J. Appl. Prob. **48,** 589–595 (2011)
Printed in England

ON THE FIRST EXIT TIME OF A NONNEGATIVE MARKOV PROCESS STARTED AT A QUASISTATIONARY DISTRIBUTION

MOSHE POLLAK,* *Hebrew University of Jerusalem*
ALEXANDER G. TARTAKOVSKY,** *University of Southern California*

Abstract

Let $\{M_n\}_{n\geq 0}$ be a nonnegative time-homogeneous Markov process. The quasistationary distributions referred to in this note are of the form $Q_A(x) = \lim_{n\to\infty} \mathrm{P}(M_n \leq x \mid M_0 \leq A, M_1 \leq A, \dots, M_n \leq A)$. Suppose that M_0 has distribution Q_A, and define $T_A^{Q_A} = \min\{n \mid M_n > A, n \geq 1\}$, the first time when M_n exceeds A. We provide sufficient conditions for $Q_A(x)$ to be nonincreasing in A (for fixed x) and for $T_A^{Q_A}$ to be stochastically nondecreasing in A.

Keywords: Changepoint problem; first exit time; Markov process; quasistationary distribution; stationary distribution

2010 Mathematics Subject Classification: Primary 60J05; 60J20
Secondary 62L10; 62L15

1. Introduction

Quasistationary distributions come up naturally in the context of first exit times of Markov processes. Of special interest—in particular in statistical applications—is the case of a nonnegative Markov process, where the first time that the process exceeds a fixed level signals that some action is to be taken. The quasistationary distribution is the distribution of the state of the process if a long time has passed and yet no crossover has occurred.

Various topics pertaining to quasistationary distributions are existence, calculation, simulation, etc. For an extensive bibliography, see [6].

The topic addressed in this paper deals with a certain aspect of the quasistationary distribution Q_A as a function of A. Pollak and Siegmund [5] have shown, under certain conditions, that if a stationary distribution Q exists then $Q_A \to Q$ as $A \to \infty$. Here we study a monotonicity property of Q_A and apply it to the behavior of the expected time of the first exceedance of A by a Markov process started at Q_A, as a function of A. Specifically, we provide conditions under which Q_A is nonincreasing.

This paper is organized as follows. We present our results and their proofs in Section 2. In Section 3 we provide examples of interesting cases where the conditions that we posit in Section 2 are satisfied. We discuss the meaning and relevance of our results in Section 4.

Received 14 September 2010; revision received 16 February 2011.
* Postal address: Department of Statistics, Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel. Email address: msmp@mscc.huji.ac.il
** Postal address: Department of Mathematics, University of Southern California, 3620 South Vermont Avenue, KAP-108, Los Angeles, CA 90089-2532, USA. Email address: tartakov@math.usc.edu

https://doi.org/10.1239/jap/1308662648 Published online by Cambridge University Press

2. Results

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, and let $\{M_n\}_{n=0}^{\infty}$ be an irreducible Markov process defined on this space taking values in $\mathcal{M} \subseteq [0, \infty)$ and having time-homogeneous transition probabilities $\rho(t, x) = \mathrm{P}(M_{n+1} \le x \mid M_n = t)$.

Let $T_A = \min\{n \mid M_n > A,\ n \ge 0\}$, and assume that the following conditions hold.

(C1) The quasistationary distribution

$$Q_A(x) = \lim_{n\to\infty} \mathrm{P}(M_n \le x \mid T_A > n)$$

exists for all $A > A_0 \ge 0$ (for some $A_0 < \infty$) and satisfies $Q_A(0) = 0$.

(C2) $\rho(s, x)$ is nonincreasing in s for all fixed $x \in \mathcal{M}$.

(C3) $\rho(ts, tx)$ is nondecreasing in t for all fixed $s, x \in \mathcal{M}$.

(C4) $\rho(s, x)/\rho(s, A)$ is nonincreasing in s for all fixed $x \in \mathcal{M},\ x \le A$.

(C5) $\rho(ts, tx)/\rho(ts, tA)$ is nondecreasing in t for all fixed $s, x \in \mathcal{M},\ x \le A$.

