Sequential Analysis, 29: 125–145, 2010

ISSN: 0747-4946 print/1532-4176 online
DOI: 10.1080/07474941003740997

State-of-the-Art in Bayesian Changepoint Detection

Alexander G. Tartakovsky[1] and George V. Moustakides[2]
[1]Department of Mathematics, University of Southern California,
Los Angeles, California, USA
[2]Department of Electrical and Computer Engineering,
University of Patras, Patras, Greece

Abstract: We provide a brief overview of the state-of-the-art in quickest (sequential) changepoint detection and present some new results on asymptotic and numerical analysis of main competitors such as the CUSUM, Shiryaev–Roberts, and Shiryaev detection procedures in a Bayesian context.

Keywords: Changepoint detection; CUSUM test; Fredholm integral equation of the second kind; Numerical analysis; Sequential analysis; Shiryaev procedure; Shiryaev–Roberts procedure.

Subject Classifications: 62L15; 60G40; 62F12; 62F15.

1. INTRODUCTION

Changepoint problems deal with detecting changes in observed stochastic processes. In a sequential setting, as long as the behavior of the observations is consistent with the target state, one is content to let the process continue. If the state changes, then one is interested in detecting the change as rapidly as possible.

When we desire to detect the change quickly, any detection policy gives rise to frequent false alarms under no change conditions. On the other hand, attempting to avoid false alarms too strenuously leads to long delays between the time of occurrence of a real change and its detection. The goal is to develop a detection policy that minimizes the average delay to detection subject to a fixed false alarm rate.

Received April 1, 2009, Revised May 20, 2009, Accepted June 1, 2009
Recommended by Nitis Mukhopadhyay
Address correspondence to Alexander G. Tartakovsky, Department of Mathematics, University of Southern California, KAP-108, Los Angeles, CA 90089-2532, USA; Fax: 213-740-2450; E-mail: tartakov@math.usc.edu or George V. Moustakides, Department of Electrical and Computer Engineering, University of Patras, Patras, Rio 26500, Greece; E-mail: moustaki@upatras.gr

In this paper we provide a brief overview of the state of the art in changepoint detection as well as present numerical and asymptotic approximations for operating characteristics (probability of false alarm, average run length to false alarm, and average detection delay) of a generic changepoint detection procedure mostly in a Bayesian context. These approximations allow for comparing various detection strategies, such as CUSUM, the Shiryaev–Roberts procedure, and the optimal Shiryaev test.

2. CHANGEPOINT MODELS

Changepoint models may differ by the structure of the monitored process (i.i.d., nonidentically distributed, dependent, etc.) and/or by the model adopted for the change point θ: an unknown deterministic parameter, a random variable independent of the observations, or a random variable completely or partially dependent on the observations.

Let $\{X_n\}_{n\geq 1}$ denote the series of random observations defined on the complete probability space $(\Omega, \mathscr{F}, \mathsf{P})$, $\mathscr{F} = \bigvee_{n\geq 0} \mathscr{F}_n$, where $\mathscr{F}_n = \sigma(X_1, \ldots, X_n)$ is the sigma-algebra generated by the first n observations ($\mathscr{F}_0 = \{\emptyset, \Omega\}$), Let P_∞ and P_0 be two probability measures defined on this probability space. We will assume that these measures are mutually locally absolutely continuous, i.e., restrictions of measures $\mathsf{P}_0^{(n)}$ and $\mathsf{P}_\infty^{(n)}$ to sigma-algebras $\mathscr{F}_n$, $n \geq 1$ are absolutely continuous with respect to each other. A more general model where both continuous and singular components may be present can be found in Shiryaev (2009). Let $\mathbf{X}_1^n = (X_1, \ldots, X_n)$ denote the vector of the first n observations and let $f_j(\mathbf{X}_1^n)$, $j = \infty, 0$ denote densities of $\mathsf{P}_j^{(n)}$ with respect to a sigma-finite measure μ_n. Suppose now that the observations $\{X_n\}_{n\geq 1}$ initially follow the measure P_∞ (nominal regime) and at some point in time $\theta = 0, 1, \ldots$ something happens and they switch to P_0 (alternative regime). For a fixed θ, the change induces a new probability measure P_θ (correspondingly a density $f_\theta(\mathbf{X}_1^n)$) which is a combination of the pre- and post-change densities. To develop the exact form of the new density, let us use the Bayes rule in order to decompose the original densities as follows

$$f_\infty(\mathbf{X}_1^n) = f_\infty(\mathbf{X}_1^\theta) \times f_\infty(\mathbf{X}_{\theta+1}^n \mid \mathbf{X}_1^\theta); \quad f_0(\mathbf{X}_1^n) = f_0(\mathbf{X}_1^\theta) \times f_0(\mathbf{X}_{\theta+1}^n \mid \mathbf{X}_1^\theta).$$

We can now combine the first component of the pre-change density with the second component of the post-change density to produce the final density

$$f_\theta(\mathbf{X}_1^n) = f_\infty(\mathbf{X}_1^\theta) \times f_0(\mathbf{X}_{\theta+1}^n \mid \mathbf{X}_1^\theta). \tag{2.1}$$

Note that in this general model the observations before the change affect the observations after the change (because the observations before and after the change are in general correlated). Furthermore, we stress that hereafter θ is considered as the last time instant under the nominal regime rather than the first instant under the alternative regime (the latter is the common practice in the literature). If E_θ denotes expectation with respect to the measure P_θ when the change is at θ, then for any $\mathscr{F}$-measurable random variable Y due to (2.1), we have the following equality

$$\mathsf{E}_\theta[Y] = \mathsf{E}_\infty[\mathsf{E}_0[Y \mid \mathscr{F}_\theta]]. \tag{2.2}$$

An alternative decomposition of the joint density that can give rise to a different changepoint model is the following:

$$f_\infty(\mathbf{X}_1^n) = f_\infty(\mathbf{X}_1^\theta) \times f_\infty(\mathbf{X}_{\theta+1}^n \mid \mathbf{X}_1^\theta); \quad f_0(\mathbf{X}_1^n) = f_0(\mathbf{X}_1^\theta \mid \mathbf{X}_{\theta+1}^n) \times f_0(\mathbf{X}_{\theta+1}^n).$$

Combining again the first component of the pre-change density with the second component of the post-change density results in

$$f_\theta(\mathbf{X}_1^n) = f_\infty(\mathbf{X}_1^\theta) \times f_0(\mathbf{X}_{\theta+1}^n). \tag{2.3}$$

Unlike the first model, here $\mathsf{E}_\theta[Y] \neq \mathsf{E}_\infty[\mathsf{E}_0[Y \mid \mathscr{F}_0]]$. This changepoint model can, for example, find application in the case where two independent processes evolve in parallel, with observations coming by sampling, initially, the first process and at some point in time θ switching to sampling the second.

It should be noted that the *same* original densities $f_\infty(\mathbf{X}_1^n)$ and $f_0(\mathbf{X}_1^n)$ gave rise to two *different* $f_\theta(\mathbf{X}_1^n)$. In order to be able to exploit the convenient formula (2.2), in the rest of this section we will adopt the first model (2.1). However, all the results presented in Section 3 are equally valid for the second model (2.3) as well.

Using the Bayes rule, we can write (2.1) in the following equivalent way

$$f_\theta(\mathbf{X}_1^n) = \left(\prod_{i=1}^{\theta} f_\infty(X_i \mid \mathbf{X}_1^{i-1})\right) \times \left(\prod_{i=\theta+1}^{n} f_0(X_i \mid \mathbf{X}_1^{i-1})\right), \tag{2.4}$$

where $f_j(X_n \mid \mathbf{X}_1^{n-1})$ denotes the conditional density of X_n given the past information $\mathbf{X}_1^{n-1}$. There are also other more complicated possibilities where the conditional densities $f_0(\mathbf{X}_i \mid \mathbf{X}_1^{i-1})$, $i = \theta+1, \ldots, n$ depend on the changepoint θ (certain state-space models and hidden Markov models fall under this category; see Tartakovsky, 2009). Model (2.4) can cover this case as well, simply by allowing $f_0^{(\theta)}(\mathbf{X}_i \mid \mathbf{X}_1^{i-1})$ to depend on θ for $i \geq \theta+1$. Note also that the densities $f_j(X_i \mid \mathbf{X}_1^{i-1})$ may depend on i.

If in (2.4) we replace $f_0(X_i \mid \mathbf{X}_1^{i-1})$ with $f_0(X_i \mid \mathbf{X}_{\theta+1}^{i-1})$ (for $i \geq \theta+1$), then obviously the representation (2.4) also covers the model in (2.3), i.e.,

$$f_\theta(\mathbf{X}_1^n) = \left(\prod_{i=1}^{\theta} f_\infty(X_i \mid \mathbf{X}_1^{i-1})\right) \times \left(\prod_{i=\theta+1}^{n} f_0(X_i \mid \mathbf{X}_{\theta+1}^{i-1})\right),$$

In the special case where the observations are i.i.d. before the change with a common pre-change density $f_\infty(X)$ and i.i.d. after the change with a common post-change density $f_0(X)$, the model in (2.4) simplifies to

$$f_\theta(\mathbf{X}_1^n) = \left(\prod_{i=1}^{\theta} f_\infty(X_i)\right) \times \left(\prod_{i=\theta+1}^{n} f_0(X_i)\right). \tag{2.5}$$

Under the i.i.d. assumption the two models in (2.1) and (2.3) coincide.

