Quickest Change Point Detection with Multiple Post-change Models

Samrat Nath and Jingxian Wu
Department of Electrical Engineering, University of Arkansas,
Fayetteville, AR 72701, USA.

Abstract: We study the sequential quickest change point detection for systems with multiple possible post-change models. A change point is the time instant at which the distribution of a random process changes. In many practical applications, the pre-change model can be easily obtained, yet the post-change distribution is unknown due to the unexpected nature of the change. In this paper, we consider the case that the post-change model is from a finite set of possible models. The objective is to minimize the average detection delay (ADD), subject to upper bounds on the probability of false alarm (PFA). Two different quickest change detection algorithms are proposed under Bayesian and non-Bayesian settings, respectively. Under the Bayesian setting, the prior probabilities of the change point and prior probabilities of possible post-change models are assumed to be known, yet this information is not available under the non-Bayesian setting. Theoretical analysis is performed to quantify the analytical performance of the proposed algorithms in terms of exact or asymptotic bounds on PFA and ADD. It is shown through theoretical analysis that when PFA is small, both algorithms are asymptotically optimal in terms of ADD minimization for a given PFA upper bound. Numerical results demonstrate that the proposed algorithms outperform existing algorithms in the literature.

Keywords: Asymptotic optimality; Average detection delay; Changepoint detection; Probability of false alarm; Sequential analysis.

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

1. INTRODUCTION

Change point detection is the process of detecting the time instants at which the distribution of a random process changes (Tartakovsky et al., 2014). It has a wide spectrum of applications in various science and engineering fields, such as quality control, anomaly detection, and seismology, etc. (Aminikhanghahi and Cook, 2017). In many applications, it is relatively easier to obtain the distribution model before the change point, which usually corresponds to normal system operations. The post-change model, on the other hand, might not be readily available due to the unexpected nature of the change. This problem is exacerbated for quickest change detection (QCD), which aims at minimizing the detection delay with only a small amount of post-change data for training post-change models (Poor and Hadjiliadis, 2009; Veeravalli and Banerjee, 2014). For many applications, the post-change model might be from a finite set of possible models, that is, there are multiple hypotheses of the post-change models. For example, for the detection of wind turbine bearing fault, the fault could be caused by a finite number of defects, such as inner race fault, outer race fault, cage fault, and roller defect (Gong and Qiao, 2013).

Quickest change point detection methods attempt to detect the change point in real time by sequentially updating a test statistics with recently observed data. They are usually designed based on the tradeoff

Address correspondence to J. Wu, Department of Electrical Engineering, University of Arkansas, Fayetteville, AR 72701, USA; E-mail: wuj@uark.edu

among several metrics, such as average detection delay (ADD), probability of false alarm (PFA), false alarm rate (FAR), and average run length (ARL) to false alarms, etc. Existing sequential change point detection methods can be classified into two categories, Bayesian and non-Bayesian (Minimax) methods. If the prior probability of the change point is known, then Bayesian procedures, such as the well-known Shiryaev procedure (Shiryaev, 1963), can be applied to minimize the ADD, under the constraint of an upper bound on the PFA. Tartakovsky and Veeravalli (2005) showed that the Shiryaev procedure is asymptotically optimal when the PFA upper bound is small. When the prior probability of the change point is unknown, non-Bayesian procedures, such as the cumulative sum (CUSUM) (Page, 1954) method and the Shiryaev-Roberts (SR) procedure (Roberts, 1966), aim at minimizing the delay with the worst-case change point distribution, under the constraint of a lower bound of ARL. The asymptotic optimality of the CUSUM and SR procedures are discussed in Lorden (1971) and Pollak and Tartakovsky (2009). The problem of QCD is studied in different fields of science and technology. Ren and Shi (2014) formulated a Bayesian QCD over wireless fading channels with energy constraints as a partially observable Markov decision problem (POMDP) and the optimal stopping rules are shown to have weak threshold structure. Raghavan and Veeravalli (2009) studied the QCD problem to detect a point of disruption in centralized multi-sensor network. A QCD algorithm is proposed in order to detect false data injection attacks (FDIA) in smart grids with time-varying dynamic models in Nath et al. (2019).

All the above procedures are developed for binary hypothesis testing, i.e. for systems with single pre-change and single post-change model, and they require precise knowledge about the distribution models before and after the change. There are limited works with unknown or uncertain post-change models. A Bayesian QCD algorithm for system with multiple candidates of post-change models is developed in Nath and Wu (2018). Lai (1998) proposed two methods for detecting a post-change distribution with an unknown parameter. In the first method, the detection is performed by using a mixture post-change distribution, which is obtained by averaging the set of possible post-change distributions with prior distributions of the unknown parameter. The second method is based on the generalized likelihood ratio test (GLRT), where the unknown parameter of the post-change model is estimated by maximizing the likelihood ratio. The GLRT-based method still requires a generous amount of post-change training data to tune the unknown parameter for the GLRT. A low-complexity adaptive-CUSUM method is presented in Nath et al. (2019) for estimating unknown statistics of post-change distributions by using a normalized Rao test statistic (De Maio, 2007).

There also have been some works related to non-Bayesian formulation of the QCD problem with multiple post-change models. For example, Tartakovsky and Veeravalli (2004) studied the non-Bayesian QCD problem in multi-channel and distributed network systems. A non-parametric sequential method is proposed using multichannel generalization of the CUSUM procedure for the detection of intrusions in information systems in Tartakovsky et al. (2006). None of the above works provide theoretical analysis to quantify the performance of the algorithms. Mei (2010) performed sequential change point detection based on the sum of local CUSUM statistics of each possible post-change model. An orthogonal matching pursuit CUSUM (OMP-CUSUM) algorithm is proposed to detect false data attack in power grid systems with unknown post-attack parameters while minimizing the detection delay in Akingeneye and Wu (2018).

The objective of this paper is to design sequential quickest change detection algorithms for systems with multiple possible post-change models under both Bayesian and non-Bayesian settings. The algorithms are developed to minimize the average detection delay (ADD), under the constraint on the upper bounds of the probability of false alarm (PFA). Under the Bayesian setting, the algorithm is developed by analyzing the likelihood ratio of the change point, the computation of which relies on the prior probabilities of change point and prior probabilities of different post-change models. Under the non-Bayesian setting, the algorithm is designed by replacing all prior probabilities with one in the Bayesian algorithm, and the resultant procedure happens to be the sum of local Shiryaev-Roberts (SR) statistics for each post-change model. The performances of the proposed algorithms are analytically quantified in terms of exact or asymptotic bounds on PFA and ADD. It is shown that when the PFA is small, the proposed algorithms are asymptotically optimal in terms of ADD minimization under a certain PFA upper bound. Numerical results demonstrate that the proposed algorithms outperform existing algorithms in the literature.

The rest of this paper is organized as follows. Section 2 presents the assumptions and problem formulation. The Bayesian and non-Bayesian detection algorithms, along with their corresponding theoretical analysis, are given in details in Sections 3 and 4, respectively. Section 5 demonstrates the performance of the algorithms through numerical results, and Section 6 concludes the paper.

2. PROBLEM FORMULATION

Consider a sequentially observed random sequence, X_n, $n = 1, 2, \cdots$. Let $\mathcal{F}_n^X = \sigma(\mathbf{X}^{1:n})$ be the σ-algebra generated by $\mathbf{X}^{1:n}$. Assume there is an unknown change point θ, such that the distributions of the random sequence are different before and after θ. Denote the probability density function (pdf) of the random sequence before the change point as $f_{0,n}(X_n|\mathbf{X}^{1:n-1})$ for $n < \theta$. The distribution after the change point could be one of a finite number of possible distribution models, denoted as $f_{i,n}(X_n|\mathbf{X}^{1:n-1})$, for $n \geq \theta$ and $i = 1, 2, \cdots, M$, with $M < \infty$. Denote the index of true post-change distribution as β, where $\beta \in \{1, \cdots, M\}$ is unknown.

