
DOI: https://doi.org/10.4213/tvp2444

Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use

http://www.mathnet.ru/eng/agreement

Download details:
IP: 54.70.40.11
March 22, 2019, 22:25:36
ASYMPTOTIC EXPONENTIALITY OF THE DISTRIBUTION OF FIRST EXIT TIMES FOR A CLASS OF MARKOV PROCESSES WITH APPLICATIONS TO QUICKEST CHANGE DETECTION

Рассматривается момент первого выхода неотрицательного возвратного по Харрису марковского процесса из интервала \([0, A]\) при \(A \to \infty\). Предлагается метод доказательства асимптотической экспоненциальной распределенности (подходящим образом нормированного) момента первого выхода, не опирающийся на обычно используемое вложение в процесс регенерации. Показано, что при определенных условиях производящая функция моментов этого нормированного момента первого выхода сходится к производящей функции моментов экспоненциального распределения с единичным параметром, и устанавливается связь между нормирующей константой и квазистационарным распределением (в предположении, что оно существует). Полученные результаты применяются для оценки распределения времени до ложной тревоги в задачах обнаружения разладки.

Ключевые слова и фразы: марковский процесс, стационарное распределение, квазистационарное распределение, момент первого выхода, асимптотическая экспоненциальная распределенность, задачи о разладке, CUSUM-процедуры, процедуры Ширяева–Робертса.

1. Introduction. Let \((\Omega, \mathcal{F}, P)\) be a probability space and \(\{X(n)\}, n = 0, 1, 2, \ldots\), a discrete-time nonnegative Harris-recurrent Markov process defined on this space. The limiting distribution as \(A \to \infty\) of the suitably standardized first exit time of the process from the interval \([0, A]\) turns out often to be exponential.

* The Hebrew University of Jerusalem, Department of Statistics, Mount Scopus, Jerusalem 91905, Israel; e-mail: msmp@mscc.huji.ac.il
** University of Southern California, Department of Mathematics, 3620 S. Vermont Ave, KAP-108, Los Angeles, CA 90089-2532, USA; e-mail: tartakov@usc.edu

1) The work of the first author was supported in part by a grant from the Israel Science Foundation, by the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem, and by the U. S. Army Research Office MURI grant W911NF-06-1-0094 at the University of Southern California. The work of the second author was supported in part by the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem and by the U. S. Office of Naval Research grant N00014-06-1-0110 and the U. S. Army Research Office MURI grant W911NF-06-1-0094 at the University of Southern California.
The standard method for proving this asymptotic exponentiality is to try to find a version of the process that is regenerative (cf. [4] and [1]). The heuristic behind this is that since the process is Harris-recurrent, it returns to a given set over and over again, and thus creates «cycles» that are «almost independent». Hence, the first cycle in which \( X(n) \) exceeds \( A \) is approximately geometrically distributed, and if the expected length of a cycle is finite and the probability of exceeding \( A \) in a given cycle tends to 0 as \( A \to \infty \), then, suitably standardized, the asymptotic distribution of the first exit time is exponential.

In this paper, we make a connection between the standardization constant and the quasi-stationary distribution. Our method of proof is a coupling argument. Although less general as a method for proving asymptotic exponentiality than the regeneration argument, we believe that our method is of interest in its own right. This notwithstanding, the regeneration argument seems to be widely unknown in the statistics community, and ought to be publicized.

The paper is organized as follows. In Section 2, we present the main result that states that the limiting distribution of the suitably standardized version of the first exit time as \( A \to \infty \) is Exponential(1) and that the moment generating function converges to that of Exponential(1), which implies that the convergence is in \( L^p \) for all \( p \geq 1 \). The proof is given in Section 3. We make a few remarks in Section 4. In Section 5, we give examples and describe applications to the evaluation of the distribution of the run length to false alarm for several change detection procedures.

2. Main result. Let \( \{X(n)\}_{n=0}^{\infty} \) be a discrete-time Harris-recurrent Markov process with state space \([0, \infty)\) and stationary transition probabilities. Let \( P^x \) denote the probability measure for the process when it starts at \( x \) (i.e., \( X(0) = x \)), and let \( P^G \) denote the probability measure when the initial state is distributed according to the distribution \( G \).

**Definition.** We call the process stochastically monotone if \( P^x \{X(1) \geq y\} \) is nondecreasing and right-continuous in \( x \) for all \( y \).

We will be interested in the behavior of the first exit time of \( X(n) \) from the interval \([0, A)\) when \( X(n) \) starts at \( x \in [0, A) \), i.e., of the stopping time

\[
N_A = \min\{n \geq 1: X(n) > A\}, \quad X(0) = x, \tag{2.1}
\]

where \( 0 \leq x < A \) and \( A \) is a positive finite threshold, assuming that the Markov process \( X(n) \) is stochastically monotone and Harris-recurrent.

The following theorem is the main result of the paper.

**Theorem 1.** Let \( X(n), n = 0, 1, 2, \ldots, \) be a stochastically monotone Harris-recurrent Markov process with state space \([0, \infty)\) and stationary transition probabilities such that:
The stationary distribution \( H(y) = \lim_{n \to \infty} P^x \{ X(n) \leq y \} \) exists and its support is \([0, \infty)\);

(C2) the quasi-stationary distribution \( H_A(y) = \lim_{n \to \infty} P^x(X(n) \leq y | N_A^x > n) \) exists for all \( 0 \leq x < A \) and for all \( 0 < A < \infty \).

Let \( p_A = P^{H_A}\{ X(1) > A \} \).

Then: (i) The distribution of \( p_A N_A^x \) is asymptotically Exponential (1) as \( A \to \infty \) for all fixed \( x \in [0, \infty) \);

(ii) the moment generating function \( E \exp\{ t p_A N_A^x \} \) of \( p_A N_A^x \) converges to \( 1/(1 - t) \) as \( A \to \infty \) for all fixed \( x \in [0, \infty) \); in particular, it follows that

\[
\lim_{A \to \infty} p_A E N_A^x = 1 \quad \text{and} \quad \lim_{A \to \infty} D\{ p_A N_A^x \} = 1.
\]

Conditions (C1) and (C2) hold in a variety of scenarios. See corresponding remarks in Section 4 and examples in Section 5.

