\eta(k)$ is decreasing in i and we denote its limit in i by $h_\infty^\eta(k)$. We use this monotonicity and a ratio test in (4.8) to see that $R_\eta = (1 - h_\infty^\eta(k))^{-1}$. Applying Theorems 4.1 and 4.2 to the chain $\{X_{2n}\}_{n=0}^\infty$ and noting that π is also the stationary measure of $\{X_{2n}\}_{n=0}^\infty$ proves the following result.

Theorem 4.3. *Consider a reversible ergodic Markov chain $\{X_n\}_{n=0}^\infty$ on the state space* $\{0, 1, \dots\}$. *Let $h_\infty^\eta(k)$ be the limiting state-k hazard rate of $\eta_{k,k}$. Then*

$$|\operatorname{P}[X_{2n} = k \mid X_0 = k] - \pi_k| \le R_\eta^{-(n+1)},$$

$$\sup_A |\operatorname{P}[X_{2n} \in A \mid X_0 = k] - \pi(A)| \le \frac{R_\eta^{-(n+1/2)}}{\sqrt{4\pi_k}}.$$

An implication of Theorem 4.3 is that total variational convergence of reversible chains holds for rates up to $\sqrt{R_\eta}$ (at least). Convergence rates along the subsequence of odd times will not 'destroy' the derived even-time rates, since the total variation distance is nonincreasing in n (see Tuominen and Tweedie (1979)). The above discourse also suggests that $\{X_{2n}\}_{n=0}^\infty$ may be easier to analyze than $\{X_n\}_{n=0}^\infty$. This is intuitive, as all eigenvalues of $\boldsymbol{P}^2$ are real, positive, and lie in $[0, 1]$, while those of $\boldsymbol{P}$ are only known to be real and to lie in $[-1, 1]$ (see Lemma 6.1, below).

Established methods exist for showing that $\operatorname{E}[r^{\eta_{k,k}}] < \infty$ for a fixed $r > 1$. For example, $\operatorname{E}[r^{\eta_{k,k}}] < \infty$ if a drift function V and a constant $b < \infty$ can be found such that

$$\operatorname{E}[V(X_2) \mid X_0 = x] \le r^{-1} V(x) + b 1_{\{k\}}(x).$$

Here it is necessary that both $V(x) \ge 1$ for all x and $r > 1$ to establish the drift inequality as a legitimate contraction. The reader is referred to Meyn and Tweedie (1993, p. 367) and Kalashnikov (1994, p. 9) for more on drift methods.

5. Examples

Example 5.1. In this example we present a renewal setting in which the optimal rate of convergence is achieved. Fix ρ, $0 < \rho \le 1$, and consider the lifetime τ with tail distribution

$$\operatorname{P}[\tau > n] = \frac{(2n)!}{n!\,(n+1)!} \left(\frac{\rho}{4}\right)^n, \qquad n \ge 0.$$

This lifetime is DHR, since

$$\frac{\operatorname{P}[\tau > n]^2}{\operatorname{P}[\tau > n-1]\operatorname{P}[\tau > n+1]} = \frac{2(2n-1)(n+2)}{(2n+2)(2n+1)} < 1,$$

and the radius of convergence of $\operatorname{E}[r^\tau]$ is ρ^{-1}, since

$$\operatorname{E}[r^\tau] = 1 + (r-1) \sum_{n=0}^\infty r^n \frac{(2n)!}{n!\,(n+1)!} \left(\frac{\rho}{4}\right)^n,$$

which is finite if and only if $r\rho < 1$ (use a ratio test).

Applying Theorem 4.1 gives the bound $|u_n - u_\infty| \le \rho^{n+1}$, and this is the best convergence bound possible. To see this, we use the fact that $u_n \downarrow u_\infty$, to obtain

$$u_n - u_\infty \ge u_n - u_{n+1} = \frac{\rho}{4} \operatorname{P}[\tau > n],$$

where the equality follows from an identity of Liggett (1989). Thus, if $r > \rho^{-1}$ then

$$\liminf_{n\to\infty} r^n(u_n - u_\infty) \geq \lim_{n\to\infty} r^n \frac{\rho}{4} \mathrm{P}[\tau > n] = \infty,$$

and the convergence rate cannot be improved upon.