In a Bayesian setting, the change point θ is random with prior probability mass function (PMF) $\mathbf{P}(\theta = k) = \pi_k$, for $k = 1, 2, \cdots$. The post-change model index is random with prior PMF $\mathbf{P}(\beta = i) = \omega_i$, for $i = 1, \cdots, M$.

Let $\mathbf{P}_{k,i}$ and $\mathbb{E}_{k,i}$ denote the probability measure and the corresponding expectation operator when the change occurs at $\theta = k < \infty$ and the post-change model index is $\beta = i$. Under $\mathbf{P}_{k,i}$, the conditional pdf of X_n is $f_{0,n}(X_n|\mathbf{X}^{1:n-1})$ for $n < k$, and it is $f_{i,n}(X_n|\mathbf{X}^{1:n-1})$ for $n \geq k$. For any $k < \infty$, we have $\mathbf{P}_k = \sum_{i=1}^M \omega_i \mathbf{P}_{k,i}$ and $\mathbb{E}_k = \sum_{i=1}^M \mathbb{E}_{k,i}$. Denote $\mathbf{P}_\infty$ and $\mathbb{E}_\infty$ as the probability measure and expectation operator for the data sequence before the change point, that is, under $\mathbf{P}_\infty$, the conditional pdf of X_n is $f_{0,n}(X_n|\mathbf{X}^{1:n-1})$. Thus $\mathbf{P}(\mathcal{E}) = \sum_{k=1}^\infty \mathbf{P}_k(\mathcal{E})$ and $\mathbb{E} = \sum_{k=1}^\infty \mathbb{E}_k$.

We need to design a test in order to detect the change point θ based on the sequentially observed data X_n. Denote $\hat{\theta}$ as the estimated value of θ. A sequential test δ can be defined as a mapping from $\mathcal{F}_n^X$ to $\hat{\theta} \in \{1, \cdots, n\}$, such that $\delta(\mathcal{F}_n^X) = \hat{\theta}$. The test needs to be designed by optimizing with respect to two performance metrics, the PFA and ADD.

For a given test δ, the PFA and ADD are defined, respectively, as

$$\text{PFA}(\delta) = \mathbf{P}(\hat{\theta} < \theta|\mathcal{F}_n^X) \tag{2.1}$$

$$\text{ADD}(\delta) = \mathbb{E}[\hat{\theta} - \theta|\hat{\theta} \geq \theta] \tag{2.2}$$

The objective is to minimize the ADD, subject to a constraint on the PFA. The problem can thus be formulated as

$$\begin{aligned} \text{(P1)} \qquad \text{minimize} \qquad & \text{ADD}(\delta) \\ \text{subject to} \qquad & \text{PFA}(\delta) < \alpha \end{aligned}$$

We propose the solution to this problem under both Bayesian and non-Bayesian setting in the following sections respectively.

3. QUICKEST CHANGE DETECTION ALGORITHM FOR BAYESIAN SETTING

In this section, we develop the algorithm that can detect the change point with minimum delay under the Bayesian setting.

3.1. Detection Algorithm

At any moment n, the detector needs to make a decision between two hypotheses

$$\mathcal{H}_1 : \theta \le n$$
$$\mathcal{H}_0 : \theta > n$$

Define the ratio of the posterior probabilities as

$$\Delta(n) = \frac{\mathbf{P}(\mathcal{H}_1 | \mathcal{F}_n^X)}{\mathbf{P}(\mathcal{H}_0 | \mathcal{F}_n^X)}. \tag{3.1}$$

Based on Bayes' rule, we have

$$\Delta(n) = \frac{\sum_{k=1}^{n} \pi_k \cdot d\mathbf{P}(\mathbf{x}^{1:n} | \theta = k)}{\Omega_n \cdot d\mathbf{P}(\mathbf{x}^{1:n} | \theta > n)} = \sum_{i=1}^{M} \omega_i \sum_{k=1}^{n} \frac{\pi_k}{\Omega_n} \prod_{t=k}^{n} \frac{f_{i,t}(X_t | \mathbf{X}^{1:t-1})}{f_{0,t}(X_t | \mathbf{X}^{1:t-1})} \tag{3.2}$$

where, $\Omega_n = \mathbf{P}(\theta > n) = \sum_{k=n+1}^{\infty} \pi_k$.

Define

$$Z_i^{k:n} = \sum_{t=k}^{n} \log \frac{f_{i,t}(X_t | \mathbf{X}^{1:t-1})}{f_{0,t}(X_t | \mathbf{X}^{1:t-1})} \tag{3.3}$$

and

$$\Delta_i(n) = \sum_{k=1}^{n} \frac{\pi_k}{\Omega_n} \exp\left(Z_i^{k:n} \right). \tag{3.4}$$

Then $\Delta(n)$ defined in (3.1) can be written as

$$\Delta(n) = \sum_{i=1}^{M} \omega_i \Delta_i(n) \tag{3.5}$$

With $\Delta(n)$ defined in (3.5), the proposed quickest change detection algorithm is a threshold-based sequential test given as follows.

Definition 3.1. (Bayesian Quickest Change Detection) For a given PFA upper bound α, the change point is detected as

$$\delta_1 : \hat{\theta}_1 = \inf \left\{ n \ge 1 : \Delta(n) \ge \frac{1-\alpha}{\alpha} \right\} \tag{3.6}$$

It should be noted that the proposed algorithm in (3.6) can be considered as an extension of the well-known Shiryaev procedure Shiryaev (1963), which only considers the case of one known post-change model. We will show next that the above algorithm is asymptotically optimal with respect to (P1).

3.2. Probability of False Alarm

We first study the PFA of the detection procedure defined in Definition 3.1.

Lemma 3.1. *For the quickest change detection algorithm in Definition 3.1, the probability of false alarm is upper bounded by α.*

Proof. Let $p(n) = \mathbf{P}(\mathcal{H}_1|\mathcal{F}_n^X) = \mathbf{P}(\theta \leq n|\mathcal{F}_n^X)$. From (3.1), we have $\Delta(n) = \dfrac{p_n}{1-p_n}$, or equivalently,

$$p(n) = \frac{\Delta(n)}{\Delta(n)+1} = 1 - \frac{1}{\Delta(n)+1} \tag{3.7}$$

It is apparent that p_n is an increasing function in $\Delta(n)$. From (3.6), we have

$$\Delta(\hat{\theta}_1) \geq \frac{1-\alpha}{\alpha} \Longrightarrow p(\hat{\theta}_1) \geq 1-\alpha. \tag{3.8}$$

From (2.1) and the definition of $p(n)$, the PFA can be calculated as

$$\text{PFA}(\delta_1) = \mathbf{P}(\hat{\theta}_1 < \theta|\mathcal{F}_n^X) = 1 - p(\hat{\theta}_1) \tag{3.9}$$

Combining (3.9) with (3.8) completes the proof. □

3.3. Average Detection Delay

In the Bayesian setting, the ADD defined in (2.2) can be computed as follows

$$\text{ADD}(\delta) = \frac{\mathbb{E}(\hat{\theta}-\theta)^+}{\mathbf{P}(\hat{\theta} \geq \theta)} = \frac{1}{\mathbf{P}(\hat{\theta} \geq \theta)} \sum\nolimits_{k=1}^{\infty} \pi_k \, \mathbf{P}_k(\hat{\theta} \geq k) \, \mathbb{E}_k(\hat{\theta} - k|\hat{\theta} \geq k) \tag{3.10}$$

where $x^+ = \max(0, x)$.

