Design and Comparison of Shiryaev-Roberts and CUSUM-Type Change-Point Detection Procedures

Alexander G. Tartakovsky[1], Aleksey S. Polunchenko[1] and George V. Moustakides[2]

[1] Center for Applied Mathematical Sciences,
Department of Mathematics,
University of Southern California,
Los Angeles, CA 90089-2532, USA
{polunche, tartakov}@usc.edu

[2] Department of Electrical & Computer Engineering,
University of Patras,
26500 Rio, Greece
moustaki@upatras.gr

Abstract. We address a simple changepoint detection problem where observations are i.i.d. before and after the change with known pre- and post-change distributions. For this setting, the CUSUM test is known to be optimal in the minimax setting for Lorden's essential supremum metric, whereas the Shiryaev-Roberts procedure is optimal for detecting a change that occurs at a distant time horizon. At the same time, a randomized extension of the Shiryaev-Roberts test proposed by Pollak, enjoys a very strong asymptotic minimax property with respect to Pollak's supremum metric. We conjecture that a deterministically initialized version of the Shiryaev-Roberts test can compete with the latter procedure very efficiently. We propose a numerical scheme for the systematic comparison of these detection procedures in both settings, i.e., minimax and for detecting changes that occur in the distant future. Our goal is accomplished by deriving a set of integral equations for the performance metrics of interest, which are solved numerically. We present numerical results for the problem of detecting a change in the mean of an exponential sequence which justify our conjecture and allow for a precise comparison of a number of changepoint detection procedures.

Keywords. CUSUM, Fredholm integral equation of the second kind, numerical analysis, sequential analysis, Shiryaev-Roberts, quickest changepoint detection.

1 Introduction

Quickest changepoint detection is concerned with the problem of detecting changes in distributions that occur at unknown points in time. The goal is to detect the change as soon as possible after its occurrence, while maintaining a prescribed false alarm level. A sequential changepoint detection procedure is a stopping time T with respect to the observed sequence at which a decision on a change occurrence is made.

In this paper we consider the simplest version of the changepoint detection problem where it is assumed that the observations are i.i.d. before and after the change with known pre- and post-change densities. The objective is to provide a comparative study of the following detection procedures: the Cumulative Sum (CUSUM) test introduced by Page (1954); the Shiryaev-Roberts (SR) test proposed by Shiryaev (1961) for the Brownian motion case and Roberts (1966) for discrete time; its randomized extension, which we refer to as the Shiryaev-Roberts-Pollak (SRP) test, suggested by Pollak (1985), whose idea was to sample the starting point from the quasi-stationary distribution; and the Shiryaev-Roberts-r_A (SR-r_A) test introduced in this paper. In the latter procedure the head start of the detection statistic is not zero as in the conventional SR procedure, but rather a deterministic point chosen in such a way so that the "best" possible performance is achieved. It is of major practical interest to compare these tests with respect to several detection speed metrics and to quantify their performance difference.

The rest of the paper is organized as follows. Section 2 gives a preliminary background in changepoint detection and presents the CUSUM, the SR, the SRP and the SR-r_A test. In Section 3 a set of integral equations for various performance characteristics is introduced, and a simple numerical scheme to solve the equations is described. Finally, Section 4 considers an example of detecting a change in the mean of an exponential sequence and reports the obtained results.

2 Problem formulation and changepoint detection procedures

Consider the simplest version of the changepoint detection problem. Let a sequence $\{X_n\}_{n\ge1}$ of i.i.d. random variables be observed sequentially. Initially the sequence is "in-control", i.e., all observations are coming from the same probability density function (pdf) $f(x)$. At an unknown point in time $\nu \ge 0$ something happens and the sequence runs "out of control" by abruptly changing its statistical properties so that from $\nu+1$ on, the pdf becomes $g(x) \not\equiv f(x)$. The objective is to detect a change as quickly as possible and with as few as possible false detections.

Given the sequence $\{X_n\}_{n\ge1}$, a sequential detection procedure is defined as a stopping time T adapted to the filtration $\{\mathcal{F}_n\}_{n\ge0}$, where $\mathcal{F}_n = \sigma(X_1,\dots,X_n)$ is the (smallest) σ-algebra generated by the observations up to time instant n and with $\mathcal{F}_0$ denoting the trivial σ-algebra. Put another way, for $n \ge 0$, the set $\{T \le n\}$ belongs to the σ-algebra $\mathcal{F}_n$. At time instant T the procedure stops and declares that a change has occurred.