We begin with a heuristic argument. A formal proof requires several auxiliary results and is given in Section 3.

Write \( N_A^{H_A} \) for the stopping time when the process \( X(n) \) starts at a random point \( X(0) = \xi \) in \([0, A]\) that has a quasi-stationary distribution \( H_A \), i.e., \( P\{ \xi \leq y \} = H_A(y) \). Then \( P^{H_A}(X(n) > A | N_A^{H_A} = n) = p_A \) for all \( n \geq 1 \), and, therefore, the distribution of \( N_A^{H_A} \) is geometric with the parameter \( p_A \) for all \( A > 0 \). Further, under conditions (C1) and (C2), the probability \( p_A \) goes to 0 as \( A \to \infty \), which implies that \( p_A N_A^{H_A} \) converges weakly to Exponential(1) as \( A \to \infty \). Intuitively, the asymptotic behavior of the stopping time \( N_A^x \) for every fixed point \( x \) is similar to that of \( N_A^{H_A} \).

Mathematical details are presented in the next section.

3. Proof. In order to prove Theorem 1 we need the following lemmas. We use the notation of the previous section, and we assume that the conditions of Theorem 1 are satisfied.

**Lemma 1.** The quasi-stationary distribution

\[ H_A(y) = \lim_{n \to \infty} P^x(X(n) \leq y | N_A^x > n) \]

converges to the stationary distribution \( H(y) \) at all continuity points \( y \) of \( H \).

**Proof.** Follows directly from [13, Theorem 1].

Recall that \( N_A^{H_A} \) is the stopping time (2.1) when the Markov process \( X(n) \) starts from the random point that has the quasi-stationary distribution \( H_A \), i.e., \( X(0) \sim H_A \).

**Lemma 2.** The distribution of \( N_A^{H_A} \) is Geometric(\( p_A \)), where \( p_A = P^{H_A}\{ X(1) > A \} \). Hence \( p_A E N_A^{H_A} = 1 \) and \( p_A N_A^{H_A} \) converges in distribution to Exponential(1) as \( A \to \infty \).

**Proof.** Since the Markov process is Harris-recurrent, there is no absorbing state, so that \( P\{ N_A^{H_A} = \infty \} = 0 \). Therefore, the geometric property
Asymptotic exponentiality of first exit times

of $N_A^{H_A}$ is obvious. Lemma 1 and the assumption that the support of $H$ is $[0, \infty)$ guarantee that $p_A \to 0$ as $A \to \infty$.

Lemma 3. Let $X^x(n)$ denote a process that starts from $x$ and has the same transition probabilities as $X(n)$. Let $0 \leq x < y < \infty$. There exists a sample space with $X^x(n)$ and $X^y(n)$ such that $X^y(n) \geq X^x(n)$ for all $n \geq 1$.

Proof. Clearly $X^y(1)$ is stochastically larger than $X^x(1)$, so that one can construct a sample space where $X^y(1) \geq X^x(1)$. To complete the proof, continue by induction on $n$.

Lemma 4. Let $0 \leq x < y < \infty$. Let $\tilde{X}^x(n)$ and $\tilde{X}^y(n)$ be independent Markov processes started at $x$ and $y$, respectively, both having the same transition probabilities as $X(n)$. Then

$$\mathsf{P}\left\{ \tilde{X}^x(n) > \tilde{X}^y(n) \text{ for at least one value of } n \right\} = 1. \quad (3.1)$$

Proof. Let $0 < \varepsilon < \frac{1}{4}$ and let $y \leq B < \infty$ be such that $H((B, \infty)) = \varepsilon$. Let $w_\varepsilon$ be such that for all $z$

$$|\mathsf{P}\{\tilde{X}^z(w_\varepsilon) \leq z\} - H(z)| < \varepsilon, \quad |\mathsf{P}\{\tilde{X}^0(w_\varepsilon) \leq z\} - H(z)| < \varepsilon.$$

By virtue of Lemma 3, $|\mathsf{P}\{\tilde{X}^z(w_\varepsilon) \leq z\} - H(z)| < \varepsilon$ for all $z$. Write $m$ for the median of the stationary distribution $H$. Obviously,

$$\mathsf{P}\left( \{B \geq \tilde{X}^x(w_\varepsilon) \vee \tilde{X}^y(w_\varepsilon)\} \setminus \{B \geq \tilde{X}^x(w_\varepsilon) \geq m, \tilde{X}^y(w_\varepsilon) \leq m\} \right)$$

$$\geq (1 - \varepsilon)^2 - \left( \frac{1}{2} - \varepsilon \right)^2,$$

$$\left( \frac{1}{2} - 2\varepsilon \right)^2 < \left( \frac{1}{2} - 2\varepsilon \right) \left( \frac{1}{2} - \varepsilon \right) \leq \mathsf{P}\{B \geq \tilde{X}^x(w_\varepsilon) \geq m, \tilde{X}^y(w_\varepsilon) \leq m\}$$

$$\leq \left( \frac{1}{2} + \varepsilon \right)^2.$$

Similarly, for any $j \geq 2$ when $0 \leq u < v \leq B$

$$\left( \frac{1}{2} + \varepsilon \right)^2 \geq \mathsf{P}\left( \tilde{X}^x(jw_\varepsilon) \geq m, \tilde{X}^y(jw_\varepsilon) \leq m \mid \tilde{X}^x((j - 1)w_\varepsilon) = u, \tilde{X}^y((j - 1)w_\varepsilon) = v \right)$$

$$\geq \left( \frac{1}{2} - 2\varepsilon \right)^2,$$

$$\mathsf{P}\left( \{B \geq \tilde{X}^x(jw_\varepsilon) \vee \tilde{X}^y(jw_\varepsilon)\} \setminus \{B \geq \tilde{X}^x(jw_\varepsilon) \geq m, \tilde{X}^y(jw_\varepsilon) \leq m\} \right)$$

$$\mid \tilde{X}^x((j - 1)w_\varepsilon) = u, \tilde{X}^y((j - 1)w_\varepsilon) = v \right)$$

$$\geq (1 - \varepsilon)^2 - \left( \frac{1}{2} - \varepsilon \right)^2 = \frac{3}{4} - \varepsilon.$$

Let $T_B = \min\{j: \tilde{X}^x(jw_\varepsilon) \vee \tilde{X}^y(jw_\varepsilon) > B\}$. 