Now regard the case where M_0 has distribution Q_A, and define

$$T_A^{Q_A} = \min\{n \mid M_n > A,\ n \ge 1;\ M_0 \sim Q_A\}.$$

Theorem 1. *Let conditions (C1)–(C5) be satisfied. Then*

(i) *M_0 is stochastically nondecreasing in A, i.e. $Q_{A_1}(x) \ge Q_{A_2}(x)$ for all x if $A_1 < A_2$;*

(ii) *$Q_{yA}(yx) \ge Q_A(x)$ for all $y \ge 1$ and all fixed $x \in \mathcal{M},\ x \le A$;*

(iii) *$T_A^{Q_A} \overset{\mathrm{st}}{\preceq} T_{yA}^{Q_{yA}}$ for all $y \ge 1$, where '$\overset{\mathrm{st}}{\preceq}$' stands for 'stochastically smaller than (or equal to)'. In particular, it follows that $\mathrm{E}[T_A^{Q_A}] \le \mathrm{E}[T_{yA}^{Q_{yA}}]$ for all $y \ge 1$.*

Although conditions (C1)–(C5) are restrictive, they are nevertheless satisfied in a number of interesting cases, some of which are provided in the next section.

Proof of Theorem 1. Let $\{U_n\}_{n\ge 0}$ be a Markov process with time-homogeneous transition probabilities

$$\mathrm{P}(U_{n+1} \le x \mid U_n = t) = \frac{\rho(t, x)}{\rho(t, A)}, \qquad x \le A,$$

where $A > 0$ is fixed and U_0 has an arbitrary distribution (possibly degenerate) on $[0, A]$.

(i) Let $y > 1$, and define $\{V_n\}$ to be a Markov process with $V_0 = yU_0$, having time-homogeneous transition probabilities

$$\mathrm{P}(V_{n+1} \le x \mid V_n = t) = \frac{\rho(t, x)}{\rho(t, yA)}, \qquad x \le yA. \tag{1}$$

Clearly, the quasistationary distribution of $\{V_n\}$ is Q_{yA}. By condition (C4),

$$\frac{\rho(U_0, x)}{\rho(U_0, A)} \ge \frac{\rho(V_0, x)}{\rho(V_0, A)} \ge \frac{\rho(V_0, x)}{\rho(V_0, yA)}.$$

It follows that $U_1 \overset{\mathrm{st}}{\preceq} V_1$. Therefore, we can construct a sample space where $U_0 \le V_0$ and $U_1 \le V_1$. Repeating this with U_1, V_1 replacing U_0, V_0 we obtain $U_2 \le V_2$. So, we can construct

https://doi.org/10.1239/jap/1308662648 Published online by Cambridge University Press

a sample space where $U_0 \le V_0$, $U_1 \le V_1$, and $U_2 \le V_2$. Continuing with this inductively we obtain a sample space where $U_n \le V_n$ for all $n \ge 0$. Consequently, $\lim_{n\to\infty} \mathrm{P}(U_n > x) \le \lim_{n\to\infty} \mathrm{P}(V_n > x)$, i.e. $Q_A(x) \ge Q_{yA}(x)$, accounting for (i).

(ii) Define $W_n = yU_n$, and (as above) let $\{V_n\}_{n\ge 0}$ be a Markov process with $V_0 = W_0 = yU_0$, having time-homogeneous transition probabilities (1). Clearly, the quasistationary distribution of $\{V_n\}$ is $Q_{yA}(x)$ and that of $\{W_n\}$ is $Q_A(x/y)$.

Since

$$\begin{aligned}
\mathrm{P}(V_1 \le x \mid V_0) &= \frac{\rho(V_0, x)}{\rho(V_0, yA)} \\
&\ge \frac{\rho(V_0/y, x/y)}{\rho(V_0/y, A)} \\
&= \mathrm{P}\bigg(U_1 \le \frac{1}{y}x \,\bigg|\, U_0 = \frac{1}{y}V_0\bigg) \\
&= \mathrm{P}(W_1 \le x \mid W_0 = V_0),
\end{aligned}$$

it follows that $V_1 \overset{\mathrm{st}}{\preceq} W_1$. Therefore, we can construct a sample space on which U_0, U_1, V_0, V_1, W_0, W_1 are all defined and such that $V_1 \le W_1$ almost surely (a.s.). Write $V_1 = s$ and $W_1 = t$, where $s \le t \le yA$, $s, t \in \mathcal{M}$. Now (by virtue of (C5))

$$\begin{aligned}
\mathrm{P}(V_2 \le x \mid V_1 = s) &= \frac{\rho(s, x)}{\rho(s, yA)} \\
&\ge \frac{\rho(t, x)}{\rho(t, yA)} \\
&\ge \frac{\rho(t/y, x/y)}{\rho(t/y, A)} \\
&= \mathrm{P}\bigg(U_2 \le \frac{1}{y}x \,\bigg|\, U_1 = \frac{1}{y}t\bigg) \\
&= \mathrm{P}(W_2 \le x \mid W_1 = t),
\end{aligned}$$

