Statistica Sinica **19** (2009), 1729-1739

OPTIMALITY PROPERTIES OF THE SHIRYAEV-ROBERTS PROCEDURE

Moshe Pollak and Alexander G. Tartakovsky

The Hebrew University of Jerusalem and the University of Southern California

Abstract: In 1961, for detecting a change in the drift of a Brownian motion, Shiryaev introduced what is now usually referred to as the Shiryaev-Roberts procedure. This procedure has a number of optimality and asymptotic optimality properties in various settings. Shiryaev (1961, 1963), and more recently Feinberg and Shiryaev (2006), established exact optimality properties in the context of monitoring a Brownian motion for a (known) change of drift. Their method of proof relies on techniques particular to Brownian motion that are not applicable in discrete time. Here we establish analogous results in a general discrete time setting, where surveillance is not relegated to a change of mean or to normal observations only. Our method of proof relies on asymptotic Bayesian analysis and on renewal theory.

Key words and phrases: Changepoint problems, CUSUM procedures, sequential detection, Shiryaev-Roberts procedures.

1. Introduction

Changepoint problems deal with detecting a change in the state of a process, where information one has about the state of affairs is in the form of observations. In the sequential setting, observations are obtained one at a time and, as long as their behavior is consistent with the initial (or target) state, one is content to let the process continue. If the state changes, then one is interested in detecting that a change is in effect, usually as soon as possible after its occurrence.

Any detection policy may give rise to false alarms. Intuitively, the desire to detect a change quickly causes one to be (relatively) trigger-happy, which will bring about many false alarms if there is no change. On the other hand, attempting to avoid false alarms too strenuously will lead to a long delay between the time of occurrence of a real change and its detection. Common operating characteristics of a sequential detection policy are ARL2FA = the Average Run Length (the expected number of observations) to False Alarm (assuming that there is no change) and the AD2D = Average Delay to Detection (the expected delay between a real change and its detection). The gist of the changepoint problem is to produce a detection policy that (at least approximately) minimizes the AD2D subject to a bound on the ARL2FA. The constitution of a good policy depends very much on what is known about the stochastic behavior of the observations, both pre- and post-change.

Let $X_1, X_2, \ldots$ denote the series of observations, and let ν be the serial number of the first post-change observation. Let $\mathbb{P}_k$ and $\mathbb{E}_k$ denote probability and expectation when $\nu = k$, and let $\mathbb{P}_\infty$ and $\mathbb{E}_\infty$ denote the same when $\nu = \infty$ (i.e., there never is a change). A sequential change detection procedure is identified with a stopping time N on $X_1, X_2, \ldots$, i.e., $\{N \le n\} \in \mathscr{F}_n$, where $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$ is the sigma-algebra generated by the first n observations ($\mathscr{F}_0 = \{\varnothing, \Omega\}$). We assume that all random objects are defined on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$, $\mathscr{F} = \vee_{n \ge 0} \mathscr{F}_n$; $\{\mathscr{F}_n\}_{n \ge 0}$ satisfies the usual conditions.

In this paper, we consider the simple changepoint problem setting where the observations $X_1, X_2, \ldots$ are independent, iid pre-change and iid post-change, with known pre- and post-change distributions. Specifically, it is assumed that X_n has density f_0 (pre-change) for $n < \nu$ and density f_1 (post-change) for $n \ge \nu$, where both f_0 and f_1 are known, and only the value of ν, the point of change, is unknown. (In practice, often f_0 is known. Realistically, f_1 is not known, but the simple setting yields a benchmark for the best one can hope for.)

In this setting, Moustakides (1986) proved that the Cusum procedure (introduced by Page (1954)) is optimal in the sense of minimizing the worst-worst case (essential supremum) expected detection delay $\sup_{k \ge 1} \operatorname{ess\,sup}_\omega \mathbb{E}_k[(N-k)^+ | \mathscr{F}_k](\omega)$ over all stopping times N for which

$$\mathrm{ARL2FA}(N) = \mathbb{E}_\infty N \ge B, \tag{1.1}$$

where $B > 1$ is a value set before the surveillance begins. See also Lorden (1971) and Ritov (1990). For detecting a change in the drift of a Brownian motion, a similar result has been established by Beibel (1996) and Shiryaev (1996).