Using previous inequalities, we obtain

\[
\Pr\left\{ B > \tilde{X}^x(jw_\varepsilon) \geq \tilde{X}^y(jw_\varepsilon) \text{ for some } 1 \leq j < T_B \right\} \\
\geq \left(\frac{1}{2} - 2\varepsilon\right)^2 \sum_{i=0}^{\infty} \left(\frac{3}{4} - \varepsilon\right)^i = \frac{(1/2 - 2\varepsilon)^2}{1 - (3/4 - \varepsilon)} = \frac{(1/2 - 2\varepsilon)^2}{1/4 + \varepsilon}.
\]

Letting \( \varepsilon \to 0 \) completes the proof.

**Lemma 5.** Using the same notation as in Lemma 4, \( \Pr\{\tilde{X}^x(\ell) \geq \tilde{X}^y(\ell) \text{ for some } \ell \leq n\} \to 1, n \to \infty, \) uniformly in \( 0 \leq x < y \leq B \).

**Proof.** This follows directly from Lemma 4 and its proof.

**Lemma 6.** Let \( \varepsilon > 0 \) and let \( 0 < B < \infty \) be such that \( H((B, \infty)) < \varepsilon \).
Let \( B \leq A < \infty \). Then \( H_A((B, A)) \leq \varepsilon \).

**Proof.** The lemma follows from the fact that \( H_A(y) \geq H(y) \) for all \( y \geq 0 \) (cf. [13, Theorem 1]).

**Proof of Theorem 1 (i).** Let \( N_A^{HA} = \min\{n: X(n) > A\} \), where \( X(0) \sim H_A \). By Lemma 2, we have \( N_A^{HA} \sim \text{Geometric}(p_A) \) and \( \lim_{A \to \infty} \Pr\{p_A N_A^{HA} > s\} = e^{-s}, s > 0 \).

Let \( \varepsilon > 0 \). Let \( 0 < B < \infty \) be such that \( H((B, \infty)) < \varepsilon \). Using the notation of Lemma 4, let \( 0 < q_B < \infty \) be such that

\[
\Pr\{\tilde{X}^0(n) \geq \tilde{X}^B(n) \text{ for some } n \leq q_B\} > 1 - \varepsilon. \tag{3.2}
\]

By virtue of Lemma 1 and Lemma 2, there exists \( A_\varepsilon \) such that for all \( A \geq A_\varepsilon \)

\[
|H_A(x) - H(x)| \leq \varepsilon \quad \text{for all } 0 \leq x \leq B, \tag{3.3}
\]
\[
|\Pr\{p_A N_A^{HA} > s\} - e^{-s}| \leq \varepsilon \quad \text{for all } 0 \leq s < \infty. \tag{3.4}
\]

Because the support of \( H \) is \([0, \infty)\), it follows from (3.3) that \( p_A q_B \to 0 \) as \( A \to \infty \).

Next, we construct the following sample space. Let \( \tilde{X}^0(n) \) be a Markov process (with transition probabilities as \( X(n) \)) starting at 0, and let \( \tilde{X}^B(n) \) be a Markov process starting at \( B \) such that they are independent until the first time that \( \tilde{X}^0(n) \geq \tilde{X}^B(n) \). Denote this time by \( \tau \). After \( \tau \), let \( \tilde{X}^0, \tilde{X}^B \) be such that \( \tilde{X}^0(n) \geq \tilde{X}^B(n) \) for all \( n \geq \tau \). (This construction is feasible by virtue of Lemma 3 and Lemma 4.)

By virtue of equation (3.2), \( \Pr\{\tau \leq q_B\} > 1 - \varepsilon \). Denote

\[
\tilde{N}_A^0 = \min\{n \geq 1: \tilde{X}^0(n) > A\}, \quad \tilde{N}_A^B = \min\{n \geq 1: \tilde{X}^B(n) > A\}.
\]

Note that \( N_A^{x} \) is stochastically larger than \( N_A^{y} \) if \( x < y \).
Now, fix $0 < s < \infty$ and let $A_B$ be large enough so that $p_A q_B < s$ for all $A \geq A_B$. Then we have the following chain of equalities and inequalities:

$$
P\{p_A N^B_A > s\} = P\{p_A N^B_A > s\} \geq P\{p_A N^B_A > s, \tau < q_B\}$$
$$= P\{p_A N^B_A > s, p_A \tau < p_A q_B < s\}$$
$$= P\{p_A \tilde{N}^B_A > s, p_A \tau < p_A q_B < s, \tilde{N}^B_A > \tau\}$$
$$\geq P\{p_A \tilde{N}^0_A > s, p_A \tau < p_A q_B < s, \tilde{N}^0_A > \tau\}$$
$$= P\{p_A \tilde{N}^0_A > s, p_A \tau < p_A q_B < s\} = P\{p_A \tilde{N}^0_A > s, \tau < q_B\}$$
$$\geq P\{p_A \tilde{N}^0_A > s\} - P\{\tau > q_B\}$$
$$\geq P\{p_A \tilde{N}^0_A > s\} - \varepsilon = P\{p_A N^0_A > s\} - \varepsilon.$$  (3.5)

On the other hand,

$$P\{p_A N^B_A > s\} = P\left(p_A N^{H_A} > s \mid X(0) = B\right)$$
$$\leq P\left(p_A N^{H_A} > s \mid X(0) = B\right) = \frac{P\{p_A N^{H_A} > s, X(0) = B\}}{P\{X(0) = B\}}$$
$$= \frac{P\{p_A N^{H_A} > s, X(0) = 0\}}{H_A([0, B])} \leq \frac{P\{p_A N^{H_A} > s\}}{H_A([0, B]).}$$

Since by the definition of $B$ and Lemma 6, $H_A([0, B]) \geq 1 - \varepsilon$, and by equation (3.4), $P\{p_A N^{H_A} > s\} \leq e^{-s} + \varepsilon$, we obtain

$$P\{p_A N^B_A > s\} \leq (e^{-s} + \varepsilon)(1 - \varepsilon)^{-1}. \quad (3.6)$$

Also, since $P\{X(0) = 0\} = H_A([0, A]) = 1$, we have

$$P\{p_A N^0_A > s\} = P\{p_A N^{H_A} > s \mid X(0) = 0\} \geq P\{p_A N^{H_A} > s \mid X(0) = 0\}$$
$$= \frac{P\{p_A N^{H_A} > s, X(0) = 0\}}{P\{X(0) = 0\}}$$
$$= P\{p_A N^{H_A} > s\} \geq e^{-s} - \varepsilon, \quad (3.7)$$

where the last inequality follows from (3.4).