so that $V_2 \overset{\mathrm{st}}{\preceq} W_2$, and we can construct a sample space on which U_0, U_1, U_2, V_0, V_1, V_2, W_0, W_1, W_2 are all defined and $V_0 = W_0$, $V_1 \le W_1$, $V_2 \le W_2$ a.s. Continuing this inductively, we obtain a sample space on which $\{U_n\}$, $\{V_n\}$, $\{W_n\}$ are all defined and $V_n \le W_n$ a.s. for all $n \ge 0$. Consequently, $\lim_{n\to\infty} \mathrm{P}(V_n > x) \le \lim_{n\to\infty} \mathrm{P}(W_n > x)$, i.e. $Q_{yA}(x) \ge Q_A(x/y)$, accounting for (ii).

(iii) Note that both first exit times $T_A^{Q_A}$ and $T_{yA}^{Q_{yA}}$ are geometrically distributed random variables, with

$$1 - \int_0^A \rho(s, A)\,\mathrm{d}Q_A(s)$$

and

$$1 - \int_0^{yA} \rho(s, yA)\,\mathrm{d}Q_{yA}(s),$$

the respective parameters of 'success'. Hence, it suffices to show that

$$\int_0^{yA} \rho(s, yA)\,\mathrm{d}Q_{yA}(s) \ge \int_0^A \rho(s, A)\,\mathrm{d}Q_A(s) \quad \text{for } y \ge 1.$$

https://doi.org/10.1239/jap/1308662648 Published online by Cambridge University Press

Note that $\rho(\mathrm{d}s, t) \le 0$. Therefore, integrating by parts yields

$$\begin{aligned}
&\int_0^{yA} \rho(s, yA)\,\mathrm{d}Q_{yA}(s) \\
&\quad = \rho(s, yA) Q_{yA}(s)\big|_0^{yA} - \int_0^{yA} Q_{yA}(s)\rho(\mathrm{d}s, yA) \\
&\quad = \rho(yA, yA) - \int_0^{yA} Q_{yA}(s)\rho(\mathrm{d}s, yA) \quad (\text{since } Q_{yA}(0) = 0 \text{ by (C1)}) \\
&\quad \ge \rho(yA, yA) - \int_0^{yA} Q_A\Big(\frac{s}{y}\Big)\rho(\mathrm{d}s, yA) \quad (\text{by (ii)}) \\
&\quad = \rho(yt, yA) Q_A(t)\big|_0^{A} - \int_0^{A} Q_A(t)\rho(\mathrm{d}(yt), yA) \\
&\quad = \int_0^{A} \rho(yt, yA)\,\mathrm{d}Q_A(t) \\
&\quad \ge \int_0^{A} \rho(t, A)\,\mathrm{d}Q_A(t) \quad (\text{by (C3)}),
\end{aligned}$$

which completes the proof.

3. Examples

Suppose that $\{M_n\}_{n\ge 0}$ obeys a recursion of the form

$$M_{n+1} = \varphi(M_n)\Lambda_{n+1}, \qquad n = 0, 1, \dots,$$

where

(D1) $\{\Lambda_i\}_{i\ge 1}$ are independent, identically distributed (i.i.d.) positive and continuous random variables;

(D2) the distribution function F of Λ_i satisfies

$$\frac{F(tx)}{F(tA)} \text{ increases in } t,\ t > 0, \text{ for fixed } x \in \mathcal{M},\ x \le A;$$

(D3) $\varphi(t)$ is continuous, positive, and nondecreasing in t;

(D4) $t/\varphi(t)$ is nondecreasing in t;

(D5) φ and F are such that $\mathrm{P}(\lim_{n\to\infty} M_n = 0) = 0$.

In this example,

$$\rho(s, x) = F\bigg(\frac{x}{\varphi(s)}\bigg).$$

Under these conditions, Theorem III.10.1 of [2] can be applied to obtain the existence of a quasistationary distribution. Conditions (D1)–(D5) are easily seen to imply conditions (C1)–(C5).

Condition (D2) is equivalent to the log of the cumulative distribution function of $\log(\Lambda_1)$ being concave. This is satisfied, for example, if $\log(\Lambda_1) = aY + b$, where a and b are real numbers and Y has a normal or an exponential distribution.

Many 'popular' Markov processes fit this model, some of which we now outline.