The previous analysis refers to the case where θ is deterministic. Since in the present paper we are mostly interested in the Bayesian setting, we now analyze the more general situation where θ is random. This may include θ being a random variable that is partially or completely dependent on the observations or completely independent of the observations. To propose a general model that can cover various cases, we will assume that there is a sequence of probabilities $\{\pi_n\}_{n\geq 0}$, where $\pi_0 = \mathsf{P}(\theta \leq 0)$ and $\pi_n = \mathsf{P}(\theta = n \,|\, \mathbf{X}_1^n)$ for $n \geq 1$. We observe that the process $\{\pi_n\}$ is $\{\mathscr{F}_n\}$-adapted, i.e., the probability whether there is a change at $\theta = k$ depends on the observations $\mathbf{X}_1^k$ accumulated up to time k. This allows for a very general modeling of the changepoint mechanisms, including the case where θ is a stopping time adapted to the filtration $\{\mathscr{F}_n\}_{n\geq 1}$ generated by the observations (see Moustakides, 2008).

Recall that P_k denotes the probability measure and E_k the corresponding expectation when the change occurs at time $\theta = k$. Consider a sequence $\{Y_n\}_{n\geq 0}$ of nonnegative, $\mathscr{F}$-measurable random variables. We are interested in computing the average of the randomly stopped sequence $Y_{\theta\vee 0}$. In other words, whenever $\theta \leq 0$ the sequence is stopped at 0, that is, we use Y_0. Write $\mathbb{1}_{\mathscr{A}}$ for the indicator of a set $\mathscr{A}$. Since $Y_{\theta\vee 0}\mathbb{1}_{\{\theta<\infty\}} = Y_0\mathbb{1}_{\{\theta\leq 0\}} + \sum_{k=1}^{\infty} Y_0\mathbb{1}_{\{\theta=k\}}$, using (2.2) and the fact that π_k is $\mathscr{F}_k$-measurable, we can write

$$\mathsf{E}[Y_{\theta\vee 0}\mathbb{1}_{\{\theta<\infty\}}] = \sum_{k=0}^{\infty} \mathsf{E}_k[Y_k \pi_k] = \sum_{k=0}^{\infty} \mathsf{E}_\infty[\mathsf{E}_0[Y_k \,|\, \mathscr{F}_k]\pi_k]. \tag{2.6}$$

This equation constitutes the basis for deriving a multitude of performance measures. If, in particular, we select $Y_n = \mathbb{1}_{\mathscr{A}}$, where $\mathscr{A} \in \mathscr{F}$, then we obtain the probability of the event $\mathscr{A}$ induced by the change

$$\mathsf{P}(\mathscr{A}) = \sum_{k=0}^{\infty} \mathsf{E}_\infty[\mathsf{P}_0(\mathscr{A} \,|\, \mathscr{F}_k)\pi_k].$$

A sequential changepoint detection procedure is a stopping time τ adapted to the filtration $\{\mathscr{F}_n\}_{n\geq 0}$. In other words, $\{\tau \leq n\} \in \mathscr{F}_n$, $\forall n \geq 0$. Since $\mathscr{F}_0$ is the trivial σ-algebra, an $\{\mathscr{F}_n\}$-adapted stopping time satisfies either $\tau > 0$ or $\tau = 0$ w.p. 1. Consequently, to avoid the latter degenerate case, from now on we will assume $\tau > 0$.

The most popular and practically interesting performance measure for τ is the average detection delay

$$\mathsf{ADD}(\tau) = \mathsf{E}[\tau - \theta \,|\, \tau > \theta] = \frac{\mathsf{E}[(\tau-\theta)^+]}{\mathsf{P}(\tau > \theta)}.$$

We observe that for the two sequences of random variables $\{\mathbb{1}_{\{\tau>n\}}\}_{n\geq 0}$ and $\{(\tau - n)^+\}_{n\geq 0}$, we can apply (2.6). This yields

$$\mathsf{E}[(\tau-\theta)^+] = \sum_{k=0}^{\infty} \mathsf{E}_\infty[\mathsf{E}_0[(\tau-k)^+ \,|\, \mathscr{F}_k]\pi_k] \tag{2.7}$$

$$\mathsf{P}(\tau > \theta) = \sum_{k=0}^{\infty} \mathsf{E}_\infty[\mathsf{E}_0[\mathbb{1}_{\{\tau>k\}} \,|\, \mathscr{F}_k]\pi_k] = \sum_{k=0}^{\infty} \mathsf{E}_\infty[\mathbb{1}_{\{\tau>k\}}\pi_k], \tag{2.8}$$

with the last equality being a direct consequence of the fact that $\{\tau > k\} \in \mathscr{F}_k$. Combining the previous two equations we end up with the following expression for the average detection delay

$$\mathsf{ADD}(\tau) = \frac{\sum_{k=0}^{\infty} \mathsf{E}_\infty[\mathsf{E}_0[(\tau-k)^+ \,|\, \mathscr{F}_k]\pi_k]}{\sum_{k=0}^{\infty} \mathsf{E}_\infty[\mathbb{1}_{\{\tau>k\}}\pi_k]}. \tag{2.9}$$

In the Bayesian approach, the false alarm rate of a detection procedure τ is usually measured by the probability of false alarm $\mathsf{PFA}(\tau) = \mathsf{P}(\tau \leq \theta)$. Taking into account that $\tau > 0$ and $\{\tau \leq k\} \in \mathscr{F}_k$, we obtain

$$\mathsf{PFA}(\tau) = \sum_{k=1}^{\infty} \mathsf{E}_\infty[\mathsf{E}_0[\mathbb{1}_{\{\tau \leq k\}} \mid \mathscr{F}_k]\pi_k] = \sum_{k=1}^{\infty} \mathsf{E}_\infty[\mathbb{1}_{\{\tau \leq k\}}\pi_k]. \tag{2.10}$$

Unlike the general formula in (2.6), for the computation of the PFA we do not include the case $\theta \leq 0$ due to the fact that $\tau > 0$.

If the sequence $\{\pi_n\}_{n\geq 0}$ does not depend on the observations, then $\pi_0 = \mathsf{P}(\theta \leq 0)$; $\pi_n = \mathsf{P}(\theta = n)$, $n = 1, 2, \ldots$ is the prior distribution of the changepoint θ. This is the model proposed by Shiryaev (1963). If we assume that the prior distribution $\{\pi_n\}$ is unknown and look for the least favorable distribution that produces the worst possible average detection delay, then we recover Pollak's performance measure (cf. Pollak, 1985). Specifically, for $\{\pi_n\}$ deterministic, we have

$$\mathscr{J}_{\mathrm{P}}(\tau) = \sup_{\{\pi_k\}} \mathsf{ADD}(\tau) = \sup_{k \geq 0} \mathsf{E}_k[\tau - k \mid \tau > k].$$

Let us now consider the most general model where $\{\pi_n\}$ is a random sequence adapted to the filtration $\{\mathscr{F}_n\}$. If we assume again that the sequence is completely unknown and maximize the average detection delay over this more general class of changepoint mechanisms, we recover Lorden's performance measure (cf. Lorden, 1971). More precisely, for $\{\pi_n\}$ being $\{\mathscr{F}_n\}$-adapted, we have

$$\mathscr{J}_{\mathrm{L}}(\tau) = \sup_{\{\pi_k\}} \mathsf{ADD}(\tau) = \sup_{k \geq 0} \operatorname{ess\,sup} \mathsf{E}_k[(\tau - k)^+ \mid \mathscr{F}_k].$$

For further details we refer to Moustakides (2008).

From the previous brief discussion we deduce that the most commonly used criteria pertain to completely different classes of changepoint mechanisms. The Shiryaev and Pollak speed of detection measures can be applied to cases where the change is imposed by a mechanism that disregards the observations. On the other hand, Lorden's measure considers a much richer class of mechanisms that take into account the observations when they decide about imposing the change or not. We must stress that both classes of changepoint mechanisms are equally important with a wide variety of applications that can fall under the first or the second category.

In the rest of this paper, we are limiting ourselves to the case where $\{\pi_k\}$ is deterministic and known beforehand. In other words, we follow the Bayesian approach proposed by Shiryaev (1963), assuming that θ is a random variable *independent* of the observations with a known prior distribution $\{\pi_k\}$.

3. BAYESIAN OPTIMALITY CRITERIA AND DETECTION PROCEDURES

3.1. Bayesian Setting and Shiryaev's Procedure

Assume that the change point θ follows the prior probability distribution $\pi_0 = \mathsf{P}(\theta \leq 0)$; $\pi_k = \mathsf{P}(\theta = k)$ for $k = 1, 2, \ldots$ ($\sum_{k=0}^{\infty} \pi_k = 1$).

