GEOMETRIC RENEWAL CONVERGENCE RATES FROM HAZARD RATES

KENNETH S. BERENHAUT* AND ROBERT LUND,* ** University of Georgia

Abstract

This paper studies the geometric convergence rate of a discrete renewal sequence to its limit. A general convergence rate is first derived from the hazard rates of the renewal lifetimes. This result is used to extract a good convergence rate when the lifetimes are ordered in the sense of new better than used or increasing hazard rate. A bound for the best possible geometric convergence rate is derived for lifetimes having a finite support. Examples demonstrating the utility and sharpness of the results are presented. Several of the examples study convergence rates for Markov chains.

Keywords: Geometric convergence; renewal sequence; Markov chains; power series; hazard rates; increasing hazard rate; new better than used

AMS 2000 Subject Classification: Primary 60K05
Secondary 60F99

1. Introduction

This paper studies geometric convergence rates for discrete renewal sequences via hazard rates. Specifically, let \( \{X_i\}_{i=1}^{\infty} \) be independent and identically distributed interrenewal lifetimes supported on some subset of \( \{1, 2, \ldots\} \). These lifetimes induce the familiar renewal sequence \( \{u_n\}_{n=0}^{\infty} \), where \( u_n \) is the probability of a renewal at time \( n \); the convention \( u_0 = 1 \) is made. We assume that \( E[X_1] < \infty \) and that \( X_1 \) is not supported on a sublattice of \( \{1, 2, \ldots\} \). Then \( u_n \to E[X_1]^{-1} := u_\infty \) as \( n \to \infty \) (cf. Smith (1958) and Feller (1968)).

It is frequently desirable, particularly in the context of Markov chains, to have an estimate of the speed of convergence of \( u_n \) to \( u_\infty \). Many authors have derived polynomial and other sub-exponential convergence rates under moment conditions such as \( E[X^\alpha] < \infty \) for some \( \alpha > 1 \) (cf. Gelfond (1964), Stone and Wainger (1967), Rogozin (1973), Pitman (1974), Lindvall (1979) and (1992), Ney (1981), Davies and Grubel (1981), Grubel (1983), Berbee (1987) and Konstantopoulos and Last (1999)). It is in general preferable to work with the ‘faster’ geometric rate; however, obtaining explicit geometric rates has proven to be difficult (cf. Meyn and Tweedie (1993) and (1994), Rosenthal (1995), Lund and Tweedie (1996) and Roberts and Tweedie (1999)). Our goal is to identify a good geometric convergence rate \( r > 1 \) and a finite first constant \( \kappa \) satisfying

\[
|u_n - u_\infty| \leq \kappa r^{-n}
\]

for each \( n \geq 0 \). Besides conceptual understanding, geometric convergence rates are useful for assessing sampled Markov chain convergence and in statistical inference problems (ergodic results for nonlinear time series, for example).

Received 25 April 2000; revision received 10 October 2000.
* Postal address: Department of Statistics, The University of Georgia, Athens, GA 30602-1952, USA.
** Email address: lund@stat.uga.edu
Some general aspects of geometric renewal convergence rates merit discussion up front. First, Kendall’s classical renewal theorem (Kendall (1959), Stone (1965)) states that an \( r > 1 \) satisfying (1.1) exists if and only if \( X \) has a geometric moment in the sense that there exists a real \( s > 1 \) such that \( F(s) = E[X^s] < \infty \) (which is henceforth assumed). Our notation here uses \( X \) as a generic lifetime whose distribution is identical to that of any \( X_i \). Proving the existence of some geometric rate is much easier than identifying explicit values of \( r > 1 \) and \( \kappa < \infty \) in (1.1).

Second, the largest \( r \) possible in (1.1) is frequently the magnitude of the smallest complex \( z \) with \( |z| > 1 \) satisfying \( F(z) = 1 \) (Heathcote (1967), Malyshev and Spieksma (1995)). To briefly see this here, define

\[
\Delta(z) = \sum_{n=0}^{\infty} (u_n - u_\infty) z^n
\]

for complex \( z \). The best geometric rate of convergence in (1.1) is the radius of convergence of \( \Delta \). Heathcote (1967) provides a key identity:

\[
\Delta(z) = \left( \frac{E[X^2] - E[X]}{2E[X]^2} \right) \frac{F^{(2)}(z)}{F^{(1)}(z)}
\]

(1.2)

for all \( |z| \leq 1 \), where \( F^{(1)}(z) = E[Z^{X^{(1)}}] \) and \( F^{(2)}(z) = E[Z^{X^{(2)}}] \) are the generating functions of the first and second distributions derived from the tails of \( X \). Specifically, \( P[X^{(1)} = k] = E[X]^{-1} P[X > k] \) and \( P[X^{(2)} = k] = E[X^{(1)}]^{-1} P[X^{(1)} > k] \) for \( k \geq 0 \). The generating functions \( F, F^{(1)} \), and \( F^{(2)} \) are related via

\[
F^{(1)}(z) = \frac{F(z) - 1}{E[X](z - 1)}, \quad F^{(2)}(z) = \frac{F^{(1)}(z) - 1}{E[X^{(1)}](z - 1)}.
\]

(1.3)

Convergence rate information can be extracted from (1.2). Let \( R_F \) denote the radius of convergence of \( F \). Then \( R_F \) is also the radius of convergence of \( F^{(1)} \) and \( F^{(2)} \). Suppose that \( F^{(1)}(z) \neq 0 \) for all \( z \) in the disc centered at zero with radius \( R \), where \( R \leq R_F \). Then \( F^{(2)}(z)/F^{(1)}(z) \) is analytic in \( |z| < R \) and (1.2) shows that \( \Delta(z) \) can be expanded into a power series which is absolutely convergent in \( |z| < R \). Taking \( R = |z_0| \), where \( z_0 \) is the smallest (in magnitude) solution to \( F^{(1)}(z) = 0 \), shows that \( |u_n - u_\infty| r^n \rightarrow 0 \) as \( n \rightarrow \infty \) for each real \( r > 0 \) with \( r < |z_0| \).

