ON OPTIMALITY OF THE SHIRYAEV–ROBERTS PROCEDURE FOR DETECTING A CHANGE IN DISTRIBUTION[1]

By Aleksey S. Polunchenko and Alexander G. Tartakovsky

University of Southern California, Los Angeles

In 1985, for detecting a change in distribution, Pollak introduced a specific minimax performance metric and a randomized version of the Shiryaev–Roberts procedure where the zero initial condition is replaced by a random variable sampled from the quasi-stationary distribution of the Shiryaev–Roberts statistic. Pollak proved that this procedure is third-order asymptotically optimal as the mean time to false alarm becomes large. The question of whether Pollak's procedure is strictly minimax for any false alarm rate has been open for more than two decades, and there were several attempts to prove this strict optimality. In this paper, we provide a counterexample which shows that Pollak's procedure is not optimal and that there is a strictly optimal procedure which is nothing but the Shiryaev–Roberts procedure that starts with a specially designed deterministic point.

1. Introduction and preliminaries. Changepoint problems deal with detecting changes in distributions of observed data that occur at unknown points in time. Let $X_1, X_2, \ldots$ be the series of observations being monitored, and let ν be the serial number of the last pre-change observation, so that $X_{\nu+1}$ is the first post-change observation. Let P_ν and E_ν denote probability and expectation when the change occurs at $\nu + 1$ for a fixed $0 \leq \nu < \infty$, and let P_∞ and E_∞ denote the same when $\nu = \infty$ (i.e., there never is a change). A sequential change detection procedure is a stopping time T adapted to the observations $X_1, X_2, \ldots$, that is, $\{T \leq n\} \in \mathscr{F}_n$, where $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$ is the sigma-algebra generated by the first n observations.

Common operating characteristics of a sequential detection procedure are the Average Run Length (ARL) to False Alarm, that is, the expected number of observations to an alarm assuming that there is no change, and the Average Delay to Detection, that is, the expected delay between a change and its detection. The goal is to find a detection procedure that minimizes the average detection delay subject to a bound on the ARL to false alarm.

In this paper, we will be interested in the simple changepoint problem setting, where the observations are independent, i.i.d. pre-change with density f_∞ and i.i.d.

post-change with density f_0. In other words, it is assumed that X_n has density f_∞ for $n \le \nu$ and density f_0 for $n > \nu$, where both f_∞ and f_0 are known but the changepoint ν is unknown. Therefore, the conditional density of the sample $(X_1, \ldots, X_n)$ for the fixed changepoint is

$$p(X_1, \ldots, X_n | \nu = k) = \prod_{i=1}^{k} f_\infty(X_i) \times \prod_{i=k+1}^{n} f_0(X_k),$$

where $\prod_{i=j}^{m} f(X_i) = 1$ when $j > m$.

In 1961, for detecting a change in the drift of a Brownian motion, Shiryaev introduced a change detection procedure, which is now usually referred to as the Shiryaev–Roberts (SR) procedure [Shiryaev (1961), (1963); Roberts (1966)]. The SR procedure calls for stopping and raising an alarm at

(1) $$T_{\mathrm{sr}}(A) = \inf\{n \ge 1 : R_n \ge A\}, \qquad \inf\{\varnothing\} = \infty,$$

where

(2) $$R_n = \sum_{k=0}^{n-1} \frac{p(X_1, \ldots, X_n | \nu = k)}{p(X_1, \ldots, X_n | \nu = \infty)} = \sum_{k=1}^{n} \prod_{i=k}^{n} \frac{f_0(X_i)}{f_\infty(X_i)}$$

is the SR statistic, and $A > 0$ is a threshold that controls the false alarm rate.

This procedure has a number of interesting optimality properties. In particular, if $A = A_\gamma$ is such that $\mathsf{E}_\infty[T_{\mathrm{sr}}(A_\gamma)] = \gamma$, then it minimizes the *integral average detection delay*

$$\mathcal{I}(T) = \frac{\sum_{\nu=0}^{\infty} \mathsf{E}_\nu (T - \nu)^+}{\mathsf{E}_\infty T}$$

over all stopping times T that satisfy

(3) $$\mathsf{E}_\infty T \ge \gamma,$$

where $\gamma > 1$ is a value set before the surveillance begins [cf. Pollak and Tartakovsky (2009) and also Feinberg and Shiryaev (2006) for the Brownian motion model].

Note that the SR statistic (2) can be written recursively as

(4) $$R_n = (1 + R_{n-1})\Lambda_n, \qquad n \ge 1, \qquad R_0 = 0,$$

where $\Lambda_n = f_0(X_n)/f_\infty(X_n)$ is the likelihood ratio. Therefore, the classical SR statistic starts from 0.