Using the general formulas (2.10) and (2.9), we obtain the following relations for the probability of false alarm and the average detection delay:

$$\mathsf{PFA}(\tau) = \sum_{k=1}^{\infty} \pi_k \mathsf{E}_\infty[\mathbb{1}_{\{\tau \le k\}}] = \sum_{k=1}^{\infty} \pi_k \mathsf{P}_\infty(\tau \le k) \tag{3.1}$$

$$\mathsf{ADD}(\tau) = \frac{\sum_{k=0}^{\infty} \pi_k \mathsf{E}_k[(\tau - k)^+]}{\sum_{k=0}^{\infty} \pi_k \mathsf{P}_\infty(\tau > k)} = \frac{\sum_{k=0}^{\infty} \pi_k \mathsf{P}_\infty(\tau > k)\mathsf{E}_k[\tau - k \mid \tau > k]}{\sum_{k=0}^{\infty} \pi_k \mathsf{P}_\infty(\tau > k)}, \tag{3.2}$$

where $\sum_{k=0}^{\infty} \pi_k \mathsf{P}_\infty(\tau > k) = 1 - \mathsf{P}(\tau \le \theta)$ for any stopping time that is finite with probability 1 (i.e., $\mathsf{P}_\infty(\tau < \infty) = 1$).

An optimal Bayesian detection strategy is a procedure for which the ADD is minimized, whereas the PFA is constrained to be below a given level $\alpha \in (0, 1)$. Specifically, define the class of changepoint detection procedures $\mathbf{C}_\alpha = \{\tau : \mathsf{PFA}(\tau) \le \alpha\}$ for which the false alarm probability does not exceed the predefined value α. The optimal changepoint detection procedure is then the stopping time

$$\tau_o = \arg\inf_{\tau \in \mathbf{C}_\alpha} \mathsf{ADD}(\tau). \tag{3.3}$$

Shiryaev (1963) considered the case of θ following a zero-modified geometric distribution

$$\mathsf{P}(\theta < 0) = \pi, \quad \mathsf{P}(\theta = k) = (1 - \pi)p(1 - p)^k, \quad k \ge 0. \tag{3.4}$$

where $\pi \in [0, 1)$, $p \in (0, 1)$. In terms of our model introduced above, this is equivalent to selecting $\pi_0 = \pi + (1 - \pi)p$ and $\pi_k = (1 - \pi)p(1 - p)^k$ for $k \ge 1$. Note that when $\alpha \ge 1 - \pi$, there is a trivial solution to the optimization problem in (3.3), since we can simply stop at 0. Indeed, this strategy produces $\mathsf{ADD} = 0$ and $\mathsf{PFA} = \mathsf{P}(\theta > 0) = 1 - \pi$, which satisfies the constraint.

Consider now the i.i.d. case (2.5), $\alpha < 1 - \pi$ and $\tau > 0$. Shiryaev (1963, 1978) proved that the optimal Bayesian detection procedure $\tau_0 = \tau_s(\varpi)$ exists and has the form

$$\tau_s(\varpi) = \inf\{n \ge 1 : \mathsf{P}(\theta < n \mid \mathscr{F}_n) \ge \varpi\}, \tag{3.5}$$

where the threshold $\varpi = \varpi_\alpha$ must be chosen to satisfy $\mathsf{PFA}(\tau_s(\varpi_\alpha)) = \alpha$. We should note that there seems to be a slight difference in our test statistic as compared to the statistic $\mathsf{P}(\theta \le n \mid \mathscr{F}_n)$ originally proposed by Shiryaev. However, this difference is only notational and it is due to the fact that in our case, θ is the last time instant under the nominal regime, whereas in Shiryaev's approach, it is the first instant under the alternative regime. Actually the two tests are exactly the same, as we realize next by expressing them in terms of an alternative test statistic.

It turns out that it is more convenient to rewrite the stopping time (3.5) in terms of the following statistic:

$$R_{n,p} = \frac{\pi}{(1 - \pi)p} \prod_{i=1}^{n} \left(\frac{\Lambda_i}{1 - p}\right) + \sum_{k=1}^{n} \prod_{i=k}^{n} \left(\frac{\Lambda_i}{1 - p}\right), \tag{3.6}$$

where $\Lambda_i = f_0(X_i)/f_\infty(X_i)$ is the likelihood ratio of the ith sample. Indeed, applying the Bayes rule it is easy to see that

$$\mathsf{P}(\theta < n \,|\, \mathscr{F}_n) = \frac{R_{n,p}}{R_{n,p} + p^{-1}}. \tag{3.7}$$

Hence, the Shiryaev procedure can be rewritten in an equivalent form as

$$\tau_{\mathrm{s}}(A) = \inf\{n \geq 1 : R_{n,p} \geq A\}, \tag{3.8}$$

where $A = \varpi/(1-\varpi)p > \pi/(1-\pi)p$ and

$$R_{n,p} = (1 + R_{n-1,p})\frac{\Lambda_n}{1-p}, \quad n \geq 1, \quad R_{0,p} = \frac{\pi}{(1-\pi)p}. \tag{3.9}$$

The stopping time τ_{s} expressed with the help of $R_{n,p}$ coincides exactly with Shiryaev's stopping time when it is expressed via the same test statistic.

The interesting point is that relations (3.6)–(3.8) continue to hold under the geometric prior model even in the general non-i.i.d. case (2.4) (with $\Lambda_i = f_0(X_i \,|\, \mathbf{X}_1^{i-1})/f_\infty(X_i \,|\, \mathbf{X}_1^{i-1})$). But for the validity of (3.9) we need Λ_i to be independent of the changepoint θ. Although under this more general setting no exact optimality properties are available (similar to the i.i.d. case), there nevertheless exist a number of interesting key asymptotic results that we offer in the next subsection.

3.2. Asymptotic Approximations for Operating Characteristics

In the sequel we briefly develop two alternative means for evaluating performance metrics in the Bayesian changepoint detection problem. The first methodology will be analytic but asymptotic (i.e., valid for small to very small values of the false alarm constraint level α), whereas the second will be numerical and applicable basically when α assumes moderate to small values. Consequently, in a sense, the two methodologies are complementary. We start with the asymptotic approach, while we postpone the more detailed presentation of the numerics for Section 4.

Let $I = \mathsf{E}_0[\log \Lambda_1]$ be the Kullback–Leibler information number and define

$$\begin{aligned} S_n &= \sum_{i=1}^{n} \log \Lambda_i, \quad S_n^p = S_n - n\log(1-p) \\ \nu_a^p &= \inf\{n : S_n^p \geq a\}, \quad \gamma_p = \lim_{a\to\infty} \mathsf{E}_0[\exp\{-(S_{\nu_a^p}^p - a)\}]. \end{aligned} \tag{3.10}$$

Note that the constant γ_p defined in (3.10) is the subject of renewal theory (see, e.g., Woodroofe, 1982). It can be computed either exactly or numerically for particular models. The following theorem describes the state-of-the-art in Bayesian changepoint detection for i.i.d. data models.

Theorem 3.1. *Assume the i.i.d. model* (2.5) *and the geometric prior distribution* (3.4).

(i) *Let* $0 < I < \infty$. *Then for all* $m \geq 1$

$$\mathsf{E}[(\tau_{\mathrm{s}}(A) - \theta)^m \,|\, \tau_{\mathrm{s}}(A) > \theta] \sim \left(\frac{\log A}{I + |\log(1-p)|}\right)^m \quad \text{as } A \to \infty. \tag{3.11}$$

(ii) *If, in addition,* $\log \Lambda_1$ *is non-arithmetic, then*

$$\mathsf{PFA}(\tau_s(A)) = \frac{\gamma_p}{Ap}(1 + o(1)) \quad \textit{as } A \to \infty, \tag{3.12}$$

so that $A = A_\alpha = \gamma_p/(p\alpha)$ *implies* $\mathsf{PFA}(\tau_s(A_\alpha)) = \alpha(1 + o(1))$ *as* $\alpha \to 0$. *With this threshold for all* $m \geq 1$ *have*

$$\inf_{\tau \in \mathbf{C}_\alpha} \mathsf{E}[(\tau - \theta)^m \mid \tau > \theta] \sim \mathsf{E}[(\tau_s(A_\alpha) - \theta)^m \mid \tau_s(A_\alpha) > \theta]$$
$$\sim \left(\frac{|\log \alpha|}{I + |\log(1-p)|}\right)^m \quad \textit{as } \alpha \to 0, \tag{3.13}$$

i.e., Shiryaev's procedure minimizes asymptotically all positive moments of the detection delay.

Proof. (i) follows from Theorem 4 and (ii) from Theorem 5(i) of Tartakovsky and Veeravalli (2005). □

For non-i.i.d. models and/or non-geometric prior distributions an optimal solution is not available. However, we can offer asymptotically optimum results. Let us define the exponential rate of convergence d of the prior distribution,

$$d = -\lim_{k \to \infty} \frac{\log P(\theta > k)}{k}, \quad d \geq 0, \tag{3.14}$$

assuming that the corresponding limit exists. If $d > 0$, then the prior distribution has (asymptotically) exponential right tail. If on the other hand $d = 0$, then this amounts to a heavy tailed distribution. For the geometric distribution (3.4) it is easily seen that $d = -\log(1-p)$.

The log-likelihood ratio for testing the hypothesis that the change occurred at the point $\theta = k$ against $\theta = \infty$ (no change) is

$$Z_n^k = \sum_{i=k+1}^{n} \log \frac{f_0(X_i \mid \mathbf{X}^{i-1})}{f_\infty(X_i \mid \mathbf{X}^{i-1})}, \quad k < n. \tag{3.15}$$

Assuming, for every $k > 0$, the validity of a strong law of large numbers, i.e., convergence of $n^{-1} Z_{k+n}^k$ to a constant $q > 0$ as $n \to \infty$, with a suitable rate, Tartakovsky and Veeravalli (2005) proved that the Shiryaev procedure (3.5) with threshold $\varpi = 1 - \alpha$ is (as $\alpha \to 0$) first-order asymptotically optimal. More precisely, we have the following theorem (cf. Theorem 3 in Tartakovsky and Veeravalli, 2005).