In 1961, for detecting a change in the drift of a Brownian motion, Shiryaev introduced what is now usually referred to as the Shiryaev-Roberts procedure (Shiryaev (1961, 1963) and Roberts (1966)). Shiryaev (1961, 1963) considered the problem of detecting a change in the mean of a Brownian motion when a stationary regime is in place, effected by a change possibly occurring in a distant future, after many false alarms have been experienced. He used features of Brownian motion to prove that, subject to a constraint on the rate of false alarms, this procedure is optimal for minimizing expected delay in detecting a change taking place at a far horizon against a stationary background of false alarms. This problem setting, called by Shiryaev "Quickest Detection of a Disorder in a Stationary Regime," has a variety of important surveillance applications. Feinberg and Shiryaev (2006) have shown that this procedure is also optimal in terms of minimizing $\int_0^\infty \mathbb{E}_t[(N-t)^+]dt$ (for the same Brownian motion model). They refer to this as "A Generalized Bayesian Setting".

For the general changepoint problem in a discrete time setting, Pollak (1985) introduced a randomized version of the Shiryaev-Roberts procedure that starts from a point sampled from the quasi-stationary distribution, and proved that it is asymptotically (as $B \to \infty$) almost optimal (within an additive term of order $o(1)$) in the sense of minimizing the supremum AD2D $\sup_{k\geq 1} \mathbb{E}_k(N-k|N \geq k)$ over all stopping times N that satisfy (1.1). It can be also deduced from Pollak (1985, 1987) that the conventional Shiryaev-Roberts procedure is asymptotically minimax for a low false alarm rate within an additive term of order $O(1)$.

In the present paper, we establish results analogous to those of Shiryaev (1961, 1963) and Feinberg and Shiryaev (2006) in a general discrete time setting where surveillance is not limited to a change of mean or to normal observations.

To be specific, in Section 2 we prove that the Shiryaev-Roberts procedure is (exactly) optimal in the sense of minimizing the "integral AD2D"$= \sum_{k=1}^\infty \mathbb{E}_k(N-k)^+$ for every $B > 1$ in the class of procedures with the ARL2FA constraint (1.1). This is instrumental in Section 3, where we consider the setting in which a change occurs in a distant future (i.e., ν is large) after being preceded by a stationary flow of false alarms. We prove that the Shiryaev-Roberts procedure is the best (exactly) that one can do in terms of minimizing the expected detection delay asymptotically when $\nu \to \infty$ in the class (1.1), for every $B > 1$.

Since Brownian motion techniques are not applicable in discrete time, our methods of proof are necessarily different from those of Shiryaev (1961, 1963) and Feinberg and Shiryaev (2006). In particular, the proof of Theorem 2 is based solely on our results in Section 2 and on renewal-theoretic considerations.

2. Minimizing Integral AD2D

Using the notation of the previous section, the Shiryaev-Roberts procedure calls for stopping and raising an alarm at

$$N_{A_B} = \min\{n \geq 1 : R_n \geq A_B\}, \tag{2.1}$$

where

$$R_n = \sum_{k=1}^n \frac{p(X_1, \ldots, X_n|\nu = k)}{p(X_1, \ldots, X_n|\nu = \infty)} = \sum_{k=1}^n \prod_{i=k}^n \frac{f_1(X_i)}{f_0(X_i)}, \tag{2.2}$$

and A_B is such that $\mathbb{E}_\infty N_{A_B} = B$.

Note that the Shiryaev-Roberts statistic R_n defined in (2.2) can be also computed recursively as

$$R_n = (1 + R_{n-1})\frac{f_1(X_n)}{f_0(X_n)}, \quad n \geq 1, \quad R_0 = 0. \tag{2.3}$$

For ease of exposition we assume throughout the paper that the likelihood ratio $f_1(X)/f_0(X)$ is $\mathbb{P}_\infty$-continuous. The case where the likelihood ratio is not continuous requires randomization on the stopping boundary A_B, i.e., deciding whether to continue or to stop when $R_n = A_B$. All our results are valid for this case as well, but presenting the fine points clutters the exposition with details that obscure the main ideas.