Pollak (1985) introduced a natural worst-case detection delay measure— *supremum average delay to detection*

$$\mathcal{J}_{\mathrm{P}}(T) = \sup_{0 \le \nu < \infty} \mathbf{E}_\nu(T - \nu | T > \nu),$$

and attempted to find an optimal procedure that would minimize $\mathcal{J}_P(T)$ over all procedures subject to constraint (3). Pollak's idea was to modify the SR statistic by randomization of the initial condition R_0 in (4) in order to make it an equalizer [i.e., to make the conditional average detection delay $\mathsf{E}_\nu(T - \nu|T > \nu)$ independent of the changepoint ν]. Pollak's version of the SR procedure starts from a random point sampled from the quasi-stationary distribution of the SR statistic R_n. He proved that this "randomized" procedure is asymptotically (as $\gamma \to \infty$) optimal within an additive term of order $o(1)$ in the sense of minimizing the supremum average detection delay $\mathcal{J}_P(T)$.

To be specific, let, for $B > 0$,

$$\mathsf{Q}_B(x) = \lim_{n \to \infty} \mathsf{P}_\infty\big(R_n \le x | T_{\mathrm{sr}}(B) > n\big)$$

denote the quasi-stationary distribution of the SR statistic, and let $R_n^{\mathsf{Q}_B}$ be given recursively

$$(5) \qquad R_n^{\mathsf{Q}_B} = (1 + R_{n-1}^{\mathsf{Q}_B})\Lambda_n, \qquad n \ge 1, \qquad R_0^{\mathsf{Q}_B} \sim \mathsf{Q}_B,$$

where $R_0^{\mathsf{Q}_B} \sim \mathsf{Q}_B$ means that $R_0^{\mathbf{Q}_B}$ is a random variable distributed according to the quasi-stationary distribution Q_B. The corresponding stopping time is given by

$$(6) \qquad T_{\mathrm{srp}}(B) = \inf\{n \ge 1 : R_n^{\mathsf{Q}_B} \ge B\}, \qquad \inf\{\varnothing\} = \infty.$$

Pollak (1985) proved that if $B = B_\gamma$ is selected so that $\mathsf{E}_\infty[T_{\mathrm{srp}}(B_\gamma)] = \gamma$, then

$$(7) \qquad \mathcal{J}_P(T_{\mathrm{srp}}(B_\gamma)) - \inf_{\{T:\mathsf{E}_\infty T \ge \gamma\}} \mathcal{J}_P(T) = o(1) \qquad \text{as } \gamma \to \infty,$$

where $o(1) \to 0$ as $\gamma \to \infty$. We will call this asymptotic optimality property *third-order asymptotic optimality* as opposed to the second-order optimality when the corresponding difference is bounded [i.e., $O(1)$] and the first-order optimality when the ratio of the corresponding values tends to 1. Therefore, the procedure given by (5) and (6), which we will refer to as the Shiryaev–Roberts–Pollak (SRP) procedure, is third-order asymptotically optimal for the low false alarm rate. Note that this result is extremely strong since the difference between the average detection delays in (7) is asymptotically small while each of them is of order $O(\log \gamma)$ (i.e., both terms go to infinity). It can be also deduced from Pollak (1985, 1987) that the conventional SR procedure is asymptotically minimax for a low false alarm rate within an additive term of order $O(1)$, that is, it is only second-order asymptotically optimal.

Since the SRP procedure is an equalizer, that is, $\mathcal{J}_P(T_{\mathrm{srp}}) = \mathsf{E}_0 T_{\mathrm{srp}} = \mathsf{E}_\nu(T_{\mathrm{srp}} - \nu|T_{\mathrm{srp}} > \nu)$ for all $\nu \ge 0$, it is tempting for one to conjecture that it may in fact be *strictly* optimal for every $\gamma > 1$. However, to date there is no proof or disproof of this conjecture [see Yakir (1997) and Mei (2006)]. Recent work of Moustakides, Polunchenko and Tartakovsky (2011) indicates that the SRP procedure may not be

exactly optimal and partially sheds light on this issue by considering a generalization of the SR procedure that starts from a specially designed deterministic point r. To emphasize the dependence on the starting point, this procedure will be referred to as the SR-r procedure. Specifically, define the stopping time

$$(8) \qquad T_{\mathrm{sr}}^r(A) = \inf\{n \geq 1 : R_n^r \geq A\}, \qquad \inf\{\varnothing\} = \infty,$$

where R_n^r obeys the recursion

$$(9) \qquad R_n^r = (1 + R_{n-1}^r)\Lambda_n, \qquad n \geq 1, \qquad R_0^r = r \geq 0.$$

Solving numerically integral equations for performance metrics for two examples that involve Gaussian and exponential models, Moustakides, Polunchenko and Tartakovsky (2011) found that the SR-r procedure (with a certain $r = r_\gamma$ that depends on γ) uniformly outperforms the SRP procedure, that is, $\mathsf{E}_\nu(T_{\mathrm{sr}}^r - \nu | T_{\mathrm{sr}}^r > \nu) < \mathsf{E}_0 T_{\mathrm{srp}}$ for all $\nu \geq 0$. We believe that these results present serious evidence against optimality of the SRP procedure. However, this may not be completely convincing since a small numerical error can be present in such experiments.