To facilitate the ADD analysis, it is assumed that $\frac{1}{n} Z_i^{k:k+n}$ almost surely converges in probability $\mathbf{P}_i$ to a positive finite number D_i (Tartakovsky and Veeravalli, 2005), that is,

$$\frac{1}{n} Z_i^{k:k+n-1} \xrightarrow[n\to\infty]{\mathbf{P}_i-a.s.} D_i \quad \forall\, k < \infty \tag{3.11}$$

In the case of identically and independently distributed (i.i.d.) data models, we have $f_{i,t}(X_t|\mathbf{X}^{1:t-1}) = f_i(X_t)$, and $D_i = \mathbb{E}\left[\log \frac{f_i(X)}{f_0(X)}\right]$ is the Kullback-Leibler (KL) divergence between $f_i(X)$ and $f_0(X)$.

The following asymptotic notations are used in the analysis. Consider two continuous functions $f(x)$ and $g(x)$ where $\lim_{x\to x_0} f(x) = \lim_{x\to x_0} g(x) = \infty$. We have the following notations.

$$f(x) \underset{x\to x_0}{\preceq} g(x) \Longleftrightarrow \lim_{x\to x_0} \frac{f(x)}{g(x)} \leq 1 \tag{3.12}$$

If both $f(x) \underset{x\to x_0}{\preceq} g(x)$ and $g(x) \underset{x\to x_0}{\preceq} f(x)$, then the two functions are called asymptotically equivalent as $x \to x_0$, and it is denoted as

$$f(x) \underset{x\to x_0}{\asymp} g(x) \Longleftrightarrow \lim_{x\to x_0} \frac{f(x)}{g(x)} = 1 \tag{3.13}$$

Theorem 3.1. *Assume the condition (3.11) holds and $\pi_k = (1-\rho)^{k-1}\rho$. As the PFA upper bound $\alpha \to 0$, we have*

$$\mathbb{E}_k[(\hat{\theta}_1 - k)^+] \underset{\alpha\to 0}{\preceq} \min_i \left[\frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log \omega_i}{D_i + |\log(1-\rho)|}\right]$$

Proof. To facilitate analysis, we first introduce a new stopping time with respect to each individual post-change distribution model as follows

$$\hat{\theta}_{1,i} = \inf\left\{n \geq 1 : \omega_i \Delta_i(n) \geq \frac{1-\alpha}{\alpha}\right\} \tag{3.14}$$

From the definition of $\Delta(n)$ in (3.5), it is evident that $\Delta(n) \geq \omega_i \Delta_i(n)$ for all i. Consequently,

$$\hat{\theta}_1 \leq \min_{i=1,2,\ldots,M} \hat{\theta}_{1,i} \tag{3.15}$$

The stopping time in (3.14) can be alternatively represented by

$$\hat{\theta}_{1,i} = \inf\left\{n \geq 1 : \log \Delta_i(n) \geq \log\left(\frac{1-\alpha}{\alpha}\right) - \log \omega_i\right\} \tag{3.16}$$

Based on the definition of $\Delta_i(n)$ in (3.4), it is easy to show that

$$\log \Delta_i(n) \geq Z_i^{k:n} + \log\left(\frac{\pi_k}{\Omega_n}\right) \triangleq V_i^{k:n} \tag{3.17}$$

Define a new stopping time

$$\zeta_{1,i} = \inf\left\{n \geq 1 : V_i^{k:n} \geq \log\left(\frac{1-\alpha}{\alpha}\right) - \log \omega_i\right\} \tag{3.18}$$

From (3.16)-(3.18), it is apparent that $\hat{\theta}_{1,i} \leq \zeta_{1,i}$, thus

$$\hat{\theta}_1 \leq \min_{i=1,2,\ldots,M} \hat{\theta}_{1,i} \leq \min_{i=1,2,\ldots,M} \zeta_{1,i} \tag{3.19}$$

For geometric pirors, we have

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\frac{\pi_k}{\Omega_{k+n-1}}\right) = |\log(1-\rho)|. \tag{3.20}$$

Combining (3.11) with (3.20) yields

$$\frac{1}{n} V_i^{k:k+n-1} \xrightarrow[n\to\infty]{\mathbf{P}_i-a.s.} D_i + |\log(1-\rho)| \triangleq q_i. \tag{3.21}$$

Define

$$T_k = \sup\left\{n \geq 1 : \left|\frac{1}{n} V_i^{k:k+n-1} - q_i\right| > \epsilon\right\}. \tag{3.22}$$

If $\zeta_{1,i} - k > T_k$, then from (3.22) we have

$$\left|\frac{1}{\zeta_{1,i} - k} V_i^{k:\zeta_{1,i}-1} - q_i\right| \leq \epsilon, \text{ if } \zeta_{1,i} - k > T_k \tag{3.23}$$

which implies

$$\zeta_{1,i} - k \leq \frac{V_i^{k:\zeta_{1,i}-1}}{q_i - \epsilon}, \text{ if } \zeta_{1,i} - k > T_k \tag{3.24}$$

Based on the definition of $\zeta_{1,i}$ in (3.18), we have

$$V_i^{k:\zeta_{1,i}-1} < \log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i \tag{3.25}$$

Combining (3.24) and (3.25) results in

$$\zeta_{1,i} - k \le \frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i}{q_i - \epsilon}, \quad \text{if } \zeta_{1,i} - k > T_k \tag{3.26}$$

When $\alpha < 0.5$ and $\epsilon < q_i$, we always have $\frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i}{q_i - \epsilon} > 0$. Therefore the following inequality is true for both $\zeta_{1,i} - k > T_k$ and $\zeta_{1,i} - k \le T_k$

$$\zeta_{1,i} - k \le \frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i}{q_i - \epsilon} + T_k, \quad \text{if } \alpha < 0.5 \text{ and } \epsilon < q_i \tag{3.27}$$

Given the convergence condition in (3.21), we have $\mathbb{E}(T_k) < \infty$. Since ϵ can be arbitrarily small, we can let $\epsilon \to 0$. Thus when $\alpha \to 0$,

$$\mathbb{E}[\zeta_{1,i} - k] \underset{\alpha\to 0}{\preceq} \frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i}{D_i + |\log(1-\rho)|}. \tag{3.28}$$

Since $\hat{\theta}_1$ is a lower bound of $\zeta_{1,i}$ as in (3.19), we have

$$\mathbb{E}[\hat{\theta}_1 - k] \underset{\alpha\to 0}{\preceq} \min_i \left[\frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i}{D_i + |\log(1-\rho)|}\right]. \tag{3.29}$$

When $\alpha \to 0$, the right hand side of (3.29) is always positive. Thus the inequality in (3.29) still holds if we replace $\hat{\theta}_1 - k$ by $(\hat{\theta}_1 - k)^+$. This completes the proof. □

From the results in Theorem 3.1, it can be seen the prior probability ω_i and the constant D_i plays an important role in determining the ADD upper bound. If ω_i is very small, that is, the i-th post-change model is very unlikely, the value of $-\log\omega_i$ will be very large, and it will not affect the delay upper bound because the minimum is performed over all M post-change models. Similarly, if D_i is very small, that is, the difference between the i-th post-change model and the pre-change model is small, then the minimum operator will exclude its impact on the delay upper bound. Consequently, the delay upper bound is dominated by the post-change models that have large ω_i and/or large D_i, that is, those models that are likely to appear and have a big difference with the pre-change model.

In addition to the asymptotic upper bound in Theorem 3.1, we also have the asymptotic lower bound for the detection delay.