In Theorem 1 we prove that the Shiryaev-Roberts procedure is exactly optimal in the sense of minimizing the integral AD2D $= \sum_{k=1}^{\infty} \mathbb{E}_k(N-k)^+$ in the class of detection procedures $\mathbf{\Delta}_B = \{N : \text{ARL2FA}(N) \ge B\}$, when the mean time to false alarm is not less than the given number $B > 1$. We begin with a sketch of the argument why one may expect this to be true.

Consider first the following Bayesian problem, denoted by $\mathcal{B}(\rho, c)$. Suppose ν is a random variable independent of the observations that has a geometric prior distribution

$$P(\nu = k) = \rho(1-\rho)^{k-1}, \quad k \ge 1, \tag{2.4}$$

and the losses associated with stopping at time N are 1 if $N < \nu$ and $c \cdot (N - \nu)$ if $N \ge \nu$, where $0 < \rho < 1$ and $c > 0$ are fixed constants. For $\mathcal{A} \in \mathscr{F}$, define the probability $\mathbb{P}^\rho(\mathcal{A}) = \sum_{k=1}^{\infty} \rho(1-\rho)^{k-1}\mathbb{P}_k(\mathcal{A})$ and let $\mathbb{E}^\rho$ denote the corresponding expectation.

Solution of $\mathcal{B}(\rho, c)$ requires minimization of the expected loss

$$\varphi_{c,\rho}(N) = \mathbb{P}^\rho(N < \nu) + c\mathbb{E}^\rho(N - \nu)^+, \tag{2.5}$$

and the Bayes rule for this problem is given by the Shiryaev procedure (cf. Shiryaev (1963, 1978))

$$T_{\rho,c} = \min\left\{n \ge 1 : \mathbb{P}^\rho(\nu \le n | \mathscr{F}_n) \ge \delta_{\rho,c}\right\}, \tag{2.6}$$

where $0 < \delta_{\rho,c} < 1$ is an appropriate threshold.

Obviously, the $\mathcal{B}(\rho, c)$ problem is equivalent to maximizing

$$\frac{1}{\rho}[1 - \varphi_{c,\rho}(N)] = \frac{\mathbb{P}^\rho(N \ge \nu)}{\rho} - c\frac{\mathbb{E}^\rho(N-\nu)^+}{\rho}.$$

In the proof of Theorem 1 we show that, for any stopping time N,

$$\frac{\mathbb{P}^\rho(N \ge \nu)}{\rho} \xrightarrow[\rho\to 0]{} \mathbb{E}_\infty N, \quad \frac{\mathbb{E}^\rho(N-\nu)^+}{\rho} \xrightarrow[\rho\to 0]{} \sum_{k=1}^{\infty} \mathbb{E}_k(N-k)^+.$$

Hence

$$\frac{1}{\rho}[1 - \varphi_{c,\rho}(N)] \xrightarrow[\rho\to 0]{} \mathbb{E}_\infty N - c\sum_{k=1}^{\infty} \mathbb{E}_k(N-k)^+,$$

which should be maximized in the class $\mathbf{\Delta}_B$.

We also show that the Shiryaev procedure $T_{\rho,c}$ converges to the Shiryaev-Roberts procedure N_{A_B} as $\rho \to 0$. Therefore, it stands to reason that the integral $\text{AD2D} = \sum_{k=1}^{\infty} \mathbb{E}_k (N-k)^+$ is minimized subject to $\mathbb{E}_\infty N \geq B$.

Theorem 1. *Let A_B be chosen so that* $\text{ARL2FA}(N_{A_B}) = B$. *Then the Shiryaev-Roberts procedure defined by* (2.1) *and* (2.3) *minimizes*

$$\sum_{k=1}^{\infty} \mathbb{E}_k (N-k)^+ \tag{2.7}$$

over all stopping times N that satisfy $\mathbb{E}_\infty N \geq B$, i.e.,

$$\inf_{N \in \mathbf{\Delta}_B} \sum_{k=1}^{\infty} \mathbb{E}_k (N-k)^+ = \sum_{k=1}^{\infty} \mathbb{E}_k (N_{A_B}-k)^+ \quad \text{for every} \quad B > 1,$$

where $\mathbf{\Delta}_B = \{N : \text{ARL2FA}(N) \geq B\}$.