From (1.3), it follows that \( F^{(1)}(z) = 0 \) whenever \( z \neq 1 \) is such that \( F(z) = 1 \). Since \( F^{(1)} \) is a probability generating function, \( F^{(1)}(1) = 1 \) and \( z = 1 \) is not a solution to \( F^{(1)}(z) = 0 \). Equation (1.3) also shows that \( F^{(2)}(z) \neq 0 \) if \( F^{(1)}(z) = 0 \); hence, zeros of \( F^{(1)} \) and \( F^{(2)} \) cannot ‘cancel’ in (1.2) and the best geometric convergence rate in (1.1) cannot exceed \( |z_0| \). There exist lifetimes for which \( F^{(1)}(z) = 0 \) has no solutions in \( 1 < |z| < R_F \) (the geometric distribution is one such example); in these cases, convergence rates can exceed \( R_F \). Heathcote (1967) and Section 2.2 of Malyshev and Spieksma (1995) are good references for more details on the above issues.

With precise knowledge of the distribution of \( X \), we can in principle compute the smallest \( z \) such that \( F^{(1)}(z) = 0 \) and identify the largest geometric convergence rate possible. Two drawbacks arise with such a ‘root-finding’ approach. First, root finding will not explicitly identify a first constant \( \kappa \). Second, and perhaps more importantly, the distribution of \( X \) may not be precisely known. For example, in a denumerable Markov chain, the renewals are typically the return times to a fixed state in the chain and the distribution of \( X \) may only be known.
to an approximation (see Meyn and Tweedie (1993) and Example 3.6 here). The difficulties of extracting root locations from an approximate distribution is demonstrated in Example 3.1 below.

In this paper, we derive a general geometric convergence rate from the hazard rates of $X$. The derived rate simplifies when $X$ has some structural ordering properties. For example, when $0 < P[X = 1] < 1$ and $X$ is new better than used (NBU) in the sense that

$$P[X > i + j | X > i] \leq P[X > j]$$

for each $i, j \geq 0$ with $P[X > i] > 0$, then

$$|u_n - u_\infty| \leq \left( \frac{1}{P[X > 1]} \right)^{(n+1)}$$

for each $n \geq 0$, thereby identifying $r = P[X > 1]^{-1}$ as a geometric convergence rate and $\kappa = P[X > 1]^{-1}$ as a first constant. Pleasant aspects of this result lie in its simplicity and ease of application.


The rest of this paper proceeds as follows. Section 2 derives a general geometric convergence rate and first constant from the hazard rates of $X$. This result is used to obtain NBU and increasing hazard rate (IHR or IFR) convergence rates. The sharpness of the rates is then addressed. Section 3 presents six examples in both renewal and Markov chain contexts.

2. Results

Let $N = \sup\{n \geq 1 : P[X > n] > 0\}$ and note that $N$ may be infinite. When $N < \infty$, $X$ is supported on a subset of $\{1, 2, \ldots, N + 1\}$. Set $\Gamma = \{1, 2, \ldots, N\} = \{n \geq 1 : P[X > n] > 0\}$. When $0 < P[X = 1] < 1$, $X$ is non-degenerate and $\Gamma$ is not empty. The hazard rates of $X$ are denoted by $h_i = P[X = i | X \geq i]$ for $i \in \Gamma$. It will be convenient to work with the ratios $r_i := P[X > i - 1]/P[X > i]$ for $i \in \Gamma$. Note that $r_i \geq 1$ and $h_i = 1 - 1/r_i$ for $i \in \Gamma$. Useful facts are that

$$r_n \geq 1 \iff P[X = n] > 0, \quad 1 \leq n \leq N,$$

and

$$P[X > n] = \prod_{j=1}^{n} r_j^{-1}, \quad 1 \leq n \leq N.$$  \hspace{1cm} (2.2)

The following power series result will prove useful in the ensuing analysis.

**Lemma 2.1.** Suppose that $\{a_k\}_{k=0}^\infty$ is a sequence of real numbers and set

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

for each fixed $n \geq 0$ and $t \in (0, 1)$. 

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

for each fixed $n \geq 0$ and $t \in (0, 1)$. 

$$S_n(t) = \sum_{k=n+1}^{\infty} (a_k - a_{k-1})t^k$$
(a) If $|a_k| \leq C\gamma^k$ for all $k \geq 0$, where $C > 0$ and $\gamma > 1$ are some constants, then

$$|S_n(t)| \leq Ct \left[ 1 + \frac{\gamma - \gamma t}{1 - \gamma t} \right] (\gamma t)^n$$

for all $n \geq 0$ and $t \in (0, \gamma^{-1})$.

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

**Proof.** We will prove (a) only—the arguments for (b) are similar. Fix $n \geq 0$ and regroup the terms in (2.3) to get

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

(2.4)

Since $t \in (0, 1)$, $t^{k+1} - t^k \leq 0$ for all $k \geq 0$ and a sign analysis of (2.4) shows that $S_n(t)$ is maximized for each fixed $t$ (over all $\{a_k\}_{k=0}^{\infty}$ with $|a_k| \leq C\gamma^k$) by taking $a_n = C\gamma^n$ and $a_k = -C\gamma^k$ for $k \geq n+1$. Similarly, $S_n(t)$ is minimized (over all $\{a_k\}_{k=0}^{\infty}$ with $|a_k| \leq C\gamma^k$) by taking $a_n = -C\gamma^n$ and $a_k = C\gamma^k$ for $k \geq n+1$. Both these choices of $\{a_k\}_{k=0}^{\infty}$ give the same value of $|S_n(t)|$. The claimed bound for $|S_n(t)|$ follows by plugging the maximizing values of the $a_k$s into (2.3):

$$|S_n(t)| = C \left| (\gamma^n + \gamma^{n+1}) t^{n+1} + \sum_{k=n+2}^{\infty} (\gamma^k - \gamma^{k-1}) t^k \right|$$

$$\leq C \left[ \gamma^n t^{n+1} (1 + \gamma) + (1 - \gamma^{-1}) \sum_{k=n+2}^{\infty} (\gamma t)^k \right]$$

$$= Ct \left[ 1 + \frac{\gamma - \gamma t}{1 - \gamma t} \right] (\gamma t)^n.$$

**Theorem 2.1.** For any discrete renewal sequence $\{u_n\}_{n=0}^{\infty}$,

$$|u_n - u_\infty| \leq r_{\min}^{-(n+1)}$$

(2.5)

for each $n \geq 0$, where $r_{\min} = \inf_{i \in \Gamma} \{r_i\} = \inf_{i \in \Gamma} \{1 - h_i\}^{-1}$.