Theorem 3.2. *Assume the condition (3.11) holds and $\pi_k = (1-\rho)^{k-1}\rho$. As the PFA upper bound $\alpha \to 0$, we have*

$$\mathbb{E}_k[(\hat{\theta}_1 - k)^+] \underset{\alpha\to 0}{\succeq} \min_i \left[\frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i}{D_i + |\log(1-\rho)|}\right]$$

Proof. To simplify notation, define

$$L_{i\alpha} = \frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i}{D_i + |\log(1-\rho)|}, \tag{3.30}$$

$$L_\alpha = \min_i \; L_{i\alpha} \tag{3.31}$$

Also, define the following events and their probabilities,

$$\mathcal{C}_{i,k} : \{k \leq \hat{\theta}_1 \leq k + (1-\epsilon)L_{i\alpha}\} \text{ and } \gamma_{i,k}^{\epsilon}(\hat{\theta}_1) = \mathbf{P}_k\{\mathcal{C}_{i,k}\}$$
$$\mathcal{C}_{k} : \{k \leq \hat{\theta}_1 \leq k + (1-\epsilon)L_{\alpha}\} \text{ and } \gamma_{k}^{\epsilon}(\hat{\theta}_1) = \mathbf{P}_k\{\mathcal{C}_{k}\}$$

where $0 \leq \epsilon < 1$ is a constant.

Since $L_\alpha = \min_i \ L_{i\alpha}$, it can be easily shown that $\mathcal{C}_k = \cap_i \mathcal{C}_{i,k}$. Thus

$$\gamma_k^{\epsilon}(\hat{\theta}_1) = \mathbf{P}_k\{\mathcal{C}_{1,k} \cap \mathcal{C}_{2,k} \cap \ldots \cap \mathcal{C}_{M,k}\} \leq \min_i \gamma_{i,k}^{\epsilon}(\hat{\theta}_1) \tag{3.32}$$

where the last inequality follows from the fact that for any events Y and Z, $\mathbf{P}(Y \cap Z) = \mathbf{P}(Y) \cdot \mathbf{P}(Z|Y) \leq \mathbf{P}(Y)$.

Given the convergence condition in (3.11), it can be proven that

$$\lim_{\alpha \to 0} \gamma_{i,k}^{\epsilon}(\hat{\theta}_1) = 0 \quad \forall\, i = 1, \ldots, M,\ 0 < \epsilon < 1 \text{ and } k \geq 1 \tag{3.33}$$

The above equation can be proved by following a similar procedure for the proof of equation (3.31) in (Tartakovsky and Veeravalli, 2005, Lemma 2). Combining (3.32) with (3.33) yields

$$\lim_{\alpha \to 0} \gamma_{k}^{\epsilon}(\hat{\theta}_1) = 0 \quad \forall\, i = 1, \ldots, M,\ 0 < \epsilon < 1 \text{ and } k \geq 1 \tag{3.34}$$

Based on the Chebyshev inequality, for any $0 \leq \epsilon < 1$, we have

$$\begin{aligned} \mathbb{E}_k[(\hat{\theta}_1 - k)^+] &\geq [(1-\epsilon)L_\alpha]\, \mathbf{P}_k\{(\hat{\theta}_1 - k)^+ \geq (1-\epsilon)L_\alpha\} \\ &= [(1-\epsilon)L_\alpha]\, \mathbf{P}_k\{(\hat{\theta}_1 - k) \geq (1-\epsilon)L_\alpha\} \end{aligned} \tag{3.35}$$

where the last equality is based on the fact that the event $\{X^+ \geq A\}$ is true if and only if $\{X \geq A\}$ is true for $A > 0$, and $(1-\epsilon)L_\alpha$ is positive when $0 \leq \epsilon < 1$.

It is also evident that,

$$\mathbf{P}_k\{\hat{\theta}_1 - k \geq (1-\epsilon)L_\alpha\} \geq \mathbf{P}_k\{\hat{\theta}_1 \geq k\} - \gamma_k^{\epsilon}(\hat{\theta}_1) \tag{3.36}$$

which is based on the fact that for two events Y and Z, $\mathbf{P}(Y \cup Z) \leq \mathbf{P}(Y) + \mathbf{P}(Z)$.

Based on Lemma 3.1, the PFA is upper bounded by α. Thus

$$\alpha \geq \text{PFA}(\hat{\theta}_1) = \sum\nolimits_{i=1}^{\infty} \pi_i \mathbf{P}_i(\hat{\theta}_1 < i) \geq \pi_k \mathbf{P}_k(\hat{\theta}_1 < k)$$

Thus $\mathbf{P}_k(\hat{\theta}_1 < k) \leq \pi_k^{-1}\alpha$, or equivalently,

$$\mathbf{P}_k\{\hat{\theta}_1 \geq k\} = 1 - \mathbf{P}_k\{\hat{\theta}_1 < k\} \geq 1 - \pi_k^{-1}\alpha \tag{3.37}$$

Combining (3.35), (3.36) and (3.37), we get,

$$\mathbb{E}_k[(\hat{\theta}_1 - k)^+] \geq [(1-\epsilon)L_\alpha]\, [1 - \pi_k^{-1}\alpha - \gamma_k^{\epsilon}(\hat{\theta}_1)].$$

Since ϵ can be arbitrarily small, we can let $\epsilon \to 0$, then

$$\lim_{\alpha \to 0} \frac{\mathbb{E}_k[(\hat{\theta}_1 - k)^+]}{L_\alpha} \geq \lim_{\alpha \to 0} [1 - \pi_k^{-1}\alpha - \gamma_k^{\epsilon}(\hat{\theta}_1)] = 1$$

where the last equality is based on the fact that $\gamma_k^{\epsilon}(\hat{\theta}_1) \to 0$ as $\alpha \to 0$ as given in (3.34). This completes the proof. $\square$

The asymptotic lower bound in Theorem 3.2 is the same as the asymptotic upper bound in Theorem 3.1. The asymptotic convergence between the lower bound and upper bound indicates that the detection method in Definition 3.1 is asymptotically optimal. That is, the algorithm can asymptotically achieve the minimum detection delay because the asymptotic lower bound is also the asymptotic upper bound.

Theorem 3.3. *Assume the condition (3.11) holds and $\pi_k = (1-\rho)^{k-1}\rho$. As the PFA upper bound $\alpha \to 0$, the quickest change detection presented in Definition 3.1 is asymptotically optimal with respect to (P1). The asymptotic ADD is*

$$ADD(\delta_1) \underset{\alpha\to 0}{\asymp} \min_i \left[\frac{\log\left(\frac{1-\alpha}{\alpha}\right) - \log\omega_i}{D_i + |\log(1-\rho)|}\right].$$

Proof. The results can be directly obtained by combining Theorems 3.1 and 3.2. □

4. QUICKEST CHANGE DETECTION ALGORITHM FOR NON-BAYESIAN SETTING

In this section, we develop the algorithm that can detect the change point with minimum delay under a non-Bayesian setting. Under a non-Bayesian setting, the prior probabilities of the change point, π_k, for $k = 1, 2, \cdots$ and the prior probabilities of the post-change model, ω_i, for $i = 1, \cdots, M$, are all unknown.

4.1. Detection Algorithm

Define

$$\Lambda(n) = \sum\nolimits_{i=1}^{M} \Lambda_i(n) \tag{4.1}$$

where,

$$\Lambda_i(n) = \sum\nolimits_{k=1}^{n} \exp\left(Z_i^{k:n}\right). \tag{4.2}$$

With $\Lambda(n)$ defined in (4.1), the proposed quickest change detection algorithm under non-Bayesian setting is a threshold-based sequential test given as follows.

