High Probability Latency Quickest Change Detection over a Finite Horizon

Yu-Han Huang and Venugopal V. Veeravalli
Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL, USA
{yuhanhh2,vvv}@illinois.edu

2024 IEEE International Symposium on Information Theory (ISIT) | 979-8-3503-8284-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISIT57864.2024.10619566

***Abstract*—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. A finite horizon variant of the quickest change detection problem is studied, in which the goal is to minimize a delay threshold (latency), under constraints on the probability of false alarm and the probability that the latency is exceeded. In addition, the horizon is not known to the change detector. A variant of the cumulative sum (CuSum) test with a threshold that increasing logarithmically with time is proposed as a candidate solution to the problem. An information-theoretic lower bound on the minimum value of the latency under the constraints is then developed. This lower bound is used to establish certain asymptotic optimality properties of the proposed test in terms of the horizon and the false alarm probability. Some experimental results are given to illustrate the performance of the test.**

I. Introduction

The problem of detecting changes or anomalies in stochastic systems and time series, often referred to as the quickest change detection (QCD) problem, arises in various engineering and scientific settings. The observations of the system are assumed to undergo a change in distribution at the change-point, and the goal is to detect this change as soon as possible, subject to false alarm constraints. See [1–4] for books and survey articles on the topic.

The QCD problem is formulated mathematically as a constrained optimization problem to minimize a measure of detection delay, subject to a constraint on an appropriate false alarm metric. The most commonly used false alarm metric is the *mean time to false alarm*, or its reciprocal, the *false alarm rate*. Detection delay is generally measured by considering the *expected detection delay*, conditioned on the change-point and possibly the history of observations before the change-point, and taking the supremum over all possible change-points over an infinite horizon. However, these metrics for false alarm and detection delay might not be appropriate in many applications. Our work is inspired by the regret analysis for piecewise stationary bandits in [5], in which the metrics used for detecting changes in the bandit environment are the *probability of false alarm* and the *probability of detection delay exceeding a threshold*. Furthermore, the false alarm and delay events are only relevant over a *finite horizon*, and the horizon is not known to the change detector.

We therefore pose a variant of the QCD optimization problem in which the goal is to minimize the delay threshold (*latency*), under constraints on the probability of false alarm and the probability that the latency is exceeded over a finite horizon, and under the assumption that the horizon is not known to the detector. We develop a variant of the cumulative sum (CuSum) test with a time-varying threshold (TVT) as a candidate solution to the optimization problem. We further develop a lower bound on the minimum value of the latency under the constraints. We use the lower bound to establish certain asymptotic optimality properties of the TVT-CuSum test in terms of the horizon and the false alarm probability. We believe that a theoretical study of this variant of the QCD problem will be of relevance to the performance analysis of algorithms for piecewise stationary bandits and reinforcement learning that employ QCD tests to detect changes in the environment (see, e.g., [6–12]).

To the best of our knowledge, there is no prior work on the variant of the QCD problem described in the previous paragraph. A related variant is investigated in [13], where it is assumed that the observations are discrete (we do not require this assumption). In [13], the desired latency is fixed and assumed to be known to the detector, and the goal is to maximize the horizon.

The remainder of the paper is organized as follows: The formulation of our QCD problem is given in Section II. Section III is devoted to the performance analysis of the TVT-CuSum test. In Section IV, we provide a lower bound, which is useful in establishing some asymptotic optimality properties of the TVT-CuSum test. Numerical results that validate the analysis are given in Section V, and some concluding remarks are give in Section VI.

II. High Probability Low Latency Quickest Change Detection

Let $\{X_n : n \in \mathbb{N}\}$ be a sequence of independent random vectors whose values are observed sequentially. For the change-point $\nu \in \mathbb{N}$,

$$X_n \sim \begin{cases} f_0, \ n < \nu \\ f_1, \ n \geq \nu \end{cases}. \tag{1}$$

In other words, before the change-point ν, the observations follow the pre-change distribution with density f_0 with respect to some dominating measure λ. The remaining observations follow the post-change distribution with density f_1 with respect to the same dominating measure λ. We use $\mathbb{P}_\nu$ to denote the probability measure when the change-point occurs at $\nu \in \mathbb{N}$, and $\mathbb{P}_\infty$ to denote the probability measure when there is no change-point (i.e., $\nu = \infty$).