Theorem 2.1 gives a convergence rate that exceeds unity and a finite first constant whenever $r_{\min} > 1$. Lemma 2.2 addresses the case where $r_{\min} = 1$ in more detail.

**Proof.** The inequality (2.5) clearly holds when $r_{\min} = 1$. Hence, suppose that $r_{\min} > 1$ and fix $1 < \rho < r_{\min}$. Let $X^*$ be a lifetime with

$$P[X^* = n] = P[X > n - 1] \rho^{n-1} - P[X > n] \rho^n,$$ 

(2.6)

so that $P[X^* > n] = P[X > n] \rho^n$ for all $n \geq 0$. We show below that (2.6) defines a legitimate discrete lifetime distribution over $\{1, 2, \ldots\}$.

The governing recurrent event recursion is

$$u_n = \sum_{k=1}^{n} P[X = k] u_{n-k}, \quad n \geq 1,$$

(2.7)
and some work will manipulate this into the tail form

\[ P[X > n] = \sum_{k=1}^{n} (u_{k-1} - u_k) P[X > n - k], \quad n \geq 1. \] (2.8)

Versions of (2.7) and (2.8) also apply to \( X^\ast \) (our notation will use \( u^\ast_n \) as the probability of a renewal at time \( n \) in a renewal process with independent lifetimes each having the same distribution as \( X^\ast \)).

Multiplying both sides of (2.8) by \( \rho^n \) and applying \( P[X^\ast > n] = P[X > n] \rho^n \) gives

\[ P[X^\ast > n] = \sum_{k=1}^{n} \rho^k (u_{k-1} - u_k) P[X^\ast > n - k], \quad n \geq 1. \] (2.9)

Comparing (2.9) with the version of (2.8) for \( X^\ast \) and inductively equating coefficients gives our key relationship:

\[ (u_{k-1} - u_k) \rho^k = u^\ast_{k-1} - u^\ast_k, \quad k \geq 1. \] (2.10)

From this, we find that

\[ |u_n - u_\infty| = \sum_{i=n+1}^{\infty} (u_{i-1} - u_i) = \sum_{i=n+1}^{\infty} (u^\ast_{i-1} - u^\ast_i) \rho^{-i}. \]

Since \( u^\ast_k \) is a probability for each \( k, 0 \leq u^\ast_k \leq 1 \), and Lemma 2.1(b) applies with \( C = 1 \) and \( t = \rho^{-1} \): \( |u_n - u_\infty| \leq \rho^{-n(1+\delta)} \). Letting \( \rho \uparrow \rho_{\min} \) gives the result in the theorem.

It remains to show that (2.6) defines a legitimate lifetime distribution on \( \{1, 2, \ldots\} \). To do this, it is sufficient to show that \( P[X > n] \rho^n \) is nonincreasing for \( n \in \Gamma \) and that \( \lim_{n \to \infty} P[X > n] \rho^n = 0 \). By the choice of \( \rho \), \( P[X > n - 1]/P[X > n] > \rho \) for all \( n \in \Gamma \). Hence,

\[ \frac{P[X > n - 1] \rho^{n-1}}{P[X > n] \rho^n} = \frac{P[X > n - 1]}{P[X > n]} \frac{1}{\rho} > 1 \]

for all \( n \in \Gamma \) and \( P[X > n] \rho^n \) is nonincreasing in \( n \in \Gamma \) as claimed.

To show that \( P[X > n] \rho^n \to 0 \) as \( n \to \infty \), it is sufficient to consider the case where \( N = \infty \) (when \( N < \infty P[X > n] = 0 \) for large \( n \)). Choose \( \delta > 0 \) so small that \( r_n \geq \rho + \delta \) for all \( n \geq 1 \). Then (2.2) gives

\[ P[X > n] \rho^n = \rho^n \prod_{j=1}^{n} r_j^{-1} \leq \left( \frac{\rho}{\rho + \delta} \right)^n, \]

which shows that \( P[X > n] \rho^n \to 0 \) as \( n \to \infty \) and completes our work.

**Remark 2.1.** When \( X \) has the geometric distribution \( P[X = n] = pq^{n-1} \) for \( n \geq 1 \) with \( 0 < p < 1 \) and \( p + q = 1 \), then \( u_0 - u_\infty = r_{\min}^{-1} \) and the rate and first constant obtained in Theorem 2.1 is exact. Hence, Theorem 2.1 cannot be improved in generality.

**Remark 2.2.** We can obtain \( |u_n - u_\infty| \leq r_{\min}^{-n} \) for \( n \geq 0 \) by a coupling-splitting argument with a uniform minorization condition (cf. Roberts and Polson (1994), Rosenthal (1995, Fact 10)). The first constant in Theorem 2.1 is better by a factor of \( r_{\min} \). Also, the proof is an elementary scale change argument which will provide us with insight into the result's optimality and improvement below (see Theorems 2.2 and 2.3 and Lemma 2.2).
A convergence rate for NBU lifetimes follows easily from Theorem 2.1. Using (1.4), we obtain $P[X > i] \leq P[X > i - 1]P[X > 1]$ for all $i \in \Gamma$. Hence,

$$r_i = \frac{P[X > i - 1]}{P[X > i]} \geq \frac{1}{P[X > 1]},$$

for all $i \in \Gamma$ and we have proved the following result.

**Corollary 2.1.** For any discrete renewal sequence $\{u_n\}_{n=0}^{\infty}$ with NBU lifetimes,

$$|u_n - u_\infty| \leq \left(\frac{1}{P[X > 1]}\right)^{(n+1)}$$

for each $n \geq 0$.