In the present paper, we construct a counterexample where all computations can be performed analytically. This example proves that the SRP procedure is not optimal while the SR-r procedure with a deterministic initialization is optimal. This result answers a long-standing question on optimality of the SRP procedure and opens a new direction in the quest for the unknown optimum.

2. The main theorem and integral equations for operating characteristics.
Theorem 1 below provides a lower bound for the infimum of Pollak's worst-case measure $\mathcal{J}_{\mathrm{P}}(T)$ which will be used to find the optimal changepoint detection procedure in Section 3. Note that a proof sketch of this lower bound has been previously given in Moustakides, Polunchenko and Tartakovsky (2011). Here we provide a complete proof.

We first need the following lemma which establishes optimality of the SR-r procedure with respect to the integral average detection delay.

LEMMA 1. *Let*

$$(10) \qquad \mathcal{I}_r(T) = \frac{r\mathsf{E}_0 T + \sum_{\nu=0}^{\infty} \mathsf{E}_\nu(T-\nu)^+}{r + \mathsf{E}_\infty T}$$

and let $T_{\mathrm{sr}}^r(A_\gamma)$ be the SR-r detection procedure with $\mathsf{E}_\infty[T_{\mathrm{sr}}^r(A_\gamma)] = \gamma$. For any $r \geq 0$, the SR-r procedure minimizes $\mathcal{I}_r(T)$ over all procedures with $\mathsf{E}_\infty T \geq \gamma$, that is,

$$(11) \qquad \inf_{\{T : \mathsf{E}_\infty T \geq \gamma\}} \mathcal{I}_r(T) = \mathcal{I}_r(T_{\mathrm{sr}}^r(A_\gamma)).$$

PROOF. The proof of (11) for $r = 0$ is given in Pollak and Tartakovsky (2009), Theorem 1 and Corollary 1. We now give its extension for an arbitrary positive r.

Consider the following Bayesian problem, which will be denoted by $\mathcal{B}(\pi, p, c)$. Suppose ν is a random variable (independent of the observations) with a zero modified geometric distribution

$$P(\nu < 0) = \pi, \qquad P(\nu = k) = (1 - \pi)p(1 - p)^k, \qquad k \geq 0,$$

and the losses associated with stopping at time T are 1 if $T \leq \nu$ and $c \cdot (T - \nu)$ if $T > \nu$, where $0 \leq \pi < 1$, $0 < p < 1$ and $c > 0$ are fixed constants. For $\mathcal{A} \in \mathscr{F}$, define the probability

$$\mathsf{P}(\mathcal{A}) = \pi \mathsf{P}_0(\mathcal{A}) + (1 - \pi) \sum_{k=0}^{\infty} p(1 - p)^k \mathsf{P}_k(\mathcal{A}),$$

and let E denote the corresponding expectation.

Solving $\mathcal{B}(\pi, p, c)$ requires minimization of the expected loss

$$\mathcal{L}_{\pi, p, c}(T) = \mathsf{P}(T \leq \nu) + c\mathbf{E}(T - \nu)^+$$

or, equivalently, maximization of the expected "gain" $\frac{1}{p}[1 - \mathcal{L}_{\pi, p, c}(T)]$, and the Bayes rule for this problem is given by the Shiryaev procedure [cf. Shiryaev (1963)]

$$T_{\pi, p, c} = \inf\{n \geq 1 : R_n^{(\pi, p)} \geq A_{\pi, p, c}\},$$

where $A_{\pi, p, c} > \pi / (1 - \pi)p$ is an appropriate threshold and

$$R_n^{(\pi, p)} = \frac{\pi}{(1 - \pi)p} \prod_{i=1}^{n} \left(\frac{\Lambda_i}{1 - p}\right) + \sum_{k=1}^{n} \prod_{i=k}^{n} \left(\frac{\Lambda_i}{1 - p}\right).$$

Let $\pi = rp$. Then, obviously, $R_n^{(\pi, p)} \xrightarrow[p \to 0]{} R_n^r$.

Now, it follows from Pollak (1985) that there are a constant $0 < c^* < \infty$ and a sequence $\{p_i, c_i\}_{i \geq 1}$ with $p_i \to 0$, $c_i \to c^*$ as $i \to \infty$, such that $T_{\mathrm{sr}}^r(A_\gamma)$ is the limit of the Bayes stopping times $T_{\pi = rp_i, p_i, c_i}$ as $i \to \infty$ and

$$(12) \qquad \limsup_{p \to 0, c \to c^*} \frac{1 - \mathcal{L}_{\pi = rp, p, c}(T_{\pi = rp, p, c})}{1 - \mathcal{L}_{\pi = rp, p, c}(T_{\mathrm{sr}}^r(A_\gamma))} = 1.$$

