ISSN 1066-5307, Mathematical Methods of Statistics, 2019, Vol. 28, No. 2, pp. 155–167. © Allerton Press, Inc., 2019.

A Multiple Hypothesis Testing Approach to Detection Changes in Distribution

G. Golubev[1*] and M. Safarian[2]**

[1] *Inst. for Information Transmission Probl., Moscow, Russia*

[2] *Dept. of Economics, Karlsruhe Inst. of Technology, Karlsruhe, Germany*

Received April 13, 2019; in final form, April 17, 2019

Abstract—Let $X_1, X_2, \ldots$ be independent random variables observed sequentially and such that $X_1, \ldots, X_{\theta-1}$ have a common probability density p_0, while $X_\theta, X_{\theta+1}, \ldots$ are all distributed according to $p_1 \neq p_0$. It is assumed that p_0 and p_1 are known, but the time change $\theta \in \mathbb{Z}^+$ is unknown and the goal is to construct a stopping time τ that detects the change-point θ as soon as possible. The standard approaches to this problem rely essentially on some prior information about θ. For instance, in the Bayes approach, it is assumed that θ is a random variable with a known probability distribution. In the methods related to hypothesis testing, this a priori information is hidden in the so-called average run length. The main goal in this paper is to construct stopping times that are free from a priori information about θ. More formally, we propose an approach to solving approximately the following minimization problem:

$$\Delta(\theta; \tau^\alpha) \to \min_{\tau^\alpha} \quad \text{subject to} \quad \alpha(\theta; \tau^\alpha) \leq \alpha \ \text{ for any } \theta \geq 1,$$

where $\alpha(\theta; \tau) = \mathsf{P}_\theta\{\tau < \theta\}$ is *the false alarm probability* and $\Delta(\theta; \tau) = \mathsf{E}_\theta(\tau - \theta)_+$ is *the average detection delay* computed for a given stopping time τ. In contrast to the standard CUSUM algorithm based on the sequential maximum likelihood test, our approach is related to a multiple hypothesis testing methods and permits, in particular, to construct universal stopping times with nearly Bayes detection delays.

Keywords: stopping time, false alarm probability, average detection delay, Bayes stopping time, CUSUM method, multiple hypothesis testing.

AMS 2010 Subject Classification: primary 62L10, 62L15; secondary 60G40.

DOI: 10.3103/S1066530719020054

1. INTRODUCTION

Let $X_1, X_2, \ldots$ be independent random variables observed sequentially. It is assumed that $X_1, \ldots, X_{\theta-1}$ have a common probability density $p_0(x)$, $x \in \mathbb{R}^d$, while $X_\theta, X_{\theta+1}, \ldots$ are all distributed according to a probability density $p_1(x)$, $x \in \mathbb{R}^d$. This paper deals with the simple change-point detection problem assuming that $p_0(\cdot)$ and $p_1(\cdot)$ are known, but the time change $\theta \in \mathbb{Z}^+$ is unknown, and the goal is to construct a stopping time $\tau \in \mathbb{Z}^+$ that detects θ as soon as possible. The existing approaches to this problem rely essentially on some prior information about θ. For instance, in the Bayes approach, it is assumed that θ is a random variable with a known probability distribution, see e.g. [12]. In methods related to sequential hypothesis testing, the prior information is hidden in the so-called average run length, see e.g. [7]. Our goal in this paper is to construct stopping times that are free from a priori information about θ, but have nearly minimal detection delays.

To be more precise, denote by P_θ the probability distribution of $\{X_1, \ldots, X_{\theta-1}, X_\theta, \ldots\}$ and by E_θ the expectation with respect to this measure. In this paper, statistical properties of τ are measured with the help of the following functions in $\theta \in \mathbb{Z}^+$:

*E-mail: golubev.yuri@gmail.com

**E-mail: mher.safarian@kit.edu

- *false alarm probability*

$$\alpha(\theta;\tau) = \mathsf{P}_\theta\big\{\tau < \theta\big\};$$

- *average detection delay*

$$\Delta(\theta;\tau) = \mathsf{E}_\theta[\tau-\theta]_+, \quad \textit{where} \;\; [x]_+ = \max\{0,x\}.$$

Heuristically, we want to find a stopping time solving the following problem:

$$\Delta(\theta;\tau^\alpha) \to \min_{\tau^\alpha} \quad \text{subject to} \;\; \alpha(\theta;\tau^\alpha) \le \alpha \;\; \text{for any} \;\; \theta \in \mathbb{Z}^+. \tag{1}$$

In other words, the goal is to construct τ^α minimizing detection delay $\Delta(\theta;\tau^\alpha)$ for any given $\theta \in \mathbb{Z}^+$ and such that $\alpha(\theta;\tau^\alpha) \le \alpha$. The main difficulty in this problem is related obviously to the fact that for a given stopping time τ^α the average detection delay $\Delta(\theta;\tau^\alpha)$ depends on θ. This means that in order to compare two stopping times τ_1^α and τ_2^α, one has to compare two functions of $\theta \in \mathbb{Z}^+$. Obviously, this is not feasible from a mathematical viewpoint and the principal objective in this paper is to propose stopping times providing good approximative solutions to (1). Notice that similar problems are common in statistics and there are well-known approaches to obtain their reasonable solutions.

In change-point detection, two standard methods are usually used for constructing stopping times.

The Bayes approach. The first Bayes change-point detection problem was stated in [4] for on-line quality control problem for continuous technological processes. This approach assumes that θ is a random variable with a known distribution

$$\pi_m = \mathsf{P}\{\theta = m\}, \quad m = 1,2,\ldots,$$

and the goal is to construct a stopping time τ_π^α that solves the averaged version of (1), i.e.,

$$\sum_{m=1}^{\infty} \pi_m \Delta(m;\tau_\pi^\alpha) \to \min_{\tau_\pi^\alpha} \quad \text{subject to} \; \sum_{m=1}^{\infty} \pi_m \alpha(m;\tau_\pi^\alpha) \le \alpha. \tag{2}$$

Let us emphasize that in contrast to (1), Problem (2) is well defined from a mathematical viewpoint, but its solution depends on the prior distribution π that is hardly known in practice.