For horizon $T \in \mathbb{N}$, we observe the random vectors $X_1, \ldots, X_T$ sequentially. Let τ be the *stopping time* of a (causal) change detector. Our goal is to minimize the latency d, for which $\mathbb{P}_\nu(\tau \geq \nu + d)$ is small for all $\nu \in \{1, \ldots, T-d\}$, while the probability of false alarm over the horizon $\mathbb{P}_\infty(\tau \leq T)$ is small as well. Therefore, the optimization problem of interest has the following form:

$$\begin{aligned} \underset{\tau}{\text{minimize}} \quad & d \\ \text{s.t.} \quad & \mathbb{P}_\infty(\tau \leq T) \leq \delta_{\mathrm{F}} \\ & \mathbb{P}_\nu(\tau \geq \nu + d) \leq \delta_{\mathrm{D}},\ \forall \nu \in \{1, \ldots, T-d\} \end{aligned} \tag{2}$$

where $\delta_{\mathrm{F}}, \delta_{\mathrm{D}} \in (0,1)$ are some (small) numbers. In addition, we require change detector (equivalently τ) to be oblivious to the knowledge of the horizon T.

III. Performance of TVT-CuSum Test

For the standard Lorden formulation of the QCD problem [14], the CuSum test is known to be optimal [15]. The CuSum test statistic is given by:

$$W_n = \max_{1 \leq k \leq n} \sum_{i=k}^{n} \log \frac{f_1(X_i)}{f_0(X_i)} \tag{3}$$

which satisfies the recursion

$$W_n = \max\{W_{n-1}, 0\} + \log \frac{f_1(X_n)}{f_0(X_n)} \tag{4}$$

with $W_0 = 0$. The CuSum stopping time is given by:

$$\tilde{\tau}_b := \inf\{n \in \mathbb{N} : W_n \geq b\}. \tag{5}$$

It is therefore natural to consider the CuSum test of (5) as a possible candidate for solving the problem of interest given in (2). However, for the CuSum test with a constant threshold b, the false alarm probability $\mathbb{P}_\infty(\tilde{\tau}_b < T)$ goes to 1 as the horizon T goes to infinity. This is because $\mathbb{E}_\infty[\tilde{\tau}_b] < \infty$ (see, e.g., [14]). Therefore, the false alarm probability cannot be controlled to a given level δ_{F} without the knowledge of T.

Taking a cue from the analysis in [16], we will show that if we let the threshold b in (5) increase logarithmically with n, the false alarm probability remains upper bounded by a constant for all $T \in \mathbb{N}$. This leads to the following modified version of the CuSum test τ_r, which we refer to as the time-varying threshold CuSum (TVT-CuSum) test:

$$\tau_r := \inf\{n \in \mathbb{N} :\ W_n \geq \beta(n, \delta_{\mathrm{F}}, r)\},\ r > 1 \tag{6}$$

where

$$\beta(n, \delta_{\mathrm{F}}, r) := \log\left(\zeta(r) \frac{n^r}{\delta_{\mathrm{F}}}\right) \tag{7}$$

with $\zeta(r) := \sum_{i=1}^{\infty} \frac{1}{i^r}$. Then, the following theorem shows that the TVT-CuSum test satisfies the false alarm probability constraint. Note that β is not a function of the horizon T.

Theorem 1 (TVT-CuSum Test: False Alarm Probability)**.** *For any horizon $T \in \mathbb{N}$ and $r > 1$,*

$$\mathbb{P}_\infty(\tau_r \leq T) \leq \delta_{\mathrm{F}}. \tag{8}$$

Proof. First, we can upper bound the probability of false alarm as follows: For all $T \in \mathbb{N}$

$$\begin{aligned} &\mathbb{P}_\infty(\tau_r \leq T) \\ &\leq \mathbb{P}_\infty(\tau_r < \infty) \\ &= \mathbb{P}_\infty(\exists\, n \in \mathbb{N} :\ W_n \geq \beta(n, \delta_{\mathrm{F}}, r)) \\ &= \mathbb{P}_\infty\left(\exists\, j \leq n :\ \sum_{i=j}^{n} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) \geq \log\left(\zeta(r) \frac{n^r}{\delta_{\mathrm{F}}}\right)\right) \\ &= \mathbb{P}_\infty\left(\exists\, j \leq n :\ \prod_{i=j}^{n} \frac{f_1(X_i)}{f_0(X_i)} \geq \zeta(r) \frac{n^r}{\delta_{\mathrm{F}}}\right) \\ &= \mathbb{P}_\infty\Bigg(\exists\, j \in \mathbb{N},\ \exists\, k \in \mathbb{N} \cup \{0\} : \\ &\qquad \prod_{i=j}^{j+k} \frac{f_1(X_i)}{f_0(X_i)} \geq \zeta(r) \frac{(j+k)^r}{\delta_{\mathrm{F}}}\Bigg) \\ &= \mathbb{P}_\infty\Bigg(\exists\, j \in \mathbb{N},\ \exists\, k \in \mathbb{N} \cup \{0\} : \\ &\qquad \frac{1}{(j+k)^r} \prod_{i=j}^{j+k} \frac{f_1(X_i)}{f_0(X_i)} \geq \frac{\zeta(r)}{\delta_{\mathrm{F}}}\Bigg). \end{aligned} \tag{9}$$