Next, for any stopping time T,

$$\begin{aligned}
\frac{\mathsf{E}(T - \nu)^+}{p} &= \frac{\pi + (1 - \pi)p}{p} \mathbf{E}_0 T + \frac{1 - \pi}{p} \sum_{k=1}^{\infty} p(1 - p)^k \mathsf{E}_k(T - k)^+ \\
&= [r + (1 - rp)]\mathsf{E}_0 T + (1 - rp) \sum_{k=1}^{\infty} (1 - p)^k \mathsf{E}_k(T - k)^+ \\
&\xrightarrow[p \to 0]{} r\mathsf{E}_0 T + \sum_{k=0}^{\infty} \mathsf{E}_k(T - k)^+
\end{aligned}$$

and

$$
\begin{aligned}
\frac{\mathsf{P}(T > \nu)}{p} &= \frac{1}{p}\left(\pi + (1-\pi)p + (1-\pi)\sum_{k=1}^{\infty} p(1-p)^k \mathsf{P}_k(T > k)\right) \\
&= \frac{rp + (1-rp)p}{p} + \frac{1-rp}{p}\sum_{k=1}^{\infty} p(1-p)^k \mathsf{P}_\infty(T > k) \\
&\xrightarrow[p\to 0]{} r + \sum_{k=0}^{\infty}\mathsf{P}_\infty(T > k) = r + \mathsf{E}_\infty T,
\end{aligned}
$$

where we used the fact that $\mathsf{P}_k(T > k) = \mathsf{P}_\infty(T > k)$ since by the definition of the stopping time the event $\{T \le k\}$ belongs to the σ-algebra $\mathscr{F}_k$ and at time instant k the distribution is still f_∞.

Since

$$
\frac{1}{p}[1 - \mathcal{L}_{\pi,p,c}(T)] = \frac{\mathsf{P}(T > \nu)}{p} - c\frac{\mathsf{E}(T - \nu)^+}{p},
$$

it follows that if $\pi = rp$, then for any stopping time T with $\mathsf{E}_\infty T < \infty$

$$
\frac{1}{p}[1 - \mathcal{L}_{\pi=rp,p,c}(T)] \xrightarrow[p\to 0]{} (r + \mathsf{E}_\infty T) - c\left(r\mathsf{E}_0 T + \sum_{k=0}^{\infty}\mathsf{E}_k(T - k)^+\right),
$$

which together with (12) establishes that the SR-r procedure minimizes $\mathcal{I}_r(T)$ over all stopping times that satisfy $\mathsf{E}_\infty T = \gamma$. In order to prove that (11) holds in the class $\{T : \mathsf{E}_\infty T \ge \gamma\}$ it suffices to apply the argument identical to that used in the proof of Corollary 1 in Pollak and Tartakovsky (2009). $\quad\square$

THEOREM 1. *Let $T_{\mathrm{sr}}^r(A)$ be defined as in (8) and let $A = A_\gamma$ be selected so that $\mathsf{E}_\infty[T_{\mathrm{sr}}^r(A_\gamma)] = \gamma$. Then for every $r \ge 0$*

$$
(13) \qquad \inf_{\{T:\mathsf{E}_\infty T \ge \gamma\}} \mathcal{J}_{\mathsf{P}}(T) \ge \frac{r\mathsf{E}_0[T_{\mathrm{sr}}^r(A_\gamma)] + \sum_{\nu=0}^{\infty}\mathbf{E}_\nu[T_{\mathrm{sr}}^r(A_\gamma) - \nu]^+}{r + \mathsf{E}_\infty[T_{\mathrm{sr}}^r(A_\gamma)]}.
$$

PROOF. Note first that for any stopping time T

$$
\begin{aligned}
\sum_{\nu=0}^{\infty}\mathsf{E}_\nu(T - \nu)^+ &= \sum_{\nu=0}^{\infty}\mathsf{P}_\nu(T > \nu)\mathsf{E}_\nu(T - \nu | T > \nu) \\
&= \sum_{\nu=0}^{\infty}\mathsf{P}_\infty(T > \nu)\mathsf{E}_\nu(T - \nu | T > \nu),
\end{aligned}
$$

where again we used the fact that $\mathsf{P}_\nu(T > \nu) = \mathbf{P}_\infty(T > \nu)$. Since

$$
\mathcal{J}_{\mathsf{P}}(T) = \sup_{k \ge 0}\mathsf{E}_k(T - k | T > k) \ge \mathsf{E}_\nu(T - \nu | T > \nu) \qquad \text{for any } \nu \ge 0
$$

and

$$\mathcal{J}_{\mathrm{P}}(T) = \frac{\mathcal{J}_{\mathrm{P}}(T)[r + \sum_{\nu=0}^{\infty} \mathsf{P}_{\infty}(T > \nu)]}{r + \sum_{\nu=0}^{\infty} \mathsf{P}_{\infty}(T > \nu)}$$

$$= \frac{r\mathcal{J}_{\mathrm{P}}(T) + \sum_{\nu=0}^{\infty} \mathcal{J}_{\mathrm{P}}(T)\mathsf{P}_{\infty}(T > \nu)}{r + \sum_{\nu=0}^{\infty} \mathsf{P}_{\infty}(T > \nu)},$$

where $\sum_{\nu=0}^{\infty} \mathsf{P}_{\infty}(T > \nu) = \mathbf{E}_{\infty} T$, we obtain that for any stopping time T with finite ARL to false alarm

$$\mathcal{J}_{\mathrm{P}}(T) \geq \frac{r\mathsf{E}_0 T + \sum_{\nu=0}^{\infty} \mathsf{E}_{\nu}(T - \nu | T > \nu)\mathsf{P}_{\infty}(T > \nu)}{r + \mathsf{E}_{\infty} T}$$

$$= \frac{r\mathsf{E}_0 T + \sum_{\nu=0}^{\infty} \mathsf{E}_{\nu}(T - \nu)^{+}}{r + \mathsf{E}_{\infty} T}.$$