(a) The exponentially weighted moving average (EWMA) processes:

$$Y_{n+1} = \alpha Y_n + \xi_{n+1}, \qquad n \geq 0,$$

where $0 \leq \alpha < 1$ and the $\{\xi_i\}$ are i.i.d. continuous random variables. Define $M_n = \mathrm{e}^{Y_n}$ and $\Lambda_n = \mathrm{e}^{\xi_n}$. Here $\varphi(t) = t^\alpha$.

(b) Let $a > 0$ and $\varphi(t) = t + a$, so that $M_{n+1} = (M_n + a)\Lambda_{n+1}$. When $a = 1$ and Λ_{n+1} is a likelihood ratio ($\Lambda_{n+1} = f_1(X_{n+1})/f_0(X_{n+1})$, where the X_i are i.i.d. with density f_0), $\{M_n\}_{n\geq 0}$ is a sequence of Shiryaev–Roberts statistics for detecting a change in distribution of X_i, from density f_0 to f_1. The standard Shiryaev–Roberts procedure calls for setting $M_0 = 0$, specifying a threshold A, and declaring at $T_A = \min\{n \mid M_n > A\}$ that a change took place. A procedure $T_A^{Q_A}$ that starts at a random point $M_0 \sim Q_A$ has asymptotic optimality properties (cf. [3], [4], and [7]). Another setting is where r_i is the return on (one unit of) investment in the ith period and $\Lambda_i = 1 + r_i$, so that an investment of m units at the beginning of the ith period will be worth $m\Lambda_i$ at its end. If one invests a units at the beginning of the first period, reinvests the $a\Lambda_i$ units and adds another a units at the beginning of the second period, and continues this way (i.e. always reinvesting and adding a units at every period), then the process $M_{n+1} = \varphi(M_n)\Lambda_{n+1}$ with $\varphi(t) = t + a$ describes the scheme.

(c) The random walk reflected from the zero barrier:

$$Y_0 = 0, \quad Y_{n+1} = (Y_n + Z_{n+1})^+, \qquad n = 0, 1, \dots,$$

where the $\{Z_i\}$ are i.i.d. and $\mathrm{P}(Z_i < 0) > 0$. Note that on the positive half-plane the trajectory of the reflected random walk $\{Y_n\}_{n\geq 0}$ is identical to the trajectory of the Markov process $\{Y_n^*\}_{n\geq 0}$ given by the recursion

$$Y_0^* = 0, \qquad Y_{n+1}^* = (Y_n^*)^+ + Z_{n+1}, \quad n = 0, 1, \dots.$$

Therefore, if $\log A > 0$, we may operate with Y_n^* instead of Y_n and all conclusions will be the same. Define $M_n = \mathrm{e}^{Y_n^*}$ and $\Lambda_i = \mathrm{e}^{Z_i}$, so that

$$M_{n+1} = \max(M_n, 1)\Lambda_{n+1}, \qquad n \geq 0.$$

Here $\varphi(t) = \max(1, t)$. This process describes a broad class of single-channel queueing systems (see, e.g. [1]). This setting can also be applied to the Cusum scheme for detecting a change in distribution, when $Z_i = \log[f_1(X_i)/f_0(X_i)]$, and X_i, f_0, and f_1 are as in (b).

4. Discussion

(a) The problem addressed in this paper sheds some light on the behavior of a Markov process constrained to a certain set. Often one starts to observe a sequence well after its initiation. If in the past the process never exited a given set then the quasistationary distribution is relevant, especially since it often kicks in fairly rapidly.

While this is of considerable interest in its own merit, our interest stems from certain aspects in changepoint detection theory. Suppose that a system yields a sequence of independent observations $X_1, X_2, \dots, X_n$. When the system is in control, the observations are i.i.d. with known density f_0. The system may go out of control, in which case a quick detection of the occurrence of the change is desirable. Consider the case where the change is abrupt, being manifest by a sustained change of the density of the observations to f_1. A trigger-happy detection scheme may set off frequent false alarms, whereas a conservative scheme may cause a long delay of detection. Common operating characteristics of a detection scheme are the average run length to false alarm (ARL2FA) and average delay to detection (ADD), and the common constraint on the false alarm rate is a requirement that ARL2FA $\geq B$ for some $B \geq 1$.