Example 5.2. Consider a chain on the two states $\{0\}$ and $\{1\}$ with transition matrix

$$\boldsymbol{P} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix}, \qquad 0 < \alpha, \beta < 1.$$

This simple chain is reversible and, as exact computations are possible, will allow us to compare the rate bounds. This example was discussed in Rosenthal (1995).

The limiting measure π has

$$\pi_0 = \frac{1-\alpha}{2-\alpha-\beta} \quad \text{and} \quad \pi_1 = \frac{1-\beta}{2-\alpha-\beta},$$

and simple calculations give

$$\mathrm{P}[\tau_{0,0} > n] = (1-\alpha)\beta^{n-1} \quad \text{and} \quad \mathrm{P}[\tau_{1,1} > n] = (1-\beta)\alpha^{n-1}, \qquad n \geq 1.$$

The state-0 hazard rates are easily identified as $h_1(0) = \alpha$ and $h_k(0) = 1-\beta$ for $k \geq 2$. The state-1 hazard rates are $h_1(1) = \beta$ and $h_k(1) = 1-\alpha$ for $k \geq 2$. Hence, $h_{\min}(0) = \min(\alpha, 1-\beta)$ and $h_{\min}(1) = \min(\beta, 1-\alpha)$. Applying Theorem 4.2 gives

$$\sup_A |\mathrm{P}[X_n \in A \mid X_0 = 0] - \pi(A)| \leq \sqrt{\frac{2-\alpha-\beta}{4(1-\alpha)}}[1-\min(\alpha, 1-\beta)]^{n+1/2},$$
$$\sup_A |\mathrm{P}[X_n \in A \mid X_0 = 1] - \pi(A)| \leq \sqrt{\frac{2-\alpha-\beta}{4(1-\beta)}}[1-\min(1-\alpha, \beta)]^{n+1/2}.$$

To obtain the optimal geometric convergence rate, we use induction to obtain

$$\boldsymbol{P}^n = \frac{1}{2-\alpha-\beta}\begin{bmatrix} 1-\beta & 1-\alpha \\ 1-\beta & 1-\alpha \end{bmatrix} + \frac{(\alpha+\beta-1)^n}{2-\alpha-\beta}\begin{bmatrix} 1-\alpha & \alpha-1 \\ \beta-1 & 1-\beta \end{bmatrix}.$$

This gives

$$\sup_A |\mathrm{P}[X_n \in A \mid X_0 = 0] - \pi(A)| = \frac{1-\alpha}{2-\alpha-\beta}|\alpha+\beta-1|^n. \tag{5.1}$$

Here, the rate bounds are informative (i.e. they exceed unity) but are not optimal. Equation (5.1) shows that the optimal geometric convergence rate is $|\alpha+\beta-1|^{-1}$. Note that the chain converges at an infinite rate when $\alpha+\beta = 1$. Indeed, when $\alpha+\beta = 1$ the X_n are independent and have distribution π for each $n \geq 1$.

In this example, $\mathrm{E}[r^{\tau_{j,0}}] < \infty$ for all initial states j when $r < \beta^{-1}$ and $\mathrm{E}[r^{\tau_{j,1}}] < \infty$ for all j when $r < \alpha^{-1}$. However, when $X_0 = 0$ the chain does not converge at geometric rates up to β^{-1}. Indeed, when $\alpha = \beta = 0.05$, the best geometric rate is $\frac{10}{9}$ but the radius of convergence of $\mathrm{E}[r^{\tau_{j,k}}]$ is 20 for every j and k. This gives a very simple example of a discrete-time reversible chain where geometric convergence does not take place out to the radii of convergence of the generating function of the first return times. Note, however, that

$$\boldsymbol{P}^2 = \begin{bmatrix} \alpha^2 + (1-\alpha)(1-\beta) & \alpha(1-\alpha) + \beta(1-\alpha) \\ \beta(1-\beta) + \alpha(1-\beta) & \beta^2 + (1-\alpha)(1-\beta) \end{bmatrix}$$

and that $\mathrm{E}[r^{\eta_{j,k}}]$ is finite for all j and k when

$$r < \min([\alpha^2 + (1-\alpha)(1-\beta)]^{-1}, [\beta^2 + (1-\alpha)(1-\beta)]^{-1}).$$

When $\alpha = \beta = 0.05$, Theorem 4.3 provides a total variational geometric convergence rate of 1.0383 (to four decimal places), which is somewhat less than the optimal rate of $\frac{10}{9}$.