Next, from (9) we obtain:

$$\begin{aligned} &\mathbb{P}_\infty(\tau_r \leq T) \\ &\overset{(a)}{\leq} \sum_{j=1}^{\infty} \mathbb{P}_\infty\left(\exists\, k \in \mathbb{N} \cup \{0\} :\ \frac{1}{(j+k)^r} \prod_{i=j}^{j+k} \frac{f_1(X_i)}{f_0(X_i)} \geq \frac{\zeta(r)}{\delta_{\mathrm{F}}}\right) \\ &\overset{(b)}{\leq} \sum_{j=1}^{\infty} \frac{\delta_{\mathrm{F}}}{\zeta(r)} \mathbb{E}_\infty\left[\frac{1}{j^r} \frac{f_1(X_j)}{f_0(X_j)}\right] = \frac{\delta_{\mathrm{F}}}{\zeta(r)} \sum_{j=1}^{\infty} \frac{1}{j^r} = \delta_{\mathrm{F}} \end{aligned}$$

where step (a) results from union bound. Step (b) stems from Doob's submartingale inequality [17] since $\left(\frac{1}{(j+k)^r} \prod_{i=j}^{j+k} \frac{f_1(X_i)}{f_0(X_i)}\right)_{k=0}^{\infty}$ is a supermartingale; this is shown in Lemma **??**, which, along with its proof, is given in Appendix **??**. ■

Next, we analyze the high probability latency for the TVT-CuSum test, which is defined as follows:

$$\begin{aligned} &d_r(T, \delta_{\mathrm{F}}, \delta_{\mathrm{D}}) \\ &:= \inf\{d \in \mathbb{N} :\ \mathbb{P}_\nu(\tau_r \geq \nu + d) \leq \delta_{\mathrm{D}}\ \forall \nu \in \{1, \ldots, T-d\}\}. \end{aligned} \tag{10}$$

For the purpose of our analysis, we define Λ to be the cumulant generating function of $\log\left(\frac{f_0(X)}{f_1(X)}\right)$ with $X \sim f_1$, i.e.,

$$\Lambda(\theta) = \log\left(\mathbb{E}_{f_1}\left[\exp\left(\theta \log\left(\frac{f_0(X)}{f_1(X)}\right)\right)\right]\right). \tag{11}$$

The following theorem gives an upper bound on d_r.

Theorem 2 (High Probability Latency for TVT-CuSum Test)**.** *For all $T \in \mathbb{N}$, $\delta_{\mathrm{F}}, \delta_{\mathrm{D}} \in (0,1)$, $r > 1$,*

$$d_r(T, \delta_{\mathrm{F}}, \delta_{\mathrm{D}})$$

$$\leq \inf_{\theta\in(0,1)} \left\{ \frac{1}{|\Lambda(\theta)|} \left[\log\left(\frac{1}{\delta_{\mathrm{D}}}\right) + \theta \log\left(\frac{1}{\delta_{\mathrm{F}}}\right) + r\theta\log(T) + \theta\log(\zeta(r)) \right] \right\}. \quad (12)$$

Proof. Fix an arbitrary $T \in \mathbb{N}$ and arbitrary $\delta_{\mathrm{F}}, \delta_{\mathrm{D}} \in (0,1)$. First, it is easy to show that $\Lambda(0) = \Lambda(1) = 0$. Since $f_1 \neq f_0$, Λ is strictly convex, and therefore, $\Lambda(\theta) < 0$ $\forall\,\theta \in (0,1)$. Next, by the definition of τ_r, $\forall\, d \in \mathbb{N}$ and $\forall\,\nu \in \{1,\ldots,T-d\}$

$$\begin{aligned}
&\mathbb{P}_\nu(\tau_r \geq \nu + d) \\
&= \mathbb{P}_\nu(\inf\{n\in\mathbb{N}:\ W_n \geq \beta(n,\delta_{\mathrm{F}},r)\} \geq \nu+d) \\
&= \mathbb{P}_\nu\Bigg(\forall\, n \in \{1,\ldots,\nu+d-1\}: \\
&\qquad \max_{1\leq j\leq n} \sum_{i=j}^{n} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) < \log\left(\zeta(r)\frac{n^r}{\delta_{\mathrm{F}}}\right)\Bigg).
\end{aligned} \quad (13)$$