Proof. Consider the Bayesian problem $\mathcal{B}(\rho, c)$ with the geometric prior distribution (2.4) and the average loss (2.5). Shiryaev (1963, 1978) proved that the expected loss (2.5) for the problem $\mathcal{B}(\rho, c)$ is minimized by the stopping time (2.6). Applying Bayes' formula, it is easy to see that

$$\mathbb{P}^\rho(\nu \leq n | \mathscr{F}_n) = \frac{R_{\rho,n}}{R_{\rho,n} + 1/\rho},$$

where

$$R_{\rho,n} = \sum_{k=1}^{n} \prod_{i=k}^{n} \left(\frac{1}{1-\rho} \frac{f_1(X_i)}{f_0(X_i)} \right).$$

Hence, the Shiryaev rule can be written in the equivalent form

$$T_{\rho,c} = \min \{ n \geq 1 : R_{\rho,n} \geq A_{\rho,c} \}, \tag{2.8}$$

where $A_{\rho,c} = (1/\rho)[\delta_{\rho,c}/(1-\delta_{\rho,c})]$.

Note first that $R_{\rho,n} \xrightarrow[\rho \to 0]{} R_n$.

By Theorem 1 of Pollak (1985), there exist a constant $0 < c^* < \infty$ and a sequence $\{\rho_i, c_i\}_{i=1}^{\infty}$ with $\rho_i \xrightarrow[i \to \infty]{} 0$, $c_i \xrightarrow[i \to \infty]{} c^*$, such that N_{A_B} is the limit of the Bayes rules T_{ρ_i, c_i} as $i \to \infty$. Furthermore,

$$\limsup_{\rho \to 0, c \to c^*} \frac{1 - \varphi_{c,\rho}(T_{\rho,c})}{1 - \varphi_{c,\rho}(N_{A_B})} = 1, \tag{2.9}$$

where $\varphi_{c,\rho}(N)$ is the expected loss associated with using the stopping time N for $\mathcal{B}(\rho, c)$.

Now, for any stopping time N,

$$\begin{aligned}\frac{1}{\rho}[1-\varphi_{c,\rho}(N)] &= \frac{1}{\rho}\left[(1-\mathbb{P}^\rho(N<\nu)) - c\mathbb{E}^\rho(N-\nu)^+\right]\\ &= \frac{\mathbb{P}^\rho(N\ge\nu)}{\rho}\left[1-c\mathbb{E}^\rho(N-\nu|N\ge\nu)\right].\end{aligned}$$

Since

$$\begin{aligned}\frac{\mathbb{P}^\rho(N\ge\nu)}{\rho} &= \frac{1}{\rho}\sum_{k=1}^\infty \mathbb{P}_k(N\ge k)\rho(1-\rho)^{k-1}\\ &= \sum_{k=1}^\infty \mathbb{P}_\infty(N\ge k)(1-\rho)^{k-1}\\ &\xrightarrow[\rho\to 0]{} \sum_{k=1}^\infty \mathbb{P}_\infty(N\ge k) = \mathbb{E}_\infty N\end{aligned}$$

(where we used the fact that $\{N<k\}\in\mathscr{F}_{k-1}$ and, hence, $\mathbb{P}_k(N\ge k) = 1-\mathbb{P}_k(N<k) = 1-\mathbb{P}_\infty(N<k) = \mathbb{P}_\infty(N\ge k)$) and

$$\begin{aligned}\frac{\mathbb{P}^\rho(N\ge\nu)\mathbb{E}^\rho(N-\nu|N\ge\nu)}{\rho} &= \frac{\mathbb{E}^\rho(N-\nu;N\ge\nu)}{\rho}\\ &= \frac{1}{\rho}\sum_{k=1}^\infty \mathbb{E}_k(N-k;N\ge k)\rho(1-\rho)^{k-1}\\ &= \sum_{k=1}^\infty \mathbb{E}_k(N-k;N\ge k)(1-\rho)^{k-1}\\ &\xrightarrow[\rho\to 0]{} \sum_{k=1}^\infty \mathbb{E}_k(N-k;N\ge k) = \sum_{k=1}^\infty \mathbb{E}_k(N-k)^+,\end{aligned}$$