Remark 5.1. The radii of convergence of $\mathrm{E}[r^{\eta_{i,i}}]$ may not be the same for each i. In Example 5.2, $\mathrm{E}[r^{\eta_{0,0}}]$ has radius of convergence $[\beta^2 + (1-\beta)(1-\alpha)]^{-1}$ and $\mathrm{E}[r^{\eta_{1,1}}]$ has radius of convergence $[\alpha^2 + (1-\beta)(1-\alpha)]^{-1}$. In general, if $\mathrm{E}[r^{\eta_{0,0}}] < \infty$ then $\mathrm{E}[r^{\eta_{k,0}}] < \infty$ for every other initial state k. To see this, note that if $\mathrm{E}[r^{\eta_{0,0}}] < \infty$, then some $k \geq 1$ can be found such that the probability of going from state 0 to state k in j steps without passing through state 0 again, denoted by $p^j_{0,k/\{0\}}$, is positive. Then, by the Markov property,

$$\mathrm{E}[r^{\eta_{0,0}}] \geq p^j_{0,k/\{0\}} r^j \, \mathrm{E}[r^{\eta_{k,0}}],$$

implying that

$$\mathrm{E}[r^{\eta_{k,0}}] \leq \frac{\mathrm{E}[r^{\eta_{0,0}}]}{p^j_{0,k/\{0\}} r^j} < \infty.$$

Remark 5.2. Since $\mathrm{P}[X_{2n} = k \mid X_0 = k]$ decreases to its limit, π_k, the entries of $\boldsymbol{P}^2$ satisfy $P^2(k, A) \geq \pi_k 1_A(k)$ for all sets A and a minorization condition holds. In fact, in a reversible chain, if $\mathrm{E}[r^{\eta_{k,k}}] < \infty$ for some $r > 1$, then conditions A1–A3 of Baxendale (2005) hold with

$$C = \{k\}, \qquad \beta = \pi_k, \qquad \nu(A) = 1_A(k), \qquad \lambda = r^{-1},$$
$$K = r^{-1}(\mathrm{E}[r^{\eta_{k,k}}] - 1), \qquad \text{and} \qquad V(x) = \mathrm{E}[r^{\eta_{x,k}}] \quad \text{for } x \neq k;$$

see Theorem 5.1 of Lund and Tweedie (1996) for construction details of the drift V. Such a minorization may prove useful in splitting arguments for reversible chains on a continuum of states.

6. Proof of Theorem 3.1

To prove Theorem 3.1, we establish a sequence of technical facts.

Lemma 6.1. *In a reversible Markov chain on the finite state space* $\{0, \dots, N\}$*, all eigenvalues of* $\boldsymbol{P}^2$ *are real and nonnegative.*

Proof. As the eigenvalues of a reversible transition probability matrix $\boldsymbol{P}$ must be real (see Theorem 3.29 of Kulkarni (1995, p. 146) or Example 2.19 of Kijima (1997, p. 62)) and the eigenvalues of $\boldsymbol{P}^2$ are the squares of those of $\boldsymbol{P}$, the result follows.

Without loss of generality, we proceed with the eigenvalues, $\{\lambda_i\}_{i=0}^N$, of $\boldsymbol{P}$ arranged in the descending absolute order

$$1 = \lambda_0 > |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_N| \geq 0. \tag{6.1}$$

The standard spectral decomposition of $\boldsymbol{P}^2$ yields

$$\mathrm{P}[X_{2n} = k \mid X_0 = k] = \pi_k + \sum_{j=0}^{N} \lambda_j^{2n} x_{j,k}^2 \tag{6.2}$$

for each state k, where $x_{j,k}$ is the jth component of the eigenvector $\boldsymbol{x}_j$. To elaborate, $\boldsymbol{x}_j$ is the eigenvector of the matrix $\boldsymbol{\Pi}_{\mathrm{d}}^{1/2} \boldsymbol{P} \boldsymbol{\Pi}_{\mathrm{d}}^{-1/2}$ corresponding to the eigenvalue λ_j (the eigenvalues of $\boldsymbol{P}$ equal those of $\boldsymbol{\Pi}_{\mathrm{d}}^{1/2} \boldsymbol{P} \boldsymbol{\Pi}_{\mathrm{d}}^{-1/2}$ (see Kijima (1997, p. 62))), chosen orthonormally in j, where $\boldsymbol{\Pi}_{\mathrm{d}}$ is the diagonal matrix whose diagonal entries are the stationary probabilities $\{\pi_i\}_{i=0}^{N}$.