Putting (3.5) and (3.7) together yields

$$P\{p_A N^B_A > s\} \geq e^{-s} - 2\varepsilon, \quad (3.8)$$

and putting (3.5) and (3.6) together obtains

$$P\{p_A N^0_A > s\} \leq (e^{-s} + \varepsilon)(1 - \varepsilon)^{-1} + \varepsilon. \quad (3.9)$$

Since for all $0 \leq x \leq B$,

$$P\{p_A N^B_A > s\} \leq P\{p_A N^0_A > s\} \leq P\{p_A N^0_A > s\}, \quad (3.10)$$
inequalities (3.8) and (3.9) imply that
\[ e^{-s} - 2\varepsilon \leq P\{p_A N_A^x > s\} \leq (e^{-s} + \varepsilon)(1 - \varepsilon)^{-1} + \varepsilon \quad \text{for all} \quad 0 \leq x \leq B. \]

Finally, fix \( x \) and let \( \varepsilon \to 0 \), so that ultimately \( B > x \). This completes the proof of Theorem 1 (i).

**Proof of Theorem 1 (ii).** Since \( N_A^x \sim \text{Geometric}(p_A) \), the random variable \( p_A N_A^x \) has a moment generating function \( M_{\lambda}^x(t) = Ee^{t p_A N_A^x} \), \( t < 1 \), and it is easy to see that as \( A \to \infty \)
\[ M_{\lambda}^x(t) \to (1 - t)^{-1} \quad \text{for} \quad t < 1. \tag{3.11} \]
Obviously, \( M_{\lambda}^x(t) = E\mathbb{E}(e^{t p_A N_A^x} | X(0)) \), where \( X(0) \) has distribution \( H_A \). It follows that for every initial state \( x \geq 0 \) and all \( t < 1 \) the value of \( p_A N_A^x \) has a moment generating function \( M_{\lambda}^x(t) = Ee^{t p_A N_A^x} \) and
\[ M_{\lambda}^x(t) = E \mathbb{E}_A X^{(0)}(t) = \int_0^A M_{\lambda}^x(t) H_A(dx). \]
For \( t \leq 0 \), by virtue of Theorem 1 (i) \( M_{\lambda}^x(t) \to (1 - t)^{-1} \) as \( A \to \infty \).
Let \( 0 < \varepsilon < 1 \) and \( C > 0 \) be such that \( H([0, C]) = \varepsilon \). For fixed \( 0 < t < 1 \), let \( A(\varepsilon) > C \) be such that
\[ 1 - \varepsilon < \frac{M_{\lambda}^x(t)}{1/(1 - t)} < 1 + \varepsilon \quad \text{whenever} \quad A \geq A(\varepsilon). \]
Recall that \( X(0) \) has distribution \( H_A \), which is a quasi-stationary distribution.

For any \( 0 < \gamma < \infty \), Markov's inequality yields \( P\{M_{\lambda}^x(t) > \gamma M_{\lambda}^x(t)\} \leq \gamma^{-1} \), so that for \( A \geq A(\varepsilon) \)
\[ P\big\{M_{\lambda}^x(t) > \gamma(1 - t)^{-1}\big\} \leq (1 + \varepsilon) \gamma^{-1}. \tag{3.12} \]
Substituting \( \gamma = (1 + \varepsilon)/\varepsilon \) in (3.12) yields
\[ P\big\{M_{\lambda}^x(t) > (1 + \varepsilon)\varepsilon^{-1}(1 - t)^{-1}\big\} \leq \varepsilon. \]
Since, by Lemma 6, \( \varepsilon = H([0, C]) \leq H_A([0, C]) \), it follows that for \( M_{\lambda}^x(t) \geq (1 + \varepsilon)\varepsilon^{-1}(1 - t)^{-1} \), the value of \( X(0) \) cannot exceed \( C \). In other words,
\[ M_{\lambda}^x(t) \leq (1 + \varepsilon)\varepsilon^{-1}(1 - t)^{-1} \quad \text{for} \quad x \geq C \quad \text{and all} \quad A \geq A(\varepsilon). \tag{3.13} \]
Let \( \beta = \min\{n: X(n) \geq C\} \). Obviously,
\[ M_{\lambda}^x(t) = Ee^{t p_A N_A^x} \leq E e^{t p_A \beta} \cdot E e^{t p_A N_A^C}. \tag{3.14} \]
Asymptotic exponentiality of first exit times

Let $\delta = P\{X^0(1) \geq C\}$. Clearly $\delta \to P\{X^0(1) > 0\} > 0$ as $\varepsilon \to 0$.

Due to the monotonicity of the process $X(n)$, the value of $\beta$ is bounded by a Geometric($\delta$)-distributed random variable, so that for $0 < t < 1$

$$1 \leq E e^{t \beta A} \leq E e^{t \beta A \text{Geometric}(\delta)} = \frac{\delta e^{pA t}}{1 - (1 - \delta e^{pA t})}.$$  

It follows that $E e^{t \beta A}$ is bounded as $A \to \infty$ (since $p_A \to 0$ as $A \to \infty$).

Since $E e^{t \beta A N^C_A} = M^C_A(t)$, relations (3.13) and (3.14) imply that $M^0_A(t)$ is also bounded as $A \to \infty$.