Then, we have $\forall\,\theta \in (0,1)$

$$\begin{aligned}
&\mathbb{P}_\nu(\tau_r \geq \nu+d) && (14)\\
&\overset{(a)}{\leq} \mathbb{P}_\nu\left(\max_{1\leq j\leq \nu+d-1} \sum_{i=j}^{\nu+d-1} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) < \log\left(\zeta(r)\frac{(\nu+d)^r}{\delta_{\mathrm{F}}}\right)\right) \\
&\overset{(b)}{\leq} \mathbb{P}_\nu\left(\sum_{i=\nu}^{\nu+d-1} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) < \log\left(\zeta(r)\frac{(\nu+d)^r}{\delta_{\mathrm{F}}}\right)\right) \\
&= \mathbb{P}_\nu\left(-\sum_{i=\nu}^{\nu+d-1} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) > -\log\left(\zeta(r)\frac{(\nu+d)^r}{\delta_{\mathrm{F}}}\right)\right) \\
&\overset{(c)}{\leq} \exp\left(-d\left(-\frac{1}{d}\theta\log\left(\zeta(r)\frac{(\nu+d)^r}{\delta_{\mathrm{F}}}\right) - \Lambda(\theta)\right)\right) \\
&= \left(\zeta(r)\frac{(\nu+d)^r}{\delta_{\mathrm{F}}}\right)^\theta \exp(d\Lambda(\theta)) && (15)\\
&\overset{(d)}{\leq} \left(\zeta(r)\frac{T^r}{\delta_{\mathrm{F}}}\right)^\theta \exp(d\Lambda(\theta)) && (16)
\end{aligned}$$

where step (a) is due to the fact that $\{\nu+d\} \subseteq \{1,\ldots,\nu+d\}$, while step (b) is owing to the fact that $\sum_{i=\nu}^{\nu+d-1}\log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) \leq \max_{1\leq j\leq\nu+d-1}\sum_{i=j}^{\nu+d-1}\log\left(\frac{f_1(X_i)}{f_0(X_i)}\right)$. Step (c) stems from the Chernoff bound [18], and step (d) is due to the fact that $\nu+d \leq T$.

Define

$$\tilde{d} \coloneqq \frac{1}{|\Lambda(\theta)|}\left[\log\left(\frac{1}{\delta_{\mathrm{D}}}\right) + \theta\log\left(\frac{1}{\delta_{\mathrm{F}}}\right) + r\theta\log(T) + \theta\log(\zeta(r))\right]. \quad (17)$$

Then, according to (16), we have:

$$\begin{aligned}
\mathbb{P}_\nu\left(\tau_r > \nu+\tilde{d}\right) &\leq \left(\zeta(r)\frac{T^r}{\delta_{\mathrm{F}}}\right)^\theta \exp\left(\tilde{d}\Lambda(\theta)\right) \\
&= \delta_{\mathrm{D}}.
\end{aligned} \quad (18)$$

By the definition of d_r, $d_r \leq \tilde{d}$, and thus (12) holds. ■

Theorem 2 shows that $d_r = \mathcal{O}(\log T)$. In the next section, we demonstrate that this growth of d_r with T is order optimal when T is large and $\delta_{\mathrm{F}} + \delta_{\mathrm{D}} < 1$.

IV. Asymptotic Information-Theoretic Lower Bound on the High Probability Detection Delay

Let τ^* be the solution to (2) and $d^*(T,\delta_{\mathrm{F}},\delta_{\mathrm{D}})$ be the corresponding minimum value. In this section, we present a lower bound for $d^*(T,\delta_{\mathrm{F}},\delta_{\mathrm{D}})$. For the purpose of our analysis, we define the following constant

$$C \coloneqq \log\left(\mathbb{E}_{f_1}\left[\frac{f_1(X)}{f_0(X)}\right]\right). \quad (19)$$

It is easy to see that $C > 0$ using Jensen's inequality. Furthermore, we assume that $C < \infty$.