As was noted in Kijima (1997, p. 63), (6.2) implies that $\mathrm{P}[X_{2n} = k \mid X_0 = k]$ converges monotonically downwards to π_k. However, more can be said; in particular, we offer the following result.

Lemma 6.2. *Consider a reversible Markov chain $\{X_n\}_{n=0}^{\infty}$ on the state space $\{0, \dots, N\}$, and fix a state k. Let $\Delta_n = u_{2n} - u_{2(n-1)}$, where $u_n = \mathrm{P}[X_n = k \mid X_0 = k]$. Then $\{\Delta_n\}_{n=1}^{\infty}$ is a positive, nonincreasing, log-convex sequence in n.*

Proof. From (6.2), we have

$$\begin{aligned}
\Delta_n &= \left[\pi_k + \sum_{j=0}^{N} (\lambda_j^2)^{n-1} x_{j,k}^2\right] - \left[\pi_k + \sum_{j=0}^{N} (\lambda_j^2)^{n} x_{j,k}^2\right] \\
&= \sum_{j=0}^{N} (\lambda_j^2)^{n-1} (1 - \lambda_j^2) x_{j,k}^2 = \sum_{j=0}^{N} (\lambda_j^2)^{n-1} \omega_j,
\end{aligned} \tag{6.3}$$

where $\omega_j = (1 - \lambda_j^2) x_{j,k}^2$ is nonnegative. The positivity and nonincrease claims now follow. To show log-convexity of $\{\Delta_n\}_{n=1}^{\infty}$, we use (6.3) and (6.1) to obtain

$$\begin{aligned}
\Delta_n^2 &= \sum_{j=0}^{N} \sum_{\ell=0}^{N} (\lambda_j^2)^{n-1} \omega_j (\lambda_\ell^2)^{n-1} \omega_\ell \\
&= \sum_{\ell=0}^{N} (\lambda_\ell^2)^{n-1} (\lambda_\ell^2)^{n-1} \omega_\ell \omega_\ell + 2 \sum_{\{\ell, j \colon \ell < j\}} \sum (\lambda_j^2)^{n-1} (\lambda_\ell^2)^{n-1} \omega_j \omega_\ell \\
&\le \sum_{\ell=0}^{N} (\lambda_\ell^2)^{2(n-1)} \omega_\ell^2 + 2 \sum_{\{\ell, j \colon \ell < j\}} \sum (\lambda_j^2)^{n-1} (\lambda_\ell^2)^{n-1} \frac{\lambda_\ell^2}{\lambda_j^2} \omega_j \omega_\ell \\
&= \sum_{\ell=0}^{N} (\lambda_\ell^2)^{2(n-1)} \omega_\ell^2 + 2 \sum_{\{\ell, j \colon \ell < j\}} \sum (\lambda_j^2)^{n-2} (\lambda_\ell^2)^{n} \omega_j \omega_\ell \\
&= \sum_{\ell=0}^{N} (\lambda_\ell^2)^{n-2} (\lambda_\ell^2)^{n} \omega_\ell \omega_\ell + 2 \sum_{\{\ell, j \colon \ell < j\}} \sum (\lambda_j^2)^{n-2} (\lambda_\ell^2)^{n} \omega_j \omega_\ell \\
&= \left[\sum_{j=0}^{N} (\lambda_j^2)^{n-2} \omega_j\right] \left[\sum_{\ell=0}^{N} (\lambda_\ell^2)^{n} \omega_\ell\right] \\
&= \Delta_{n-1} \Delta_{n+1}.
\end{aligned}$$

Now consider a sequence $\{\kappa_n\}_{n=1}^{\infty}$ in the form of (2.2) defined recursively by

$$\kappa_n = \sum_{\ell=1}^{n} C_\ell \kappa_{n-\ell}, \qquad n \geq 1, \tag{6.4}$$

with $\kappa_0 = 1$. Here $\{C_\ell\}_{\ell=1}^{\infty}$ is assumed to be positive, nondecreasing, and log-convex. The following result is similar to Lemma 2.4 of Hansen and Frenk (1991).