Denote $\varphi(t) = \limsup_{A \to \infty} M^0_A(t) < \infty$. Let $\{A_i\}_{i=1}^{\infty}$ be a sequence such that $\lim_{i \to \infty} M^0_{A_i}(t) = \varphi(t)$. Construct a set $\{t_j\}_{j=1}^{\infty}$ dense in $(0, t)$. Because $M^0_{A_i}(u)$ is monotone in $u$, one can obtain a subsequence $\{A_{i_j}\}$ of $\{A_i\}$ such that $M^0_{A_{i_j}}(u)$ converges as $j \to \infty$ for all $0 < u < t$. Since the limit is a moment generating function, by Theorem 1 (i) it must be $1/(1 - t)$. The same argument can be applied to $\liminf_{A \to \infty} M^0_A(t)$.

It follows that the limit $\lim_{A \to \infty} M^0_A(t)$ exists and is equal to $(1 - t)^{-1}$ for all $t < 1$. Because $M^0_A(t)$ is monotone in $x$ and because of (3.11), $\lim_{A \to \infty} M^0_A(t)$ necessarily equals $1/(t - 1)$ for all $t < 1$ and every fixed $x \in [0, \infty)$. This completes the proof of Theorem 1 (ii).

4. Remarks. 1. Let $G$ be a distribution with support $[0, A]$ and define the operator $T$ as

$$T(G) = \text{the distribution of } X(1) \text{ conditioned on } \{X(1) \leq A, \ X(0) \sim G\}.$$  

If $T$ is a continuous operator (in the weak* topology on the distribution functions over $[0, A]$), then a quasi-stationary distribution exists, i.e., condition (C2) in Theorem 1 is satisfied (cf. [5, Theorem III.10.1]).

2. Even if $T$ is not a continuous operator, sometimes condition (C2) can be verified by solving for $T(G) = G$ and arguing that this is the quasi-stationary distribution. For an example, see [11].

3. The proof can be modified easily to extend Theorem 1 to the case where the support of the stationary distribution $H$ is $[c, \infty)$ for some $c > 0$ (i.e., the set $[0, c)$ is not in the state space or is transient).

5. Examples and applications. Theorem 1 can be applied to a number of popular Harris-recurrent Markov processes. Below we present two examples. These are of interest when applying certain changepoint detection procedures.

5.1. Example 1: An additive-multiplicative Markov process. Let $\Lambda_1, \Lambda_2, \ldots$ be nonnegative continuous independent and identically distributed (i.i.d.) random variables with $\beta = E \Lambda_i$ and $\mu = E \ln \Lambda_i$. For $x \geq 0$, define recursively:

$$X(0) = x, \quad X(n) = (1 + X(n - 1)) \Lambda_n, \quad n = 1, 2, \ldots.$$  

(5.1)
This process is of interest in a number of applications (cf. [6], [11], [12]). For example, in the problem of detecting a change in distribution, the Shiryaev-Roberts statistic can be written as (cf. [11], [12])

$$R(n) = (1 + R(n - 1)) \frac{f_{\theta_1}(Y_n)}{f_{\theta_0}(Y_n)}, \quad R(0) = 0,$$

where \(\{Y_n, n \geq 1\}\) are independent, having probability density \(f_{\theta_0}\) before a change and putative density \(f_{\theta_1}\) after a change; \(\theta_0\) and \(\theta_1\) are fixed parameters, and one stops and declares that the change is in effect at \(N_A = \min\{n: R(n) > A\}\).

When \(\mu < 0\), the process \(\{X(n)\}\) is Harris-recurrent and has a stationary distribution (for any \(x \geq 0\)). To see this, note that \(X(n)\) can be written as

$$X(n) = \sum_{k=0}^{n} \prod_{i=k}^{n} \Lambda_i = \sum_{k=0}^{n} \exp \left\{ \sum_{i=k}^{n} \ln \Lambda_i \right\},$$

where \(\Lambda_0 = x\). Obviously,

$$\sum_{k=0}^{n} \exp \left\{ \sum_{i=k}^{n} \ln \Lambda_i \right\} \overset{d}{=} \sum_{k=1}^{n} \exp \left\{ \sum_{i=1}^{k} \ln \Lambda_i \right\} + x \exp \left\{ \sum_{i=1}^{n} \ln \Lambda_i \right\},$$

where the right-hand side converges (for every \(x \geq 0\) as \(n \to \infty\)) to the random variable

$$\sum_{k=1}^{\infty} \exp \left\{ \sum_{i=1}^{k} \ln \Lambda_i \right\},$$

which is a.s. finite when \(\mu < 0\). Since we assumed above that \(\Lambda_1\) is continuous, the quasi-stationary distribution exists (see Remark 1 in Section 4). It follows from Theorem 1 that a suitably standardized version of the first exceedance time over \(A\) (i.e., \(p_A N_A^x\)) is asymptotically exponentially distributed.

Note that while using the conventional regeneration argument is perhaps possible, embedding the Markov process (5.1) into «regenerative cycles» by no means is either straightforward or obvious, which is especially true when \(1 \leq \beta = E \Lambda_i < \infty\) and \(\mu = E \ln \Lambda_i < 0\). This case does have meaning for applications. For example, regard the aforementioned changepoint detection problem. When there never is a change, the observations \(Y_i, i \geq 1\), have density \(f_{\theta_0}\), so that \(\beta = \int [f_{\theta_1}(y)/f_{\theta_0}(y)] f_{\theta_0}(y) \, dy = 1\) while by Jensen’s inequality \(\mu = \int \ln[f_{\theta_1}(y)/f_{\theta_0}(y)] f_{\theta_0}(y) \, dy < 0\). If there is a change — for argument’s sake let it be in effect from the very beginning — the observations \(Y_i, i \geq 1\), have density \(f_{\theta}\) (not necessarily \(f_{\theta_1}\); the post-change parameter is seldom known in advance, and the putative \(\theta_1\) is merely a representation of a «meaningful» change). For \(\theta\) close to \(\theta_0\), one would obtain \(\beta = \int [f_{\theta_1}(y)/f_{\theta_0}(y)] f_{\theta}(y) \, dy > 1\) and \(\mu = \int \ln[f_{\theta_1}(y)/f_{\theta_0}(y)] f_{\theta}(y) \, dy < 0\).
Before going into further details, we discuss an issue related to computing $p_A$, the standardizing factor. If $p_A$ were amenable to direct calculation, one could use this to approximate $\mathbb{E}N^x_A \approx 1/p_A$. Unfortunately, in most cases direct evaluation of $p_A$ is not an easy task, and evaluation of $\mathbb{E}N^x_A$ has to be done by other methods. (But see [11] and [8] for examples that allow some tractability; and [9] for numerical techniques that allow for computing $p_A$ almost precisely.) Nonetheless, evaluation of $p_A$ is of interest on its own merits (cf. [16]), as $p_A$ is an approximation of the probability that there will be a first upcrossing of the threshold $A$ at a specified time $n$, and $1 - (1 - p_A)^m$ is an approximation of the probability that there will be a first upcrossing of $A$ in a given stretch of $m$ observations (i.e., for the «local false alarm probability» $P(n \leq N^x_A \leq n + m - 1 \mid N^x_A \geq n)$). Therefore, if $\mathbb{E}N^x_A$ can be evaluated, $p_A$ can be approximated by $1/\mathbb{E}N^x_A$.