Therefore,

(14) $$\inf_{\{T:\mathsf{E}_{\infty} T \geq \gamma\}} \mathcal{J}_{\mathrm{P}}(T) \geq \inf_{\{T:\mathsf{E}_{\infty} T \geq \gamma\}} \mathcal{I}_r(T),$$

where $\mathcal{I}_r(T)$ is defined in (10).

By Lemma 1, the infimum on the right-hand side in (14) is attained for the SR-r detection procedure $T_{\mathrm{sr}}^r(A_\gamma)$, which completes the proof. $\square$

Notice that if r can be chosen so that the SR-r procedure becomes an equalizer [i.e., $\mathsf{E}_0 T_{\mathrm{sr}}^r = \mathsf{E}_{\nu}(T_{\mathrm{sr}}^r - \nu | T_{\mathrm{sr}}^r > \nu)$ for $\nu \geq 0$], then it is optimal since the right-hand side in (13) is equal to $\mathsf{E}_0 T_{\mathrm{sr}}^r$ which in turn is equal to $\sup_{\nu \geq 0} \mathsf{E}_{\nu}(T_{\mathrm{sr}}^r - \nu | T_{\mathrm{sr}}^r > \nu) = \mathcal{J}_{\mathrm{P}}(T_{\mathrm{sr}}^r)$. This observation will be used in Section 3 for proving that the SR-r procedure with a specially designed $r = r_A$ is strictly optimal for an exponential model.

Introduce the following notation:

$$\delta_{\nu}(r) = \mathsf{E}_{\nu}(T_{\mathrm{sr}}^r - \nu)^{+}, \qquad \rho_{\nu}(r) = \mathsf{P}_{\infty}(T_{\mathrm{sr}}^r > \nu), \qquad \nu \geq 0,$$

$$\phi(r) = \mathsf{E}_{\infty} T_{\mathrm{sr}}^r, \qquad \psi(r) = \sum_{\nu=0}^{\infty} \mathsf{E}_{\nu}(T_{\mathrm{sr}}^r - \nu)^{+},$$

where, obviously, $\rho_0(T_{\mathrm{sr}}^r) = 1$ and $\delta_0(r) = \mathsf{E}_0 T_{\mathrm{sr}}^r$.

In the rest of the paper we will assume for simplicity that Λ_1 is continuous. For $i = 0, \infty$, let $F_i(x) = \mathsf{P}_i(\Lambda_1 \leq x)$ denote the distribution functions of the likelihood ratio under the change and no-change hypotheses.

Moustakides, Polunchenko and Tartakovsky (2009, 2011) used the Markov property of the SR-r statistic (9) to obtain the following integral equations for

performance metrics:

$$(15) \qquad \phi(r) = 1 + \int_0^A \phi(x) \frac{\partial}{\partial x} F_\infty \left(\frac{x}{1+r} \right) dx,$$

$$(16) \qquad \delta_0(r) = 1 + \int_0^A \delta_0(x) \frac{\partial}{\partial x} F_0 \left(\frac{x}{1+r} \right) dx,$$

$$(17) \qquad \delta_\nu(r) = \int_0^A \delta_{\nu-1}(x) \frac{\partial}{\partial x} F_\infty \left(\frac{x}{1+r} \right) dx, \qquad \nu = 1, 2, \ldots,$$

$$(18) \qquad \rho_\nu(r) = \int_0^A \rho_{\nu-1}(x) \frac{\partial}{\partial x} F_\infty \left(\frac{x}{1+r} \right) dx, \qquad \nu = 1, 2, \ldots,$$

$$(19) \qquad \psi(r) = \delta_0(r) + \int_0^A \psi(r) \frac{\partial}{\partial x} F_\infty \left(\frac{x}{1+r} \right) dx.$$

The conditional average delay to detection of the SR-r procedure is computed as

$$\mathsf{E}_\nu(T_{\text{sr}}^r - \nu | T_{\text{sr}}^r > \nu) = \frac{\delta_\nu(r)}{\rho_\nu(r)}, \qquad \nu \geq 0,$$

and the lower bound as

$$\mathcal{I}_r(T_{\text{sr}}^r) = \frac{r \delta_0(r) + \psi(r)}{r + \phi(r)}.$$