A hypothesis testing approach. The first non-Bayesian change-point detection algorithm based on sequential hypothesis testing was proposed in [7]. Denote by $X^n = \{X_1,\ldots,X_n\}$ the observations till moment n. The main idea in this approach is to test sequentially

$$\begin{aligned} &\textit{simple hypothesis} \\ &\mathrm{H}_0^n\colon X^n \sim \prod_{i=1}^n p_0(x_i) \\ &\textit{vs. composite alternative} \\ &\mathrm{H}_1^n\colon X^n \sim \prod_{i=1}^{m-1} p_0(x_i) \prod_{i=m}^{n} p_1(x_i), \quad \text{for some} \;\; m \in [1,n], \end{aligned} \tag{3}$$

and to compute stopping time τ as follows:

- if H_1^n is accepted, then $\tau = n$;
- if H_0^n is accepted, then H_0^{n+1} and H_1^{n+1} are tested.

In order to motivate stopping times proposed in this paper, let us discuss briefly the basic statistical properties of the above mentioned approaches.

1.1. The Bayes Approach

Usually in this approach the geometric prior distribution

$$\pi_m = \gamma(1-\gamma)^{m-1}, \quad m = 1, 2, \ldots, \quad \gamma > 0,$$

is used. The parameter γ is assumed to be known. In this case, the optimal stopping time is given by the following famous theorem [12].

Theorem 1. *The optimal Bayes stopping time* (*see* (2)) *is given by*

$$\tau_\gamma^\alpha = \min\{k\colon \bar{\pi}_\gamma(k) \geq A_\gamma^\alpha\}, \tag{4}$$

where A_γ^α *is a threshold and*

$$\bar{\pi}_\gamma(k) = \mathsf{P}\{\theta \leq k \mid X^k\}.$$

Remark. Computing the exact value of A_γ^α is difficult. For instance, this can be done with the help of the Monte Carlo method. However it is also well known (see Remark 1, p. 200 in [12]) that $A_\gamma^\alpha \leq 1-\alpha$.

Notice that the geometric prior distribution results in the following recursive formula for the posterior probability (see, e.g., [12]):

$$\bar{\pi}_\gamma(k) = \frac{[\gamma + (1-\gamma)\bar{\pi}_\gamma(k-1)]p_1(X_k)}{[\gamma + (1-\gamma)\bar{\pi}_\gamma(k-1)]p_1(X_k) + [1-\bar{\pi}_\gamma(k-1)](1-\gamma)p_0(X_k)}. \tag{5}$$

So, if we denote

$$\rho_\gamma(k) = \frac{\bar{\pi}_\gamma(k)}{1-\bar{\pi}_\gamma(k)},$$

then (5) may be rewritten in the following equivalent form:

$$\rho_\gamma(k) = \frac{\gamma + \rho_\gamma(k-1)}{1-\gamma} \times \frac{p_1(X_k)}{p_0(X_k)}, \quad \rho_\gamma(0) = 0. \tag{6}$$

From this equation it is clear, in particular, that the Bayes stopping time depends on γ that is hardly known in practice. In statistics, in order to avoid such dependence, the uniform prior distribution is usually used. Let us see how this idea works in the change-point detection. The uniform prior distribution assumes that $\gamma = 0$ and in this case we immediately obtain from (6)

$$\rho_0(k) = \rho_0(k-1) \times \frac{p_1(X_k)}{p_0(X_k)}.$$

Therefore, for

$$L_0(k) = \log[\rho_0(k)],$$

we get

$$L_0(k) = \sum_{i=1}^k \log \frac{p_1(X_i)}{p_0(X_i)}.$$

Hence the optimal stopping time is given by

$$\tau_\circ^\alpha = \min\bigl\{k\colon L_0(k) \geq t^\alpha\bigr\}, \tag{7}$$

where t^α is a threshold which we will compute later on. Figure 1 shows a typical trajectory of $L_0(X^k)$, $k = 1, 2, \ldots$, in detecting change $\theta = 0$ to $\theta = 80$ in the Gaussian distribution $\mathcal{N}(\theta, 1)$.

Computing the false alarm probability for this stopping time is not difficult with the help of the following simple fact. Let

$$\varphi(\lambda) = \mathsf{E}_\infty \exp\biggl[\lambda \log \frac{p_1(X_1)}{p_0(X_1)}\biggr].$$

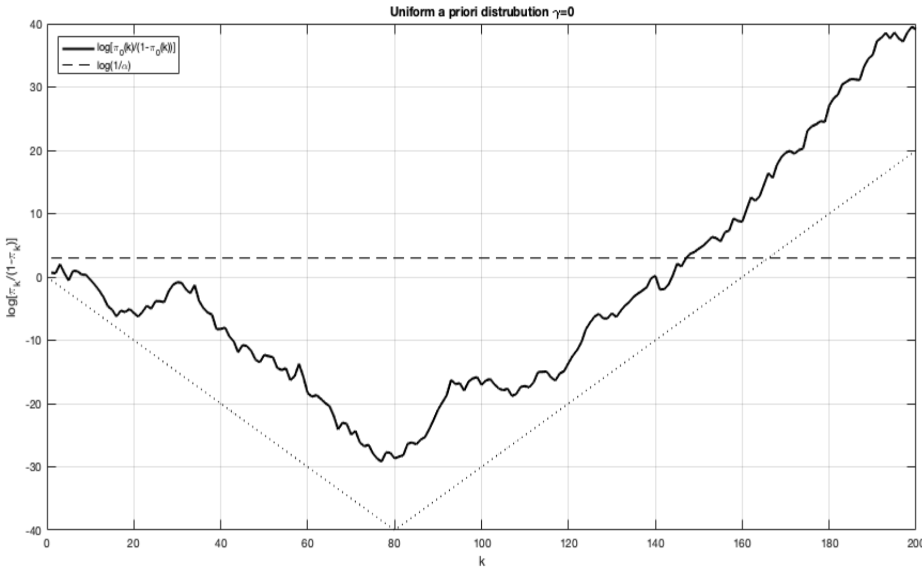

Fig. 1. Detecting change $\theta = 0$ to $\theta = 80$ in the mean of $\mathcal{N}(\theta, 1)$ with the help of $\tau_\circ^\alpha$.