Theorem 3 (Lower Bound for High Probability Latency)**.** *For all $\delta_{\mathrm{F}}, \delta_{\mathrm{D}} \in (0,1)$ such that $\delta_{\mathrm{F}} + \delta_{\mathrm{D}} < 1$*

$$\begin{aligned}
d^*(T,\delta_{\mathrm{F}},\delta_{\mathrm{D}}) &\geq \left(\frac{1}{C} + o(1)\right) \\
&\cdot \left[\log(T) + \log\left(\frac{1}{\delta_{\mathrm{F}}}\right) + \log(1-\delta_{\mathrm{F}}-\delta_{\mathrm{D}}) + o(1)\right]
\end{aligned}$$

as $T \to \infty$.

Proof. For $c > C$, define the events

$$\mathcal{A} \coloneqq \left\{\nu \leq \tau^* < \nu + d^*,\ \sum_{i=\nu}^{\nu+d^*-1}\log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) \geq d^* c\right\} \quad (20)$$

and

$$\mathcal{B} \coloneqq \left\{\nu \leq \tau^* < \nu + d^*,\ \sum_{i=\nu}^{\nu+d^*-1}\log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) < d^* c\right\}. \quad (21)$$

We note that $\mathcal{A}\cap\mathcal{B} = \emptyset$ and $\mathcal{A}\cup\mathcal{B} = \{\nu \leq \tau^* < \nu+d^*\}$, which we will use later in the proof.

From the problem formulation (2) we have that $\forall\,\nu \in \{1,\ldots,T-d^*\}$,

$$\begin{aligned}
\delta_{\mathrm{D}} &\geq \mathbb{P}_\nu(\tau^* \geq \nu+d^*) \\
&= 1 - \mathbb{P}_\nu(\tau^* < \nu+d^*) \\
&= 1 - \mathbb{P}_\nu(\tau^* < \nu) - \mathbb{P}_\nu(\nu \leq \tau^* < \nu+d^*) \\
&= 1 - \mathbb{P}_\infty(\tau^* < \nu) - \mathbb{P}_\nu(\nu \leq \tau^* < \nu+d^*) \\
&\geq 1 - \mathbb{P}_\infty(\tau^* \leq T) - \mathbb{P}_\nu(\mathcal{B}) - \mathbb{P}_\nu(\mathcal{A}).
\end{aligned} \quad (22)$$

Next, since τ^* satisfies the false alarm probability constraint $\mathbb{P}_\infty(\tau^* \leq T) \leq \delta_{\mathrm{F}}$, by Lemma **??** in Appendix **??**, there exists a change-point $\tilde{\nu} \in \{1,\ldots,T-d^*\}$ such that

$$\mathbb{P}_\infty(\tilde{\nu} \leq \tau^* < \tilde{\nu}+d^*) \leq \frac{\delta_{\mathrm{F}}}{\lfloor T/d^* \rfloor}. \quad (23)$$

For this choice of $\tilde{\nu}$, we have

$$\begin{aligned}
\mathbb{P}_{\tilde{\nu}}(\mathcal{B}) &= \int_{\mathcal{B}} \prod_{i=1}^{\tilde{\nu}-1} f_0(x_i) \prod_{i=\tilde{\nu}}^{\tilde{\nu}+d^*-1} f_1(x_i) \otimes_{i=1}^{\tilde{\nu}+d^*-1} d\lambda(x_i) \\
&= \int_{\mathcal{B}} \prod_{i=\tilde{\nu}}^{\tilde{\nu}+d^*-1} \frac{f_1(x_i)}{f_0(x_i)} \prod_{i=1}^{\tilde{\nu}+d^*-1} f_0(x_i) \otimes_{i=1}^{\tilde{\nu}+d^*-1} d\lambda(x_i) \\
&\overset{(a)}{\leq} \int_{\mathcal{B}} \exp(d^*c) \prod_{i=1}^{\tilde{\nu}+d^*-1} f_0(x_i) \otimes_{i=1}^{\tilde{\nu}+d^*-1} d\lambda(x_i) \\
&= \exp(d^*c) \int_{\mathcal{B}} \prod_{i=1}^{\tilde{\nu}+d^*-1} f_0(x_i) \otimes_{i=1}^{\tilde{\nu}+d^*-1} d\lambda(x_i) \\
&\overset{(b)}{=} \exp(d^*c)\, \mathbb{P}_\infty(\mathcal{B}) \\
&\overset{(c)}{\leq} \exp(d^*c)\, \mathbb{P}_\infty(\tilde{\nu} \leq \tau^* < \tilde{\nu} + d^*) \\
&\overset{(d)}{\leq} \frac{\exp(d^*c)\,\delta_{\mathrm{F}}}{\lfloor T/d^* \rfloor}
\end{aligned} \tag{24}$$

where step (a) results from the definition of $\mathcal{B}$, and step (b) stems from the fact that under $\mathbb{P}_\infty$, every X_i follows the density f_0. Step (c) is owing to the fact that $\mathcal{B} \subseteq \{\tilde{\nu} \leq \tau^* < \tilde{\nu} + d^*\}$, and step (d) is due to (23).