Suppose now that $\beta = \mathbb{E}A_i = 1$. Let $f_0$ be the density of $A_i$ and define $f_i(A) = A f_0(A)$. (Since $\mathbb{E}A = 1$, it follows that $f_i$ is a bona fide probability density.) Note that $A$ is a likelihood ratio, $A = f_i(A)/f_0(A)$. It follows from [12] (see also [18]) that

$$
\mathbb{E}_{f_0} N^x_A = \gamma^{-1} A(1 + o(1)) \quad \text{as} \quad A \to \infty, \tag{5.3}
$$

where $\mathbb{E}_{f_0}$ is the expectation with respect to the density $f_0$ and $\gamma$ is a constant that can be calculated by renewal theory (cf. [19], [14]), so that $p_A \approx \gamma/A$. See Remark in the end of Section 5.2 for evaluation of $p_A$ when $\mathbb{E}A_i \neq 1$.

5.2. Example 2: A reflected random walk. Let $\{Z_n\}_{n=1}^\infty$ be a sequence of i.i.d. continuous random variables with a negative mean $\mu = \mathbb{E}Z_n < 0$. For $n \geq 1$, define

$$
X(n) = \max\{0, X(n-1) + Z_n\}, \quad X(0) = x > 0. \tag{5.4}
$$

Since $\mu < 0$, the Markov process $\{X(n)\}$ is Harris-recurrent and has a stationary distribution. To see this, note that

$$
X(n) = \max\{0, Z_1 + \cdots + Z_n + x, \ldots, Z_2 + \cdots + Z_n, \ldots, Z_n\}.
$$

Write $S_i = \sum_{k=1}^i Z_k$, $S_0 = 0$. Since the vector $(Z_1, \ldots, Z_n)$ has the same distribution as $(Z_n, \ldots, Z_1)$, it follows that

$$
X(n) \overset{d}{=} \max\{\max\{0, S_1, S_2, \ldots, S_{n-1}\}, x + S_n\},
$$

where the right-hand side converges (as $n \to \infty$ for any $x \geq 0$) to the random variable $\max_{i \geq 0} S_i$, which is a.s. finite whenever $\mu = \mathbb{E}Z_i < 0$.

The process (5.4) describes a broad class of single-channel queuing systems (see, e.g., [2]) as well as a popular cumulative sum decision statistic for detecting a change in distribution [10] and has been studied extensively by
itself, outside the framework of general Markov processes. For instance, for 
\(x = 0\), the asymptotic exponentiality of the stopping time

\[
N^x_a = \min \{ n \geq 1: X(n) > a \}, \quad a > 0
\]

(5.5)

(as \(a \to \infty\)) has been proven in [7], which can be generalized easily for any 
\(x > 0\). (The process \(\{X(n)\}\) obviously is a renewal process, so, although 
our Theorem 1 covers this example when the conditions (C1) and (C2) are 
 satisfied, it is not needed to prove asymptotic exponentiality of \(N^x_A\), as it can 
be derived from general results; cf. [1, Chap. VI].)

Assume for simplicity that \(x = 0\). If there exists a positive \(\omega\) such 
that \(\mathbb{E}e^{\omega Z_i} = 1\), let \(f_0(z)\) be the density of \(Z_i\) and define \(f_1(z) = e^{\omega z} f_0(z)\). 
Since \(\mathbb{E}e^{\omega Z_i} = 1\), it follows that \(f_1\) is a bona fide probability density, 
and \(f_1(Z)/f_0(Z) = e^{\omega z}\) is a likelihood ratio. Hence, assuming that \(\mu_1 = \int \ln[f_1(z)/f_0(z)] f_1(z) \, dz < \infty\) and letting

\[
N^0_a = \min \{ n \geq 1: \max (0, \omega X(n) - 1) + \omega Z_n > \omega a \},
\]

standard renewal-theoretic methods (cf. [19], [14]) readily apply to obtain that

\[
\mathbb{E}N^0_a = \delta^{-1} e^{\omega A}(1 + o(1)) \quad \text{as} \quad a \to \infty,
\]

so that \(p_A = \delta e^{-\omega a}\), where \(A = e^a\). Here \(0 < \delta < 1\) is a constant that can be 
computed explicitly by a renewal-theoretic argument (cf. [16]).

Remark. Clearly, \(N^x_a\) of Example 2 is larger than \(N^x_A\) of Example 1 
(with \(A = e^a\)), so that \(\mathbb{E}N^x_A \leq \delta^{-1} A^\omega (1 + o(1))\). Theorem 5 of [6] as well as 
Theorem 4 of [3] imply that

\[
P\{X(\infty) > y\} = C y^{-\omega}(1 + o(1)) \quad \text{as} \quad y \to \infty,
\]

where \(X(\infty)\) is a random variable that has the stationary distribution 
of \(\{X(n)\}\) and \(C\) is a positive finite constant. Note that \(X(\infty)\) is stochastically 
larger than a random variable that has the quasi-stationary distribution. Therefore, the first upcrossing over \(A\) of the process \(X(n)\) starting at a random \(X(0)\) distributed like \(X(\infty)\) will occur no later than the first 
upcrossing over \(A\) of the process \(X(n)\) starting at a random \(X(0)\) that has 
the quasi-stationary distribution. The proportion of times that the former 
exceeds \(A\) is \(P\{X(\infty) > A\}\). It follows that \(\mathbb{E}N^x_A = C^{-1} A^\omega (1 + o(1))\), so 
that \(p_A\) has an order of magnitude \(1/A^\omega\).