Theorem 3.2. *Assume condition* (3.14) *is satisfied. Furthermore, let the following condition*

$$\sum_{k=0}^{\infty} \left[\pi_k \sum_{n=1}^{\infty} n^{r-1} \mathsf{P}_k\big(|Z_{k+n}^k - q| > n\varepsilon\big)\right] < \infty \tag{3.16}$$

hold for all $\varepsilon > 0$ *and some* $q > 0$ *and* $r \geq 1$. *If* $\varpi = 1 - \alpha$, *then the detection procedure* $\tau_s(\varpi)$ *defined in* (3.5) *belongs to the class* $\mathbf{C}_\alpha$ *and for all* $m \leq r$,

$$\inf_{\tau \in \mathbf{C}_\alpha} \mathsf{E}[(\tau - \theta)^m \mid \tau > \theta] \sim \mathsf{E}[(\tau_s - \theta)^m \mid \tau_s > \theta] \sim \left(\frac{|\log \alpha|}{q + d}\right)^m \quad \textit{as } \alpha \to 0. \tag{3.17}$$

i.e., Shiryaev's procedure minimizes asymptotically moments of the detection delay of order $m \leq r$.

From Theorem 3.2 we conclude that the Shiryaev procedure is asymptotically optimal (with respect to positive moments of the detection delay) for a wide class of prior distributions and non-i.i.d. models under very general conditions. Baron and Tartakovsky (2006) established similar results for general continuous time models. Note that q can be treated as the effective local Kullback–Leibler information number coinciding, in the i.i.d. case, with $I = \mathsf{E}_0[\log \Lambda_1]$, the classical Kullback–Leibler information number.

Let us now consider the Shiryaev–Roberts and CUSUM procedures in the Bayesian context. We recall that these two popular tests are minimax optimal regardless of the knowledge of a prior distribution for the changepoint θ. It is clear that, under the Bayesian setup, both detection procedures exhibit performance which is inferior to the (asymptotically) optimal Shiryaev's test. What we would like to examine next is whether this loss in performance is in fact essential as $\alpha \to 0$, that is, when the false alarm probability is small.

The Shiryaev–Roberts (SR) statistic $R_n = \sum_{k=1}^{n} \prod_{i=k}^{n} \Lambda_i$ is the limiting form of Shiryaev's statistic $R_{n,p}$ when we select $\pi = 0$ and let $p \to 0$. It can be computed recursively

$$R_n = (1 + R_{n-1})\Lambda_n, \quad n \geq 1, \;\; R_0 = 0. \tag{3.18}$$

The corresponding stopping time is defined as

$$\tau_{\mathrm{sr}}(A) = \inf\{n \geq 1 : R_n \geq A\}, \quad A > 0. \tag{3.19}$$

In contrast to the SR statistic, which is based on averaging, the CUSUM procedure is motivated by the maximum likelihood argument, i.e., $U_n = \max_{1 \leq k \leq n} \prod_{j=k}^{n} \Lambda_j$. The CUSUM stopping time is defined as

$$\tau_{\mathrm{cs}}(A) = \inf\{n \geq 1 : U_n \geq A\}, \quad A > 0. \tag{3.20}$$

The statistic U_n can be computed recursively

$$U_n = \max\{1, U_{n-1}\}\Lambda_n, \quad n \geq 1, \;\; U_0 = 1. \tag{3.21}$$

It is worth pointing out that the two recursions in (3.18) and (3.21) developed for the i.i.d. case are also applicable in the general non-i.i.d. case as long as the likelihood ratio $\Lambda_n = f_0(X_n \mid \mathbf{X}^{n-1}) / f_\infty(X_n \mid \mathbf{X}^{n-1})$ does not depend on the changepoint θ.

Regarding comparison of the SR and CUSUM tests against the optimum performance delivered by the Shiryaev procedure, there exists a strong evidence that the two tests are asymptotically inferior under the Bayesian setup, unless $d = 0$, that is, unless the prior distribution is heavy-tailed. It follows from the proof of Theorem 6 in Tartakovsky and Veeravalli (2005) that as long as condition (3.16) is satisfied for some $\tau \geq 1$, the following asymptotic approximations hold:

$$\begin{aligned} \mathsf{E}[(\tau_{\mathrm{sr}}(A) - \theta)^m \mid \tau_{\mathrm{sr}}(A) > \theta] &\sim \mathsf{E}[(\tau_{\mathrm{cs}}(A) - \theta)^m \mid \tau_{\mathrm{cs}}(A) > \theta] \\ &\sim \left(\frac{\log A}{q}\right)^m \quad \text{as } A \to \infty \end{aligned} \tag{3.22}$$

for $m \leq r$.

In order to be able to compare the asymptotic performance of the SR and CUSUM tests with the optimum performance given in (3.13) and (3.17), we need a good estimate of the PFA of each test versus the corresponding threshold A (similar to (3.12) for Shiryaev's procedure). Unfortunately, no such estimate is currently available. In particular, the upper bound $\mathsf{PFA}(\tau_{\mathrm{sr}}) \le O(1)/A$ suggested in Tartakovsky and Veeravalli (2005), which can be easily derived from Doob's submartingale inequality, is not accurate unless d/q is small. We *conjecture* that the asymptotically (as $A \to \infty$) accurate relations are of the form

$$\mathsf{PFA}(\tau_{\mathrm{sr}}(A)) \sim \frac{O(1)}{A^{s(d)}}, \quad \mathsf{PFA}(\tau_{\mathrm{cs}}(A)) \sim \frac{O(1)}{A^{s(d)}} \tag{3.23}$$

(with different constants $O(1)$), where $s(d) > 1$ and $s(d) \to 1$ as $d \to 0$. Although we are unable to prove this conjecture at the moment, it is justified in the Brownian motion case (detection of a change in the drift from 0 to ρ) and an exponential prior distribution with parameter β (i.e., $\mathsf{P}(\theta > t) = e^{-\beta t}$). Then it is possible to obtain exact formulas showing that the asymptotic approximations (3.23) are valid with $s(d) = 0.5[1 + (1 + 4d/q)^{1/2}]$, $q = \rho^2/2$ and $d = \beta$.

From (3.23) setting $\mathsf{PFA} = \alpha$ and solving for A, then substituting in (3.22) produces

$$\mathsf{E}[(\tau_{\mathrm{sr}} - \theta)^m \mid \tau_{\mathrm{sr}} > \theta] \sim \mathsf{E}[(\tau_{\mathrm{cs}} - \theta)^m \mid \tau_{\mathrm{cs}} > \theta] \sim \left(\frac{|\log \alpha|}{qs(d)}\right)^m \quad \text{as } \alpha \to 0.$$

Comparing with (3.17) shows that the asymptotic relative efficiency of the asymptotically optimal Shiryaev procedure compared to the SR and CUSUM tests is given by $[qs(d)/(q+d)]^m$. Because $q + d$ corresponds to the optimum performance, we certainly have $q + d \ge qs(d)$. The question is whether this inequality is in fact strict. Again we conjecture that this is indeed the case, provided that $d > 0$. When $d = 0$ or tends to 0, then $s(d) \to 1$ and the asymptotic relative efficiency is 1, i.e., the SR and CUSUM procedures are asymptotically optimal. This claim is supported by our numerical computations presented in Section 4.2 and from the Brownian motion example where $q + d = \rho^2/2 + \beta$, which is strictly greater than $qs(\beta) = \rho^2(1 + \sqrt{1 + 8\beta/\rho^2})/4$, as one can easily verify. However, if $\beta/\rho^2 \to 0$, then the ratio $qs(\beta)/(q + \beta)$ approaches 1.

Finally, consider the i.i.d. case and a likelihood-based version of the well-celebrated Shewhart chart, which is given by the stopping time

$$\tau_{\mathrm{sw}}(A) = \inf\{n \ge 1 : \Lambda_n \ge A\}, \tag{3.24}$$

where $\Lambda_n = f_0(X_n)/f_\infty(X_n)$. In other words, this procedure raises an alarm the first time the instantaneous likelihood ratio Λ_n exceeds a threshold $A > 0$.

Note that whenever $A \le 1$ the CUSUM procedure (3.20) coincides with the Shewhart test (3.24). Therefore, by Moustakides (1986), Shewhart's procedure is optimal with respect to Lorden's worst-case metric $\mathcal{J}_{\mathrm{L}}(\tau) = \sup_k \operatorname{ess\,sup} \mathsf{E}_k[(\tau - k)^+ \mid \mathcal{F}_k]$ subject to the constraint on the mean time to false alarm $\mathsf{E}_\infty[\tau] \ge T > 1$ for the range of values of T that correspond to threshold values in the interval (0, 1]. This range of false alarm values can be significant when we detect large changes.