Then, for any $\nu \in \{1, \ldots, T - d^*\}$,

$$\begin{aligned}
&\mathbb{P}_\nu(\mathcal{A}) \\
&= \mathbb{P}_\nu\left(\nu \leq \tau^* < \nu + d^*, \sum_{i=\nu}^{\nu+d^*-1} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) \geq d^*c\right) \\
&\leq \mathbb{P}_\nu\left(\sum_{i=\nu}^{\nu+d^*-1} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) \geq d^*c\right) \\
&= \mathbb{P}_\nu\left(\prod_{i=\nu}^{\nu+d^*-1} \frac{f_1(X_i)}{f_0(X_i)} \geq e^{d^*c}\right) \\
&\overset{(a)}{\leq} e^{-d^*c}\mathbb{E}_\nu\left[\prod_{i=\nu}^{\nu+d^*-1} \frac{f_1(X_i)}{f_0(X_i)}\right] \\
&= e^{-d^*c} \prod_{i=\nu}^{\nu+d^*-1} \mathbb{E}_\nu\left[\frac{f_1(X_i)}{f_0(X_i)}\right] \\
&\overset{(b)}{\leq} e^{-d^*(c-C)}
\end{aligned} \tag{25}$$

where step (a) results from the Markov inequality. Step (b) is due to the definition of C, which is in the theorem statement.

Next, by plugging (24) and (25) into (22) with $\nu = \tilde{\nu}$, we have

$$\begin{aligned}
\delta_{\mathrm{D}} &\geq 1 - \delta_{\mathrm{F}} - \frac{\exp(d^*c)\,\delta_{\mathrm{F}}}{\lfloor T/d^* \rfloor} - e^{-d^*(c-C)} \\
&\geq 1 - \delta_{\mathrm{F}} - \frac{d^* \exp(d^*c)\,\delta_{\mathrm{F}}}{T - d^*} - e^{-d^*(c-C)}.
\end{aligned} \tag{26}$$

By Theorem 1 and 2, $d^* \leq d_r$ for any $r > 1$. Since d_r grows logarithmically with T, $d^* = \mathcal{O}(\log(T))$; therefore, we can observe that for any $b \in (0,1)$, $d^* \leq bT$ for T large enough. In addition, this b approaches 0 as T goes to infinity. Thus, for any $s \in (0,1)$, $T - d^* \geq sT$ for T large enough, and s approaches 1 as T goes to infinity.

Furthermore, $\delta_{\mathrm{F}} + \delta_{\mathrm{D}} < 1$, $d^*(T, \delta_{\mathrm{F}}, \delta_{\mathrm{D}}) \to \infty$ as $T \to \infty$ by Lemma **??** in Appendix **??**. Now, we choose $c \to C$ such that $e^{-d^*(c-C)} \to 0$ as $T \to \infty$.

Again, since $d^* \to \infty$ as $T \to \infty$, for any $\tilde{c} > c$, $d^* \exp(cd^*) < \exp(\tilde{c}d^*)$ for T large enough, and $\tilde{c}$ goes to c as T goes to infinity. Hence, by (26), for T large enough, we have:

$$\delta_{\mathrm{D}} \geq 1 - \delta_{\mathrm{F}} - \frac{\delta_{\mathrm{F}}}{sT} \exp(\tilde{c}d^*) - \epsilon \tag{27}$$

where $\epsilon \in (0, 1 - \delta_{\mathrm{F}} - \delta_{\mathrm{D}})$, $\epsilon \to 0$ as $T \to \infty$. Rearranging the terms in (27), we obtain

$$\frac{\delta_{\mathrm{F}}}{sT} \exp(\tilde{c}d^*) \geq 1 - \delta_{\mathrm{F}} - \delta_{\mathrm{D}} - \epsilon. \tag{28}$$

By taking $\log$ on both sides, we have

$$\begin{aligned}
\tilde{c}d^* \geq& \log(T) \\
&+ \log\left(\frac{1}{\delta_{\mathrm{F}}}\right) + \log(s) + \log(1 - \delta_{\mathrm{F}} - \delta_{\mathrm{D}} - \epsilon).
\end{aligned} \tag{29}$$

