Quickest Change Detection With Non-Stationary Post-Change Observations

Yuchen Liang, Graduate Student Member, IEEE, Alexander G. Tartakovsky, Senior Member, IEEE, and Venugopal V. Veeravalli, Fellow, IEEE

Abstract—The problem of quickest change detection in a change in the distribution of a sequence of independent observations is considered. The pre-change observations are assumed to be stationary with a known distribution, while the post-change observations are allowed to be non-stationary with some possible parametric uncertainty in their distributions. In particular, it is assumed that the cumulative Kullback-Leibler divergence between the post-change and the pre-change distributions grows in a certain manner with time after the change-point. For the case where the post-change distributions are known, a universal asymptotic lower bound on the delay is derived, as the false alarm rate goes to zero. Furthermore, a window-limited Cumulative Sum (CuSum) procedure is developed, and shown to achieve the lower bound asymptotically. For the case where the post-change distributions have parametric uncertainty, a window-limited (WL) generalized likelihood-ratio (GLR) CuSum procedure is developed and is shown to achieve the universal lower bound asymptotically. Extensions to the case with dependent observations are discussed. The analysis is validated through numerical results on synthetic data. The use of the WL-GLR-CuSum procedure in monitoring pandemics is also demonstrated.

Index Terms—Quickest change detection, non-stationary observations, CuSum procedure, generalized likelihood-ratio CuSum procedure, pandemic monitoring.

I. INTRODUCTION

The problem of quickest change detection (QCD) is of fundamental importance in a variety of applications and has been extensively studied in mathematical statistics (see, e.g., [1], [2], [3], [4] for overviews). Given a sequence of observations whose distribution changes at some unknown change point, the goal is to detect the change in distribution as quickly as possible after it occurs, while not making too many false alarms.

Manuscript received 18 April 2022; revised 11 September 2022; accepted 8 December 2022. Date of publication 19 December 2022; date of current version 21 April 2023. This work was supported in part by the National Science Foundation under Grant ECCS-2033900 and in part by the Army Research Laboratory through the University of Illinois at Urbana-Champaign under Cooperative Agreement W911NF-17-2-0196. An earlier version of this paper was presented in part at the 2022 IEEE International Conference on Acoustics, Speech, and Signal Processing (virtual) [DOI: 10.1109/ICASSP43922.2022.9747602] and in part at the 2022 Conference on Information Sciences and Systems (virtual). (Corresponding author: Venugopal V. Veeravalli.)

Yuchen Liang and Venugopal V. Veeravalli are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA (e-mail: yliang35@illinois.edu; vvv@illinois.edu).

Alexander G. Tartakovsky is with AGT StatConsult, Los Angeles, CA 90274 USA (e-mail: alex.tartakovsky@gmail.com).

Communicated by Lifeng Lai, Associate Editor for Signal Processing and Source Coding.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2022.3230583.

Digital Object Identifier 10.1109/TIT.2022.3230583

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

In the classical formulations of the QCD problem, it is assumed that observations are independent and identically distributed (i.i.d.) with known pre- and post-change distributions. In many practical situations, while it is reasonable to assume that we can accurately estimate the pre-change distribution, the post-change distribution is rarely completely known. Furthermore, in many cases, it is reasonable to assume that the system is in a steady state before the change point and produces i.i.d. observations, but in the post-change mode the observations may be substantially non-identically distributed, i.e., non-stationary. For example, in the pandemic monitoring problem, the distribution of the number of people infected daily might have achieved a steady (stationary) state before the start of a new wave, but after the onset of a new wave, the post-change observations may no longer be stationary. Indeed, during the early phase of the new wave, the mean of the post-change distribution grows approximately exponentially. We will address the pandemic monitoring problem in detail in Section V.

In this paper, our main focus is on the QCD problem with independent observations1, where the pre-change observations are assumed to be stationary with a known distribution, while the post-change observations are allowed to be non-stationary with some possible parametric uncertainty in their distributions.