5.3. Applications to sequential changepoint detection and a 
Monte Carlo experiment. The importance of the asymptotic exponentiality of the run length in sequential changepoint detection methods is 
twofold. First, it shows that the mean time to false alarm (the so-called
average run length), which is a popular measure of the false alarm rate, is indeed an exhaustive performance metric. Second, the result can be used for the evaluation of the local false alarm probabilities of the corresponding detection schemes (see Example 1 above; see [16] for a more detailed discussion of the importance of local false alarm probabilities in a variety of applications).

To be more specific, assume that there is a sequence of i.i.d. variables (observations) \( Y_1, Y_2, \ldots \) that follow the density \( f_0(y) \) under the no-change hypothesis (the in-control mode) and the density \( f_1(y) \) after the change occurs (the out-of-control mode). The change occurs at an unknown point in time \( \nu; 1 \leq \nu < \infty \). Therefore, conditioned on \( \nu = k \), \( Y_n \sim f_0(y) \) for \( n < k \) and \( Y_n \sim f_1(y) \) for \( n \geq k \). We write \( P_\infty (E_k) \) and \( P_k (E_k) \), respectively, for probability measures (expectations) when there is no change (i.e., \( \nu = \infty \)) and when the change occurs at point \( k \). Let \( Z_n = \ln\left[ \frac{f_1(Y_n)}{f_0(Y_n)} \right] \) be the corresponding log-likelihood ratio and let \( S_n = \sum_{i=1}^{n} Z_i \). Let \( I_1 = E_1 Z_1 \) and \( I_0 = E_\infty (-Z_1) \) be the Kullback–Leibler information numbers, which are assumed finite.

We begin with the cumulative sum (CUSUM) test. The CUSUM statistic is given by the recursion \( (5.4) \) and the corresponding stopping time is defined in \( (5.5) \). The difference from the previous section is that \( Z_n, n = 1, 2, \ldots, \) are not arbitrary random variables with negative mean, but rather log-likelihood ratios with mean \( \mu = -I_0 \). This simplifies most of the calculations, since \( E e^{Z_n} = 1 \). Recall that in this section we denote this expectation by \( E^\infty \).

Rewrite the corresponding stopping time in the following form:

\[
N_A = \min \{n \geq 1: \max\{1, W(n - 1) e^{Z_n}\} > A\}, \tag{5.7}
\]

where \( W(0) = 1 \) and \( A = e^a \). The asymptotic approximation for the average run length to false alarm \( (5.6) \) holds with \( \omega = 1, e^a = A, \) and \( \delta = I_1 \gamma^2 \) (cf. [16]), which implies that \( p_A \sim I_1 \gamma^2 / A \). Here \( \gamma = \lim_{y \to \infty} E_1 \exp\{- (S_{\tau_y} - y)\} \), where \( \tau_y = \min\{n: S_n > y\} \) is the first time when the random walk \( S_n = \sum_{i=1}^{n} Z_i \) crosses the level \( y \). The constant \( \gamma \) is the subject of renewal theory (cf. [19] or [14]) and can be computed explicitly.

We now proceed with the Shiryaev–Roberts detection test. The Shiryaev–Roberts statistic is defined by \( (5.2) \), where \( f_{\theta_1}(Y_n)/f_{\theta_0}(Y_n) = e^{Z_n} \) and \( R(0) = 0 \). The corresponding stopping time is

\[
\hat{N}_A = \min \{n \geq 1: R(n) > A\}.
\]

We now denote it by \( \hat{N}_A \) to distinguish from the CUSUM stopping time in the following calculations and comparison.

Since \( E_\infty e^{Z_n} = 1 \), the process \( R(n) - n \) is a zero-mean martingale, which allows us to approximate the average run length to false alarm:

\[
E_\infty \hat{N}_A \sim \gamma^{-1} A \quad \text{as} \quad A \to \infty.
\]
This approximation follows from (5.3) above. The distribution of the
Shiryaev–Roberts stopping time is approximately Exponential($p_A$) with
$p_A \sim \gamma/A$. (The asymptotic exponentiality of the suitably standardized
run length to false alarm has been shown in [20].)

In order to verify the accuracy of asymptotic approximations for reason­
able values of the threshold $A$, we performed Monte Carlo (MC) simulations
for the following example. Consider the case where observations are inde­
pendent, originally having an Exponential(1) distribution, changing at an
unknown time to Exponential(1/(1 + q)), i.e.,

$$f_0(y) = e^{-y} \mathbb{I}\{y \geq 0\}, \quad f_1(y) = (1 + q)^{-1}e^{-y/(1+q)} \mathbb{I}\{y \geq 0\}, \quad q > 0. \quad (5.8)$$

In this case

$$I_1 = q - \ln(1 + q) \quad \text{and} \quad \gamma = (1 + q)^{-1}.$$ 

Applying Example 1, the likelihood ratio is

$$\Lambda_n = e^{Z_n} = (1 + q)^{-1}e^{qY_n/(1+q)}$$

and the average run length (ARL) to false alarm of the Shiryaev–Roberts
procedure is

$$\text{ARL}_{SR}(A) = E_\infty N_A \approx (1 + q) A. \quad (5.9)$$

Applying Example 2, an approximation of the ARL to false alarm of
the CUSUM test is

$$\text{ARL}_{CU}(A) = E_\infty N_A \approx \frac{(1 + q)^2}{q - \ln(1 + q)} A. \quad (5.10)$$

We simulated the CUSUM and Shiryaev–Roberts procedures under the
assumption of no change (i.e., all simulated observations are Exponential(1)).
Each combination of (test, threshold) was simulated 100,000 times. The re­
results are reported in Tables 1 and 2. We present the results of simulations
when the parameter $q$ equals 3, which is a reasonable value in certain applica­
tions such as detection of a randomly appearing target in noisy measure­
ments, in which case $q$ is the signal-to-noise ratio (see, e.g., [15] and [17]).