We will retain the notation introduced in the proof of Theorem 2.1 below. Specifically, for a fixed $\rho > 1$, $X^*$ denotes a 'lifetime' with tail 'probabilities' $P[X^* > n] = P[X > n]\rho^n$ for $n \geq 0$. The quotations qualify lifetime and probability as $\{P[X^* > n]\rho^n\}_{n=0}^{\infty}$ may not be a proper tail probability sequence for all values of $\rho$; however, (2.7) and (2.8) remain valid and this is all that will be needed. Superscripts of * will refer to quantities in a renewal process with lifetimes each having the same distribution as that of $X^*$; for instance,

$$r^*_n = \frac{P[X^* > n - 1]}{P[X^* > n]} = \frac{r_n}{\rho}, \quad 1 \leq n \leq N. \quad (2.11)$$

Note that starred versions of (2.1) and (2.2) hold for $1 \leq i \leq N$.

Our next result investigates the optimality of Theorem 2.1.

**Theorem 2.2.** Suppose that $\{u_n\}_{n=0}^{\infty}$ is any renewal sequence with $N < \infty$. Set $\Gamma' = \{n \leq N : r_n > r_{\text{min}}\} \cup \{N + 1\}$. If the greatest common divisor of elements in $\Gamma'$ exceeds unity, then $r_{\text{min}}$ is the largest geometric convergence rate possible in (1.1).

**Proof.** In the proof of Theorem 2.1, a rate was obtained by rescaling the renewal probability differences via (2.10). Sometimes, $|u^*_{n-1} - u^*_n| \to 0$ geometrically as $n \to \infty$ and this fact can be used in (2.10) to obtain a better convergence rate. Under the above assumptions, we will show that this is not the case.

Take $\rho = r_{\text{min}}$. Since $N < \infty$, the support set of $X$ is finite, $P[X^* > n] \to 0$ as $n \to \infty$, and the arguments in the proof of Theorem 2.1 show that $X^*$ is a legitimate lifetime random variable.

We now show that $\Gamma'$ is the set of $n$ where $P[X^* = n] > 0$. Note that $\Gamma'$ contains $N + 1$ (by assumption) and that $P[X^* = N + 1] = P[X = N]\rho^N > 0$ follows from (2.6) with $P[X > N + 1] = 0$. Now for $1 \leq n \leq N$, $n \in \Gamma'$ if and only if $r_n > r_{\text{min}}$. Using $\rho = r_{\text{min}}$, (2.11) and the starred version of (2.1) shows that $n \in \Gamma'$ if and only if $P[X^* = n] > 0$. Hence, $\Gamma'$ is the support set of $X^*$ as claimed.

Let $L > 1$ be the greatest common divisor of all elements in $\Gamma'$. By hypothesis, $L > 1$ and $X^*$ is supported on a lattice with span $L$. Standard renewal results for lattice lifetimes (cf. Feller (1968)) show that $u^*_{nL} \to L(\sum_{k=1}^{\infty} kP[X^* = k])^{-1} > 0$ as $n \to \infty$ and that $u^*_n = 0$ for every $n$ not evenly divisible by $L$. Combining this with (2.10) gives

$$|u_{nL-1} - u_{nL}|r^*_n = |u^*_{nL-1} - u^*_{nL}| = |u^*_{nL}| > C > 0 \quad (2.12)$$

for some positive constant $C$ and all sufficiently large $n$. 

Suppose now that \(|u_n - u_\infty|r_{\text{min}}^n \to 0\) as \(n \to \infty\). Then
\[
|u_{n-1} - u_n|r_{\text{min}}^n \leq r_{\text{min}}|u_{n-1} - u_\infty|r_{\text{min}}^{n-1} + |u_n - u_\infty|r_{\text{min}}^n \to 0
\]
as \(n \to \infty\), which contradicts (2.12). Thus, the rate in Theorem 2.1 is optimal.

**Remark 2.3.** The finiteness of \(N\) assumed in Theorem 2.2 is only used to show that \(r_{\text{min}}^n P[X > n] \to 0\) as \(n \to \infty\). Theorem 2.2 remains true when \(N = \infty\) if \(r_{\text{min}}^n P[X > n] \to 0\) as \(n \to \infty\).

A lifetime \(X\) is said to have an increasing hazard rate (IHR) if \(h_i\) (or equivalently \(r_i\)) is increasing in \(i \in \Gamma\). All IHR lifetimes are NBU (Kijima (1997, Theorem 3.7)), but there exist NBU lifetimes that are not IHR (cf. Bryson and Siddiqui (1969)). Parametric examples of IHR lifetimes include the uniform distribution over \(\{1, 2, \ldots, N + 1\}\) and the negative binomial distribution. Since convolutions of IHR lifetimes are again IHR (see Kijima (1997, Theorem 3.6)), lifetimes which are independent sums of \(k\) IHR lifetimes are also IHR. Such convolution lifetimes arise in systems with \(k - 1\) spares.

Corollary 2.1 shows that \(P[X > 1]^{-1}\) is a geometric convergence rate for the IHR class. We will improve this rate for strictly IHR lifetimes below in Theorem 2.3. Before that, the case where \(P[X = 1] = 0\) (i.e., \(r_1 = 1\)) is studied. Lifetimes with \(P[X = 1] = 0\) arise within the IHR class. The proof of the following result is given in Berenhaut and Lund (2000).

**Lemma 2.2.** For any renewal sequence \(\{u_n\}_{n=0}^{\infty}\) with \(P[X = 1] = 0\) and \(P[X = 2] > 0\) (equivalently, \(1 = r_1 < r_2\)) and \(r_2 \leq \inf_{3 \leq n \leq N} \{r_n\}\),
\[
|u_n - u_\infty| \leq \frac{K(r_2)}{\sqrt{r_2}} \left[ 1 + \frac{\sqrt{r_2} - 1}{\sqrt{r_2} \eta(r_2) - 1} \right] [\sqrt{r_2} \eta(r_2)]^{-n}, \quad n \geq 3, \tag{2.13}
\]
where \(\eta(r_2)\) is the smallest (in magnitude) root of the cubic polynomial
\[
z^3 + (\sqrt{r_2} - 1)z^2 + (\sqrt{r_2} - 1)z - 1. \tag{2.14}
\]
The other constants in (2.13) are \(r_2 = P[X > 2]^{-1}\) and \(K(r_2) = 3 \max_{1 \leq i \leq 3} \kappa_i(r_2)\), where
\[
\kappa_i(r_2) = \begin{cases}
1, & r_2 \geq 4, \\
\frac{a^2 + a + 1}{a + 2a^2 + 1} & 2.618034 \leq r_2 < 4, \\
\frac{a^3 + 2a^2 + 1}{a + 2a^2 + 1} & 1 < r_2 < 2.618034
\end{cases}
\]
(here \(\lambda_i(r_2), 1 \leq i \leq 3\), are the three roots of (2.14) and \(a = 1 - \sqrt{r_2}\)).