There have been extensions of the classical formulation to the case where the pre- and/or post-change distributions are not fully known and observations may be non-i.i.d., i.e., dependent and nonidentically distributed. For the i.i.d. case with parametric uncertainty in the post-change regime, Lorden [5] proposed a generalized likelihood ratio (GLR) Cumulative Sum (CuSum) procedure, and proved its asymptotic optimality in the minimax sense as the false alarm rate goes to zero, for one-parameter exponential families. An alternative to the GLR-CuSum, the mixture-based CuSum, was proposed and studied by Pollak [6] in the same setting as in [5]. The GLR approach has been studied in detail for the problem of detecting the change in the mean of a Gaussian i.i.d. sequence with an unknown post-change mean by Siegmund [7]. Both the mixture-based and GLR-CuSum procedures have been studied by Lai [8] in the pointwise setting in the non-i.i.d. case of possibly dependent and non-identically distributed observations, with parametric uncertainty in the post-change regime. More specifically, in [8], Lai assumed that the log-likelihood ratio process (between post- and pre-change distributions)

1The extension to the case of dependent observations is discussed in Section IV.
normalized by the number of observations \( n \) converges to a positive and finite constant as \( n \to \infty \), which can be interpreted as a Kullback-Leibler (KL) information number. In the case of independent (but non-identically distributed observations) this means that the expected value of the log-likelihood ratio process grows approximately linearly in the number of observations \( n \), for large \( n \). Tartakovsky \[9\] and Tartakovsky et al. \[1\] refer to such a case as “asymptotically homogeneous” (or stationary) case. In \[8\], Lai also developed a universal lower bound on the worst-case expected delay as well as on the expected delay to detection for every change point and proved that a specially designed window-limited (WL) CuSum procedure asymptotically achieves the lower bound as the maximal probability of false alarm approaches 0, when both pre- and post-change distributions are completely known, i.e., that the designed WL-CuSum procedure is asymptotically pointwise optimal to first order. For the case where the post-change distribution has parametric uncertainty, Lai proposed and analyzed a WL-GLR-CuSum procedure. A general Bayesian theory for non-i.i.d. asymptotically stationary stochastic models has been developed by Tartakovsky and Veeravalli \[10\] and Tartakovsky \[11\] for the discrete-time scenario, and by Baron and Tartakovsky \[12\] for the continuous-time scenario, when both pre- and post-change models are completely known. It was shown in these works that a Shiryaev-type change detection procedure minimizes not only average detection delay but also higher moments of the detection delay asymptotically, as the weighted probability of false alarm goes to zero, under very general conditions for the prior distribution of the change point. Extensions of these results to the case of the parametric composite post-change hypothesis have been provided by Tartakovsky \[2\], \[13\] where it has been shown that the mixture Shiryaev-type detection rule is asymptotically first-order optimal in the Bayesian setup and by Pergamenchchikov and Tartakovsky \[14\] where it was shown that the mixture Shiryaev-Roberts-type procedure pointwise and minimax asymptotically optimal in the non-Bayesian setup, but in the asymptotically stationary case where the cumulative KL divergence between post- and pre-change distributions \( g(n) \) grows linearly in \( n \) as \( n \to \infty \).

Note that all the previously cited works have focussed on the asymptotically stationary case. To the best of our knowledge, the asymptotically non-stationary case where the expected value of the log-likelihood ratio process normalized to some nonlinear function \( g(n) \) converges to a positive and finite (information) number has never been considered.\(^2\) Our contributions are as follows:

1. We develop a universal asymptotic (as the false alarm rate goes to zero) lower bound on the worst-case expected delay for our problem setting with non-stationary post-change observations.
2. We develop a window limited CuSum (WL-CuSum) procedure that asymptotically achieves the lower bound on the worst-case expected delay when the post-change distribution is fully known.
3. We develop and analyze a WL-GLR-CuSum procedure that asymptotically achieves the worst-case expected delay when the post-change distribution has parametric uncertainty.
4. We validate our analysis through numerical results and demonstrate the use of our approach in monitoring pandemics.

The rest of the paper is structured as follows. In Section II, we derive the information bounds and propose an asymptotically optimal WL-CuSum procedure when the post-change distributions are completely known. In Section III, we propose an asymptotically optimal WL-GLR-CuSum procedure when the post-change distributions have unknown parameters. In Section IV, we discuss possible extensions to the general non-i.i.d. case where the observations can be dependent and non-stationary. In Section V, we present some numerical results, including results on monitoring pandemics. We conclude the paper in Section VI. In the Appendix, we provide proofs of certain results.