This implies that

$$\begin{aligned}
d^*(T, \delta_{\mathrm{F}}, \delta_{\mathrm{D}}) \geq& \frac{1}{\tilde{c}} \log(T) + \frac{1}{\tilde{c}} \log\left(\frac{1}{\delta_{\mathrm{F}}}\right) \\
&+ \frac{1}{\tilde{c}} \log(s) + \frac{1}{\tilde{c}} \log(1 - \delta_{\mathrm{F}} - \delta_{\mathrm{D}} - \epsilon).
\end{aligned} \tag{30}$$

Since $\tilde{c} \to c$, $c \to C$, $s \to 1$, and $\epsilon \to 0$, all as $T \to \infty$, we have the theorem. ■

From Theorem 3, it is straightforward to see the following corollary.

Corollary 1. *When $\delta_{\mathrm{D}} + \delta_{\mathrm{F}} < 1$, as T goes to infinity, $d^*(T, \delta_{\mathrm{F}}, \delta_{\mathrm{D}}) \geq \frac{1}{C} \log(T)(1 + o(1))$*

This corollary, along with Theorem 2, shows that the growth of d_r with T of the TVT-CuSum test in (6) is order optimal when T is large and $\delta_{\mathrm{D}} + \delta_{\mathrm{F}} < 1$.

V. Experimental Results

In this section, we present some simulation results for the performance of the TVT-CuSum test and compare it with the upper bound in Theorem 2 and lower bound in Theorem 3. The number of trials used was 200000. The pre-change distribution was taken to be $\mathcal{N}(0,1)$, while the post-change distribution was taken to be $\mathcal{N}(1,1)$. According to (15), an upper bound on the probability of the delay exceeding the high probability latency d increases[1] with ν. Hence, to simulate the worst-case scenario for the high probability latency across ν, we should pick a value of ν closer to the horizon T than 0. We therefore set ν to be difference between the horizon T and the upper bound in Theorem 2. The parameter r of the TVT-CuSum test is set to be 2 in the experiments. If the test failed to detect

[1]This is in contrast with the standard CuSum test with a constant threshold, for which the worst-case delay occurs when the change-point is at $\nu = 0$.

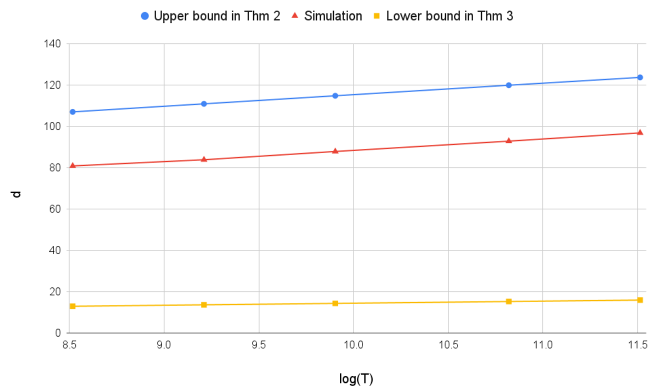

Fig. 1. High probability latency d_r as a function of the horizon T, with $\delta_{\mathrm{F}} = \delta_{\mathrm{D}} = 0.01$.

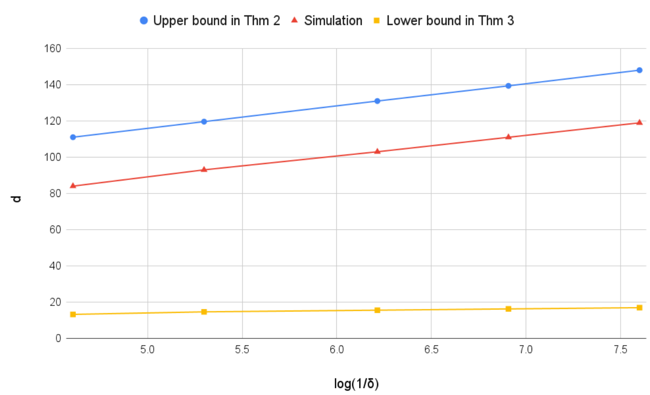

Fig. 2. High probability latency d_r as a function of $\delta = \delta_{\mathrm{F}} = \delta_{\mathrm{D}}$, with the horizon $T = 10000$

the change before the horizon, which was very rare in our simulations, we set the delay to be $T - \nu$. We recorded the detection delay for each trial and picked the $100(1-\delta_{\mathrm{D}})^{\text{th}}$ percentile out of all the values to be the empirical value of the high probability latency d. Then, we compared it with the upper bound in Theorem 2 and lower bound in Theorem 3. In the experiments, the quantities δ_{F} and δ_{D} are set to the same value, denoted by δ. Figure 1 and 2 exhibit how the high probability latency grows with T and δ

As shown in Figure 1 and 2, the upper bound is closer to the simulation results than the lower bound, and the slope of the upper bound and that of the simulation results are approximately the same.