Finally, for \(\rho = \sqrt{r_2}\), \(u_n^*\) admits the geometric bound
\[
|u_n^*| \leq K(r_2) \eta(r_2)^{-n}, \quad n \geq 3. \tag{2.15}
\]

Table 1 numerically lists the rate and first constant of Lemma 2.2. Note that the rate exceeds unity for all \(r_2 > 1\).

The results in Lemma 2.2 also hold when \(\sqrt{r_2} < \inf_{3 \leq n \leq N} \{r_n\}\) (Berenhaut and Lund (2000)). This fact is now used to improve the Theorem 2.1 rate for ‘strictly’ IHR lifetimes.
TABLE 1: Rates and first constants of Lemma 2.2.

<table>
<thead>
<tr>
<th>Rate</th>
<th>First Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_2$</td>
<td>$r_1$</td>
</tr>
<tr>
<td>1.01</td>
<td>11.901118</td>
</tr>
<tr>
<td>1.05</td>
<td>11.527072</td>
</tr>
<tr>
<td>1.10</td>
<td>11.104149</td>
</tr>
<tr>
<td>1.50</td>
<td>8.995885</td>
</tr>
<tr>
<td>2.00</td>
<td>8.179491</td>
</tr>
<tr>
<td>4.00</td>
<td>11.543913</td>
</tr>
<tr>
<td>10.00</td>
<td>28.7401234</td>
</tr>
<tr>
<td>100.00</td>
<td>297.5916478</td>
</tr>
</tbody>
</table>

**Theorem 2.3.** Suppose that $\{u_n\}_{n=0}^\infty$ is any discrete renewal sequence with $r_1 < r_2$ and 

$$\sqrt{r_1 r_2} \leq \inf_{3 \leq n \leq N} \{r_n\}. \quad (2.16)$$

Then

$$|u_n - u_\infty| \leq \frac{K(r_2/r_1)}{\sqrt{r_1 r_2}} \left[ 1 + \frac{\sqrt{r_1 r_2} - 1}{\sqrt{r_1 r_2} \eta(r_2/r_1) - 1} \right] \left[ \frac{r_2}{r_1} \right]^n \quad (2.17)$$

for each $n \geq 3$. In particular, if $X$ is strictly IHR ($r_n$ or $h_n$ is strictly increasing over $n \geq 1$), then the rate and first constant in (2.17) hold.

**Proof.** Take $\rho = r_1$ and note that $P[X^* = 1] = 0$. We will show that Lemma 2.2 applies to $X^*$. Equations (2.11) and (2.16) give $r_1^* = 1, r_2^* > 1$, and

$$\inf_{3 \leq n \leq N} \{r_n^*\} = \inf_{3 \leq n \leq N} \left( \frac{r_n}{r_1} \right) > \frac{\sqrt{r_1 r_2}}{r_1} = \sqrt{r_2^*}. \quad (2.17)$$

The comment after Lemma 2.2 shows that the conclusions of that result apply to $X^*$. In particular, (2.15) provides the geometric bound

$$|u_n^{**}| \leq K(r_2^*)^n \eta(r_2^*)^{-n}, \quad n \geq 3, \quad (2.18)$$

where $\{u_n^{**}\}_{n=0}^\infty$ are the renewal probabilities induced by a ‘lifetime’ $X^{**}$ with $P[X^{**} > n] = P[X^* > n]/(r_2^*)^{n/2}$ for $n \geq 0$. As noted above, $\{P[X^{**} > n]\}_{n=0}^\infty$ may not be a legitimate tail probability sequence (but this is not needed).

The scale change arguments in (2.10) give

$$u_n - u_n = (u_n - u_n^*) r_1^n = (u_n^{**} - u_n^{**}) [r_1 \sqrt{r_2^*}]^{-n}, \quad n \geq 1. \quad (2.19)$$

By (2.18), Lemma 2.1(a) applies to $\{u_k^{**}\}_{k=3}^\infty$ with $C = K(r_2^*)$ and $\gamma = \eta(r_2^*)^{-1}$. Using (2.19) and Lemma 2.1 with $t = (r_1 \sqrt{r_2^*})^{-1}$ gives

$$|u_n - u_\infty| = \left| \sum_{i=n+1}^{\infty} (u_{i-1} - u_i) \right| = \left| \sum_{i=n+1}^{\infty} (u_i^{**} - u_i^{**}) [r_1 \sqrt{r_2^*}]^{-i} \right| \leq \frac{K(r_2^*)}{r_1 \sqrt{r_2^*}} \left[ 1 + \frac{r_1 \sqrt{r_2^*} - 1}{r_1 \sqrt{r_2^*} \gamma - 1} \right] [r_1 \sqrt{r_2^*} \eta(r_2^*)]^{-n}, \quad (2.20)$$

for $n \geq 3$. Applying $r_2^* = r_2/r_1$ in (2.20) proves (2.17) and finishes our work.
Our final result bounds the largest convergence rate possible in (1.1) when $N < \infty$. We stress that the result is not true when $N$ is infinite.