Lemma 6.3. *Suppose that $\{\kappa_\ell\}_{\ell=0}^{\infty}$ and $\{C_n\}_{n=1}^{\infty}$ are related by (6.4). If $\{C_n\}_{n=1}^{\infty}$ is positive, nonincreasing, and log-convex, then $\{\kappa_\ell\}_{\ell=0}^{\infty}$ is also log-convex.*

Proof. First, (6.4) and $C_2 \geq 0$ give

$$\kappa_1^2 - \kappa_0\kappa_2 = C_1^2 - (\kappa_1 C_1 + C_2) = -C_2 \leq 0.$$

Hence, $\kappa_1^2 \leq \kappa_0\kappa_2$. Similarly (6.4) gives

$$\kappa_0\kappa_3 - \kappa_2\kappa_1 = (\kappa_2 C_1 + \kappa_1 C_2 + \kappa_0 C_3) - \kappa_2 C_1 = \kappa_1 C_2 + \kappa_0 C_3 \geq 0. \tag{6.5}$$

An identity that will allow us to exploit log-convexity is

$$\begin{aligned}
&\sum_{\ell=1}^{n} (\kappa_{n-\ell}\kappa_{n+2} - \kappa_{n+1}\kappa_{n+1-\ell})(C_{n+2}C_\ell - C_{\ell+1}C_{n+1}) \\
&\quad = \kappa_{n+2}C_{n+2}\sum_{\ell=1}^{n}\kappa_{n-\ell}C_\ell - \kappa_{n+2}C_{n+1}\sum_{\ell=1}^{n}\kappa_{n-\ell}C_{\ell+1} - \kappa_{n+1}C_{n+2}\sum_{\ell=1}^{n}\kappa_{n+1-\ell}C_\ell \\
&\qquad + \kappa_{n+1}C_{n+1}\sum_{\ell=1}^{n}\kappa_{n+1-\ell}C_{\ell+1} \\
&\quad = \kappa_{n+2}C_{n+2}\kappa_n - \kappa_{n+2}C_{n+1}(\kappa_{n+1} - \kappa_n C_1) - \kappa_{n+1}C_{n+2}(\kappa_{n+1} - \kappa_0 C_{n+1}) \\
&\qquad + \kappa_{n+1}C_{n+1}(\kappa_{n+2} - \kappa_{n+1}C_1 - \kappa_0 C_{n+2}) \\
&\quad = (C_{n+2} + C_1 C_{n+1})(\kappa_n\kappa_{n+2} - \kappa_{n+1}^2).
\end{aligned} \tag{6.6}$$

Taking $n = 1$ in (6.6) yields

$$(C_3 + C_1C_2)(\kappa_1\kappa_3 - \kappa_2^2) = (\kappa_0\kappa_3 - \kappa_2\kappa_1)(C_3C_1 - C_2^2) \geq 0.$$

By (6.5) and nonnegativity and convexity of the C_ℓ, we must have $\kappa_2^2 \leq \kappa_1\kappa_3$.

To complete the proof via induction, suppose that

$$\kappa_{n+1}^2 \leq \kappa_n\kappa_{n+2}$$

for all $n \leq p-2$, where $p > 2$. Taking $n = p-1$ in (6.6) yields

$$\begin{aligned}
&(C_{p+1} + C_1C_p)(\kappa_{p-1}\kappa_{p+1} - \kappa_p^2) \\
&\quad = \sum_{\ell=1}^{p-1}(\kappa_{p-1-\ell}\kappa_{p+1} - \kappa_p\kappa_{p-\ell})(C_{p+1}C_\ell - C_{\ell+1}C_p).
\end{aligned} \tag{6.7}$$

https://doi.org/10.1239/jap/1152413736 Published online by Cambridge University Press

The induction hypothesis and nonnegativity and convexity of the C_ℓ imply that each term in parentheses on the right-hand side of (6.7) is nonnegative. Since $C_{p+1} + C_1 C_p$ is also nonnegative, we conclude that

$$\kappa_p^2 \le \kappa_{p-1}\kappa_{p+1},$$

as required.