Next, we present integral equations for the operating characteristics of the randomized SRP procedures (5) and (6). Here the most crucial problem is the computation of the quasi-stationary distribution $\mathsf{Q}_B(x)$ of the SR statistic. By Harris (1963), Theorem III.10.1, in the continuous case the quasi-stationary distribution exists. Its density $q_B(x) = d\mathsf{Q}_B(x)/dx$ satisfies the following integral equation:

$$(20) \qquad \lambda_B q_B(x) = \int_0^B q_B(r) \frac{\partial}{\partial x} F_\infty \left(\frac{x}{1+r} \right) dr$$

[see Pollak (1985)], where λ_B is the leading eigenvalue of the linear operator associated with the kernel

$$K_\infty(x, r) = \frac{\partial}{\partial x} F_\infty \left(\frac{x}{1+r} \right), \qquad x, r \in [0, B).$$

Thus, $q_B(x)$ is the corresponding (left) eigenfunction. It also satisfies the constraint

$$(21) \qquad \int_0^B q_B(x)\, dx = 1.$$

Equations (20) and (21) uniquely define λ_B and $q_B(x)$. The equations have unique solutions, since $\lambda_B < 1$, as follows from Moustakides, Polunchenko and Tartakovsky (2011).

Once $q_B(x)$ is available we can compute the ARL to false alarm and the average detection delay of the SRP procedure T_{srp}

$$(22) \qquad \mathsf{E}_\infty[T_{\text{srp}}(B)] = \int_0^B \phi(r) q_B(r) \, dr,$$

$$(23) \qquad \mathsf{E}_0[T_{\text{srp}}(B)] = \int_0^B \delta_0(r) q_B(r) \, dr.$$

We recall that the SRP procedure is an equalizer: $\mathsf{E}_\nu(T_{\text{srp}} - \nu | T_{\text{srp}} > \nu) = \mathsf{E}_0 T_{\text{srp}}$.

The integral equations derived above are Fredholm equations of the second kind. Usually, they do not allow for an analytical solution and should be solved numerically. However, in the next section, we provide an example where analytical solutions can be obtained.

3. An example. Consider the exponential model with the pre-change mean 1 and the post-change mean θ^{-1}, $\theta > 1$, that is, $f_\infty(x) = e^{-x} \mathbb{1}_{\{x \geq 0\}}$ and $f_0(x) = \theta e^{-\theta x} \mathbb{1}_{\{x \geq 0\}}$. We will call this model the $\mathcal{E}(1, \theta)$-model. In the sequel we will assume that $\theta = 2$ and the thresholds in both procedures SR-r and SRP do not exceed 2.

THEOREM 2. *Assume the $\mathcal{E}(1, 2)$-model. Let in the SR-r procedure $T_{\text{sr}}^{r_A}$ the initializing value be chosen as $r_A = \sqrt{1 + A} - 1$ and let the threshold $A = A_\gamma$ be selected from the transcendental equation*

$$(24) \qquad A + (\gamma - 1)\sqrt{1 + A} \log(1 + A) - 2(\gamma - 1)\sqrt{1 + A} = 0.$$

Then, for every $1 < \gamma < \gamma_0$, where $\gamma_0 = (1 - 0.5 \log 3)^{-1} \approx 2.2188$, the ARL to false alarm $\mathsf{E}_\infty[T_{\text{sr}}^{r_A}(A)] = \gamma$ and the SR-r procedure is minimax, that is,

$$(25) \qquad \mathcal{J}_{\text{P}}(T_{\text{sr}}^{r_A}) = \inf_{\{T: \mathsf{E}_\infty T \geq \gamma\}} \mathcal{J}_{\text{P}}(T).$$

Let, in the SRP procedure, the threshold $B = B_\gamma$ be selected as

$$(26) \qquad B = \exp\left\{ \frac{2(\gamma - 1)}{\gamma} \right\} - 1.$$

Then $\mathsf{E}_\infty[T_{\text{srp}}(B)] = \gamma$ and $\mathcal{J}_{\text{P}}(T_{\text{srp}}(B)) > \mathcal{J}_{\text{P}}(T_{\text{sr}}^{r_A}(A))$ for all $1 < \gamma < \gamma_0$. Therefore, the SRP procedure is suboptimal.

PROOF. Consider first the SRP procedure. As it will become apparent later, threshold $B = B_\gamma$ in this procedure does not exceed 2 when $\gamma < \gamma_0$. By (20), for $B < 2$ the quasi-stationary density $q_B(x) = d\mathsf{Q}_B(x)/dx$ satisfies the integral equation

$$\lambda_B q_B(x) = \frac{1}{2} \int_0^B q_B(r) \frac{dr}{1 + r},$$

which due to the constraint (21) yields $\lambda_B = \frac{1}{2}\log(1+B)$ and $q_B(x) = B^{-1} \times \mathbb{1}_{\{x \in [0,B)\}}$. Thus, for $B < 2$ the quasi-stationary distribution $Q_B(x) = x/B$ is uniform, and, moreover, it is attained already for $n = 1$ when the very first observation becomes available.