it follows that for any stopping time N that has finite ARL2FA

$$\frac{1}{\rho}\Big[1-\varphi_{c,\rho}(N)\Big] \xrightarrow[\rho\to 0]{} \mathbb{E}_\infty N - c\sum_{k=1}^\infty \mathbb{E}_k(N-k)^+, \tag{2.10}$$

which together with (2.9) establishes that the Shiryaev-Roberts procedure minimizes (2.7) over all stopping times that satisfy $\mathbb{E}_\infty N = B$. Note that if $B_1 > B$, then $N_{A_{B_1}}$ is stochastically larger than N_{A_B}, i.e., all expectations in (2.7) become larger. This implies that the Shiryaev-Roberts procedure minimizes (2.7) in the class $\boldsymbol{\Delta}_B$. This completes the proof of the theorem.

Corollary 1. *The Shiryaev-Roberts procedure defined by* (2.1) *and* (2.2) *minimizes*

$$\frac{\sum_{k=1}^\infty \mathbb{E}_k(N-k|N\ge k)\mathbb{P}_\infty(N\ge k)}{\sum_{j=1}^\infty \mathbb{P}_\infty(N\ge j)} \tag{2.11}$$

over all stopping times N that satisfy $\mathbb{E}_\infty N \geq B$, i.e.,

$$\inf_{N \in \mathbf{\Delta}_B} \sum_{k=1}^{\infty} w_k(N)\mathbb{E}_k(N-k|N \geq k) = \sum_{k=1}^{\infty} w_k(N_{A_B})\mathbb{E}_k(N_{A_B}-k|N_{A_B} \geq k),$$

where

$$w_k(N) = \frac{\mathbb{P}_\infty(N \geq k)}{\sum_{j=1}^{\infty}\mathbb{P}_\infty(N \geq j)},$$

and the threshold A_B is selected so that $\mathbb{E}_\infty N_{A_B} = B$.

Proof. Obviously, $\sum_{j=1}^{\infty}\mathbb{P}_\infty(N \geq j) = \mathbb{E}_\infty N = B$, so the denominator in (2.11) is constant over all stopping times with exact equality $\mathbb{E}_\infty N = B$. As for the numerator,

$$\begin{aligned}\mathbb{E}_k(N-k|N \geq k)\mathbb{P}_\infty(N \geq k) &= \mathbb{E}_k(N-k|N \geq k)\mathbb{P}_k(N \geq k)\\ &= \mathbb{E}_k(N-k; N \geq k) = \mathbb{E}_k(N-k)^+. \quad (2.12)\end{aligned}$$

Application of Theorem 1 concludes the proof for stopping times $\{N : \mathbb{E}_\infty N = B\}$. It remains to prove that this is true for all stopping times $N \in \mathbf{\Delta}_B$.

Write

$$\mathcal{J}(N) = \frac{\sum_{k=1}^{\infty}\mathbb{E}_k(N-k|N \geq k)\mathbb{P}_\infty(N \geq k)}{\sum_{j=1}^{\infty}\mathbb{P}_\infty(N \geq j)} = \frac{\sum_{k=1}^{\infty}\mathbb{E}_k(N-k)^+}{\mathbb{E}_\infty N},$$

and let N be such that $\mathbb{E}_\infty N = B_1 > B$.

Define a randomized stopping time

$$T = \begin{cases} N & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases}$$

where $p = B/B_1$.

Note that $\mathcal{J}(T) = \mathcal{J}(N)$ for every $0 < p \leq 1$, since $\mathbb{E}_k(T-k)^+ = p\mathbb{E}_k(N-k)^+$ and $\mathbb{E}_\infty T = p\mathbb{E}_\infty N$. Therefore, for any stopping time N such that $\mathbb{E}_\infty N > B$, we can find another stopping time T with $\mathbb{E}_\infty T = B$ and such that $\mathcal{J}(T) = \mathcal{J}(N)$, which means that it is sufficient to optimize over stopping times that satisfy exact equality $\mathbb{E}_\infty N = B$. Finally, since the optimum over stopping times with $\mathbb{E}_\infty N = B$ is the Shiryaev-Roberts procedure that does not randomize 0, it follows that this procedure is optimal in the class $\mathbf{\Delta}_B$.