Lemma 1. *For any* $\lambda > 0$

$$\mathsf{E}_\infty \exp\{-\tau_\circ^\alpha \log[\varphi(\lambda)]\}\mathbf{1}(\tau_\circ^\alpha < \infty) \le \exp(-\lambda t^\alpha).$$

It follows immediately from the definition of $\varphi(\lambda)$ that if $\lambda = 1$, then $\varphi(\lambda) = 1$. So, by this Lemma we get

$$\mathsf{P}_\infty\{\tau_\circ^\alpha < \infty\} \le \exp(-t^\alpha).$$

As to the average detection delay, it can be easily computed with the help of the famous Wald identity [2, 14]. The following proposition summarizes principal properties of $\tau_\circ^\alpha$. Let us assume that

$$\mu_0 \stackrel{\text{def}}{=} \int \log \frac{p_0(x)}{p_1(x)} p_0(x)\, dx > 0 \quad \text{and} \quad \mu_1 \stackrel{\text{def}}{=} \int \log \frac{p_1(x)}{p_0(x)} p_1(x)\, dx > 0.$$

Proposition 1. *Let* $t^\alpha = \log(1/\alpha)$ *in* (7). *Then*

$$\alpha(\theta; \tau_\circ^\alpha) \le \alpha, \quad \Delta(\theta; \tau_\circ^\alpha) = \frac{\log(1/\alpha) + \theta\mu_0}{\mu_1}.$$

We would like to emphasize that $\Delta(\theta; \tau_\circ^\alpha)$ is linear in θ. Unfortunately, this is not good both from practical and from theoretical viewpoints. In order to understand why it is so, let us turn back to the Bayes setting assuming that $\gamma > 0$. This case is described by the following proposition.

Proposition 2. *Suppose* $\gamma > 0$. *Then for* τ_γ^α *defined by* (4)*, as* $\gamma \to 0$*,*

$$\sup_{\theta \in \mathbb{Z}^+} \alpha(\theta; \tau_\gamma^\alpha) = 1 + o(1), \qquad \Delta(\theta; \tau_\gamma^\alpha) = \frac{\log[1/(\gamma\alpha)]}{\mu_1} + O(1). \tag{8}$$

Let us explain heuristically some simple ideas in the proof of this proposition. Its formal proof may be obtained with the help of the standard technique (see, e.g., [1]).

For

$$L_\gamma(k) = \log[\rho_\gamma(k)] + \log \frac{1}{\gamma}$$

we obtain from (6)

$$L_\gamma(k) = \log\bigl\{1 + \exp\bigl[L_\gamma(k-1)\bigr]\bigr\} + \log\frac{p_1(X_k)}{p_0(X_k)} - \log(1-\gamma) \tag{9}$$

and

$$\tau_\gamma^\alpha = \min\left\{k\colon L_\gamma(k) \ge \log\frac{1}{\alpha\gamma}\right\}.$$

Therefore it is clear from (9) that if $k < \theta$ and

$$\mathsf{E}_0 \log\frac{p_1(X_k)}{p_0(X_k)} < \log(1-\gamma),$$

then $L_\gamma(k)$ is a stationary Markov process with a bounded mean. Next notice that

$$\log[1+\exp(x)] = x + O(\exp(-x)), \quad x \to \infty.$$

Therefore, when $k > \theta$,

$$\mathsf{E}\left[\log\frac{p_1(X_k)}{p_0(X_k)} - \log(1-\gamma)\right] > 0$$

and hence we obtain the following approximation

$$L_\gamma(k) \approx L_\gamma(\theta-1) + \sum_{s=\theta}^{k}\left[\log\frac{p_1(X_k)}{p_0(X_k)} - \log(1-\gamma)\right].$$

So, (8) follows from Wald's identity.

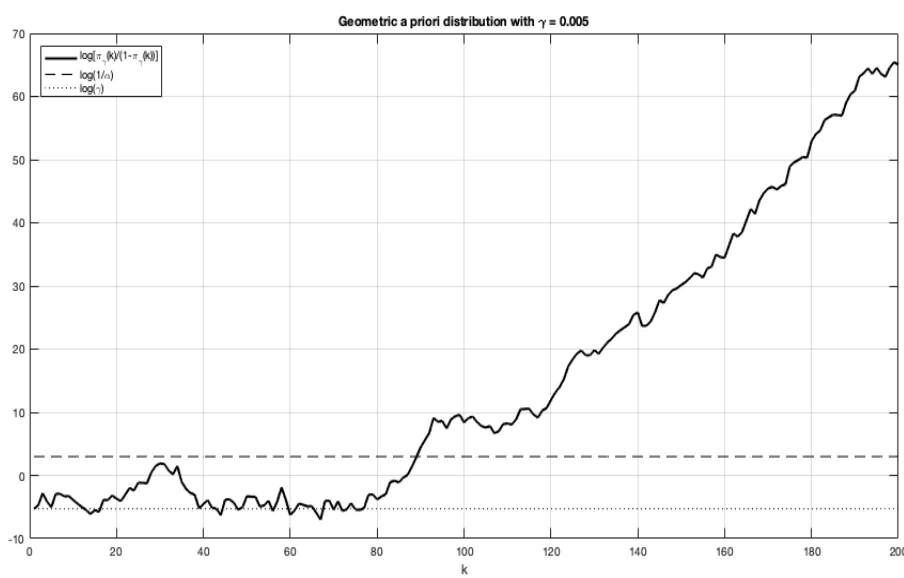

Fig. 2. Detecting change $\theta = 0$ to $\theta = 80$ in the mean of $\mathcal{N}(\theta, 1)$ with the help of τ_γ^α ($\gamma = 0.005$).