Clearly, the P_∞-distribution of the SRP stopping time T_{srp} is geometric with the parameter $1 - \lambda_B$, so that the ARL to false alarm is

$$(27) \qquad \mathsf{E}_\infty[T_{\mathrm{srp}}(B)] = \frac{1}{1 - \lambda_B} = \frac{1}{1 - (1/2)\log(1+B)}.$$

It follows that $\mathsf{E}_\infty[T_{\mathrm{srp}}(B)] = \gamma$ when the threshold $B = B_\gamma$ is chosen as in (26) and that $B < 2$ whenever $\gamma < \gamma_0$.

By (23), the average detection delay of the SRP procedure is equal to

$$(28) \qquad \mathsf{E}_0[T_{\mathrm{srp}}(B)] = \frac{1}{B}\int_0^B \delta_0(r)\,dr,$$

so that we need to compute the ARL to detection $\delta_0(r) = \mathsf{E}_0 T_{\mathrm{sr}}^r$ of the SR-r procedure which also has to be computed for the evaluation of the performance of the SR-r procedure itself.

Assume that $A < 2$. By (16), we have

$$\delta_0(r) = 1 + \frac{1}{2(1+r)^2}\int_0^A \delta_0(x)x\,dx,$$

so that

$$\int_0^A \delta_0(r)r\,dr = \int_0^A r\,dr + \frac{1}{2}\left[\int_0^A \frac{x\,dx}{(1+x)^2}\right]\left[\int_0^A \delta_0(r)r\,dr\right]$$

$$= \frac{A^2}{2} + \frac{1}{2}\left[\log(1+A) - \frac{A}{1+A}\right]\left[\int_0^A \delta_0(r)r\,dr\right],$$

which implies that

$$\int_0^A r\delta_0(r)\,dr = A^2\left[\frac{A}{1+A} + 2\left(1 - \frac{1}{2}\log(1+A)\right)\right]^{-1}.$$

Consequently,

$$(29) \qquad \delta_0(r) = 1 + \frac{A^2}{2(1+r)^2}\left[\frac{A}{1+A} + 2\left(1 - \frac{1}{2}\log(1+A)\right)\right]^{-1}.$$

Using (28) and (29), we find

$$(30) \quad \mathsf{E}_0[T_{\mathrm{srp}}(B)] = \bar{\delta}_0(B) = 1 + \frac{B^2}{2(1+B)}\left[\frac{B}{1+B} + 2\left(1 - \frac{1}{2}\log(1+B)\right)\right]^{-1}.$$

Consider now the SR-r procedure. By (15), for the ARL to false alarm $\phi(r) = \mathsf{E}_\infty[T_{\mathrm{sr}}^r(A)]$ we have

$$\phi(r) = 1 + \frac{1}{2(1+r)}\int_0^A \phi(x)\,dx,$$

so that

$$\int_0^A \phi(r)\,dr = \int_0^A dr + \frac{1}{2}\left[\int_0^A \frac{dr}{1+r}\right]\left[\int_0^A \phi(x)\,dx\right],$$

and therefore

$$\int_0^A \phi(r)\,dr = A\left[1 - \frac{1}{2}\log(1+A)\right]^{-1}.$$

Consequently,

$$(31) \qquad \phi(r) = 1 + \frac{A}{2(1+r)}\left[1 - \frac{1}{2}\log(1+A)\right]^{-1}.$$

Recall that for $A < 2$ the statistic R_n^r already kicks in the uniform quasi-stationary distribution for $n = 1$ and any $0 \le r < A$, so that T_{sr}^r is an equalizer for $\nu \ge 1$ and any $r \in [0, A)$, that is, $\delta_\nu(r) = \bar{\delta}_0(A)$ for all $\nu \ge 1$ and $r < A$ with $\bar{\delta}_0(A)$ given by (30). This implies that

$$(32) \qquad \mathcal{J}_{\mathrm{P}}(T_{\mathrm{sr}}^r) = \sup_{\nu \ge 0} \mathsf{E}_\nu(T_{\mathrm{sr}}^r - \nu | T_{\mathrm{sr}}^r > \nu) = \max\{\bar{\delta}_0(A), \delta_0(r)\}.$$

Let $r = r_A = \sqrt{1+A} - 1$, in which case $\bar{\delta}_0(A) = \delta_0(r_A)$, that is, for this value of the head start the SR-r procedure is an equalizer for all $\nu \ge 0$. Therefore, by Theorem 1 the procedure $T_{\mathrm{sr}}^{r_A}$ that starts from the deterministic point $r_A = \sqrt{1+A} - 1$ is optimal, and (25) holds if threshold $A = A_\gamma$ is selected so that $\mathsf{E}_\infty T_{\mathrm{sr}}^{r_A} = \gamma$. Substituting $r = \sqrt{1+A} - 1$ in (31) and equalizing the result to γ, yields transcendental equation (24). It is easily verified that $A_\gamma < 2$ for $\gamma < \gamma_0$. This completes the proof of optimality of the SR-r procedure for all $1 < \gamma < \gamma_0$.