**Theorem 2.4.** Suppose that $\{u_n\}_{n=0}^{\infty}$ is any discrete renewal sequence with $N < \infty$. Then the largest convergence rate possible in (1.1) cannot exceed

$$r_{\text{max}} = \max_{i \in \Gamma} \{r_i\} = \max_{i \in \Gamma} (1 - h_i)^{-1}.$$

**Proof.** Recall from Section 1 that the largest convergence rate in (1.1) cannot exceed the magnitude of the smallest zero of $F^{(1)}$. Hence, we must show that a zero of $F^{(1)}$ exists in the region $1 < |z| \leq r_{\text{max}}$. Since $F^{(1)}$ is a polynomial of degree $N \geq 1$, it is sufficient to show that $F^{(1)}(z) \neq 0$ for $|z| > r_{\text{max}}$, as the fundamental theorem of algebra would then guarantee that a zero of $F^{(1)}$ exists in $1 < |z| \leq r_{\text{max}}$. Recall that $F^{(1)}$ cannot have a zero in $|z| \leq 1$ (see (1.2)). Let $q_n = P[X > n]$ for $n \geq 0$ and set

$$Q(z) = \sum_{k=0}^{N} q_k z^k. \tag{2.21}$$

Since $F^{(1)}(z) = E[X|X|^{-1} Q(z)$, our work reduces to proving that $Q$ has at least one zero in $1 < |z| \leq r_{\text{max}}$. We will show that $Q$ has no zeroes in $|z| > r_{\text{max}}$. Because $q_0 = 1$ and $q_n \geq 0$ for $n \geq 0$, $Q$ cannot have a positive real root and attention is confined to those $z$ that are not positive reals. We rescale our work into showing that $Q(r_{\text{max}} z)$ has no roots in $|z| > 1$.

Suppose that $|z| > 1$ and that $z$ is not a positive real. Take $\rho = r_{\text{max}}$ and use (2.21) and $q^*_k = q_k r_{\text{max}}^{k}$ to get

$$|(1 - z^{-1}) Q(r_{\text{max}} z)| = \left| \sum_{k=0}^{N} q^*_k z^k - \sum_{k=0}^{N} q^*_k z^{k-1} \right| = \left| q^*_N z^N - q^*_0 z^{-1} - \sum_{k=0}^{N-1} (q^*_{k+1} - q^*_k) z^k \right| \geq |q^*_N z^N| - |q^*_0 z^{-1} + \sum_{k=0}^{N-1} (q^*_{k+1} - q^*_k) z^k|, \tag{2.22}$$

where the last line is merely the triangle inequality.

Since $r_i \leq \rho$ for each $i \in \Gamma$,

$$q^*_0 \leq q^*_1 \leq \cdots \leq q^*_N. \tag{2.23}$$

Because $z$ is not positive and real, $\arg(z^k)$ is nonconstant as $k$ ranges over $\{-1, 0, \ldots, N - 1\}$ and a strict version of the triangle inequality holds for the sum in (2.22). Using this, (2.23) and $|z^k| < |z^N|$ for $k \in \{-1, 0, \ldots, N - 1\}$ give

$$\left| q^*_0 z^{-1} + \sum_{k=0}^{N-1} (q^*_{k+1} - q^*_k) z^k \right| < |q^*_0||z^{-1}| + \sum_{k=0}^{N-1} |q^*_{k+1} - q^*_k||z^k| \leq |z^N| \left[ q^*_0 + \sum_{k=0}^{N-1} (q^*_{k+1} - q^*_k) \right] = q^*_N |z|^N. \tag{2.24}$$

Combining (2.22) and (2.24) gives $|(1 - z^{-1}) Q(r_{\text{max}} z)| > 0$, which implies that $|Q(r_{\text{max}} z)| > 0$ since $1 - z^{-1} \neq 0$. 


3. Examples

This section presents examples demonstrating the applicability and optimality of the Section 2 results. Our first three examples consider parametric lifetime distributions where the largest convergence rates can be computed and compared to the bounds obtained. Subsequent examples move to Markov chain applications.

Example 3.1. Suppose that $X$ has the geometric distribution $P[X = n] = pq^{n-1}$, $n = 1, 2, \ldots$, where $0 < q < 1$ and $p + q = 1$. In this case, $u_n = p$ for $n \geq 1$, $|u_n - u_\infty| = 0$ for $n \geq 1$, and the largest geometric convergence rate is infinity. Here, $r_i = q^{-1}$ for all $i \geq 1$ and Theorem 2.1 shows that $q^{-1}$ is a geometric convergence rate with first constant $q^{-1}$:

$$|u_n - u_\infty| \leq q^{n+1}.$$  \hspace{1cm} (3.1)

Equality is achieved in (3.1) when $n = 0$.

Now consider the truncated geometric lifetime $\min(X, N + 1)$. This distribution counts the number of games that a team plays in a single elimination tournament of $2^{N+1}$ equally capable teams ($p = \frac{1}{2}$). Here, $P[X = n] = pq^{n-1}$ for $1 \leq n \leq N$, $P[X = N + 1] = q^N$, and $P[X = k] = 0$ for $k \geq N + 2$. For this distribution, $r_n = q^{-1}$ for each $n \in \Gamma = \{1, 2, \ldots, N\}$. Applying Theorem 2.1 shows that (3.1) again holds; however, the limiting renewal probability is now $u_\infty = p(1 - q^{N+1})^{-1} = \text{E}[\min(X, N + 1)]^{-1}$. Theorem 2.4 shows that the largest geometric convergence rate for truncated geometric lifetimes cannot exceed $q^{-1}$. Hence, the rate we obtain is exact and the results have identified the optimal geometric rate of convergence (note that Theorem 2.2 is also applicable here).

When $N = 1$, $X$ is supported on $\{1, 2\}$ with $P[X = 1] = p$ and $P[X = 2] = 1 - p$. Some tedious computations produce the exact relations

$$u_n - u_\infty = \begin{cases} 
\frac{1}{1 + q}q^{n+1}, & n \text{ even,} \\
-\frac{1}{1 + q}q^{n+1}, & n \text{ odd.}
\end{cases}$$

Our results again identify the largest geometric convergence rate (the first constant is suboptimal by a factor of $1 + q$).

It may be surprising that the convergence rates for $X$ and $\min(X, N + 1)$ are so different, as the two distributions are close for large $N$ in terms of total variation separation. This demonstrates the sensitivity of the convergence rate problem and why location of roots of power series are difficult to quantify. It is also worth noting that convergence rates are not entirely tied to large values of $X$ as $\min(X, N + 1) \leq X$, $\min(X, N + 1)$ has the finite convergence rate $q^{-1}$, while $X$ has an infinite convergence rate.