Figure 2 illustrates typical behavior of $\log[\rho_\gamma(k)]$ with $\gamma = 0.005$. Notice that if the stopping time $\tau_\circ^\alpha$ is used in the considered case, then by (8) we get

$$\mathsf{E}\Delta(\theta; \tau_\circ^\alpha) = \frac{\log(1/\alpha)}{\mu_1} + \frac{\mu_0}{\mu_1} \times \frac{1}{\gamma}.$$

So, we see that for small γ this mean detection delay may be far away from the Bayes one given by

$$\mathsf{E}\Delta(\theta;\tau_\gamma^\alpha) = \frac{\log(1/\alpha)}{\mu_1} + \frac{1}{\mu_1}\times\log\frac{1}{\gamma} + O(1) \quad \text{as} \quad \gamma\alpha\to 0.$$

Let us now summarize briefly some facts related to the classical Bayes approach.

- If $\gamma = 0$, then the average detection delay of the Bayes stopping time grows linearly in θ.
- If $\gamma > 0$, then the maximal false alarm probability of the Bayes stopping time is 1.

So, it is clear that the standard Bayes technique cannot provide good solutions to (1).

1.2. Sequential hypothesis testing approach

The main idea of this approach is based on the well-known sequential test for two simple hypothesis [15]. However, in contrast to the standard setting in [15], in the change-point detection, this approach has a solely heuristic motivation since here we deal with testing a simple hypothesis versus a composite alternative whose complexity grows with new observations.

In the classical sequential hypothesis testing there are two well-known methods:

- maximum likelihood (ML);
- Bayes.

The ML test accepts H_1^n (see (3)) if

$$\max_{k\le n}\frac{\prod_{i=1}^{k-1}p_0(X_i)\prod_{i=k}^{n}p_1(X_i)}{\prod_{i=1}^{n}p_0(X_i)} \ge t'^{\alpha}$$

or, equivalently, if

$$M(n)\ge t^\alpha,$$

where

$$M(n) = \max_{k\le n}\sum_{i=k}^{n}\log\frac{p_1(X_i)}{p_0(X_i)}.$$

The threshold t^α is defined by

$$t^\alpha = \min\Big\{t\colon \mathsf{P}_\infty\big\{M(n)\ge t\big\}\le\alpha\Big\},$$

where α is the type I error probability. Notice that by Lemma 1

$$\mathsf{P}_\infty\big\{M(n)\ge x\big\}\le\exp(-x).$$

Therefore the ML test results in the following stopping time:

$$\tau_{\mathrm{ml}}^\alpha = \min\Big\{n\colon M(n)\ge\log\frac{1}{\alpha}\Big\}. \qquad (10)$$

Notice also that the test statistic $M(n)$ admits a simple recursive computation [7]. Indeed,

$$\max_{k\le n}\sum_{i=k}^{n}\log\frac{p_1(X_i)}{p_0(X_i)} = \max\bigg\{\log\frac{p_1(X_n)}{p_0(X_n)}, \log\frac{p_1(X_n)}{p_0(X_n)} + \max_{k\le n-1}\sum_{i=k}^{n-1}\log\frac{p_1(X_i)}{p_0(X_i)}\bigg\}$$

$$= \log\frac{p_1(X_n)}{p_0(X_n)} + \max\bigg\{0, \max_{k\le n-1}\sum_{i=k}^{n-1}\log\frac{p_1(X_i)}{p_0(X_i)}\bigg\}.$$

Therefore

$$M(n) = \log \frac{p_1(X_n)}{p_0(X_n)} + \big[M(n-1)\big]_+. \tag{11}$$

This method is usually called the CUSUM algorithm. It is well known that it is optimal in the Lorden [5] sense, i.e., for a properly chosen α, $\tau^{\alpha}_{\mathrm{ml}}$ minimizes

$$\sup_{\theta\in\mathbb{Z}^+} \operatorname{ess\,sup} \mathsf{E}_\theta\big[(\tau-\theta)_+ \mid X_1,\dots,X_{\theta-1}\big]$$

in the class of stopping times $\big\{\tau\colon \mathsf{E}_\infty\tau \ge T\big\}$, see [6].

However with this method one cannot control the false alarm probability as the following proposition shows.

Proposition 3. *For any* $\alpha \in (0,1)$

$$\sup_{\theta\in\mathbb{Z}^+} \alpha(\theta;\tau^{\alpha}_{\mathrm{ml}}) = 1,$$

and as $\alpha \to 0$

$$\Delta(\theta;\tau^{\alpha}_{\mathrm{ml}}) = \frac{1}{\mu_1}\log\frac{1}{\alpha} + O(1).$$

The proof of this proposition is standard and therefore omitted.

The Bayes test is based on the assumption that θ is a random variable with uniform distribution on $[1,n]$. So, this test accepts H_1^n if

$$S(n) \stackrel{\mathrm{def}}{=} \sum_{k=1}^{n} \frac{\prod_{i=1}^{k-1} p_0(X_i)\prod_{i=k}^{n} p_1(X_i)}{\prod_{i=1}^{n} p_0(X_i)} \ge t^{\alpha}. \tag{12}$$

Since

$$S(n) = \sum_{k=1}^{n}\prod_{i=k}^{n}\frac{p_1(X_i)}{p_0(X_i)}$$

and

$$\sum_{k=1}^{n}\prod_{i=k}^{n}\frac{p_1(X_i)}{p_0(X_i)} = \sum_{k=1}^{n-1}\prod_{i=k}^{n}\frac{p_1(X_i)}{p_0(X_i)} + \frac{p_1(X_n)}{p_0(X_n)} = \bigg[1+\sum_{k=1}^{n-1}\prod_{i=k}^{n-1}\frac{p_1(X_i)}{p_0(X_i)}\bigg]\frac{p_1(X_n)}{p_0(X_n)},$$

the test statistic in (12) admits the following recursive computation:

$$S(n) = \big[1+S(n-1)\big]\times\frac{p_1(X_n)}{p_0(X_n)}. \tag{13}$$

The corresponding stopping time is defined by

$$\tau^{\alpha}_{\mathrm{S}} = \min\big\{n\colon S(n)\ge t^{\alpha}\big\}.$$

In the literature, this method is known as Shiryaev–Roberts (SR) algorithm. It was first proposed in [11] and [10]. In [8] and [3] it is shown that it minimizes the integral average delay

$$\frac{1}{\mathsf{E}_\infty\tau}\sum_{\theta=1}^{\infty}\mathsf{E}_\theta(\tau-\theta)_+$$

over all stopping times τ with $\mathsf{E}_\infty\tau \ge T$. More detailed statistical properties of SR procedure can be found in [9].