In order to show that for every given $\gamma \in (1, \gamma_0)$ the SRP procedure is inferior it suffices to show that $\mathsf{E}_\infty[T_{\mathrm{sr}}^{r_A}(A)] > \mathsf{E}_\infty[T_{\mathrm{srp}}(A)]$. By (31), the ARL to false alarm of the SR-r procedure is equal to

$$(33) \qquad \mathsf{E}_\infty[T_{\mathrm{sr}}^{r_A}(A)] = \phi(r_A) = 1 + \frac{A}{2\sqrt{A+1}}\left[1 - \frac{1}{2}\log(1+A)\right]^{-1}.$$

Comparing (33) with (27), we obtain that we have only to show that

$$1 + \frac{A}{2\sqrt{A+1}}\left[1 - \frac{1}{2}\log(1+A)\right]^{-1} > \left[1 - \frac{1}{2}\log(1+A)\right]^{-1},$$

that is, that $A/\sqrt{A+1} > \log(A+1)$, which holds for any $A > 0$. Thus, we conclude that the SRP procedure is suboptimal and the proof is complete. $\square$

Let, for example, $\gamma = 2$. Then, by (26) and (30), the threshold in the SRP procedure is equal to $B = e - 1 \approx 1.71828$ and the average detection delay $\mathsf{E}_0[T_{\mathrm{srp}}(B)] = \mathcal{J}_{\mathrm{P}}(T_{\mathrm{srp}}(B)) \approx 1.33275$.
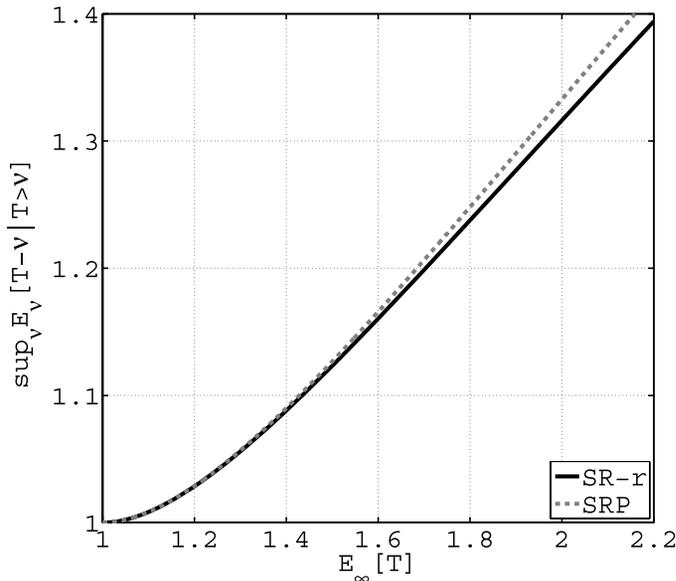
FIG. 1. *Supremum average detection delay versus the ALR to false alarm for $A \in (0, 2)$.*

For $\gamma = 2$, solving the transcendental equation (24) yields $A \approx 1.66485$ and the initialization point $r_A \approx 0.63244$. By (32), the average detection delay of the SR-*r* procedure $\mathsf{E}_0[T_{sr}^{r_A}(A)] = \mathcal{J}_P(T_{sr}^{r_A}(A)) \approx 1.31622$.

Figure 1 depicts the supremum average detection delays versus the ARL to false alarm for the two changepoint detection procedures for the entire range of $A \in (0, 2)$.

REMARK. At an additional effort, the same conclusion can be reached in the more general case where the parameter of the post-change distribution $\theta > 1$ and $A, B < \theta$.

REFERENCES

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. MR2392106

HARRIS, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin. MR0163361

MEI, Y. (2006). Comments on, "A note on optimal detection of a change in distribution," by Benjamin Yakir. *Ann. Statist.* **34** 1570–1576. MR2278368

MOUSTAKIDES, G. V., POLUNCHENKO, A. S. and TARTAKOVSKY, A. G. (2009). Numerical comparison of CUSUM and Shiryaev–Roberts procedures for detecting changes in distributions. *Comm. Statist. Theory Methods* **38** 3225–3239. MR2568216

MOUSTAKIDES, G. V., POLUNCHENKO, A. S. and TARTAKOVSKY, A. G. (2011). A numerical approach to comparative efficiency analysis of quickest change-point detection procedures. *Statist. Sinica*. **21**. To appear.

POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13** 206–227. MR0773162

POLLAK, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15** 749–779. MR0888438

POLLAK, M. and TARTAKOVSKY, A. G. (2009). Optimality properties of the Shiryaev–Roberts procedure. *Statist. Sinica* **19** 1729–1739. MR2589207

ROBERTS, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8** 411–430. MR0196887

SHIRYAEV, A. N. (1961). The problem of the most rapid detection of a disturbance in a stationary process. *Dokl. Math.* **2** 795–799.

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

YAKIR, B. (1997). A note on optimal detection of a change in distribution. *Ann. Statist.* **25** 2117–2126. MR1474086

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
3620 S. VERMONT AVE, KAP-416F
LOS ANGELES, CALIFORNIA 90089-2532
USA
E-MAIL: polunche@usc.edu
           tartakov@usc.edu