Definition 4.1. (Non-Bayesian Quickest Change Detection) For a given PFA upper bound α, the change point is detected as

$$\delta_2 : \hat{\theta}_2 = \inf\left\{n \geq 1 : \Lambda(n) \geq \frac{M\overline{\theta}}{\alpha}\right\} \tag{4.3}$$

where

$$\overline{\theta} = \sum\nolimits_{k=1}^{\infty} k\pi_k \tag{4.4}$$

is the prior mean of the change point.

Note that the statistic $\Lambda_i(n)$ is the Shiryaev-Roberts (SR) statistic (Roberts, 1966) for detecting a change corresponding to the i-th post-change model. The detection procedure δ_2 is therefore an extension of the SR procedure adapted to detect changes in system with multiple post-change models (Tartakovsky and Veeravalli, 2004). The non-Bayesian detection algorithm in (4.3) has a similar flavor as the algorithm proposed in Mei (2010). The test statistic in (Mei, 2010, Equation. (9)) is the sum of the CUSUM statistics for each post-change hypothesis. In (4.3), the test statistic is the sum of the SR statistics for each post-change hypothesis. Next, we will show that the proposed detection method in Definition 4.1 is asymptotically optimum with respect to (P1).

4.2. Probability of False Alarm

We first study the PFA of the detection procedure defined in Definition 4.1.

Lemma 4.1. *For the quickest change detection algorithm in Definition 4.1, the probability of false alarm is upper bounded by α.*

Proof. The statistic $\Lambda_i(n)$ defined in (4.2) can be written in a recursive form as

$$\Lambda_i(n+1) = \lambda_i(n+1)\left[1 + \Lambda_i(n)\right] \tag{4.5}$$

where $\lambda_i(n)$ is the likelihood ratior (LR) of the i-th post-change model at time n

$$\lambda_i(n) = \frac{f_{i,n}(X_n|\mathbf{X}^{1:n-1})}{f_{0,n}(X_n|\mathbf{X}^{1:n-1})} \tag{4.6}$$

It is straightforward that $\mathbb{E}_\infty[\lambda_i(n)] = 1$, and from (4.5)

$$\mathbb{E}_\infty[\Lambda_i(n+1)|\mathcal{F}_n^X] = 1 + \Lambda_i(n). \tag{4.7}$$

Thus $\Lambda_i(n)$ is a submartingale with respect to the probability measure $\mathbf{P}_\infty$. In addition, since $\mathbb{E}[\Lambda_i(1)] = \mathbb{E}[\lambda_i(1)] = 1$, we have $\mathbb{E}_\infty[\Lambda_i(n)] = n$.

Combining the definition of $\Lambda(n)$ in (4.1) with (4.7), we get

$$\mathbb{E}_\infty\left[\Lambda(n+1)|\mathcal{F}_n^X\right] = M + \Lambda(n). \tag{4.8}$$

Thus $\Lambda(n)$ is also a submartingale with respect to $\mathbf{P}_\infty$ and $\mathbb{E}_\infty\left[\Lambda(n)\right] = Mn$. Using Doob's submartingale inequality, we get

$$\mathbf{P}_\infty\{\hat{\theta}_2 < n\} = \mathbf{P}_\infty\left\{\max_{1\le k\le n}\Lambda(k) \ge \frac{M\overline{\theta}}{\alpha}\right\} \le \frac{n\alpha}{\overline{\theta}}$$

Therefore,

$$\text{PFA}(\delta_2) = \sum_{k=1}^{\infty}\pi_k\mathbf{P}_\infty\{\hat{\theta}_2 < k\} \le \sum_{k=1}^{\infty}\frac{\pi_k k\alpha}{\overline{\theta}} = \alpha.$$

□

4.3. False Alarm Rate

For a given test δ, the FAR is defined as

$$\text{FAR}(\delta) = \frac{1}{\mathbb{E}_\infty(\hat{\theta})} \tag{4.9}$$

where $\mathbb{E}_\infty(\hat{\theta})$ is known as the ARL to false alarm.

Lemma 4.2. *For the quickest change detection algorithm in Definition 4.1, the false alarm rate is upper bounded by $\alpha/\overline{\theta}$.*

Proof. From (4.8), it is obvious that $\Lambda(n) - Mn$ forms a Martingale with respect to $\mathbf{P}_\infty$.

If $\mathbb{E}_\infty[\hat{\theta}_2] = \infty$, then $\text{FAR}(\delta_2) = 0$ and it is bounded by $\alpha/\bar{\theta}$.

If $\mathbb{E}_\infty[\hat{\theta}_2] < \infty$, then based on the optional stopping theorem, we have

$$\mathbb{E}_\infty[\Lambda(\hat{\theta}_2) - M\hat{\theta}_2] = \mathbb{E}_\infty[\Lambda(1) - M] = 0. \tag{4.10}$$

Thus

$$\mathbb{E}_\infty[\hat{\theta}_2] = \frac{1}{M}\mathbb{E}_\infty[\Lambda(\hat{\theta}_2)] \tag{4.11}$$

Combining (4.11) with (4.3) yields

$$\mathbb{E}_\infty[\hat{\theta}_2] \geq \frac{\bar{\theta}}{\alpha} \tag{4.12}$$

which implies $\text{FAR}(\delta_2) \leq \alpha/\bar{\theta}$. □

4.4. Average Detection Delay

The asymptotic upper and lower bounds for the ADD of the detection method in Definition 4.1 are derived respectively in a similar manner as in Subsection 3.3.

Theorem 4.1. *Assuming the condition (3.11) holds, as the PFA upper bound $\alpha \to 0$, we have*

$$\mathbb{E}_k[(\hat{\theta}_2 - k)^+] \underset{\alpha \to 0}{\preceq} \min_i \frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i}$$

Proof. The proof follows a similar procedure as the proof of Theorem 3.1. First define a new stopping time with respect to each individual post-change distribution model as follows

$$\zeta_{2,i} = \inf\left\{n \geq 1 : Z_i^{k:n} \geq \log\left(\frac{M\bar{\theta}}{\alpha}\right)\right\} \tag{4.13}$$

From (4.1) and (4.2), it is straightforward that $\log\Lambda(n) \geq \log\Lambda_i(n) \geq Z_i^{k:n}$. Thus

$$\hat{\theta}_2 \leq \min_{i=1,2,\ldots,M} \zeta_{2,i}. \tag{4.14}$$

Given the fact that $\frac{1}{n}Z_i^{k:k+n-1}$ almost surely converges to D_i in probability $\mathbf{P}_i$ as $n \to \infty$ as in (3.11), we can define

$$\Gamma_k = \sup\left\{n \geq 1 : \left|\frac{1}{n}Z_i^{k:k+n-1} - D_i\right| > \epsilon\right\}. \tag{4.15}$$

If $\zeta_{2,i} - k > \Gamma_k$, then from (4.15) we have

$$\left|\frac{1}{\zeta_{2,i} - k}Z_i^{k:\zeta_{2,i}-1} - D_i\right| \leq \epsilon, \ \text{ if } \zeta_{2,i} - k > \Gamma_k \tag{4.16}$$

which implies

$$\zeta_{2,i} - k \leq \frac{Z_i^{k:\zeta_{2,i}-1}}{D_i - \epsilon}, \ \text{ if } \zeta_{2,i} - k > \Gamma_k \tag{4.17}$$

From (4.13), we have

$$Z_i^{k:\zeta_{2,i}-1} < \log\left(\frac{M\bar{\theta}}{\alpha}\right) \tag{4.18}$$

Combining (4.17) and (4.18) results in

$$\zeta_{2,i} - k \le \frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i - \epsilon}, \text{ if } \zeta_{2,i} - k > T_k \tag{4.19}$$