II. INFORMATION BOUNDS AND OPTIMAL DETECTION

Let \( \{X_n\}_{n \geq 1} \) be a sequence of independent random variables (generally vectors), and let \( \nu \) be a change point. Assume that \( X_1, \ldots, X_{\nu-1} \) all have density \( p_0 \) with respect to some non-degenerate, sigma-finite measure \( \mu \) and that \( X_{\nu}, X_{\nu+1}, \ldots \) have densities \( p_{1,\nu}, p_{1,\nu+1}, \ldots \), respectively, with respect to \( \mu \). Note that the observations are allowed to be non-stationary after the change point and the post-change distributions may generally depend on the change point.

Let \( \{F_n\}_{n \geq 0} \) be the filtration, i.e., \( F_0 = \{\Omega, \emptyset\} \) and \( F_n = \sigma\{X_\ell, 1 \leq \ell \leq n\} \) is the sigma-algebra generated by the vector of \( n \) observations \( X_1, \ldots, X_n \) and let \( F_\infty = \sigma(X_1, X_2, \ldots) \). In what follows we denote by \( \mathbb{P}_\nu \), the probability measure on the entire sequence of observations when the change-point is \( \nu \). That is, under \( \mathbb{P}_\nu \), the random variables \( X_1, \ldots, X_{\nu-1} \) are i.i.d. with the common (pre-change) density \( p_0 \) and \( X_\nu, X_{\nu+1}, \ldots \) are independent with (post-change) densities \( p_{1,\nu}, p_{1,\nu+1}, \ldots \). Let \( \mathbb{E}_\nu \) denote the corresponding expectation. For \( \nu = \infty \) this distribution will be denoted by \( \mathbb{P}_\infty \) and the corresponding expectation by \( \mathbb{E}_\infty \). Evidently, under \( \mathbb{P}_\infty \) the random variables \( X_1, X_2, \ldots \) are i.i.d. with density \( p_0 \). In the sequel, we denote by \( \tau \) Markov (stopping) times with respect to the filtration \( \{F_n\}_{n \geq 0} \), i.e., the event \( \{\tau = n\} \) belongs to \( F_n \).

The change-time \( \nu \) is assumed to be unknown but deterministic. The problem is to detect the change quickly while not causing too many false alarms. Let \( \tau \) be a stopping time defined on the observation sequence associated with the detection rule, i.e., \( \tau \) is the time at which we stop taking observations and declare that the change has occurred. The problem is to detect the change quickly, minimizing the delay to detection \( \tau - \nu \), while not causing too many false alarms.

A. Classical Results Under i.i.d. Model

A special case of the model described above is where both the pre- and post-change observations are i.i.d.,
i.e., \( p_{1,n,\nu} \equiv p_1 \) for all \( n \geq \nu \geq 1 \). In this case, Lorden [5] proposed solving the following optimization problem to find the best stopping time \( \tau \):

\[
\inf_{\tau \in \mathcal{C}_a} \text{WADD} (\tau) \tag{1}
\]

where

\[
\text{WADD} (\tau) := \sup_{\nu \geq 1} \text{ess sup}_\nu \mathbb{E}_\nu \left[ (\tau - \nu + 1)^+ | \mathcal{F}_{\nu - 1} \right] \tag{2}
\]

characterizes the worst-case expected delay, and \( \text{ess sup} \) stands for essential supremum. The constraint set is

\[
\mathcal{C}_a := \{ \tau : \text{FAR} (\tau) \leq \alpha \} \tag{3}
\]

with

\[
\text{FAR} (\tau) := \frac{1}{\mathbb{E}_\infty [\tau]} \tag{4}
\]

which guarantees that the false alarm rate of the algorithm does not exceed \( \alpha \). Recall that \( \mathbb{E}_\infty [\cdot] \) is the expectation operator when the change never happens, and we use the conventional notation \((\cdot)^+ := \max\{0,\cdot\}\) for the nonnegative part. The mean time to a false alarm (MTFA) \( \mathbb{E}_\infty [\tau] \) is sometimes referred to as the average run length to false alarm.