Remark 1. It is not feasible to obtain an analytic closed form for A_B (to guarantee the exact equality $\mathbb{E}_\infty N_{A_B} = B$) and, if precision is absolutely necessary, the calculation is done best by (very tedious) Monte Carlo or by solving integral equations numerically. An approximation (cf. Pollak (1987) and Yakir (1995))

$$\mathbb{E}_\infty N_A \approx \frac{A}{\gamma} \quad (2.13)$$

can be obtained by noticing that $R_n - n$ is a $\mathbb{P}_\infty$ - martingale with zero expectation, so that by the optional sampling theorem $\mathbb{E}_\infty(R_{N_A} - N_A) = 0$. Hence $\mathbb{E}_\infty N_A = \mathbb{E}_\infty R_{N_A}$, and, since R_{N_A} is the first excess over A, renewal theory can be applied to the "overshoot" $\log(R_{N_A}) - \log A$. The constant γ in (2.13) depends on the model, satisfies $0 < \gamma < 1$, and can be computed numerically. Therefore, $A = B\gamma$ guarantees $\mathbb{E}_\infty N_A \approx B$. This approximation is asymptotically accurate when $B \to \infty$ and is fairly accurate already for small values of B (say $B \geq 10$). See, e.g., Pollak and Tartakovsky (2008).

While Theorem 1 and Corollary 1 are of interest in their own right, they are useful for proving another interesting optimality result, as will become apparent in the next section.

3. Optimality for a Change Appearing after Many Re-Runs

Consider a context in which it is of utmost importance to detect a real change as quickly as possible after its occurrence, even at the price of raising many false alarms (using a repeated application of the same stopping rule) before the change occurs. This essentially means that the changepoint ν is very large compared to the constant B which, in this case, defines the mean time between consecutive false alarms.

To be more specific, let $N_{A_B}^{(1)}, N_{A_B}^{(2)}, \ldots$ be sequential independent repetitions of N_{A_B} defined in (2.1) and let, for $j \geq 1$,

$$Q_j = N_{A_B}^{(1)} + N_{A_B}^{(2)} + \cdots + N_{A_B}^{(j)} \tag{3.1}$$

be the time of the j-th alarm, i.e.,

$$N_{A_B}^{(i)} = \min\left\{n \geq Q_{i-1} + 1 : R_n^{(i)} \geq A_B\right\} - Q_{i-1}, \tag{3.2}$$

where $N_{A_B}^{(0)} = Q_0 = 0$ and

$$R_n^{(i)} = \left(1 + R_{n-1}^{(i)}\right) \frac{f_1(X_n)}{f_0(X_n)} \quad \text{for} \quad Q_{i-1} + 1 \leq n \leq Q_i, \quad R_{Q_{i-1}}^{(i)} = 0. \tag{3.3}$$

Thus $R_n^{(i)}$, $n \geq Q_{i-1} + 1$, is nothing but the Shiryaev-Roberts statistic that is renewed from scratch after the $(i-1)$st false alarm (under $\mathbb{P}_\infty$) and is applied to the segment of data $X_{Q_{i-1}+1}, X_{Q_{i-1}+2}, \ldots$.

Note that $\mathbb{E}_\infty N_{A_B}^{(i)} = B$ for $i \geq 1$.

Let $J_\nu = \min\{j \geq 1 : Q_j \geq \nu\}$, i.e., Q_{J_ν} is the time of detection of a true change that occurs at ν after $J_\nu - 1$ false alarms have been raised.

Our next theorem states that the Shiryaev-Roberts procedure defined by Q_{J_ν} is asymptotically (as $\nu \to \infty$) optimal with respect to the expected delay

$\mathbb{E}_\nu(Q_{J_\nu}-\nu)$ in the class of detection procedures for which the mean time between false alarms is not less than B. Note that this result is not asymptotic with respect to the ARL2FA, it holds for every $B>1$.