The design of such procedures involves optimizing a trade-off between two types of performance metrics, one being a measure of the detection delay and the other being the rate of false alarms. Let us denote with $\mathbb{P}_k$, $\mathbb{E}_k$ the probability and the corresponding expectation induced by a change occurring at time $\nu = k \ge 0$. We will adopt the convention that if the change occurs at $\nu = k$, then the first post-change observation is X_{k+1}. According to this definition $\mathbb{P}_\infty$ ($\mathbb{E}_\infty$) denotes the probability (expectation) when there is no change ($\nu=\infty$), while $\mathbb{P}_0$ ($\mathbb{E}_0$) the corresponding quantities when the change takes place before surveillance begins.

We are interested in two different mathematical setups. First is the minimax approach proposed by Lorden (1971) who suggested to use $\mathcal{J}_{\mathrm{L}}(T) = \sup_{k\ge0} \operatorname{ess\,sup} \mathbb{E}_k[(T-k)^+|X_1,\dots,X_k]$ as the measure of performance (worst average detection delay). Lorden (1971) proposed to minimize $\mathcal{J}_{\mathrm{L}}(T)$ in the class $\varDelta(\gamma) = \{T\colon\ \mathbb{E}_\infty[T] \ge \gamma\}$, where $\gamma > 1$ is a prescribed false alarm level. The value of $\mathbb{E}_\infty[T]$ is usually called the average run length (ARL) to false alarm. An alternative, appropriate for a significantly narrower class of changepoint problems (for a discussion see Moustakides, 2009), was proposed by Pollak (1985) where the detection speed is expressed via the supremum average (conditional) detection delay $\mathcal{J}_{\mathrm{P}}(T) = \sup_{k\ge0}\mathbb{E}_k[T-k|T>k]$.

The second formulation aims at minimizing the relative integral average detection delay $\mathrm{RIADD}(T) = \sum_{k=0}^{\infty}\mathbb{E}_k[(T-k)^+]/\,\mathbb{E}_\infty[T]$ in the class $\varDelta(\gamma)$. Following Shiryaev (1963) who considered this problem for the Brownian motion, Pollak and Tartakovsky (2009) argue that this is instrumental in detecting a change that occurs in the distant future (large ν) and is preceded by a stationary flow of false alarms. Specifically, consider a context in which it is of utmost importance to detect a real change quickly even at the expense of raising many false alarms (using a repeated application of the same stopping rule) before the change occurs. That is, the changepoint ν is substantially larger than the ARL to false alarm γ which, in this case, defines the mean time between false alarms. Let $T_1, T_2, \dots$ denote sequential independent copies of the stopping time T and let $\mathcal{T}_j = T_1 + T_2 + \cdots + T_j$ be the time of the j-th alarm. Define $I_\nu = \min\{j \ge 1\colon\ \mathcal{T}_j > \nu\}$. In other words, $\mathcal{T}_{I_\nu}$ is the time of detection of a true change that occurs at ν after $I_\nu - 1$ false alarms have been raised. Denote $\mathrm{STADD}(T) = \lim_{\nu\to\infty}\mathbb{E}_\nu[\mathcal{T}_{I_\nu} - \nu]$ the limiting value of the average detection delay that we will refer to as the *stationary average detection delay* (STADD). It follows from Theorem 2 in Pollak and Tartakovsky (2009) that $\mathrm{STADD}(T) = \mathrm{RIADD}(T) = \sum_{k=0}^{\infty}\mathbb{E}_k[(T-k)^+]/\,\mathbb{E}_\infty[T]$. $\mathrm{STADD}(T)$ is the second performance measure we will adopt for our comparisons in the context of a multi-cycle change detection.