Evidently, the distribution of the Shewhart stopping time is geometric under P_∞ and P_k for any $k \geq 0$, which immediately implies that

$$\mathsf{E}_\infty[\tau_{\mathrm{sw}}] = \frac{1}{\mathsf{P}_\infty(\Lambda_1 > A)}; \quad \mathsf{PFA}(\tau_{\mathrm{sw}}) = \frac{(1-\pi)(1-p)\mathsf{P}_\infty(\Lambda_1 > A)}{p + (1-p)\mathsf{P}_\infty(\Lambda_1 > A)}$$
$$\mathsf{ADD}(\tau_{\mathrm{sw}}) = \mathsf{E}_k[\tau_{\mathrm{sw}} - k \mid \tau_{\mathrm{sw}} > k] = \mathsf{E}_0[\tau_{\mathrm{sw}}] = \frac{1}{\mathsf{P}_0(\Lambda_1 > A)},$$

where for $\mathsf{PFA}(\tau_{\mathrm{sw}})$ we assumed the geometric prior distribution (3.4).

Assume now that instead of minimizing the average delay to detection we are interested in maximizing the instantaneous probability of detection $\mathsf{P}(\tau = \theta)$, i.e., we wish to find a stopping time that delivers $\sup_\tau \mathsf{P}(\tau = \theta)$, where the supremum is taken over all $\{\mathcal{F}_n\}$-adapted stopping times. It follows from Bojdecki (1979) that in the i.i.d. case and for the geometric prior distribution the Shewhart stopping time (3.24) is optimal.

3.3. Optimality of the Shiryaev–Roberts Procedure in a Generalized Bayesian Setting

Let us return to the i.i.d. case (2.5) and the geometric prior distribution (3.4) with $\pi = 0$. Recall that the Shiryaev rule is given by (3.8) and (3.9). Note first that, as $p \to 0$, i.e., when the prior distribution becomes improper uniform, then $R_{n,p}$ tends to the SR statistic R_n defined in (3.18). It can be easily shown that, for any stopping time τ,

$$\frac{\mathsf{P}(\tau > \theta)}{p} \underset{p\to 0}{\longrightarrow} \mathsf{E}_\infty[\tau], \quad \frac{\mathsf{E}[(\tau-\theta)^+]}{p} \underset{p\to 0}{\longrightarrow} \sum_{k=0}^{\infty} \mathsf{E}_k[(\tau - k)^+].$$

Thus, one may conjecture that the SR test defined in (3.19) minimizes the integral average detection delay

$$\mathcal{I}(\tau) = \frac{\sum_{k=0}^{\infty} \mathsf{E}_k[(\tau-k)^+]}{\mathsf{E}_\infty[\tau]} \tag{3.25}$$

over all detection procedures with $\mathsf{E}_\infty[\tau] \geq T$, i.e., over all stopping times in the class $\mathbf{C}_T = \{\tau : \mathsf{E}_\infty[\tau] \geq T\}$ for which the average run length (ARL) to false alarm is no less than $T > 1$.

The exact result is given in the next theorem, which also provides an asymptotic approximation for the integral average delay to detection for large threshold values. Before stating the desired theorem we need to introduce some additional notation. Recall that $S_n = \sum_{i=1}^n \log \Lambda_i$; define the one-sided test $\nu_a = \inf\{n \geq 1 : S_n \geq a\}$ and the average (limiting) overshoot $\varkappa = \lim_{a\to\infty} \mathsf{E}_0[S_{\nu_a} - a]$. Define also the random variable $V_n = \sum_{i=1}^n e^{-S_i}$ and the constant

$$C = \int_0^\infty \int_0^\infty \log(1 + y + x)\, d\mathsf{Q}_\infty(x)\, d\mathsf{Q}_0(y) \tag{3.26}$$

where $\mathsf{Q}_\infty(x) = \lim_{n\to\infty} \mathsf{P}_\infty(R_n \leq x)$ is the P_∞-stationary distribution of the SR statistic R_n and $\mathsf{Q}_0(y) = \lim_{n\to\infty} \mathsf{P}_0(V_n \leq y)$ is the P_0-stationary distribution for V_n.

The constant $\varkappa$ is the subject of renewal theory and can be computed either exactly or numerically (see, e.g., Woodroofe, 1982). The constant C can be computed either by Monte Carlo or numerically (see Section 4).

Theorem 3.3. *Let A_T be chosen so that $\mathsf{E}_\infty[\tau_{sr}(A_T)] = T$. Then the Shiryaev–Roberts procedure defined by* (3.18) *and* (3.19) *minimizes $\mathcal{J}(\tau)$ over all stopping times τ that satisfy $\mathsf{E}_\infty[\tau] \geq T$, i.e.,*

$$\inf_{\tau \in \mathbf{C}_T} \mathcal{J}(\tau) = \mathcal{J}(\tau_{sr}(A_T)) \quad \text{for every } T > 1. \tag{3.27}$$

If $\mathsf{E}_0[(\log \Lambda_1)^2] < \infty$ and Λ_1 is non-arithmetic, then

$$\mathcal{J}(\tau_{sr}(A)) = \frac{1}{I}(\log A + \varkappa - C) + o(1) \quad \text{as } A \to \infty. \tag{3.28}$$

Proof. Optimality of the SR procedure, that is, (3.27) follows from Theorem 1 of Pollak and Tartakovsky (2009). Although the asymptotic approximation (3.28) is intuitively appealing, the proof is highly nontrivial and fairly long and will be presented elsewhere. □

Feinberg and Shiryaev (2006) obtained the same result as (3.27) for the Brownian motion model (for detecting a change in the constant drift). They refer to this problem as "A Generalized Bayesian Setting." It is interesting that the generalized Bayesian setting is closely related to another, completely different problem where the changepoint θ is assumed to be an unknown deterministic number and the change occurs at a far time horizon and has to be detected by applying a repeated sequential procedure that starts from scratch after each alarm (cf. Pollak and Tartakovsky, 2009; Shiryaev, 1961, 1963).

Finally, we would like to mention that for $\pi > 0$, the statistic $R_{n,p}$ defined in (3.6) and (3.9) is initialized not from zero but from the point $R_{0,p} = \pi/[(1-\pi)p]$. Selecting $\pi = rp$ and then letting $p \to 0$, we arrive at the generalization of the SR statistic that starts not necessarily from zero but from any point $r \geq 0$:

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

This test (called the SR$-r$ test) has been recently introduced by Moustakides et al. (2010) and shown to have certain interesting minimax properties (see also Polunchenko and Tartakovsky, 2010).

4. NUMERICAL PERFORMANCE EVALUATION

4.1. Integral Equations for Operating Characteristics

In this section, we provide integral equations for the operating characteristics of the changepoint detection procedures considered in the previous sections. Using numerical techniques one can then obtain efficient numerical approximations. The proposed approach borrows ideas from Moustakides et al. (2009, 2010).

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

$$\tau_A^r = \inf\{n \geq 1 : \mathscr{S}_n^r \geq A\}, \quad A > 0, \tag{4.1}$$

where $\mathscr{S}_n^r$ is a Markov statistic satisfying the recursion

$$\mathscr{S}_n^r = \Phi(\mathscr{S}_{n-1}^r)\Lambda_n, \quad n \geq 1, \;\; \mathscr{S}_0^r = r \geq 0, \tag{4.2}$$

and $\Phi(\mathscr{S})$ is a positive-valued function. Indeed, for the Shiryaev procedure we have $\Phi(\mathscr{S}) = (1+\mathscr{S})/(1-p)$ (see (3.9)), for CUSUM $\Phi(\mathscr{S}) = \max\{1, \mathscr{S}\}$ (see (3.21)) and for the SR procedure $\Phi(\mathscr{S}) = 1 + \mathscr{S}$ (see (3.18)).

We now derive a set of equations for the performance metrics of the generic detection procedure given by (4.1) and (4.2), which can then be easily adapted to all aforementioned special cases by substituting the proper function $\Phi(\mathscr{S})$. For simplicity we assume that the likelihood ratio Λ_1 is continuous and let $F_i(x) = \mathsf{P}_i(\Lambda_1 \leq x)$ denote the distribution function of Λ_1 for $i = \{\infty, 0\}$.

We begin with the performance evaluation of the generic stopping time within the Bayesian context and for the geometric prior model defined in (3.4). The two quantities that need to be computed are the false alarm probability $\mathsf{PFA}(\tau_A^r)$ from (3.1) and the average detection delay $\mathsf{ADD}(\tau_A^r)$ from (3.2). For $k \geq 0$, let $\delta_k(r) = \mathsf{E}_k[(\tau_A^r - k)^+]$ and $\rho_k(r) = \mathsf{P}_\infty(\tau_A^r > k)(\rho_0^r = 1)$.