The figures also indicate that the lower bound is very loose. There may two reasons why this the case: (i) in deriving the lower bound, we did not impose the condition that test τ is oblivious to the knowledge of the horizon T; and (ii) the TVT-CuSum test could be improved considerably. We believe that the first reason is more likely to be true.

VI. Conclusions

We posed a new variant of the QCD problem with a finite horizon, in which the goal is to minimize the latency, under constraints on the probability of false alarm and the probability of the latency being exceeded. We proposed the TVT-CuSum test with a threshold that increases logarithmically with time as a candidate solution. We show that the growth of the high probability latency d_r of the TVT-CuSum test with T is order optimal, when T is large and $\delta_{\mathrm{F}} + \delta_{\mathrm{D}} < 1$. As indicated in Section V, there might still be room for improvement for the lower bound given in Theorem 3. We leave this refinement of the lower bound for future work. It is also of interest to generalize the analysis in this paper to the more practical scenario in which there is uncertainty in the pre- and post-change distributions.

Acknowledgement

The work of Y.-H. Huang and V.V. Veeravalli was supported by the National Science Foundation under grant ECCS-2033900

References

[1] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge University Press, 2009.
[2] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Change-Point Detection* (Statistics). CRC Press, 2014.
[3] V. V. Veeravalli and T. Banerjee, "Quickest change detection," in *Academic press library in signal processing: Array and statistical signal processing*, Cambridge, MA: Academic Press, 2013.
[4] L. Xie, S. Zou, Y. Xie, and V. V. Veeravalli, "Sequential (quickest) change detection: Classical results and new directions," *IEEE Journal on Selected Areas in Information Theory*, vol. 2, no. 2, pp. 494–514, 2021. DOI: 10.1109/JSAIT.2021.3072962.
[5] L. Besson, E. Kaufmann, O.-A. Maillard, and J. Seznec, "Efficient change-point detection for tackling piecewise-stationary bandits," *The Journal of Machine Learning Research*, vol. 23, no. 1, pp. 3337–3376, 2022.
[6] F. Liu, J. Lee, and N. Shroff, "A change-detection based framework for piecewise-stationary multi-armed bandit problem," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.
[7] Y. Cao, W. Zheng, B. Kveton, and Y. Xie, "Nearly optimal adaptive procedure for piecewise-stationary bandit: A change-point detection approach," *AISTATS, Okinawa, Japan*, 2019.
[8] S. Padakandla, P. KJ, and S. Bhatnagar, "Reinforcement learning algorithm for non-stationary environments," *Applied Intelligence*, vol. 50, pp. 3590–3606, 2020.
[9] N. Dahlin, S. Bose, and V. V. Veeravalli, "Controlling a markov decision process with an abrupt change in the transition kernel," in *2023 American Control Conference (ACC)*, IEEE, 2023, pp. 3401–3408.
[10] L. Wang, H. Zhou, B. Li, L. R. Varshney, and Z. Zhao, "Near-optimal algorithms for piecewise-stationary cascading bandits," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 3365–3369.
[11] H. Zhou, L. Wang, L. Varshney, and E.-P. Lim, "A near-optimal change-detection based algorithm for piecewise-stationary combinatorial semi-bandits," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 6933–6940.
[12] H. Zhou, J. Chen, L. R. Varshney, and A. Jagmohan, "Nonstationary reinforcement learning with linear function approximation," *arXiv preprint arXiv:2010.04244*, 2020.
[13] V. Chandar, A. Tchamkerten, and G. Wornell, "Optimal sequential frame synchronization," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3725–3728, 2008.

[14] G. Lorden, “Procedures for reacting to a change in distribution,” *The annals of mathematical statistics*, pp. 1897–1908, 1971.

[15] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986. DOI: 10.1214/aos/1176350164.

[16] E. Kaufmann and W. M. Koolen, “Mixture martingales revisited with applications to sequential tests and confidence intervals,” *The Journal of Machine Learning Research*, vol. 22, no. 1, pp. 11 140–11 183, 2021.

[17] J. L. Doob, *Stochastic Processes*. John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1953, pp. viii+654.

[18] H. Chernoff, “A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations,” *The Annals of Mathematical Statistics*, pp. 493–507, 1952.