Proof of Theorem 3.1. Let

$$u_n = \mathrm{P}[X_{2n} = k \mid X_0 = k]$$

for a fixed state k. Then, by (2.2),

$$\mathrm{P}[\eta_{k,k} > n] = \sum_{\ell=1}^{n} (u_{\ell-1} - u_\ell)\, \mathrm{P}[\eta_{k,k} > n - \ell]. \tag{6.8}$$

By Lemma 6.2, $u_{\ell-1} - u_\ell$ is positive, nonincreasing, and log-convex in ℓ. Applying Lemma 6.3 to (6.8) proves the first part of the theorem.

To prove the second statement of the theorem, simply use the representation in (6.2) and argue as in the proof of Lemma 6.2.

Acknowledgements

The authors acknowledge support from the National Science Foundation, grant number DMS 0304407. Our thanks go to the referee for several insightful comments that improved the paper.

References

Baxendale, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Prob.* **15,** 700–738.

Berenhaut, K. S. and Lund, R. (2001). Geometric renewal convergence rates from hazard rates. *J. Appl. Prob.* **38,** 180–194.

Berenhaut, K. S. and Lund, R. (2002). Renewal convergence rates for DHR and NWU lifetimes. *Prob. Eng. Inf. Sci.* **16,** 67–84.

Brown, M. (1980). Bounds, inequalities, and monotonicity properties for some specialized renewal processes. *Ann. Prob.* **8,** 227–240.

Chen, M.-F. (2000). Equivalence of exponential ergodicity and L^2-exponential convergence of Markov chains. *Stoch. Process. Appl.* **87,** 281–297.

Chen, M.-F. (2005). *Eigenvalues, Inequalities and Ergodic Theory*. Springer, London.

Diaconis, P. and Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Prob.* **1,** 36–61.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, 3rd edn. John Wiley, New York.

Hansen, B. G. and Frenk, J. B. G. (1991). Some monotonicity properties of the delayed renewal function. *J. Appl. Prob.* **28,** 811–821.

Kalashnikov, V. V. (1994). *Topics on Regenerative Processes*. CRC Press, Boca Raton, FL.

Keilson, J. and Kester, A. (1978). Unimodality preservation in Markov chains. *Stoch. Process. Appl.* **7,** 179–190.

Kendall, D. G. (1959). Unitary dilations of Markov transition operators, and the corresponding integral representations for transition-probability matrices. In *Probability and Statistics*, ed. U. Grenander, John Wiley, New York, pp. 139–161.

Kijima, M. (1997). *Markov Processes for Stochastic Modeling*. Chapman and Hall, London.

Kulkarni, V. G. (1995). *Modeling and Analysis of Stochastic Systems*. Chapman and Hall, London.

Liggett, T. M. (1989). Total positivity and renewal theory. In *Probability, Statistics, and Mathematics*, eds T. W. Anderson, K. B. Athreya and D. L. Iglehart, Academic Press, Boston, MA, pp. 141–162.

Lindvall, E. T. (1992). *Lectures on the Coupling Method*. John Wiley, New York.

Lund, R. B. and Tweedie, R. L. (1996). Geometric convergence rates for stochastically ordered Markov chains. *Math. Operat. Res.* **21,** 182–194.

MEYN, S. P. AND TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, Berlin.

ROSENTHAL, J. S. (1995). Convergence rates for Markov chains. *SIAM Rev.* **37,** 387–405.

ROSS, S. M. (1996). *Stochastic Processes*, 2nd edn. John Wiley, New York.

SENGUPTA, D., CHATTERJEE, A. AND CHAKRABORTY, B. (1995). Reliability bounds and other inequalities for discrete life distributions. *Microelectron. Reliab.* **35,** 1473–1478.

SHAKED, M. AND SHANTHIKUMAR, J. G. (1987). IFRA properties of some Markov jump processes with general state space. *Math. Operat. Res.* **12,** 562–568.

SHAKED, M. AND SHANTHIKUMAR, J. G. (1994). *Stochastic Orders and Their Applications*. Academic Press, Boston, MA.

STOYAN, D. (1983). *Comparison Methods for Queues and other Stochastic Models*. John Wiley, Chichester.

STROOCK, D. W. (2005). *An Introduction to Markov Processes*. Springer, Berlin.

TUOMINEN, P. AND TWEEDIE, R. L. (1979). Exponential ergodicity in Markovian queueing and dam models. *J. Appl. Prob.* **16,** 867–880.