When $\alpha < M\bar{\theta}$ and $\epsilon < D_i$, we always have $\frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i-\epsilon} > 0$. Therefore the following inequality is true for both $\zeta_{2,i} - k > \Gamma_k$ and $\zeta_{2,i} - k \le \Gamma_k$

$$\zeta_{2,i} - k \le \frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i - \epsilon} + \Gamma_k, \text{ if } \alpha < M\bar{\theta} \text{ and } \epsilon < D_i \tag{4.20}$$

Given the convergence condition in (3.11), we have $\mathbb{E}(\Gamma_k) < \infty$. Setting $\epsilon \to 0$ and $\alpha \to 0$, we have

$$\mathbb{E}[\zeta_{2,i} - k] \underset{\alpha\to 0}{\preceq} \frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i}. \tag{4.21}$$

Since $\hat{\theta}_2$ is a lower bound of $\zeta_{2,i}$ as in (3.19), we have

$$\mathbb{E}[\hat{\theta}_2 - k] \underset{\alpha\to 0}{\preceq} \min_i \frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i}. \tag{4.22}$$

When $\alpha \to 0$, the right hand side of (3.29) is always positive. Thus the inequality in (3.29) still holds if we replace $\hat{\theta}_2 - k$ by $(\hat{\theta}_2 - k)^+$. This completes the proof. $\square$

From the results in Theorem 4.1, it can be seen that the delay upper bound is dominated only by the post-change models that have large D_i, that is, those models that have a big difference with the pre-change model.

Theorem 4.2. *Assuming the condition (3.11) holds, as the PFA upper bound $\alpha \to 0$, we have*

$$\mathbb{E}_k[(\hat{\theta}_2 - k)^+] \underset{\alpha\to 0}{\succeq} \min_i \frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i}$$

Proof. The proof follows a similar procedure as the proof of Theorem 3.2. First define

$$Y_{i\alpha} = \frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i}, \text{ and } Y_\alpha = \min_i Y_{i\alpha} \tag{4.23}$$

Define the following events and their respective probabilities,

$$\mathcal{D}_{i,k} : \{k \le \hat{\theta}_2 \le k + (1-\epsilon)Y_{i\alpha}\} \text{ and } \phi_{i,k}^{\epsilon}(\hat{\theta}_2) = \mathbf{P}_k\{\mathcal{D}_{i,k}\}$$
$$\mathcal{D}_k : \{k \le \hat{\theta}_2 \le k + (1-\epsilon)Y_\alpha\} \text{ and } \phi_k^{\epsilon}(\hat{\theta}_2) = \mathbf{P}_k\{\mathcal{D}_k\}$$

where $0 \le \epsilon < 1$ is a constant.

Since $Y_\alpha = \min_i \ Y_{i\alpha}$, we have $\mathcal{D}_k = \cap_i \ \mathcal{D}_{i,k}$, which implies

$$\phi_k^\epsilon(\hat{\theta}_2) \leq \min_i \phi_{i,k}^\epsilon(\hat{\theta}_2). \tag{4.24}$$

Similar to (4.25), it can be shown that

$$\lim_{\alpha \to 0} \phi_{i,k}^\epsilon(\hat{\theta}_2) = 0, \ \forall i = 1, \ldots, M, \ 0 < \epsilon < 1 \text{ and } k \geq 1. \tag{4.25}$$

Combining (4.24) with (4.25) yields

$$\lim_{\alpha \to 0} \phi_k^\epsilon(\hat{\theta}_2) = 0, \ \forall i = 1, \ldots, M, \ 0 < \epsilon < 1 \text{ and } k \geq 1 \tag{4.26}$$

Based on the Chebyshev inequality, for any $0 \leq \epsilon < 1$, we have

$$\mathbb{E}_k[(\hat{\theta}_2 - k)^+] \geq [(1-\epsilon)Y_\alpha] \, \mathbf{P}_k\{(\hat{\theta}_2 - k) \geq (1-\epsilon)Y_\alpha\} \tag{4.27}$$

Since $\{\hat{\theta}_2 \geq k\} = \{\hat{\theta}_2 - k \geq (1-\epsilon)Y_\alpha\} \cup \mathcal{D}_k$, we have

$$\mathbf{P}_k\{\hat{\theta}_2 - k \geq (1-\epsilon)Y_\alpha\} \geq \mathbf{P}_k\{\hat{\theta}_2 \geq k\} - \phi_k^\epsilon(\hat{\theta}_2). \tag{4.28}$$

Based on Lemma 4.1, the PFA is upper bounded by α. Thus

$$\alpha \geq \text{PFA}(\hat{\theta}_2) = \sum_{i=1}^{\infty} \pi_i \mathbf{P}_i(\hat{\theta}_2 < i) \geq \pi_k \mathbf{P}_k(\hat{\theta}_2 < k),$$

and consequently,

$$\mathbf{P}_k\{\hat{\theta}_2 \geq k\} = 1 - \mathbf{P}_k\{\hat{\theta}_2 < k\} \geq 1 - \pi_k^{-1}\alpha \tag{4.29}$$

Combining (4.27), (4.28), and (4.29), we get,

$$\mathbb{E}_k[(\hat{\theta}_2 - k)^+] \geq [(1-\epsilon)Y_\alpha] \, [1 - \pi_k^{-1}\alpha - \phi_k^\epsilon(\hat{\theta}_2)].$$

Since ϵ can be arbitrarily small, we can let $\epsilon \to 0$, then

$$\lim_{\alpha \to 0} \frac{\mathbb{E}_k[(\hat{\theta}_2 - k)^+]}{Y_\alpha} \geq 1$$

This completes the proof. □

The asymptotic lower bound in Theorem 4.2 is the same as the asymptotic upper bound in Theorem 4.1. The asymptotic convergence between the lower bound and upper bound indicates that the detection method in Definition 4.1 is asymptotically optimal.

Theorem 4.3. *Assuming the condition (3.11) holds, as the PFA upper bound $\alpha \to 0$, the quickest change detection presented in Definition 4.1 is asymptotically optimal with respect to (P1). The asymptotic ADD is*

$$ADD(\delta_2) \underset{\alpha \to 0}{\asymp} \min_i \frac{\log\left(\frac{M\bar{\theta}}{\alpha}\right)}{D_i}$$

Proof. The results can be directly obtained by combining Theorems 4.1 and 4.2. □

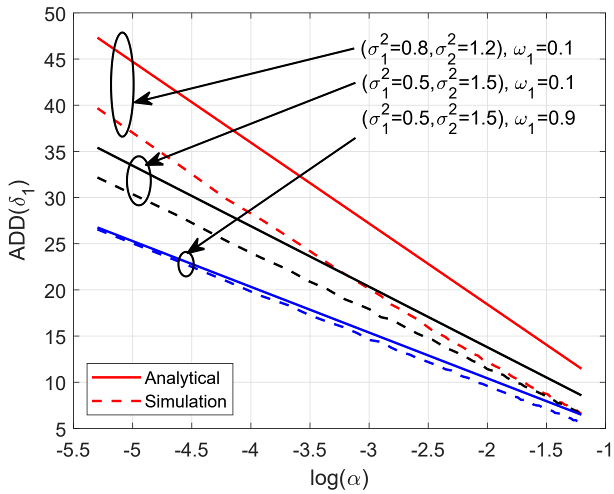

Figure 1. Average detection delay of the proposed Bayesian algorithm.

5. NUMERICAL RESULTS

Numerical results are presented in this section to demonstrate the performance of the proposed change detection algorithms under both Bayesian and non-Bayesian settings. All simulation results are obtained by averaging over 10,000 Monte-Carlo trials. The change point follows a geometric distribution with $\rho = 0.1$ in all simulations.