For $n \ge 1$, define $\Lambda_n = g(X_n)/f(X_n)$, to be the "instantaneous" likelihood ratio between the post-change and pre-change hypotheses. To avoid complications we shall assume that Λ_1 is continuous. Yet, if need be, the case where Λ_1 is non-arithmetic can also be covered with a certain additional effort. The SR procedure stops and raises an alarm at

$$T_A^{\mathrm{SR}} = \inf\{n \ge 1\colon\ R_n \ge A\}$$

assuming $\inf\{\varnothing\} = \infty$, where R_n is the SR detection statistic defined as $R_n = \sum_{k=1}^{n}\prod_{j=k}^{n}\Lambda_j$ and $A = A_\gamma > 0$ is a threshold chosen so that the false alarm constraint $\mathbb{E}_\infty[T_A^{\mathrm{SR}}] = \gamma$ is met. It is easily verified that the SR statistic allows for the following convenient recursive representation

$$R_n = (1 + R_{n-1})\,\Lambda_n, \quad R_0 = 0. \tag{1}$$

As we mentioned above, Pollak and Tartakovsky (2009) showed that the SR procedure $T^{\mathrm{SR}}_{A_\gamma}$ is *exactly* optimal in the sense of minimizing the relative integral average detection delay $\mathrm{RIADD}(T)$ and hence the stationary average detection delay $\mathrm{STADD}(T)$ of a multi-cycle detection procedure for every $\gamma > 1$.

Having the SR test as a prototype we now propose to consider initializing the test from any value $R_0 = r \geq 0$, either random or deterministic. Let us define the modified SR statistic R^r_n as in (1), but with the initial condition $R^r_0 = r$, and the corresponding stopping time as

$$\mathcal{S}^r_A = \inf\{n \geq 1\colon R^r_n \geq A\}, \tag{2}$$

where A is selected so that $\mathbb{E}_\infty[\mathcal{S}^r_A] = \gamma$. We shall refer to this test as SR-r, emphasizing its relation to the initializing value r. Clearly, the threshold A and the starting point r are related through $\mathbb{E}_\infty[\mathcal{S}^r_A] = \gamma$. To satisfy this equality one can either assume that A is a function A_r of r, or the opposite, that is, that r is a function r_A of the threshold A. Additionally we will usually assume that $r < A$.

It can be shown that

$$\begin{aligned}\mathcal{J}_{\mathrm{P}}(\mathcal{S}^r_A) \geq \inf_{T\in\Delta(\gamma)} \mathcal{J}_{\mathrm{P}}(T) \geq \inf_{T\in\Delta(\gamma)} \frac{r\,\mathbb{E}_0[T] + \sum_{k=0}^\infty \mathbb{E}_k[(T-k)^+]}{r + \mathbb{E}_\infty[T]} \\ = \frac{r\,\mathbb{E}_0[\mathcal{S}^r_A] + \sum_{k=0}^\infty \mathbb{E}_k[(\mathcal{S}^r_A-k)^+]}{r + \mathbb{E}_\infty[\mathcal{S}^r_A]} = \mathcal{J}_{\mathrm{B}}(\mathcal{S}^r_A).\end{aligned}$$

Therefore, it is reasonable to choose the starting point r of the SR-r statistic such that it would minimize the difference between the upper bound $\mathcal{J}_{\mathrm{P}}(\mathcal{S}^r_A)$ and the lower bound $\mathcal{J}_{\mathrm{B}}(\mathcal{S}^r_A)$, i.e., $r_A = \arg\min_r\{\mathcal{J}_{\mathrm{P}}(\mathcal{S}^r_A) - \mathcal{J}_{\mathrm{B}}(\mathcal{S}^r_A)\}$.

The SRP procedure is defined similarly to (1) and (2). Only instead of $R_0 = r$ being deterministic it is now a random variable distributed according to the quasi-stationary distribution of the SR statistic R_n, i.e.,

$$\mathbb{Q}_A(x) = \mathbb{P}[R_0 \leq x] = \lim_{n\to\infty} \mathbb{P}_\infty[R^0_n \leq x | \mathcal{S}^0_A > n], \quad x \in [0, A]. \tag{3}$$

The SRP procedure stops at

$$\mathcal{S}^{\mathbb{Q}}_A = \inf\{n \geq 1\colon R^{\mathbb{Q}}_n \geq A\},$$

where $R^{\mathbb{Q}}_n$ satisfies the recursion (1) with $R^{\mathbb{Q}}_0 \sim \mathbb{Q}_A$, i.e., the initializing variable is random and distributed according to the quasi-stationary distribution $\mathbb{Q}_A(x)$. The threshold A is selected so that the false alarm constraint is satisfied with equality, i.e., $\mathbb{E}_\infty[\mathcal{S}^{\mathbb{Q}}_A] = \gamma$.