Notice that for

$$V(n) = \log[S(n)]$$

we obtain obviously from (13)

$$V(n) = \log\{1 + \exp[V(n-1)]\} + \log \frac{p_1(X_n)}{p_0(X_n)}.$$

Comparing this equation with (9) we see that the SR algorithm may be viewed as the limiting case ($\gamma \to 0$) of the standard Bayes change-point detection method and it is not surprising that the fact similar to Proposition 3 holds for SR algorithm.

As one can see in Fig. 3, in practice there is no significant difference between CUSUM and SR algorithms.

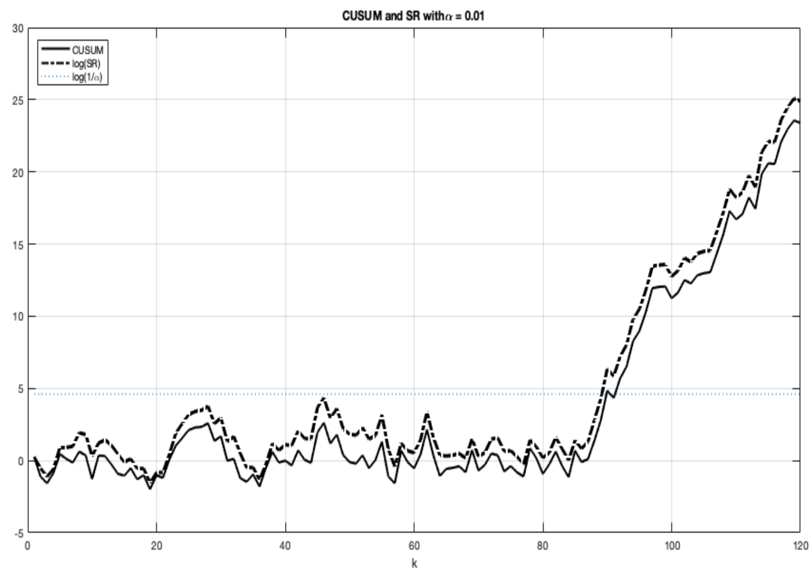

Fig. 3. Detecting change $\theta = 0$ to $\theta = 80$ in the mean of $\mathcal{N}(\theta, 1)$ with the help of CUSUM and SR procedures.

Summarizing, the standard hypothesis testing methods result in stopping times with false alarm probability 1 and thus they cannot provide reasonable approaches to solving (1).

2. A MULTIPLE HYPOTHESIS TESTING APPROACH

The main idea in this approach is to replace the constant threshold in the ML test (10) by the one depending on k. So, we will consider the following stopping times:

$$\tau^\alpha = \min\{k\colon M(k) \ge t^\alpha(k)\}.$$

In order to control the false alarm probability and to obtain a nearly minimal average detection delay, we are looking for a *minimal deterministic function* $t^\alpha(k)$, $k = 1, 2, \ldots$, such that

$$\mathsf{P}_\infty\Big\{\max_{k \ge \mathbb{Z}^+}\big[M(k) - t^\alpha(k)\big] \ge 0\Big\} \le \alpha.$$

We begin the construction of $t^\alpha(\cdot)$ with the following function:

$$\varphi_0(x) = 1 + \log(x),\ x \in \mathbb{R}^+,$$

and recurrently iterate it m times, i.e., compute

$$\varphi_k(x) = \varphi_0\big[\varphi_{k-1}(x)\big],\ k = 1, \ldots, m.$$

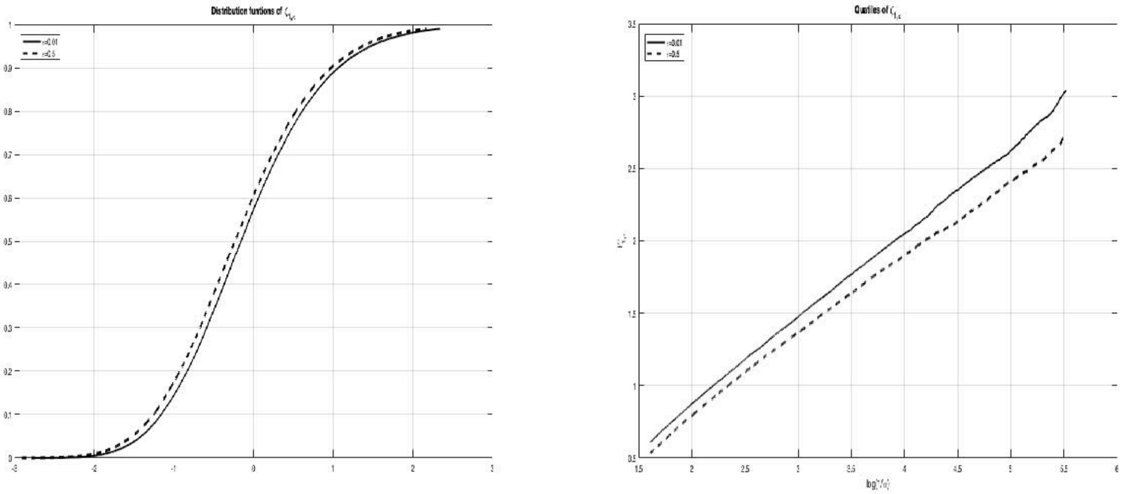

Fig. 4. Distribution functions and α-quantiles of $\zeta_{1,\epsilon}$ for $\epsilon = 0.01$ and $\epsilon = 0.5$.