We first study the performance of the algorithm in Definition 3.1 under the Bayesian setting. The algorithm utilizes the prior probabilities of change points as well as the prior probabilities of post-change models. In the first example, we consider $M = 2$ possible post-change models. The pre-change and post-change distributions are zero-mean Gaussian distributions with variance σ_i^2, that is, $f_i \sim \mathcal{N}(0, \sigma_i^2)$. We have $\sigma_0^2 = 1$ for the pre-change distribution, and we will consider different combinations of the post-change parameters (σ_1^2, σ_2^2). Figure 1 shows the average detection delay, $\text{ADD}(\delta_1)$, as a function of the PFA upper bound α under various combinations of (σ_1^2, σ_2^2) and model prior probability ω_1. When $(\sigma_1^2 = 0.5, \sigma_2^2 = 1.5)$, we have $D_1 = 0.0966$ and $D_2 = 0.0473$. When $(\sigma_1^2 = 0.8, \sigma_2^2 = 1.2)$, we have $D_1 = 0.0116$ and $D_2 = 0.0088$. Under all configurations, the asymptotic analytical ADDs have the same slopes as their simulation counterparts. Thus the asymptotic results provide very good predictions regarding the trend of the detection delay. The performance difference of the three cases becomes smaller as the PFA increases.

Figure 2 shows the PFA of the proposed Bayesian detection algorithm as a function of the PFA upper bound α. There are $M = 2$ post-change models. The pre- and post-change data follow exponential distributions with the parameters λ_i, for $i = 0, 1, 2$. We have $\lambda_0 = 1$ for the pre-change distribution, and $\lambda_1 = 0.5$ and $\lambda_2 = 1.5$ for the post-change distributions. It can be clearly observed that PFA obtained from numerical simulations is always below its upper bound as proved in Lemma 3.1. The analytical upperbound has the same trend as the simulated PFA under all system configurations.

Figure 3 compares the ADD of the proposed Bayesian algorithm with an adaptation of the well-known Shiryaev procedure. The adapted procedure exploits the mixture post-change distribution, which is obtained as $h(x) = \sum_i \omega_i f_i(x)$ (Lai, 1998). The Shiryaev procedure is then employed by using f_0 and h as the pre-change and post-change models, respectively. In this example, there are four post-change models, $f_i \sim \mathcal{N}(\mu_i, 1)$, with $\mu_1 = 0.6$, $\mu_2 = 0.8$, $\mu_3 = 1.2$, and $\mu_4 = 1.4$. The prior probabilities of these models are $\omega_1 = 0.1$, $\omega_2 = 0.2$, $\omega_3 = 0.3$, and $\omega_4 = 0.4$, respectively. The pre-change model follows the distribution $f_0 \sim \mathcal{N}(1, 1)$. The proposed algorithm outperforms the Shiryaev-Mixture algorithm under the entire range

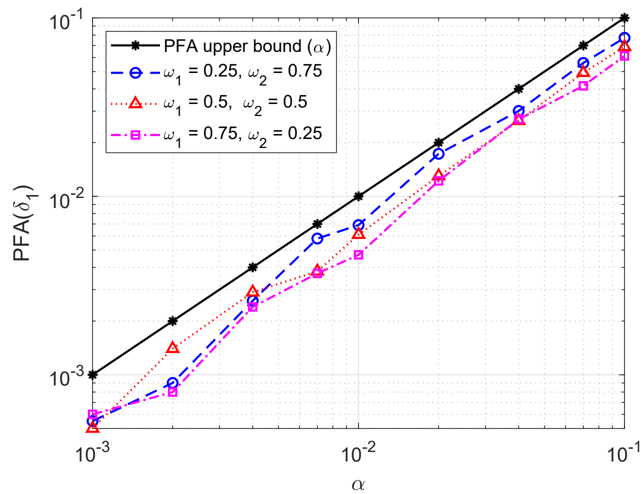

Figure 2. Probability of false alarm for the proposed Bayesian algorithm.

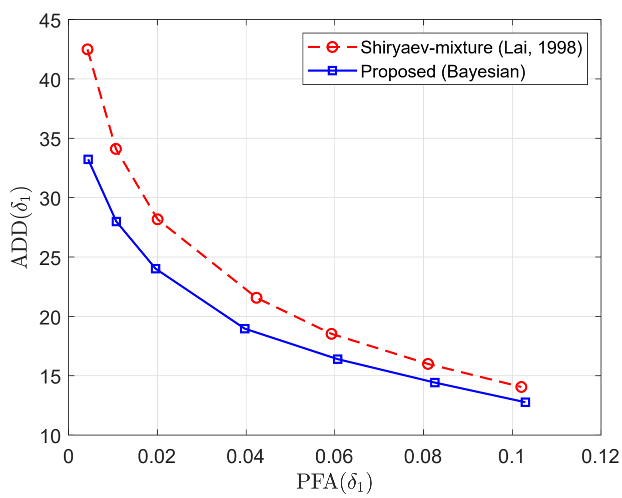

Figure 3. Comparison of the proposed Bayesian algorithm with adapted Shiryaev algorithms.

of PFA. At PFA $= 0.02$, the ADDs of the Shiryaev-Mixture and the proposed algorithms are 28 and 24, which corresponds to an improvement of 14% over the Shiryaev-Mixture algorithm.

Next, we study the performance of the algorithm in Definition 4.1 under the non-Bayesian setting. There are $M = 2$ equiprobable post-change models, i.e., $\omega_1 = \omega_2 = 0.5$. The pre-change and post-change distributions are zero-mean Gaussian distributions with variance σ_i^2. We have $\sigma_0^2 = 1$ for the pre-change distribution, and we will consider different combinations of the post-change parameters (σ_1^2, σ_2^2). Figure 4 shows ADD(δ_2) as a function of the PFA upper bound α. When $(\sigma_1^2 = 0.6, \sigma_2^2 = 1.4)$, we have $D_1 = 0.0554$ and $D_2 = 0.0318$. When $(\sigma_1^2 = 0.55, \sigma_2^2 = 1.45)$, we have $D_1 = 0.0739$ and $D_2 = 0.0392$. When $(\sigma_1^2 = 0.5, \sigma_2^2 = 1.5)$, we have $D_1 = 0.0966$ and $D_2 = 0.0473$. Similar to the Bayesian case, the asymptotic analytical ADDs for non-Bayesian method have similar slopes as their simulation counterparts under different configurations. Thus the asymptotic results are very good predictors for the trend of the detection delay.

Figure 5 shows the PFA of the non-Bayesian algorithm as a function of the PFA upper bound α under

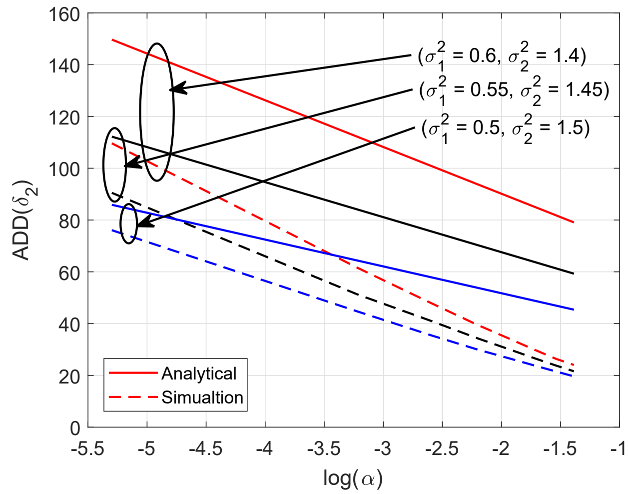

Figure 4. Average detection delay of the proposed non-Bayesian algorithm.