We observe that for the PFA we can write

$$\begin{aligned}\mathsf{PFA}(\tau_A^r) &= (1-\pi)p\sum_{k=1}^{\infty}(1-p)^k \mathsf{P}_\infty(\tau_A^r \leq k)\\ &= (1-\pi)\left\{1 - p\sum_{k=0}^{\infty}(1-p)^k\rho_k(r)\right\}.\end{aligned} \tag{4.3}$$

Similarly for the numerator $\mathsf{N}(\tau_A^r)$ and the denominator $\mathsf{D}(\tau_A^r)$ of $\mathsf{ADD}(\tau_A^r)$, we have

$$\begin{aligned}\mathsf{N}(\tau_A^r) &= [\pi + (1-\pi)p]\mathsf{E}_0[\tau_A^r] + (1-\pi)p\sum_{k=1}^{\infty}(1-p)^k\mathsf{E}_k[(\tau_A^r - k)^+]\\ &= \pi\delta_0(r) + (1-\pi)p\sum_{k=0}^{\infty}(1-p)^k\delta_k(r)\end{aligned} \tag{4.4}$$

$$\mathsf{D}(\tau_A^r) = 1 - \mathsf{PFA}(\tau_A^r) = \pi + (1-\pi)p\sum_{k=0}^{\infty}(1-p)^k\rho_k(r). \tag{4.5}$$

From (4.3)–(4.5) we realize that we need to find suitable equations for the evaluation of the two series

$$\psi_p(r) = \sum_{k=0}^{\infty}(1-p)^k\delta_k(r) \quad \text{and} \quad \chi_p(r) = \sum_{k=0}^{\infty}(1-p)^k\rho_k(r),$$

as well as for the average run length to detection $\delta_0(r)$. Using the Markov property of the statistic $\mathscr{S}_n^r$, it is readily seen that $\delta_0(r)$, $\psi_p(r)$, and $\chi_p(r)$ satisfy the following integral equations

$$\delta_0(r) = 1 + \int_0^A \delta_0(x)\left[\frac{\partial}{\partial x}F_0\left(\frac{x}{\Phi(r)}\right)\right]dx \tag{4.6}$$

$$\psi_p(r) = \delta_0(r) + (1-p)\int_0^A \psi_p(x)\left[\frac{\partial}{\partial x}F_\infty\left(\frac{x}{\Phi(r)}\right)\right]dx \tag{4.7}$$

$$\chi_p(r) = 1 + (1-p)\int_0^A \chi_p(x)\left[\frac{\partial}{\partial x}F_\infty\left(\frac{x}{\Phi(r)}\right)\right]dx. \tag{4.8}$$

Using the solutions of equations (4.6)–(4.8), we can compute the two quantities of interest as follows:

$$\mathsf{PFA}(\tau_A^r) = (1-\pi)\{1 - p\chi_p(r)\} \quad \text{and} \quad \mathsf{ADD}(\tau_A^r) = \frac{\pi\delta_0(r) + (1-\pi)p\psi_p(r)}{\pi + (1-\pi)p\chi_p(r)}.$$

Next, in order to compute the integral average detection delay $\mathcal{J}(\tau_A^r)$ given in (3.25), we have to compute the ARL to false alarm $\chi_0(r) = \mathsf{E}_\infty[\tau_A^r]$ and the series $\psi_0(r) = \sum_{k=0}^\infty \delta_k(r)$. These two functions satisfy the integral equations

$$\chi_0(r) = 1 + \int_0^A \chi_0(x)\left[\frac{\partial}{\partial x}F_\infty\left(\frac{x}{\Phi(r)}\right)\right]dx$$

$$\psi_0(r) = \delta_0(r) + \int_0^A \psi_0(x)\left[\frac{\partial}{\partial x}F_\infty\left(\frac{x}{\Phi(r)}\right)\right]dx$$

that constitute special cases of (4.7) and (4.8) with $p = 0$. The integral average detection delay is then computed as $\mathcal{J}(\tau_a^r) = \psi_0(r)/\chi_0(r)$.

In addition, we may be interested in computing the conditional average detection delays $\mathsf{E}_k[\tau_A^r - k \mid \tau_A^r > k] = \delta_k(r)/\rho_k(r)$ for any $k \geq 1$. The functions $\delta_k(r)$ and $\rho_k(r)$ can be computed recursively (cf. Moustakides et al., 2009, 2010).

In order to implement the asymptotic approximations for the integral average detection delay (3.28), we should be able to compute the constant C defined in (3.26). For this computation we need the two densities $q_\infty(x) = d\mathsf{Q}_\infty(x)/dx$ and $q_0(x) = d\mathsf{Q}_0(x)/dx$. Let R_∞ and V_∞ be the limiting (as $n \to \infty$) random variables of R_n and V_n, respectively, which have densities $q_\infty(x)$ and $q_0(x)$. To find the desired densities, observe that, by recursion (3.18), R_∞ and $(1 + R_\infty)\Lambda_1$ have the same density $q_\infty(x)$ under P_∞. Similarly V_∞ and $(1 + V_\infty)\Lambda_1^{-1}$ have the same density $q_0(x)$ under P_0. To see this, note that, by the i.i.d. property of the data, V_n has the same P_0-distribution as the random variable $\widetilde{V}_n = \sum_{i=1}^n \prod_{j=i}^n \Lambda_i^{-1}$ that follows the recursion $\widetilde{V}_n = (1 + \widetilde{V}_{n-1})\Lambda_n^{-1}$. Consequently, we have the following integral equations:

$$q_\infty(x) = \int_0^\infty q_\infty(y)\left[\frac{\partial}{\partial x}F_\infty\left(\frac{x}{1+y}\right)\right]dy;$$

$$q_0(x) = -\int_0^\infty q_0(y)\left[\frac{\partial}{\partial x}F_0\left(\frac{1+y}{x}\right)\right]dy.$$

Thus, $q_\infty(x)$ and $q_0(x)$ are the eigenfunctions corresponding to the unit eigenvalues of the linear operators defined, respectively, with the following kernels:

$$\mathcal{K}_\infty(x, y) = \frac{\partial}{\partial x}F_\infty\left(\frac{x}{1+y}\right), \quad \mathcal{K}_0(x, y) = -\frac{\partial}{\partial x}F_0\left(\frac{1+y}{x}\right).$$

The constant C is then found by numerical integration.

This completes our presentation of the equations that are necessary for the computation of the metrics we introduced in the previous sections. We observe that the totality of these equations are Fredholm equations of the second kind. Since generally no analytical solutions are possible, we must resort to numerical techniques. A simple numerical scheme using a quadrature rule with $N \gg 1$ breakpoints to approximate the integrals can be used to provide an approximate solution. The resulting systems of linear equations can be solved either directly or iteratively. The accuracy of such numerical schemes is directly related to the number N of breakpoints we use, and one expects accuracy to improve with increasing number of points. Clearly, when the false alarm constraint becomes very stringent, i.e., threshold A takes large values, this will require exceedingly large number of points resulting in unrealistic processing times. This is why these techniques can be used for moderate to small values of the false alarm constraint, as we have mentioned earlier.

4.2. Example

Consider an i.i.d. exponential example with a mean before the change equal to 1 and after the change equal to $1 + \lambda$,

$$f_\infty(x) = e^{-x}\mathbb{1}_{\{x\geq 0\}}, \quad f_0(x) = (1+\lambda)^{-1}e^{-x/(1+\lambda)}\mathbb{1}_{\{x\geq 0\}}, \quad \lambda > 0.$$

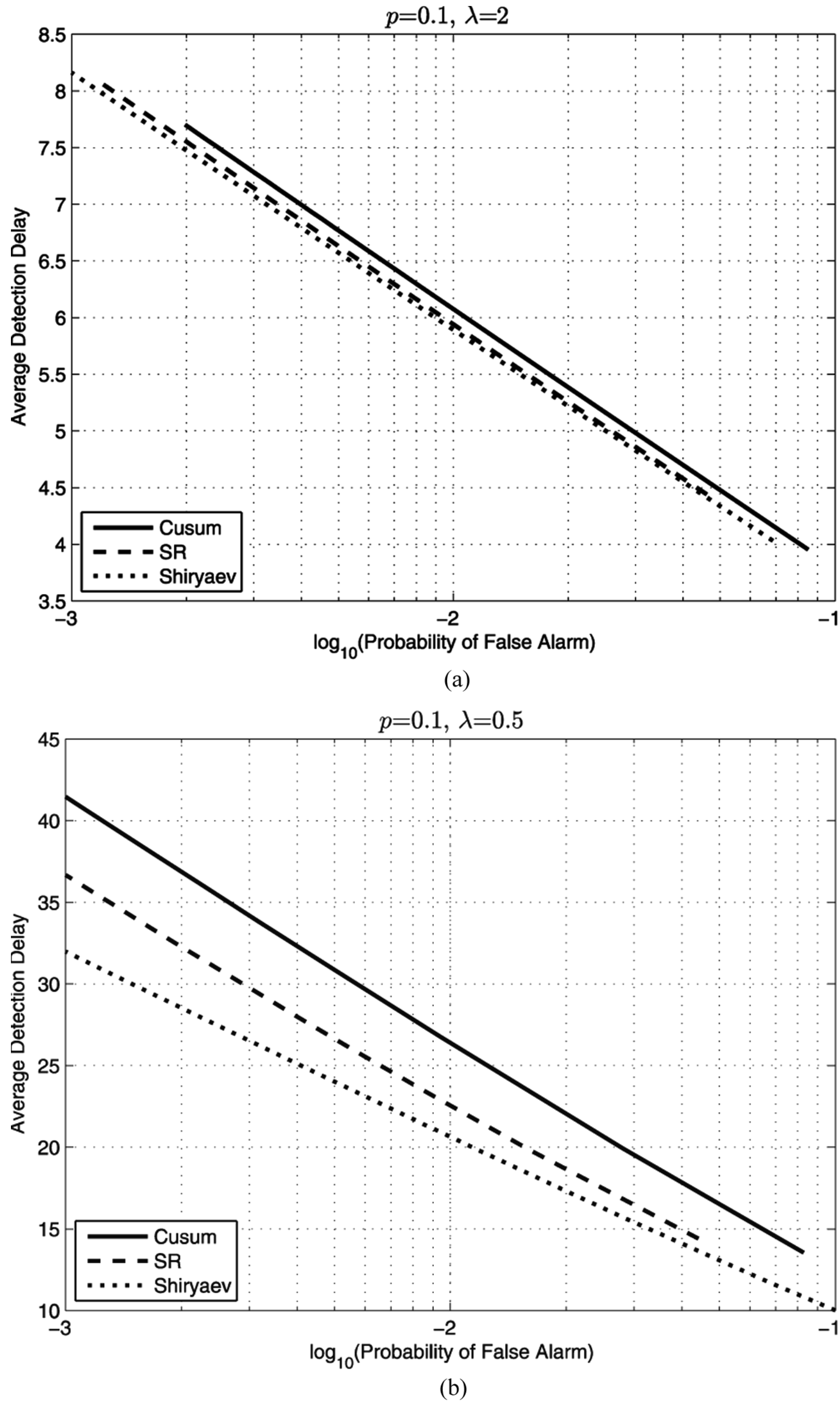
In Figure 1 we depict the operating characteristics of the SR, CUSUM, and Shiryaev tests in terms of ADD versus PFA for $p = 0.1$ and $\lambda = 2$ and 0.5. These operating characteristics were computed by solving numerically the corresponding integral equations presented in the previous subsection. Note that for the exponential case, the Kullback–Leibler information number is equal to $I = \lambda - \log(1+\lambda)$. It is seen from Figure 1(a) that for $\lambda = 2$ (i.e., when $I \gg |\log(1-p)|$), the SR test performs as good as Shiryaev's test (the difference in performance is negligible), as expected. The CUSUM test performs somewhat worse but the difference is not dramatic. On the other hand, for $\lambda = 0.5$ (i.e., when the values of I and $|\log(1-p)|$ are comparable), Shiryaev's test performs much better, also as expected.