Next, for a given $\epsilon \in (0,1)$, define

$$b_{m,\epsilon}(x) = -\log\left[\frac{1}{\epsilon[\varphi_m(x)]^{\epsilon}} - \frac{1}{\epsilon[\varphi_m(x+1)]^{\epsilon}}\right], \quad x \in \mathbb{R}^+. \tag{14}$$

Consider the following random variable:

$$\zeta_{m,\epsilon} = \max_{n\in\mathbb{Z}^+}\bigl\{M(n) - b_{m,\epsilon}(n)\bigr\},$$

where $M(n)$ is defined by (11). The following lemma shows that this random variable is nondegenerate.

Lemma 2. *For any* $\epsilon \in (0,1)$, $m \geq 1$, *and* $x > -\log(1 - 0.2075/2) \approx 0.11$

$$\mathsf{P}_\infty\bigl\{\zeta_{m,\epsilon} \geq x\bigr\} \leq 1 - \exp\bigl\{-\mathrm{e}^{-x}\bigl[\epsilon^{-1} + \mathrm{e}^{-x}\bigr]\bigr\}.$$

Therefore we can define the α-quantile of $\zeta_{m,\epsilon}$ by

$$t^{\alpha}_{m,\epsilon} = \min\bigl\{x\colon \mathsf{P}_\infty\bigl\{\zeta_{m,\epsilon} \geq x\bigr\} \leq \alpha\bigr\}.$$

Figure 4 shows distribution functions and α-quantiles of $\zeta_{1,\epsilon}$ for $\epsilon = \{0.01, 0.5\}$ computed with the help of Monte Carlo method with $5 \cdot 10^4$ replications.

The following lemma describes principal statistical properties of the stopping time

$$\tau^{\alpha}_{m,\epsilon} = \min\bigl\{n\colon M(n) \geq b_{m,\epsilon}(n) + t^{\alpha}_{m,\epsilon}\bigr\}.$$

Lemma 3. *For any* $\epsilon \in (0,1]$

$$\alpha\bigl(\theta; \tau^{\alpha}_{m,\epsilon}\bigr) \leq \alpha, \qquad \Delta\bigl(\theta; \tau^{\alpha}_{m,\epsilon}\bigr) \leq d^{\alpha}_{m,\epsilon}(\theta),$$

where $d^{\alpha}_{m,\epsilon}(\theta)$ *is a solution to*

$$\mu_1 d^{\alpha}_{m,\epsilon}(\theta) = b_{m,\epsilon}\bigl[\theta + d^{\alpha}_{m,\epsilon}(\theta)\bigr] + t^{\alpha}_{m,\epsilon}. \tag{15}$$

The following theorem summarizes the principal statistical properties of $\tau_{m,\epsilon}$.

Theorem 2. *For any* $\epsilon \in (0,1]$, *uniformly in* $\theta \in \mathbb{Z}^+$,

$$\begin{aligned} &\alpha\bigl(\theta; \tau^{\alpha}_{m,\epsilon}\bigr) \leq \alpha, \\ &\Delta\bigl(\theta; \tau^{\alpha}_{m,\epsilon}\bigr) \leq \frac{1}{\mu_1}\log\frac{\theta}{\alpha} + O(1) + \frac{1}{\mu_1}\Biggl\{\sum_{j=1}^{m}\log[\varphi_j(\theta)] + \epsilon\log[\varphi_m(\theta)] + \log\frac{1}{\epsilon}\Biggr\}. \end{aligned} \tag{16}$$

Remark. It is easy to check with a simple algebra that for any given $\theta > 1$

$$\lim_{j\to\infty} j \log[\varphi_j(\theta)] = 2.$$

In order to explain why $\tau_{m,\epsilon}$ is a good stopping time, suppose θ is a random variable with the geometric distribution, i.e.,

$$\mathsf{P}\big\{\theta = k\big\} = \gamma(1-\gamma)^{k-1}, \quad k \in \mathbb{Z}^+.$$

Then, averaging (16) w.r.t. this distribution, we obtain

$$\begin{aligned}\mathsf{E}\Delta\big(\theta;\tau^\alpha_{m,\epsilon}\big) \le \frac{1}{\mu_1}\log\frac{1}{\alpha\gamma} + O(1)\\ + \frac{1}{\mu_1}\bigg\{\sum_{j=1}^m \log\bigg[\varphi_j\bigg(\frac{1}{\gamma}\bigg)\bigg] + \epsilon\log\bigg[\varphi_m\bigg(\frac{1}{\gamma}\bigg)\bigg] + \log\frac{1}{\epsilon}\bigg\}\end{aligned} \tag{17}$$

as $\gamma \to 0$, and with (8) we arrive at

Theorem 3. *As* $\gamma \to 0$,

$$\begin{aligned}&\mathsf{E}\alpha\big(\theta;\tau^\alpha_{m,\epsilon}\big) \le \mathsf{E}\Delta\big(\theta;\tau^\alpha_\gamma\big) = \alpha,\\ &\mathsf{E}\Delta\big(\theta;\tau^\alpha_{m,\epsilon}\big) \le \mathsf{E}\Delta\big(\theta;\tau^\alpha_\gamma\big) + O(1) + \frac{1}{\mu_1}\bigg\{\sum_{j=1}^m \log\bigg[\varphi_j\bigg(\frac{1}{\gamma}\bigg)\bigg] + \epsilon\log\bigg[\varphi_m\bigg(\frac{1}{\gamma}\bigg)\bigg] + \log\frac{1}{\epsilon}\bigg\}\\ &\qquad = (1+o(1))\mathsf{E}\Delta\big(\theta;\tau^\alpha_\gamma\big),\end{aligned}$$

where τ^α_γ *is the Bayes stopping time (see Theorem 1).*

Remark. This theorem demonstrates that the stopping time $\tau_{m,\epsilon}$ has a nearly Bayes detection delay. From a formal mathematical viewpoint, it is clear also that the larger m, the better the upper bound for the average detection delay for small γ. However, from a practical viewpoint $m = 1$ would be a reasonable choice.