The problem with a long history is finding a procedure that minimizes ADD subject to ARL2FA $\geq B$, and in tandem what this minimum is. One popular procedure that has certain optimality properties is the Shiryaev–Roberts procedure, which is based on the sequence of statistics

$$R_n = \sum_{k=1}^{n} \prod_{i=k}^{n} \frac{f_1(X_i)}{f_0(X_i)},$$

and raises an alarm at

$$T_A^{\mathrm{sr}} = \min\{n \mid R_n \geq A,\ n \geq 1\},$$

where A is suitably chosen so that ARL2FA $= B$. It is easy to see that the sequence $\{R_n\}_{n\geq 1}$ is a Markov process, satisfying the recursion

$$R_{n+1} = (R_n + 1)\frac{f_1(X_{n+1})}{f_0(X_{n+1})}, \qquad R_0 = 0.$$

Pollak [4] (see also [7]) showed that if R_n is started off at R_0 that has a quasistationary distribution Q_A (where A is selected in such a way that ARL2FA $= B$), then the rule that announces a detection the first time that the resulting Markov process exceeds A attains the minimum ADD up to an additive term $o(1)$, where $o(1) \to 0$ as $B \to \infty$. It is intuitive to expect that the run length to false alarm of such a procedure is stochastically increasing in A, something that implies monotonicity in A of the ARL2FA. This is important in the sense of guaranteeing (for given B) a unique solution to the equation $B =$ ARL2FA of $T_A^{Q_A}$.

(b) Some of the conditions (C1)–(C5) are 'natural', others less so. Condition (C2) is known in the literature as 'stochastic monotonicity', a formulation of a situation where one would expect the next observation to be larger when the present observation is large than when it is small. Condition (C4) means the same when the process is constrained to $[0, A]$. Condition (C3) is a brake on the rate of decrease in (C2) and condition (C5) does the same for (C4). In a similar vein, Theorem 1(ii) is a brake on the rate of increase (in A) of Theorem 1(i).

(c) At a first glance, the monotonicity properties in A of Q_A and $T_A^{Q_A}$ are 'obvious'. At a second glance, however, they are not easy to prove. As a matter of fact, monotonicity does not hold in full generality, as the following counterexample indicates.

Let $\{U_n\}$ be a time-homogeneous Markov process defined by the transition probabilities

$$\mathrm{P}(U_n \leq x \mid U_{n-1} = t) = \rho(t, x) = \begin{cases} \frac{1}{2}x & \text{if } t \leq 1, \\ 1 & \text{if } t > 1, \end{cases}$$

where $0 \leq x \leq 2$. Clearly, Q_A is a $U[0, A]$ distribution when $A \leq 1$. When $1 < A \leq 2$, obviously $Q_A(0) = 1 - Q_A(1)$, so that $Q_A(x) = Q_A(1)x/A + [1 - Q_A(1)]$ for $0 \leq x \leq A$. Plugging in $x = 1$ yields

$$Q_A(1) = \frac{A}{2A-1}, \qquad Q_A(0) = \frac{A-1}{2A-1}, \qquad Q_A(x) = \frac{x+A-1}{2A-1}.$$

So, for example, Q_1 and Q_2 intersect, meaning that there is no monotonicity relation between them.

Acknowledgements

The authors thank the anonymous referee for many constructive comments that improved this paper. This work was supported by the US Army Research Office MURI grant W911NF-06-1-0094, by the US Defense Threat Reduction Agency grant HDTRA1-10-1-0086, and by the US National Science Foundation grants CCF-0830419 and EFRI-1025043 at the University of Southern California. The work of Moshe Pollak was also supported by a grant from the Israel Science Foundation and by the Marcy Bogen Chair of Statistics at the Hebrew University of Jerusalem.

References

[1] Borovkov, A. A. (1976). *Stochastic Processes in Queuing Theory*. Springer, New York.

[2] Harris, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin.

[3] Moustakides, G. V., Polunchenko, A. S. and Tartakovsky, A. G. (2011). A numerical approach to performance analysis of quickest change-point detection procedures. *Statistica Sinica* **21,** 571–598.

[4] Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13,** 206–227.

[5] Pollak, M. and Siegmund, D. (1986). Convergence of quasistationary to stationary distributions for stochastically monotone Markov processes. *J. Appl. Prob.* **23,** 215–220.

[6] Pollett, P. K. (2008). Quasi-stationary distributions: a bibliography. Available at www.maths.uq.edu.au/~pkp/papers/qsds/qsds.pdf.

[7] Tartakovsky, A. G., Pollak, M. and Polunchenko, A. S. (2011). Third-order asymptotic optimality of the generalized Shiryaev–Roberts changepoint detection procedures. To appear in *Theory Prob. Appl.*