In Figure 2 we present the false alarm probability PFA of the Shiryaev test as a function of the threshold A. Solid curves correspond to the numerical computation of the probability using integral equations and dashed ones using the asymptotic formula (3.12), i.e., $\mathsf{PFA}(\tau_s) \approx \gamma_p/(Ap)$. For the exponential case γ_p can be computed analytically as $\gamma_p = 1/(1+\lambda)$; consequently, $\mathsf{PFA}(\tau_s) \approx [Ap(1+\lambda)]^{-1}$. As we can see, the asymptotic formula provides a very efficient approximation as long as $\mathsf{PFA}(\tau_s) \leq 0.1$. Since for most practical applications larger values of the false alarm probability are of no interest, we can conclude that the approximation (3.12) provides an excellent fit.

Figure 3 depicts the ADD as a function of PFA of the SR procedure. We can see that the slope of the ADD depends on p, which, according to our conjecture, comes from the exponent $s(p)$ in approximations (3.23). This is particularly visible for $\lambda = 0.5$.

Finally, in Figure 4 we test the correctness of the asymptotic formula (3.11) for the average detection delay of Shiryaev's procedure

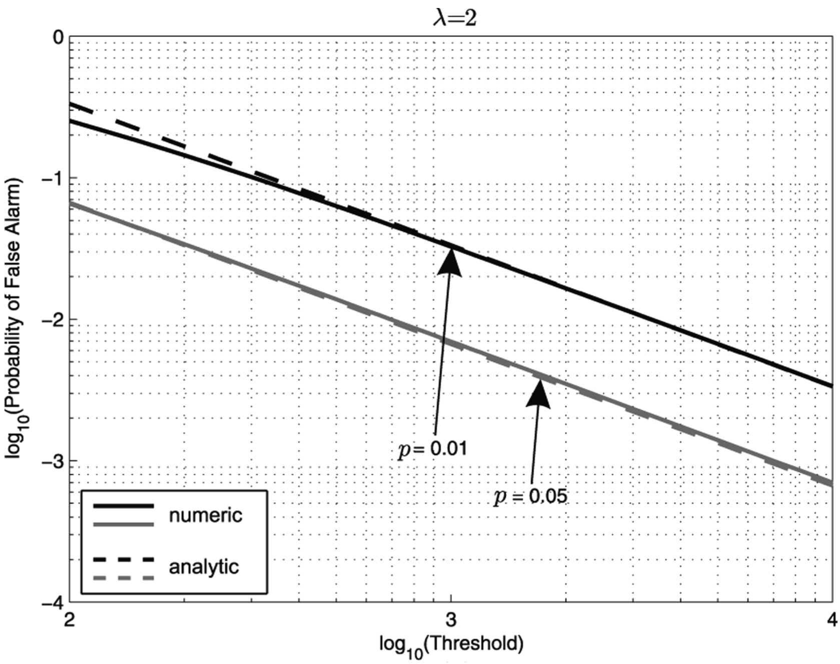
$$\mathsf{ADD}(\tau_s) \approx \frac{\log(A(\mathsf{PFA}))}{\lambda - \log(1+\lambda) - \log(1-p)}$$



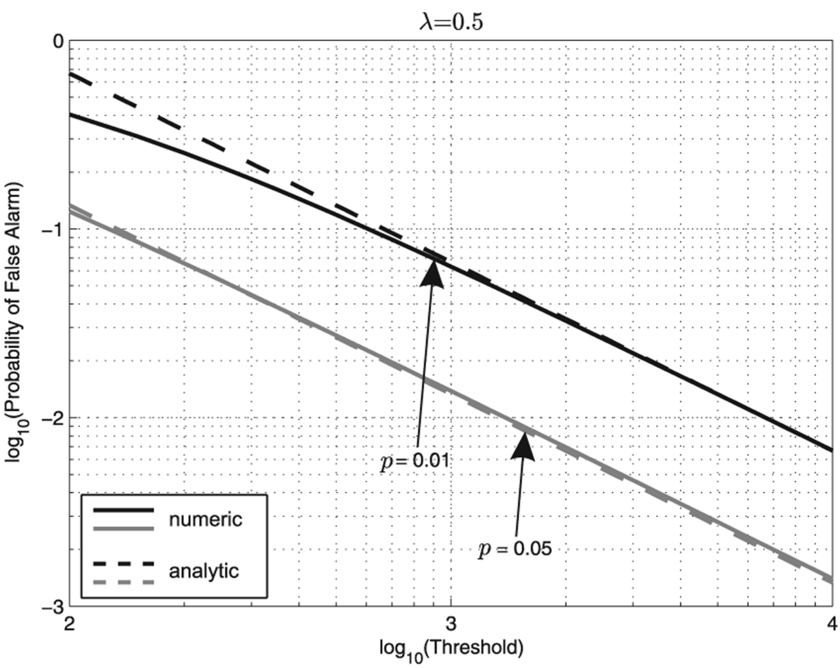
(a)

(b)

Figure 1. Operating characteristics of the three tests for $p = 0.1$ and (a) $\lambda = 2$, (b) $\lambda = 0.5$.

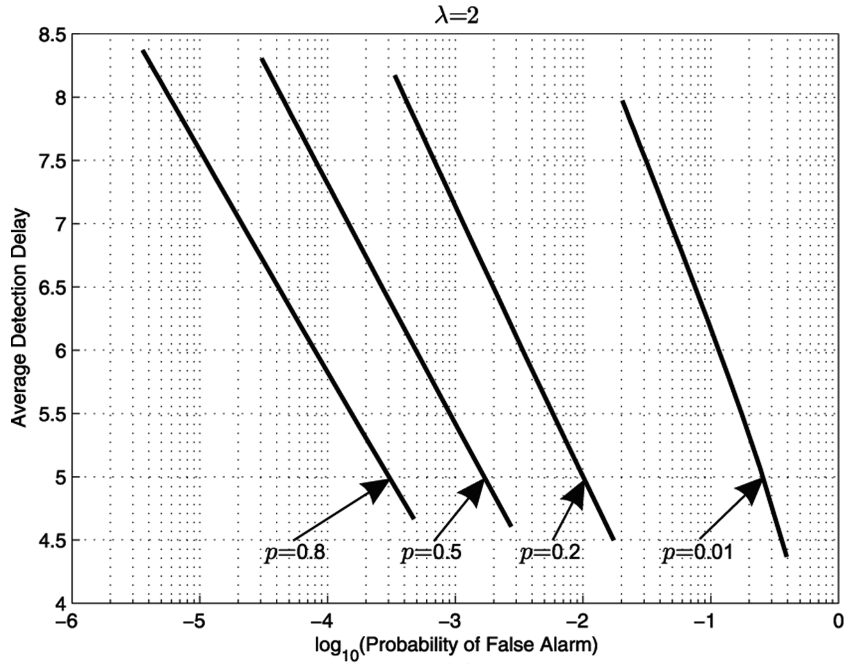


(a)

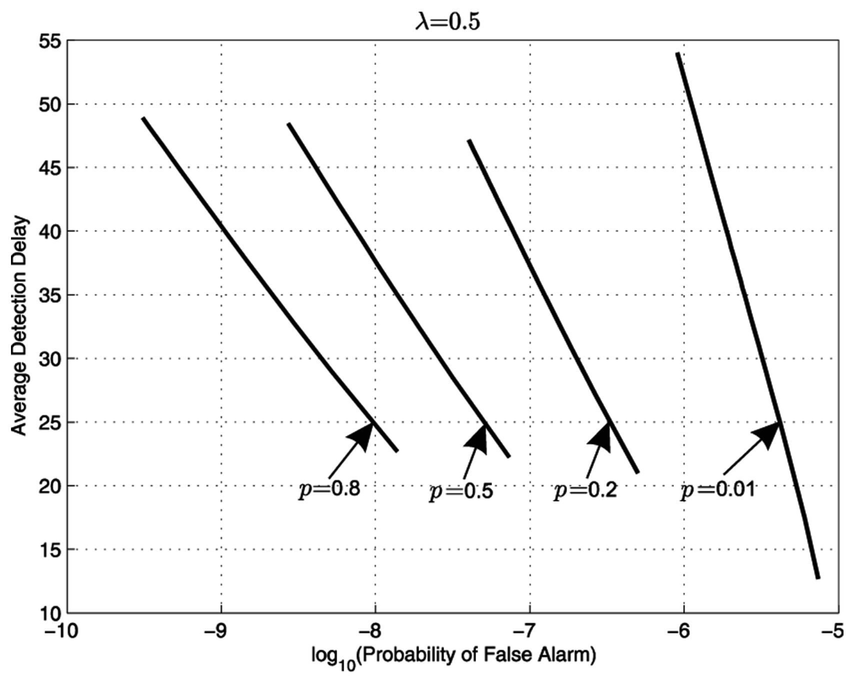


(b)

Figure 2. The probability of false alarm of Shiryaev's test as a function of the threshold A for $p = 0.01, 0.05$ and (a) $\lambda = 2.0$, (b) $\lambda = 0.5$.