3. APPENDIX

Proof of Lemma 1. Since

$$Y_k = \exp\big\{-k\log[\varphi(\lambda)] + \lambda L_0(k)\big\}$$

is a martingale with $\mathsf{E}_\infty Y_k = 1$, we have

$$\begin{aligned}1 &= \mathsf{E}_\infty Y_{\tau^\alpha_\circ} = \mathsf{E}_\infty Y_{\tau^\alpha_\circ}\mathbf{1}(\tau^\alpha_\circ < \infty) + \mathsf{E}_\infty Y_{\tau^\alpha_\circ}\mathbf{1}(\tau^\alpha_\circ = \infty)\\ &\ge \mathsf{E}_\infty Y_{\tau^\alpha_\circ}\mathbf{1}(\tau^\alpha_\circ < \infty) = \mathsf{E}_\infty \exp\big\{-\tau^\alpha_\circ\log[\varphi(\lambda)] + \lambda A\big\}\mathbf{1}(\tau^\alpha_\circ < \infty).\end{aligned}$$

□

In what follows, we denote by e_k i.i.d. standard exponential random variables.

Lemma 4. *For any* $m \ge 1$ *and* $x > -\log(1 - 0.2075/2) \approx 0.11$

$$\mathsf{P}\Big\{\max_{k\in\mathbb{Z}^+}[e_k - b_{m,\epsilon}(k)] \ge x\Big\} \le 1 - \exp\Big\{-\mathrm{e}^{-x}\big[\epsilon^{-1} + \mathrm{e}^{-x}\big]\Big\},$$

where $b_{m,\epsilon}(\cdot)$ *is defined by* (14).

Proof. It is easy to check with a simple algebra that for any $u \in [0,1)$

$$\log(1-u) \geq -u - \frac{u^2}{2(1-u)}.$$

Therefore with this inequality we obtain

$$\begin{aligned}
\mathsf{P}\Big\{\max_{k\in\mathbb{Z}^+}[e_k - b_{m,\epsilon}(k)] \geq x\Big\} &= 1 - \prod_{k=1}^{\infty}\Big\{1 - \mathsf{P}\big\{e_k \geq x + b_{m,\epsilon}(k)\big\}\Big\} \\
&= 1 - \exp\biggl\{\sum_{k=1}^{\infty}\log\Big[1 - \mathrm{e}^{-x-b_{m,\epsilon}(k)}\Big]\biggr\} \\
&\leq 1 - \exp\biggl\{-\mathrm{e}^{-x}\sum_{k=1}^{\infty}\mathrm{e}^{-b_{m,\epsilon}(k)} - \frac{\mathrm{e}^{-2x}}{2(1-\mathrm{e}^{-x})}\sum_{k=1}^{\infty}\mathrm{e}^{-2b_{m,\epsilon}(k)}\biggr\}. \qquad (18)
\end{aligned}$$

It follows immediately from the definition of $b_{m,\epsilon}$ (see (14)) that

$$\sum_{k=1}^{\infty}\mathrm{e}^{-b_{m,\epsilon}(k)} = \frac{1}{\epsilon\varphi_m(1)} = \frac{1}{\epsilon}.$$

It is also easy to check numerically that for any $m \geq 1$ and $\epsilon > 0$

$$\sum_{k=1}^{\infty}\mathrm{e}^{-2b_{m,\epsilon}(k)} < 0.2075.$$

Therefore, substituting the above equations in (18), we complete the proof. □

Lemma 5. *For any $x > 0$*

$$\mathsf{P}_\infty\Big\{\max_{k\in\mathbb{Z}^+}[M(k) - b_{m,\epsilon}(k)] \geq x\Big\} \leq \mathsf{P}\Big\{\max_{k\in\mathbb{Z}^+}[e_k - b_{m,\epsilon}(k)] \geq x\Big\},$$

where the random process $M(k)$ is defined by (11).

Proof. Define random integers $\varkappa_1 < \varkappa_2 < \ldots$ by

$$\varkappa_k = \min\big\{s > \varkappa_{k-1} \colon M(s) \leq 0\big\}, \quad \varkappa_0 = 0,$$

It is clear (see (11)) that these random variables are renovation points for the random process $M(k)$ and therefore the random variables

$$\mu_k = \max_{\varkappa_k < s \leq \varkappa_{k+1}} M(s), \quad k = 1, 2, \ldots,$$

are independent. Since $b_{m,\epsilon}(k)$ is nondecreasing in k and obviously $\varkappa_k \geq k$, we get

$$\max_{k\in\mathbb{Z}^+}[M(k) - b_{m,\epsilon}(k)] \leq \max_{k\in\mathbb{Z}^+}\max_{\varkappa_k < s \leq \varkappa_{k+1}}[M(s) - b_{m,\epsilon}(t_k)] \leq \max_{k\in\mathbb{Z}^+}[\mu_k - b_{m,\epsilon}(k)].$$

Therefore, to finish the proof, it suffices to notice that by (11) and Lemma 1

$$\mathsf{P}_\infty\{\mu_k \geq x\} \leq \mathsf{P}_\infty\biggl\{\max_{k>\theta}\sum_{s=\theta}^{k}\log\frac{p_0(X_s)}{p_1(X_s)} \geq x\biggr\} \leq \exp(-x).$$

□

Lemma 2 follows now immediately from Lemmas 4 and 5.