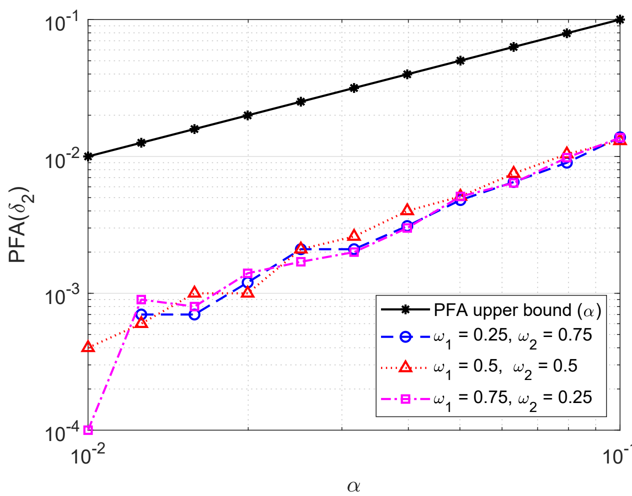

Figure 5. Probability of false alarm for the proposed non-Bayesian algorithm.

different system configurations. There are $M = 2$ post-change models. The pre-change and post-change data follow Gaussian distributions with unit variance and mean μ_i. We have $\mu_0 = 1$ for the pre-change distribution, and $\mu_1 = 0.5$ and $\mu_2 = 1.5$ for the post-change distributions. The simulated PFAs are always under the theoretical upper bounds as proved in Lemma 4.1. Under the non-Bayesian setting, the theoretical PFA upper bound is not as tight as its Bayesian counterpart. The PFA upper bound is about one order of magnitude higher than the results obtained from numerical simulations.

Figure 6 illustrates the performance of the non-Bayesian change detection algorithm described in Definition 4.1. In this example, there are $M = 3$ post-change models. The data follow a two-dimensional multivariate Gaussian distribution with zero-mean and covariance matrix

$$R = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}.$$

The coefficient r is set to 0 before the change point. After the change point, we set $r = 0.1, 0.5$, and 0.9 for

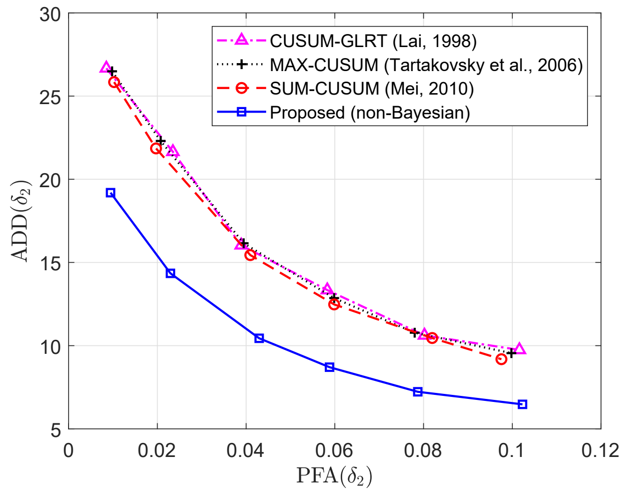

Figure 6. Comparison of the proposed non-Bayesian algorithm with adapted CUSUM algorithms.

the three post-change models, respectively. The prior probabilities of the post-change models are $\omega_1 = 0.1$, $\omega_2 = 0.3$, and $\omega_3 = 0.6$, respectively. Figure 6 compares the ADD of the proposed non-Bayesian algorithm with the ADD of three different non-Bayesian algorithms based on CUSUM procedure. The CUSUM-GLRT procedure (Lai, 1998) uses GLRT by estimating the unknown parameter, which corresponds to the post-change model in this example. The SUM-CUSUM procedure (Mei, 2010) exploits the sum of the local CUSUM statistics corresponding to the individual post-change models. The MAX-CUSUM procedure (Tartakovsky et al., 2006) uses the maximum of the local CUSUM statistics. The proposed non-Bayesian algorithm achieves significant performance gains over the existing CUSUM-based algorithms.

6. CONCLUSION

Quickest change point detection with multiple possible post-change models has been studied in this paper. We have proposed two quickest change detection algorithms under the Bayesian and non-Bayesian settings, respectively. Theoretical analysis has been performed to obtain the PFA upper bounds and asymptotic bounds on ADD when the PFA is small. It has been shown that both algorithms are asymptotically optimal in terms of average detection delay. Numerical results have shown that the proposed algorithms outperform existing algorithms in terms of average detection delay under the same PFA constraints.

FUNDING

The work was supported in part by the U.S. National Science Foundation (NSF) under Award Number ECCS-1711087.

REFERENCES

Akingeneye, I. and Wu, J. (2018). Low Latency Detection of Sparse False Data Injections in Smart Grids, *IEEE Access* 6: 58564–58573.

Aminikhanghahi, S. and Cook, D. J. (2017). A Survey of Methods for Time Series Change Point Detection, *Knowledge and Information Systems* 51: 339-367.

De Maio, A. (2007). Rao Test for Adaptive Detection in Gaussian Interference with Unknown Covariance Matrix, *IEEE Transactions on Signal Processing* 55: 3577–3584.

Gong, X. and Qiao, W. (2013). Bearing Fault Diagnosis for Direct-drive Wind Turbines via Current-demodulated Signals. *IEEE Transactions on Industrial Electronics* 60: 3419-3428.

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, *The Annals of Mathematical Statistics* 42: 1897-1908.

Mei, Y. (2010). Efficient Scalable Schemes for Monitoring a Large Number of Data Streams, *Biometrika* 97: 419-433.

Nath, S., Akingeneye, I, Wu, J., and Han, Z. (2019). Quickest Detection of False Data Injection Attacks in Smart Grid with Dynamic Models, *IEEE Journal of Emerging and Selected Topics in Power Electronics*. Early Access. https://doi.org/10.1109/JESTPE.2019.2936587.

Nath, S. and Wu, J. (2018). Bayesian Quickest Change Point Detection with Multiple Candidates of Post-change Models, *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, November 26-29, 2018, pp. 51-55, Anaheim, CA, USA.

Page, E. S. (1954). Continuous Inspection Schemes, *Biometrika* 41: 100–115.

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

Poor, H. V. and Hadjiliadis, O. (2009). *Quickest Detection*, Cambridge: Cambridge University Press.

Raghavan, V. and Veeravalli, V. V. (2009). Bayesian Quickest Change Process Detection, *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, June 28- July 3, 2009, pp. 644–648, Seoul, South Korea.

Ren, X. and Shi, L. (2014). Bayesian Quickest Change Detection under Energy Constraints over Wireless Sensor Networks with Correlated Fading Channels, *Proceedings of 11th IEEE International Conference on Control & Automation (ICCA)*, June 18-20, 2014, pp. 1192-1197, Taichung, Taiwan.

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 of Probability and Its Applications* 8: 22-46.

Tartakovsky, A., Nikiforov, I., and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, Florida: Chapman and Hall/CRC.

Tartakovsky, A. G., Rozovskii, B. L., Blažek, R. B., and Kim, H. (2006). Detection of Intrusions in Information Systems by Sequential Change-point Methods. *Statistical Methodology* 3: 252–293.

Tartakovsky, A. G. and Veeravalli, V. V. (2004). Change-Point Detection in Multichannel and Distributed Systems with Applications, *Applied Sequential Methodologies: Real-World Examples with Data Analysiss*, N. Mukhopadhyay, S. Datta, and S. Chattopadhyay, eds., pp. 339-370, New York: Marcel Dekker.

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.

Veeravalli, V. V. and Banerjee, T. (2014). Quickest Change Detection, in *Academic Press Library in Signal Processing*, chapter 6, A. M. Zoubir, M. Viberg, R. Chellappa, and S. Theodoridis, eds., pp. 209-255, Amsterdam: Elsevier.