(a)



(b)

Figure 3. Operating characteristics of the SR test for various p and (a) $\lambda = 2.0$, (b) $\lambda = 0.5$.

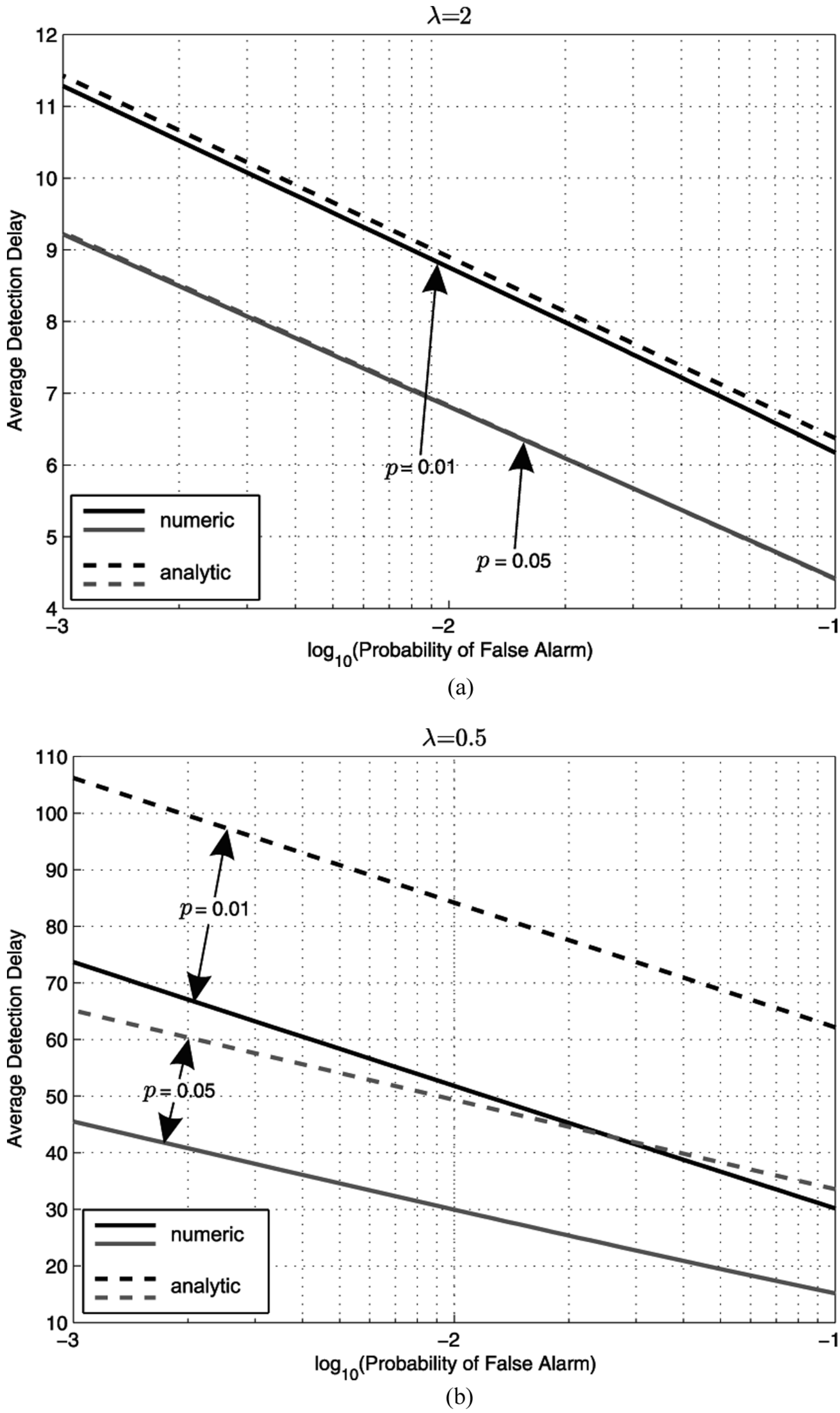


Figure 4. Operating characteristics of the Shiryaev test computed numerically and using the asymptotic formula for various p and (a) $\lambda = 2$, (b) $\lambda = 0.5$.

with threshold $A(\mathsf{PFA})$ related to the false alarm probability PFA by the asymptotic formula (3.12), i.e., $A(\mathsf{PFA}) = 1/[p(1+\lambda)\mathsf{PFA}]$. Recall that according to Figure 2, the latter formula is fairly accurate. As we can see, the slopes exhibit a very good match between the numerical and the analytical values for different values of p, λ. There is, however, a constant shift (especially for $\lambda = 0.5$) that can be explained by the fact that first order approximations neglect constants. The difference increases when the Kullback–Leibler number decreases. These constants seem to be difficult (if at all possible) to compute either analytically or numerically. Therefore, the integral equations and the numerical techniques proposed in subsection 4.1 are valuable tools for achieving accurate performance evaluation. Similar graphs are observed for the SR and CUSUM tests.

ACKNOWLEDGMENTS

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. We thank Greg Sokolov for performing numerical computations.

REFERENCES

Baron, M. and Tartakovsky, A. G. (2006). Asymptotic Optimality of Change-Point Detection Schemes in General Continuous-Time Models, *Sequential Analysis* 25: 257–296.

Bojdecki, T. (1979). Probability Maximizing Approach to Optimal Stopping and Its Application to a Disorder Problem, *Stochastics* 3: 61–71.

Feinberg, E. A. and Shiryaev, A. N. (2006). Quickest Detection of Drift Change for Brownian Motion in Generalized Bayesian and Minimax Settings, *Statistics & Decisions* 24: 445–470.

Lai, T. L. (1998). Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems, *IEEE Transactions on Information Theory* 44: 2917–2929.

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. (2008). Sequential Change Detection Revisited, *Annals of Statistics* 36: 787–807.

Moustakides, G. V., Polunchenko, A. S., and Tartakovsky, A. G. (2009). Numerical Comparison of CUSUM and Shiryaev-Roberts Procedures for Detecting Changes in Distributions, *Communications in Statistics: Theory and Methods* 38: 3225–3239.

Moustakides, G. V., Polunchenko, A. S., and Tartakovsky, A. G. (2010). A Numerical Approach to Comparative Efficiency Analysis of Quickest Change-Point Detection Procedures, *Statistica Sinica* 20: in press.

Pollak, M. (1985). Optimal Detection of a Change in Distribution, *Annals of Statistics* 13: 206–227.

Pollak, M. (1987). Average Run Lengths of an Optimal Method of Detecting a Change in Distribution, *Annals of Statistics* 15: 749–779.

Pollak, M. and Tartakovsky, A. G. (2009). Optimality Properties of the Shiryaev–Roberts Procedure, *Statistica Sinica* 19: 1729–1739.

Polunchenko, A. S. and Tartakovsky, A. G. (2010). Optimality of the Shiryaev–Roberts Procedure for Detecting a Change in Distribution, *Annals of Statistics* (accepted in Nov. 2009). Available at ArXiv:0904.3370v2 [math.ST], 25 August 2009.

Shiryaev, A. N. (1961). The Problem of the Most Rapid Detection of a Disturbance in a Stationary Process, *Soviet Mathematics—Doklady* 2: 795–799.

Shiryaev, A. N. (1963). On Optimum Methods in Quickest Detection Problems, *Theory of Probability and Its Applications* 8: 22–46.

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

Shiryaev, A. N. (2009). On the Stochastic Models and Optimal Methods in the Quickest Detection Problems, *Theory of Probability and Its Applications*, 53: 385–401 (Translation from Russian *Teoriya Veroyatnostej i ee Primeneniya* 53: 417–436, 2008).

Tartakovsky, A. G. (2009). Asymptotic Optimality in Bayesian Changepoint Detection Problems under Global False Alarm Probability Constraint, *Theory of Probability and Its Applications*, 53: 443–466 (Translation from Russian *Teoriya Veroyatnostej i ee Primeneniya* 53: 472–499, 2008).

Tartakovsky, A. G. and Veeravalli, V. V. (2005). General Asymptotic Bayesian Theory of Quickest Change Detection, *Theory of Probability and Its Applications* 49: 458–497.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Philadelphia: SIAM.