The CUSUM test is motivated by the maximum likelihood argument and is based on the comparison of the maximum likelihood ratio $V_n = \max_{1\leq k\leq n}\prod_{j=k}^n \Lambda_k$ with a positive detection threshold A, i.e., the CUSUM stopping time is

$$T^{\mathrm{CS}}_A = \inf\{n \geq 1\colon V_n \geq A\}. \tag{4}$$

It is easily verified that the statistic V_n can be computed recursively as

$$V_n = \max\{1, V_{n-1}\}\,\Lambda_n, \quad V_0 = 1. \tag{5}$$

Note that the conventional Page's CUSUM statistic is given by $W_n = \max\{0, W_{n-1} + \log\Lambda_n\}$ where $W_0 = 0$. Clearly, the trajectories of this statistic coincide with the trajectories of $\log V_n$ *on the positive half plane* and, therefore, the CUSUM stopping time defined in (4) is equivalent to the familiar Page's stopping time $T^{\mathrm{PG}}_A = \inf\{n \geq 1\colon W_n \geq \log A\}$ as long as $A > 1$. Note also that, although not crucial for most practical purposes, the CUSUM procedure given by (4) and (5) is more general than the classical Page rule since it allows for thresholds $A \leq 1$ (the classical test with such thresholds stops in one step). The threshold $A = A_\gamma$ is chosen in such a way so that the ARL to false alarm meets the constraint $\mathbb{E}_\infty[T^{\mathrm{CS}}_{A_\gamma}] = \gamma$ exactly. While we use the same notation A for the thresholds in both the CUSUM and the SR-r procedure, we emphasize that the thresholds are in fact fairly different for achieving the same false alarm rate.

3 The methodology for performance evaluation

Note that all aforementioned tests are particular cases of the following stopping time

$$T_A^s = \inf\{n \ge 1\colon S_n^s \ge A\} \tag{6}$$

with the corresponding Markov detection statistic following the recursion

$$S_n^s = \xi(S_{n-1}^s)\,\Lambda_n, \quad n = 1, 2, \dots, \tag{7}$$

where $S_0^s = s \ge 0$ is a preset (fixed) starting point, A is a positive (detection) threshold and $\xi(s)$ is a positive-valued function. Indeed, for CUSUM $\xi(s) = \max\{1, s\}$ and for the SR-r tests $\xi(s) = 1 + s$.

We now derive a set of equations for the performance metrics of the generic detection procedure given by (6) and (7), which then can be easily adapted to the CUSUM and SR-r procedures by appropriately choosing $\xi(s)$.

For fixed A and s, define $\phi_i(s) = \mathbb{E}_i[T_A]$, where $i = \{\infty, 0\}$, so that $\phi_\infty(s) = \mathbb{E}_\infty[T_A]$ is the ARL to false alarm and $\phi_0(s) = \mathbb{E}_0[T_A]$ is the ARL to detection. For $k \ge 0$, define $\delta_k(s) = \mathbb{E}_k[(T_A - k)^+]$, $\rho_k(s) = \mathbb{P}_\infty(T > k)$ and let $F_i(x) = \mathbb{P}_i[\Lambda_1 \le x]$ denote the cumulative distribution function of the likelihood ratio Λ_1 for $i = \{\infty, 0\}$. Using the Markov property of the statistic S_n it can be shown that

$$\phi_i(s) = 1 + \int_0^A \phi_i(x) \left[\frac{\partial}{\partial x} F_i\left(\frac{x}{\xi(s)}\right)\right] dx, \tag{8}$$

and for $k \ge 1$

$$\delta_k(s) = \int_0^A \delta_{k-1}(x) \left[\frac{\partial}{\partial x} F_\infty\left(\frac{x}{\xi(s)}\right)\right] dx,\ \rho_k(s) = \int_0^A \rho_{k-1}(x) \left[\frac{\partial}{\partial x} F_\infty\left(\frac{x}{\xi(s)}\right)\right] dx, \tag{9}$$

with the initial conditions $\delta_0(s) = \phi_0(s)$ and $\rho_0(s) = 1$. The integral equation (8) gives both the ARL to false alarm $\mathbb{E}_\infty[T_A^s] = \phi_\infty(s)$ and the ARL to detection $\mathbb{E}_0[T_A^s] = \phi_0(s)$. Also, from (9) one can recursively compute $\delta_k(s)$ and $\rho_k(s)$, quantities that are necessary for the evaluation of the conditional average detection delay $\mathbb{E}_k[T_A^s - k | T_A^s > k] = \delta_k(s)/\rho_k(s)$.