Example 3.2. Suppose that $X$ is uniformly distributed over $\{1, 2, \ldots, N + 1\}$. Then $P[X = n] = 1/(N + 1)$ for $1 \leq n \leq N + 1$ and $P[X > n] = (N + 1 - n)/(N + 1)$ for $0 \leq n \leq N$. Hence, $r_n = (N - n + 2)/(N - n + 1)$ for $n \in \Gamma = \{1, 2, \ldots, N\}$ and $X$ is strictly IH. Note that $r_{\min} = (N + 1)/N$ and $r_{\max} = 2$. By Theorems 2.1 and 2.4, the best geometric convergence rate is between $(N + 1)/N$ and 2. Theorem 2.1 gives the bound

$$|u_n - u_\infty| \leq \left(\frac{N + 1}{N}\right)^{-(n+1)},$$

where $u_\infty = 2/(N + 2) = \text{E}[X]^{-1}$. 
Since \( r_n \) is strictly increasing in \( n \), Theorem 2.3 also applies, so

\[
|u_n - u_\infty| \leq K(\beta_N) \sqrt{\frac{N - 1}{N + 1}} \left[ 1 + \frac{\sqrt{N + 1} - \sqrt{N - 1}}{\sqrt{N + 1}\eta(\beta_N) - \sqrt{N - 1}} \right] \left[ \frac{N + 1}{\sqrt{N - 1}} \eta(\beta_N) \right]^{-n},
\]

where \( \eta(\beta_N) \) is the smallest root of the cubic \( z^3 + (\beta_N^{-1/2} - 1)z^2 + (\beta_N^{-1/2} - 1)z - 1, \beta_N = N^2/(N^2 - 1), \) and \( K(\beta_N) \) is the constant discussed in Lemma 2.2.

Table 2 numerically compares the rates in Theorems 2.1 and 2.4 with the best convergence rate. The best convergence rates were obtained by numerically finding the smallest root of \( F^{(1)}(z) = 0 \). Overall, the rate bounds appear quite good.

**Example 3.3.** Our final parametric example studies the negative binomial waiting time where \( X \) is the number of trials needed to obtain two successes from independent trials each having success probability \( p \). Specifically, \( P[X = n] = (n - 1)(1 - p)^{n-2}p^2 \) for \( n \geq 2 \) and \( P[X = 1] = 0 \). This distribution is IHR, as it is the convolution of two independent geometric lifetimes, each of which are IHR. Straightforward computations give

\[
P[X > n] = (1 - p)^{n-1}[1 + p(n - 1)], \quad n \geq 1,
\]

with \( P[X > 0] = 1 \), and

\[
r_n = \frac{1 + p(n - 2)}{(1 - p)[1 + p(n - 1)]}, \quad n \geq 2,
\]

with \( r_1 = 1 \).

Lemma 2.2 provides a geometric renewal convergence rate and first constant. The results are numerically summarized in Table 3 (the algebraic computations, similar to the last example, are omitted). The best geometric convergence rate to \( u_\infty = p/2 \), obtained by finding the smallest root of \( F^{(1)}(z) = 0 \), is \( |1 - 2p|^{-1} \). Overall, the rate bounds are not as good as in the previous two examples. Good convergence rates for lifetimes with \( P[X = 1] = 0 \) appear to be harder to obtain.

We now consider applications to Markov chains.

**Example 3.4.** Consider a Markov chain \( \{X_n\}_{n=0}^{\infty} \) on the state space \( \{1, 2, \ldots\} \) with transition probability matrix \( P = [p_{i,j}]_{i,j=1}^{\infty} \) where only transitions up one state or to the bottom state are allowed: \( p_{n,n+1} = \alpha_n \) and \( p_{n,1} = 1 - \alpha_n \) for each \( n \geq 1 \). All other transitions have zero probability.
Table 3: Comparison of convergence rates.

<table>
<thead>
<tr>
<th>p</th>
<th>Lemma 2.2 rate</th>
<th>Actual rate</th>
<th>First constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>1.000417</td>
<td>1.111111</td>
<td>11.975008</td>
</tr>
<tr>
<td>0.10</td>
<td>1.001668</td>
<td>1.250000</td>
<td>11.900131</td>
</tr>
<tr>
<td>0.25</td>
<td>1.010463</td>
<td>2.000000</td>
<td>11.380832</td>
</tr>
<tr>
<td>0.50</td>
<td>1.041831</td>
<td>∞</td>
<td>9.649191</td>
</tr>
<tr>
<td>0.75</td>
<td>1.084683</td>
<td>2.000000</td>
<td>8.185113</td>
</tr>
<tr>
<td>0.90</td>
<td>1.079731</td>
<td>1.250000</td>
<td>15.070778</td>
</tr>
<tr>
<td>0.95</td>
<td>1.054720</td>
<td>1.111111</td>
<td>29.490192</td>
</tr>
</tbody>
</table>

Suppose that $X_0 = 1$ and consider the first return time $\tau_1 = \inf\{n \geq 1 : X_n = 1\}$. The structure of the chain gives

$$P[\tau_1 > n] = \prod_{i=1}^{n} \alpha_i, \quad n \geq 0,$$

(3.2)

and it follows that $\alpha_n = \alpha_{n}^{-1}$ for $n \geq 1$.

To avoid trite work, assume that $0 < \alpha_n < 1$ for all $n \geq 1$ and $E[\tau_1] = \sum_{n=0}^{\infty} \prod_{i=1}^{n} \alpha_i < \infty$. Then $E[\tau_1] < \infty$ and the state 1 renewal probabilities $u_n = P[X_n = 1 | X_0 = 1]$ are aperiodic. Theorem 2.1 gives

$$|u_n - u_\infty| \leq \alpha_{\text{max}}^{(n+1)},$$

(3.3)

where $\alpha_{\text{max}} = \sup_{n \geq 1} \alpha_n$ and $u_\infty = E[\tau_1]^{-1}$ is the limiting probability of being in state 1. Equation (3.3) gives an 'informative' convergence rate and finite first constant whenever $\sup_{n \geq 1} \alpha_n < 1$. Whereas the bound $|u_n - u_\infty| \leq \alpha_{\text{max}}^n$ can also be obtained from a coupling–minorization argument (cf. Rosenthal (1995, Fact 10)), it is not evident how to obtain the sharper bound in (3.3) with such methods.