Proof of Lemma 3. It follows from (11) that for all $k \ge \theta$

$$M(k) \ge \sum_{s=\theta}^{k} \log \frac{p_0(X_s)}{p_1(X_s)}$$

and therefore

$$\Delta(\theta; \tau_{m,\epsilon}) \le \mathsf{E}_\theta \tau^+,$$

where

$$\tau^+ = \min\left\{k \ge 1 : \sum_{s=\theta}^{\theta+k} \log \frac{p_0(X_s)}{p_1(X_s)} \ge b_{m,\epsilon}(\theta + k) + t_{m,\epsilon}^{\alpha}\right\}.$$

The computation of $\mathsf{E}_\theta \tau^+$ is based on the famous Wald's identity [14] (see also [2]). For given $\theta \in \mathbb{Z}^+$ and $\epsilon > 0$ define

$$B(k) = b_{m,\epsilon}(\theta + k) + t_{m,\epsilon}^{\alpha}, \quad k \in \mathbb{Z}^+.$$

It is clear that $B(\cdot)$ is a convex function and hence for any $k_0 \in \mathbb{Z}^+$

$$B(k) \le B(k_0) + B'(x_0)(k - k_0).$$

Hence

$$\tau^+ \le \tau^{++} = \min\left\{k \ge 1 \colon \sum_{s=\theta}^{\theta+k} \log \frac{p_0(X_s)}{p_1(X_s)} \ge B(k_0) + B'(k_0)(k - k_0)\right\}.$$

Next, we obtain by Wald's identity

$$\mu_1 \mathsf{E}_\theta \tau^{++} \le B(k_0) + B'(k_0)\big(\mathsf{E}_\theta \tau^{++} - k_0\big)$$

and thus

$$\mathsf{E}_\theta \tau^{++} \le \frac{B(k_0) - B'(k_0)k_0}{\mu_1 - B'(k_0)}. \tag{19}$$

To complete the proof, let us choose $k_0 = d_{m,\epsilon}^{\alpha}(\theta)$ (see (15)) and notice that $B(k_0) = \mu_1 k_0$. Hence by (19)

$$\mathsf{E}_\theta \tau^{++} \le k_0 = d_{m,\epsilon}^{\alpha}(\theta).$$

□

Proof of Theorem 2. It follows immediately from Lemma 2 that

$$t_{m,\epsilon}^{\alpha} \le \log \frac{1}{\alpha\epsilon} + o(1), \quad \alpha\epsilon \to 0. \tag{20}$$

Next, by convexity of $b_{m,\epsilon}(\cdot)$ we obtain for any x, x_0

$$b_{m,\epsilon}(\theta + x) \le b_{m,\epsilon}(\theta + x_0) + b'_{m,\epsilon}(\theta + x_0)(x - x_0).$$

Therefore choosing

$$x_0 = \frac{b_{m,\epsilon}(\theta) + t_{m,\epsilon}^{\alpha}}{\mu_1}$$

we have with (15)

$$d_{m,\epsilon}^{\alpha}(\theta) \le \frac{b_{m,\epsilon}(\theta + x_0) + t_{m,\epsilon}^{\alpha}}{\mu_1 - b'_{m,\epsilon}(\theta + x_0)}. \tag{21}$$

So, our next step is to upper bound $b_{m,\epsilon}(\cdot)$. To do this, notice that

$$-\frac{1}{\epsilon}\frac{d\varphi_m^{-\epsilon}(x)}{dx} = \varphi_m^{-1-\epsilon}(x)\varphi_m'(x) = \frac{\varphi_m^{-\epsilon}(x)}{x}\prod_{j=1}^{m}\frac{1}{\varphi_j(x)},$$

and thus

$$-\log\left[-\frac{1}{\epsilon}\frac{d\varphi_m^{-\epsilon}(x)}{dx}\right] = \log(x) + \sum_{j=1}^{m}\log[\varphi_j(x)] + \epsilon\log[\varphi_m(x)].$$

Therefore with this equation and (14) we obtain

$$b_{m,\epsilon}(k) = \log(k) + \sum_{j=1}^{m}\log[\varphi_j(k)] + \epsilon\log[\varphi_m(k)] + o(1), \quad k\to\infty. \tag{22}$$

It is also easy to check that

$$b'_{m,\epsilon}(k) = O\big(k^{-1}\big). \tag{23}$$

Finally, substituting (20), (22), and (23) in (21), we complete the proof. □

REFERENCES

1. M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application* (Prentice-Hall, 1996).
2. D. Blackwell, "On an Equation of Wald", Ann. Math. Statist. **17**, 84–87 (1946).
3. E. A. Feinberg and A. N. Shiryaev, "Quickest Detection of Drift Change for Brownian Motion in Generalized and Bayesian Settings", Statist. Decisions **24**, 445–470 (2006).
4. M. A. Girshick and H. Rubin, "A Bayes Approach to a Quality Control Model", Ann. Math. Statist. **23**, 114–125 (1952).
5. G. Lorden, "Procedures for Reacting to a Change in Distribution", Ann. Math. Statist. **42** (6), 1897–1908 (1971).
6. G. V. Moustakides, "Optimal Stopping Times for Detecting Changes in Distributions", Ann. Statist. **14** (4), 1370–1387 (1986).
7. E. S. Page, "Continuous Inspection Schemes", Biometrika **41**, 100–115 (1954).
8. M. Pollak and A. G. Tartakovsky, "Optimality Properties of Shirayev–Roberts Procedure", Statist. Sinica **19**, 1729–1739 (2009).
9. A. S. Polunchenko and A. G. Tartakovsky, "On Optimality of the Shirayev–Roberts Procedure for Detecting a Change in Distribution", Ann. Statist. **38** (6), 3445–3457 (2010).
10. S. W. Roberts, "A Comparison of Some Control Chart Procedures", Technometrics **8**, 411–430 (1966).
11. A. N. Shiryaev, "The Problem of the Most Rapid Detection Disturbance in a Stationary Process", Dokl. Math. **2**, 795–799 (1961).
12. A. N. Shiryaev, *Optimal Stopping Rules* (Springer-Verlag, Berlin, Heidelberg, 1978).
13. A. N. Shiryaev, "On Optimum Methods in Quickest Detection Problems", Theory Probab. Appl. **8**, 22–46 (1963).
14. A. Wald, "On Cumulative Sums of Random Variables", Ann. Math. Statist. **15** (3), 283–296 (1944).
15. A. Wald, "Sequential Tests of Statistical Hypotheses", Ann. Math. Statist. **16** (2), 117–186 (1945).