Let $\psi(s) = \sum_{k=0}^\infty \mathbb{E}_k[(T_A^s - k)^+] = \sum_{k=0}^\infty \delta_k(s)$. It can be shown that $\psi(s)$ satisfies the following equation

$$\psi(s) = \delta_0(s) + \int_0^A \psi(x) \left[\frac{\partial}{\partial x} F_\infty\left(\frac{x}{\xi(s)}\right)\right] dx. \tag{10}$$

Consequently, knowing $\psi(s)$ and $\phi_\infty(s)$ from (10) and (8) we can compute the stationary average delay to detection $\mathrm{STADD}(T_A^s) = \psi(s)/\phi_\infty(s)$.

Let $q_A(x) = d\,\mathbb{Q}_A(x)/dx$ denote the density of the quasi-stationary distribution (3) (recall that we assume that Λ_1 and therefore R_n are continuous). For the quasi-stationary density we have

$$\lambda_A q_A(x) = \int_0^A q_A(s) \left[\frac{\partial}{\partial x} F_\infty\left(\frac{x}{\xi(s)}\right)\right] ds, \tag{11}$$

where $0 < \lambda_A < 1$ is the leading eigenvalue of the linear operator and $q_A(x)$ the corresponding eigenfunction. We also conclude that for the SRP test, the ARL to false alarm is $1/(1-\lambda_A)$ and the conditional average detection delay is $\mathbb{E}_0[T_A^q] = \mathbb{E}_k[T_A^q - k | T_A^q > k] = \int_0^A q_A(x)\delta_0(x)\,dx$, i.e., it does not depend on the point of change k.

Observe that (8) and (10) are Fredholm equations of the second kind. Since generally (except in trivial cases) no analytical solutions to such equations are possible, numerical techniques may be in order. A simple and efficient numerical scheme is based on using a quadrature rule with $N \gg 1$ breakpoints to approximate the integrals in the right-hand side of (8) and (10), thereby turning each of these equations into a system of linear equations, which can then be solved either directly or iteratively. A similar approach can be employed to recover the quasi-stationary distribution from Eq. (11). Clearly, accuracy rides on the number of sample points N: the larger it is, the finer the partition and the more accurate the approximation.

4 A numerical example: Exponential scenario

Consider the case where observations are independent, originally having an Exponential(1) distribution, changing at an unknown time ν to Exponential$(1+\theta)$, i.e., $f(x) = e^{-x}\,\mathbb{1}_{\{x\geq 0\}}$ and $g(x) = \frac{1}{1+\theta}\exp\{-\frac{x}{1+\theta}\}\,\mathbb{1}_{\{x\geq 0\}}$, where $\theta > 0$.

We performed extensive numerical computations for various values of the parameter. Below we present sample results for $\theta = 0.1$ corresponding to a relatively small, not easily detectable change. For all of the procedures we consider the ARL to false alarm is let to go up to $\mathbb{E}_\infty[T] = 10^4$, i.e., low false alarm rate. The integration interval $[0, A]$ is sampled from $N = 10^5$ equidistant points. We are confident that such sampling is sufficiently fine and provides a very high numerical precision, since the results of Monte Carlo experiments for the conventional SR procedure (with 10^6 replications) matched our numerical results within 0.5%.