Theorem 2. *Let ν be the time of the change. Let $N_{A_B}^{(1)}, N_{A_B}^{(2)}, \ldots$ be sequential independent repetitions of N_{A_B} as defined in* (3.2) *and* (3.3)*, and let $Q_1, Q_2, \ldots$ be as in* (3.1)*. Let $J_\nu = \min\{j : Q_j \geq \nu\}$.*

(i) $\lim_{\nu\to\infty} \mathbb{E}_\nu(Q_{J_\nu}-\nu)$ *exists.*

(ii) *Suppose a detection procedure N with* $\mathrm{ARL2FA}(N) \geq B$ *is applied repeatedly. Let $N_1, N_2, \ldots$ be sequential repetitions of N, let $W_j = \sum_{i=1}^{j} N_i$, and let $K_\nu = \min\{j : W_j \geq \nu\}$. Then*

$$\lim_{\nu\to\infty} \mathbb{E}_\nu(Q_{J_\nu}-\nu) \leq \lim_{\nu\to\infty} \mathbb{E}_\nu(W_{K_\nu}-\nu), \tag{3.4}$$

i.e., the Shiryaev-Roberts procedure is optimal for every $B>1$ in the class $\boldsymbol{\Delta}_B = \{N : \mathbb{E}_\infty N \geq B\}$.

Proof. Proof of (i). By renewal theory, the distribution of $\nu - Q_{J_\nu-1}$ has a limit

$$\lim_{\nu\to\infty} \mathbb{P}_\nu\left(\nu - Q_{J_\nu-1} = k\right) = \frac{\mathbb{P}_\infty(N_{A_B} \geq k)}{\sum_{j=1}^{\infty} \mathbb{P}_\infty(N_{A_B} \geq j)} \quad \text{for} \quad k = 1, 2, \ldots \tag{3.5}$$

(see, e.g., Feller (1966, p.356)).

When conditioning on $\nu - Q_{J_\nu-1} = k$, the observations $X_{Q_{J_\nu-1}+1}$, $X_{Q_{J_\nu-1}+2}$, $\ldots$ $X_{\nu-1}$, X_ν, $\ldots$ behave exactly like $X_1, X_2, \ldots, X_{\nu-1}, X_\nu, \ldots$ when $\nu = k$. Therefore, by conditioning on $\nu - Q_{J_\nu-1}$, using (3.5) and letting N_{A_B} be independent of $N_{A_B}^{(1)}, N_{A_B}^{(2)}, \ldots$, we obtain

$$\begin{aligned}
\mathbb{E}_\nu(Q_{J_\nu}-\nu) &= \mathbb{E}_\nu\left[\mathbb{E}_\nu\left(Q_{J_\nu}-\nu | \nu - Q_{J_\nu-1}\right)\right] \\
&= \sum_{k=1}^{\nu} \mathbb{E}_k\left(N_{A_B}-k | \nu - Q_{J_\nu-1} = k, N_{A_B} \geq k\right) \mathbb{P}_\infty\left(\nu - Q_{J_\nu-1} = k\right) \\
&= \sum_{k=1}^{\nu} \mathbb{E}_k\left(N_{A_B}-k | N_{A_B} \geq k\right) \mathbb{P}_\infty\left(\nu - Q_{J_\nu-1} = k\right) \\
&\xrightarrow[\nu\to\infty]{} \frac{\sum_{k=1}^{\infty} \mathbb{E}_k\left(N_{A_B}-k | N_{A_B} \geq k\right) \mathbb{P}_\infty\left(N_{A_B} \geq k\right)}{\sum_{j=1}^{\infty} \mathbb{P}_\infty\left(N_{A_B} \geq j\right)} \\
&= \frac{\sum_{k=1}^{\infty} \mathbb{E}_k(N_{A_B}-k)^+}{\mathbb{E}_\infty N_{A_B}} = \frac{\sum_{k=1}^{\infty} \mathbb{E}_k(N_{A_B}-k)^+}{B},
\end{aligned}$$

which completes the proof of (i).