In principle, we can attempt to improve the above rates with the methods of Roberts and Tweedie (1999). We address this further in the next example. Another avenue of improvement follows from Theorem 2.3: when $\alpha_n$ is strictly decreasing in $n$, (3.2) shows that $\tau_1$ is IHR and, hence, also NBU. Shanthikumar (1984) and Marshall and Shaked (1986) give other examples of Markov chains with a 'NBU state'. Theorem 2.3 provides a better convergence rate than $\alpha_{\text{max}}$ (the first constant may be larger however). The parameters in (2.17) are $r_1 = \alpha_{-1}$ and $r_2 = \alpha_2^{-1}$.

The above example is quite general. Indeed, any discrete lifetime distribution is the return time $\tau_1$ of some chain with the above structure. The connection is that $X_n$ is the age of the lifetime in place at time $n$ increased by one unit. Specifically, given a lifetime $X$, set $\alpha_n = 1 - h_n$. Then straightforward manipulations give

$$P[\tau_1 = k] = p_{k,1} \left( \prod_{n=1}^{k-1} p_{n,n+1} \right) = \frac{P[X = k]}{P[X \geq k]} \prod_{n=1}^{k-1} \frac{P[X \geq n + 1]}{P[X \geq n]}$$

$$= P[X = k]$$

as claimed.
Finally, suppose that
\[
p_{n,1} = \begin{cases} \alpha_*, & n \text{ odd,} \\ \alpha^*, & n \text{ even,} \end{cases}
\]
with \( p_{n,n+1} = 1 - p_{n,1} \), where \( 0 < \alpha_* < \alpha^* < 1 \). Then the convergence rate in (3.3) is \( \alpha_{\text{max}} = 1 - \alpha_* \) and Theorem 2.2 and Remark 2.3 show that this rate is optimal.

**Example 3.5.** It is instructive to compare the above rates to the recent Markov chain rates in Roberts and Tweedie (1999). Since the rates in Example 3.1 are optimal, our comparison will concentrate on the Example 3.2 lifetime where \( X \) is uniformly distributed over \( \{1, 2, \ldots, N + 1\} \) and the rates could, in principle, be improved. We will only compare rates; the first constants in Section 2 are usually smaller and easier to compute than those in Roberts and Tweedie (1999). Because of this, the rates from the two methods are not strictly ordered for all \( n \), regardless of which rate is superior.

Let \( \{X_n\} \) be the age Markov chain defined in Example 3.4. Roberts and Tweedie (1999) derive a total variational convergence rate for \( \{X_n\} \) to its invariant measure; this rate must also be a renewal convergence rate. The methods of Roberts and Tweedie (1999) require that we identify a small set \( C \) of the state space, a splitting probability \( \varepsilon \in (0, 1) \), and a minorizing probability measure \( \varphi \) satisfying \( P(x, A) \geq \varepsilon \varphi(A) \) for each \( x \in C \) and all measurable subsets \( A \). We must also find a solution to the drift inequality
\[
E[V(X_1) \mid X_0 = x] \leq \lambda V(x) + b1_C(k),
\]
for some finite \( b, \lambda < 1 \), and some drift function \( V \) with \( V(x) \geq 1 \) for all \( x \) in the state space.

For the uniform lifetime in Example 3.2, we can reproduce the \( (N + 1)/N \) convergence rate in Theorem 2.1 by selecting \( C = \{1, \ldots, N + 1\} \) (the whole space is small), \( \varphi \) to be the point mass at unity, and \( \varepsilon = (N + 1)^{-1} \) (see Roberts and Tweedie (1999, Theorem 5.2)). For any choice of \( \lambda < 1 \), the drift function is \( V(x) \equiv 1 \) with \( b = 1 - \lambda \). Whereas the simultaneous selection of the above parameters is extremely intensive, numerical optimization of the rates was investigated when \( N = 3 \). Table 4 summarizes these computations.

Approximate optimization of the drift parameters was accomplished by a grid search—absolute optimality is not claimed. No patterns of good drift and minorizing parameter choices were clear. For example, the rate for \( C = \{1, 2, 3\} \) was obtained with \( \varepsilon = 0.25, \varphi(\{1\}) = 1, V(1) = 1, V(2) = 1, V(3) = 2, V(4) = 3, \lambda = 0.4 \) and \( b = 1.267 \). For the uniform lifetime, the Section 2 rates are superior to any of the above ‘drift–minorizing’ rates.
Example 3.6. Consider a Markov disease progression model on the severity categories \{1, 2, \ldots, N\}. State 1 represents the onset (new patient) of the disease and state N signifies imminent death. The transition matrix is denoted by \( P = [p_{i,j}]_{i,j=1}^{N} \).

If the disease is in \( \{k\} \) at time \( n, 1 \leq k < N \), then death occurs at time \( n+1 \) with probability \( p_{k,1} \) and \( X_{n+1} = 1 \), signifying that a new patient is under examination. Transitions are uniform to all worsening states: \( p_{k,j} = (N-k)^{-1}(1-p_{k,1}) \) for \( k+1 \leq j \leq N \). The disease progresses monotonically in the sense that transitions to better states of health are not allowed: \( p_{k,j} = 0 \) when \( j < k \) and \( j > 1 \). Transitions from state \( N \) must go to state 1 (\( p_{N,1} = 1 \)), signifying death and that a new patient is being examined.

Suppose that \( p_{k,1} \) is increasing in \( k \). Then death is more likely as the disease progresses and \( \tau_{1} = \inf\{n \geq 1 : X_{n} = 1\} \) (with \( X_{0} = 1 \)) is IHR (we omit formal algebraic justification). Apply Theorem 2.1 to get that

\[
|P[X_{n} = 1 | X_{0} = 1] - u_{\infty}| \leq (1 - p_{1,1})^{(n+1)},
\]

where \( u_{\infty} \) is the state 1 limiting probability. Better convergence rates can again be obtained from Theorem 2.3. In closing, we note that the majority of 'structure' in the transitions is not used to obtain the above convergence rate. Also, the distribution of \( \tau_{1} \) is not readily accessible in closed form.

Acknowledgements

The authors thank the National Science Foundation (Grant DMS-9703838) for financial support. We are very thankful to a referee for comments and insights that substantially improved this manuscript.

References