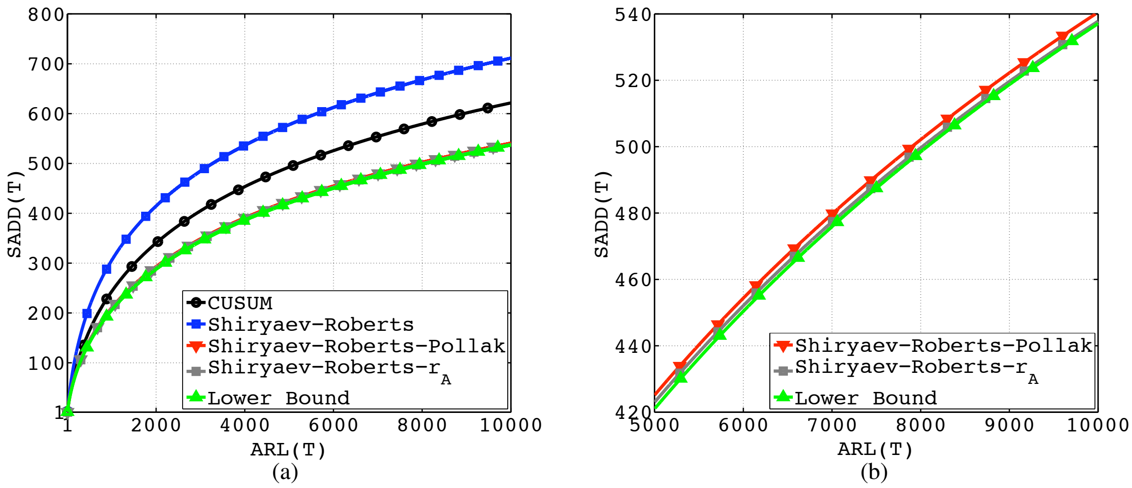

Fig. 1. $\mathrm{SADD}(T)$ vs. $\mathrm{ARL}(T)$ for all of the procedures of interest and the lower bound for $\theta = 0.1$.

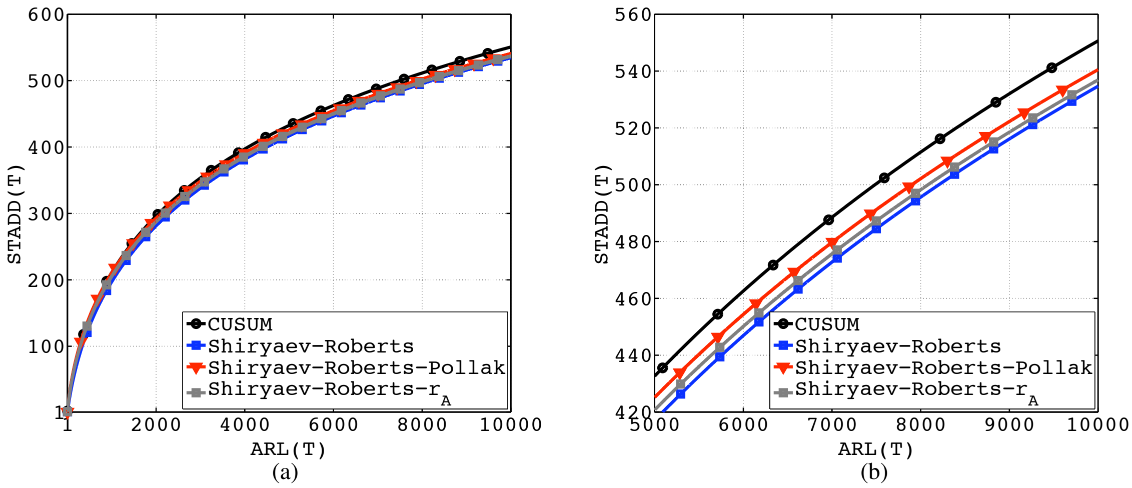

Fig. 2. $\mathrm{STADD}(T)$ vs. $\mathrm{ARL}(T)$ for the procedures of interest for $\theta = 0.1$.

Shown in Fig. 1(a) is the supremum average detection delay $\mathrm{SADD}(T) = \mathcal{J}_{\mathrm{P}}(T)$ as a function of the ARL to false alarm $\mathbb{E}_\infty[T]$ for all of the detection procedures of interest, plus the lower bound $\mathcal{J}_{\mathrm{B}}(T)$. CUSUM outperforms the classical Shiryaev-Roberts test but SRP and SR-r_A are more efficient. Fig. 1(b) is a magnified version of the $\mathrm{SADD}(T)$-vs-$\mathrm{ARL}(T)$ curve for the SR test, the SR-r_A test and the lower bound for relatively high values of the false alarm rate, $\mathrm{ARL}(T) \in [5000, 10000]$. It can be seen that the best minimax performance is offered by the SR-r_A test: performance-wise this test is very close to the lower bound $\mathcal{J}_{\mathrm{B}}(T)$. This suggests that the unknown optimal test can offer only a practically insignificant improvement over SR-r_A with respect to Pollak's $\mathcal{J}_{\mathrm{P}}(T)$ measure. Although the difference in performance between the SRP and SR-r_A procedures is very small, we may conclude that the SRP procedure is not exactly optimal but only order-3 asymptotically optimal, as has been proven by Pollak

(1985). This conclusion is important since the question of whether the SRP procedure is exactly optimal or not was an open question for two decades.