Proof of (ii). The same argument as in the proof of (i) yields

$$\lim_{\nu\to\infty} \mathbb{E}_\nu(W_{K_\nu}-\nu) = \frac{\sum_{k=1}^{\infty} \mathbb{E}_k(N-k)^+}{\mathbb{E}_\infty N} = \mathcal{J}(N).$$

By Corollary 1, $\mathcal{J}(N_{A_B}) \le \mathcal{J}(N)$ for any $N \in \mathbf{\Delta}_B$, which concludes the proof.

Remark 2. It is worth noting that Theorem 2 is important in a variety of surveillance applications such as target detection and tracking, rapid detection of intrusions in computer networks, and environmental monitoring, to name a few. In all of these applications, it is of utmost importance to detect very rapidly changes that may occur in a distant future, in which case the true detection of a real change may be preceded by a long interval with frequent false alarms that are being filtered by a separate mechanism or algorithm. For example, falsely initiated target tracks are usually filtered by a track confirmation/deletion algorithm; false detections of attacks in computer networks in anomaly-based Intrusion Detection Systems (IDS) may be filtered by Signature-based IDS algorithms, etc. See, e.g., Tartakovsky (1991), Tartakovsky and Veeravalli (2004) and Tartakovsky, Rozovskii, Blažek and Kim (2006). The practical implication of Theorem 2 is that in these circumstances one has reason to prefer the Shiryaev-Roberts procedure to other surveillance schemes.

Acknowledgements

Moshe Pollak is Marcy Bogen Professor of Statistics at the Hebrew University of Jerusalem. His work 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 Alexander Tartakovsky was supported in part 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.

We are grateful to George Moustakides for useful discussions that improved the proof of Corollary 1.

References

Beibel, M. (1996). A note on Ritov's Bayes approach to the minimax property of the Cusum procedure. *Ann. Statist.* **24**, 1804-1812.

Feinberg, E. A. and Shiryaev, A. N. (2006). Quickest detection of drift change for Brownian motion in generalized Bayesian and minimax settings. *Statist. Decisions* **24**, 445-470.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications.* Vol. II, Wiley, New York.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42**, 1897-1908.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14**, 1379-1387.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100-115.

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

Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15**, 749-779.

Pollak, M. and Tartakovsky, A. G. (2008). Asymptotic exponentiality of the distribution of first exit times for a class of Markov processes with applications to quickest change detection. *Teoriya Veroyatn. Primen.* **53**, 500-515.

Ritov, Y. (1990). Decision theoretic optimality of the Cusum procedure. *Ann. Statist.* **18**, 1466-1469.

Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8**, 411-430.

Shiryaev, A. N. (1961). The problem of the most rapid detection of a disturbance in a stationary process. *Soviet Math. Dokl.* **2**, 795-799 (translation from Dokl. Akad. Nauk SSSR **138**, 1039-1042, 1961).

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **8**, 22-46.

Shiryaev, A. N. (1978). *Optimal Stopping Rules.* Springer-Verlag, New York.

Shiryaev, A. N. (1996). Minimax optimality of the method of cumulative sum (cusum) in the case of continuous time. *Russian Math. Surveys* **51**, 750-751.

Tartakovsky, A. G. (1991). *Sequential Methods in the Theory of Information Systems.* Radio i Svyaz', Moscow (in Russian).

Tartakovsky, A. G., Rozovskii, B. L., Blažek, R. and Kim, H. (2006). Detection of intrusions in information systems by sequential change-point methods (with discussion). *Statistical Methodology* **3**, 252-340.

Tartakovsky, A. G. and Veeravalli, V. V. (2004). Change-point detection in multichannel and distributed systems with applications. In *Applications of Sequential Methodologies* (Edited by N. Mukhopadhyay, S. Datta and S. Chattopadhyay), 331-363. Marcel Dekker, Inc., New York.

Yakir, B. (1995). A note on the run length to false alarm of a change-point detection policy. *Ann. Statist.* **23**, 272-281.

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 Avenue, Los Angeles, CA 90089-2532, U.S.A.

E-mail: tartakov@math.usc.edu

(Received November 2007; accepted June 2008)