It is seen that the approximation (5.9) for the Shiryaev–Roberts test is very
accurate for all threshold values, even when the ARL is small. On the other
hand, the approximation (5.10) for the CUSUM test (given in the row «FO
ARL$_{CU}$» in Table 1, where FO stands for «first order») is not especially
accurate. This happens primarily because the first order approximation
takes into account only the first term of expansion and ignores the second
Asymptotic exponentiality of first exit times

term $O(\ln A)$ as well as constants. An accurate, higher order (HO) approximation can be obtained using the results of [17] which give:

$$\text{ARL}_{CU}(A) \approx \frac{(1 + q)^2}{q - \ln(1 + q)} A - \frac{1}{\ln(1 + q) - q/(1 + q)} \ln A - \frac{1}{1 + q} \frac{q}{q - \ln(1 + q)} \frac{1}{(1 + q) \ln(1 + q) - q}.$$ 

In Table 1 the row «HO ARL$_{CU}$» corresponds to this latter approximation, which perfectly fits the MC estimates (denoted by «MC ARL$_{CU}$») for all tested threshold values $A \geq 1.2$.

**Table 1.** The ARL versus threshold for the CUSUM test for $q = 3$

<table>
<thead>
<tr>
<th>$A$</th>
<th>1.2</th>
<th>1.7</th>
<th>2.5</th>
<th>4.6</th>
<th>9.2</th>
<th>13.0</th>
<th>17.1</th>
<th>21</th>
<th>41</th>
</tr>
</thead>
<tbody>
<tr>
<td>FO ARL$_{CU}$</td>
<td>11.90</td>
<td>16.86</td>
<td>24.79</td>
<td>45.61</td>
<td>91.22</td>
<td>128.90</td>
<td>169.55</td>
<td>208.22</td>
<td>406.52</td>
</tr>
<tr>
<td>HO ARL$_{CU}$</td>
<td>7.96</td>
<td>12.36</td>
<td>19.69</td>
<td>35.96</td>
<td>84.07</td>
<td>121.21</td>
<td>161.43</td>
<td>199.77</td>
<td>397.02</td>
</tr>
<tr>
<td>ARL$_{CU}$</td>
<td>8.04</td>
<td>12.45</td>
<td>19.79</td>
<td>35.97</td>
<td>84.33</td>
<td>121.23</td>
<td>161.88</td>
<td>200.44</td>
<td>397.16</td>
</tr>
<tr>
<td>MC SD($N_A$)</td>
<td>7.49</td>
<td>11.88</td>
<td>19.18</td>
<td>38.61</td>
<td>83.21</td>
<td>119.73</td>
<td>159.91</td>
<td>198.97</td>
<td>396.84</td>
</tr>
</tbody>
</table>

**Table 2.** The ARL versus threshold for the Shiryaev–Roberts test for $q = 3$

<table>
<thead>
<tr>
<th>$A$</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARL$_{SR}$</td>
<td>4.01</td>
<td>8.03</td>
<td>20.00</td>
<td>39.94</td>
<td>79.99</td>
<td>119.82</td>
<td>159.17</td>
<td>200.42</td>
<td>399.46</td>
</tr>
<tr>
<td>MC ARL$_{SR}$</td>
<td>4.01</td>
<td>8.03</td>
<td>20.00</td>
<td>39.94</td>
<td>79.99</td>
<td>119.82</td>
<td>159.17</td>
<td>200.42</td>
<td>399.46</td>
</tr>
<tr>
<td>MC SD($\tilde{N}_A$)</td>
<td>3.00</td>
<td>6.78</td>
<td>18.34</td>
<td>37.92</td>
<td>77.33</td>
<td>117.39</td>
<td>157.19</td>
<td>197.90</td>
<td>396.94</td>
</tr>
</tbody>
</table>

In these tables we also present the MC estimates of standard deviations SD($N_A$) and SD($\tilde{N}_A$) of the stopping times. As one would expect, the standard deviations are the same (approximately) as the means, and the similarity grows as $A$ increases. The fit is slightly better for the CUSUM test.

Figures 1(a) and 1(b) show the logarithm of the empirical (MC estimates) survival functions $\ln P_\infty(\tau_A > y)$ and $\ln P_\infty(\tau_A > y)$ for the CUSUM and Shiryaev–Roberts procedures, where $\tau_A = N_A/\text{ARL}_{CU}$ and $\tilde{\tau}_A = \tilde{N}_A/\text{ARL}_{SR}$ are the corresponding standardized stopping times, along with the logarithm of the exponential probability plot $\ln e^{-y} = -y$. The quantile-quantile plots (QQ-plots) for the stopping times are shown in Figures 2(a) and 2(b). The QQ-plots display sample quantiles of $N_A$ and $\tilde{N}_A$ versus theoretical quantiles from the exponential distribution. If the distributions of the stopping times are exponential, the plots will be close to linear. These figures show that, for the chosen putative value of the post-change parameter ($q = 3$), the exponential distribution approximates the distributions of the stopping times very well. It is seen that the exponential
approximation works very well already for \( A = 13 \) (\( \text{ARL}_{\text{CU}} \approx 120 \)) for the CUSUM test and for \( A = 40 \) (\( \text{ARL}_{\text{SR}} \approx 160 \)) for the Shiryaev–Roberts test. When considering that in practical applications the values of the ARL to false alarm usually range from 300 and upwards, the exponential distribution seems to be a perfect fit.

![Fig. 1. Empirical estimates of \( \ln[\Pr(\tau_A > y)] \) and \( \ln[\Pr(\tilde{\tau}_A > y)] \) for the CUSUM and Shiryaev–Roberts procedures.](image1)

![Fig. 2. QQ-plots for the stopping times of the CUSUM and Shiryaev–Roberts procedures.](image2)

**Acknowledgments.** We are grateful to Aleksei Polunchenko for the help with Monte Carlo simulations.
REFERENCES

2. Боровков А. А. Вероятностные процессы в теории массового обслуживания. М.: Наука, 1972, 367 c.

Поступила в редакцию 16.11.2007
Исправленный вариант 23.IV.2008