Fig. 2(a) shows the behavior of the stationary average detection delay $\mathrm{STADD}(T)$ against the ARL to false alarm $\mathbb{E}_\infty[T]$ for all of the procedures under discussion. Since for detecting distant changes using a repeated application of the same stopping rule the SR test is exactly optimal, it can be seen that it performs better than the CUSUM test. For the SRP test the $\mathrm{STADD}(T)$-vs-$\mathrm{ARL}(T)$ curve is exactly the same as the $\mathrm{SADD}(T)$-vs-$\mathrm{ARL}(T)$ curve since for this test $\mathrm{STADD}(T) = \mathrm{SADD}(T)$ due to its equalizer property. Fig. 2(b) is a magnified version of Fig. 2(a) for $\mathrm{ARL}(T) \in [5000, 10000]$. The obtained results for selected values of $\mathrm{ARL}(T)$ are also summarized in Table 1.

Table 1. Summary of operating characteristics for the procedures of interest for $\theta = 0.1$

Test	γ	50	100	250	500	750	1000	2500	5000	7500	10000
CUSUM	A	1.59	1.98	2.85	4.19	5.38	6.52	12.93	23.14	33.14	43.06
	ARL	49.71	100.49	249.35	500.36	750.59	1001.05	2499.28	5001.24	7499.61	9999.37
	STADD	28.08	48.48	90.44	143.07	180.61	210.93	326.84	432.57	500.25	550.55
	SADD	31.8	55.31	104.02	165.12	208.52	243.39	375.11	492.8	566.94	621.46
Shiryaev-Roberts	A	46.0	91.0	228.0	455.0	682.0	909.0	2273.0	4546.0	6818.0	9091.0
	ARL	50.6	100.1	250.8	500.5	750.2	999.9	2500.29	5000.53	7499.86	9999.84
	STADD	21.89	39.67	81.36	130.38	167.13	196.93	311.91	416.97	484.42	534.59
	SADD	41.92	72.88	140.3	213.41	265.17	305.63	452.56	577.81	655.23	711.31
Shiryaev-Roberts-Pollak	A	104.0	173.0	353.0	626.0	885.0	1138.0	2597.0	4957.0	7286.0	9601.0
	ARL	49.84	99.91	249.65	500.00	749.69	1000.05	2499.71	4999.82	7499.66	9999.63
	STADD	30.09	50.71	94.73	144.46	180.87	210.24	322.51	425.13	491.22	540.48
	SADD	30.09	50.71	94.73	144.46	180.87	210.24	322.51	425.13	491.22	540.48
Shiryaev-Roberts-r_A	A	105.0	173.0	347.0	612.0	862.0	1106.0	2526.0	4839.0	7132.0	9419.0
	r_A	66.0	90.2	132.0	172.7	197.9	216.7	278.31	322.37	345.31	361.07
	ARL	49.5	100.1	249.7	500.5	750.2	999.9	2500.29	5000.53	7499.86	9999.84
	STADD	28.5	48.22	91.25	139.9	176.23	205.31	317.58	420.72	487.26	536.88
	SADD	29.85	50.45	94.23	143.78	180.0	209.25	320.9	422.88	488.74	537.8

We would like to underline that the criteria evaluated in this article are tailored towards a special class of changepoint mechanisms in which the decision about the change is made without taking into account the observations (see Moustakides 2009). This means that they refer to a less general class than Lorden's performance measure. Consequently the numerical findings tend to display a biased view, favoring the SR test and its variants as compared to the CUSUM test which is optimum in the Lorden sense and therefore capable of confronting, efficiently, a much richer class of changepoint mechanisms.

Acknowledgements

This work was supported in part by the U.S. Army Research Office MURI grant W911NF-06-1-0094 and by the U.S. National Science Foundation grant CCF-0830419 at the University of Southern California.

References

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Annals of Mathematical Statistics*, **42**, 1897–1908.

Moustakides, G.V. (1986). Optimal stopping times for detecting changes in distributions. *Annals of Statistics*, **14**, 1379–1387.

Moustakides, G.V. (2009). Change-point models and performance measures for sequential change detection. *Proc. IWSM 2009*, Troyes, France.

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

Pollak, M. (1985). Optimal detection of a change in distribution. *Annals of Statistics*, **13**, 206–227.

Pollak, M. and Tartakovsky, A. G. (2009). Optimality properties of the Shiryaev-Roberts procedure. *Statistica Sinica*, in press.

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

Shiryaev, A.N. (1963). On optimum methods in quickest detection problems. *Theory Probability and Its Applications*, **8**, 22–46.