Lorden also showed that Page’s CuSum detection algorithm [15] solves the problem in (1) asymptotically as \( \alpha \to 0 \), which is given by:

\[
\tau_{\text{Page}} (b) := \inf \left\{ n : \max_{1 \leq k \leq n + 1} \sum_{i=k}^{n} \log \frac{p_1(X_n)}{p_0(X_n)} \geq b \right\}. \tag{5}
\]

It was shown by Moustakides [16] that the CuSum algorithm is exactly optimal for the problem in (1) if threshold \( b = b_\alpha \) is selected so that \( \text{FAR} (\tau_{\text{Page}} (b_\alpha)) = \alpha \). If threshold \( b_\alpha \) is selected in a special way that accounts for the overshoot of the CuSum statistic \( \max_{1 \leq k \leq n + 1} \sum_{i=k}^{n} \log \frac{p_1(X_n)}{p_0(X_n)} \) over \( b_\alpha \) at stopping, which guarantees the approximation \( \text{FAR} (\tau_{\text{Page}} (b_\alpha)) \sim \alpha \) as \( \alpha \to 0 \), then we have the following third-order asymptotic approximation (as \( \alpha \to 0 \)) for the worst-case expected detection delay of the optimal procedure:

\[
\inf_{\tau \in \mathcal{C}_a} \text{WADD} (\tau) = \text{WADD} (\tau_{\text{Page}} (b_\alpha)) + o(1),
\]

\[
\text{WADD} (\tau_{\text{Page}} (b_\alpha)) = \frac{1}{D(p_1||p_0)} (|\log \alpha| - \text{const} + o(1))
\]

(see, e.g., [1]), which also implies the first-order asymptotic approximation (as \( \alpha \to 0 \)):

\[
\inf_{\tau \in \mathcal{C}_a} \text{WADD} (\tau) \sim \text{WADD} (\tau_{\text{Page}} (|\log \alpha|)) = \frac{|\log \alpha|}{D(p_1||p_0)} (1 + o(1)), \tag{6}
\]

where \( Y_\alpha \sim G_\alpha \) is equivalent to \( Y_\alpha = G_\alpha (1 + o(1)) \). Here \( D(p_1||p_0) \) is the Kullback-Leibler (KL) divergence between \( p_1 \) and \( p_0 \). Also, in the following we use a standard notation \( o(x) \) as \( x \to x_0 \) for the function \( f(x) \) such that \( f(x)/x \to 0 \) as \( x \to x_0 \), i.e., \( o(1) \to 0 \) as \( \alpha \to 0 \), and \( O(x) \) for the function \( f(x) \) such that \( f(x)/x \) is bounded as \( x \to 0 \), i.e., \( O(1) \) is a finite constant.

Along with Lorden’s worst average detection delay \( \text{WADD} (\tau) \), defined in (2), we can also consider the less pessimistic Pollak’s performance measure [17]:

\[
\text{SADD} (\tau) := \sup_{\nu \geq 1} \mathbb{E}_\nu [\tau - \nu + 1 | \tau \geq \nu].
\]

Pollak suggested the following minimax optimization problem in class \( \mathcal{C}_a \):

\[
\inf_{\tau \in \mathcal{C}_a} \text{SADD} (\tau). \tag{7}
\]

An alternative to CuSum is the Shiryaev-Roberts (SR) change detection procedure \( \tau_{\text{SR}} \) based not on the maximization of the likelihood ratio over the unknown change point but on summation of likelihood ratios (i.e., on averaging over the uniform prior distribution). As shown in [18], the SR procedure is second-order asymptotically minimax with respect to Pollak’s measure:

\[
\inf_{\tau \in \mathcal{C}_a} \text{SADD} (\tau) = \text{SADD} (\tau_{\text{SR}}) + O(1) \quad \text{as} \quad \alpha \to 0.
\]

The CuSum procedure with a certain threshold \( b_\alpha \) also has a second-order optimality property with respect to the risk \( \text{SADD} (\tau) \). A detailed numerical comparison of CuSum and SR procedures for i.i.d. models was performed in [19].

B. Information Bounds for Non-Stationary Post-Change Observations

In the case where both the pre- and post-change observations are independent and the post-change observations are non-stationary, the log-likelihood ratio is:

\[
Z_{n,k} = \log \frac{p_{1,n,k}(X_n)}{p_0(X_n)} \tag{8}
\]

where \( n \geq k \geq 1 \). Here \( k \) is a hypothesized change-point and \( X_n \) is drawn from the true distribution \( \mathbb{P}_\nu (\nu \in [1, \infty) \) or \( \nu = \infty) \).

In the classical i.i.d. model described in Section II-A, the cumulative KL-divergence after the change point increases linearly in the number of observations. We generalize this condition as follows. Let \( g_\nu : \mathbb{R}^+ \to \mathbb{R}^+ \) be an increasing and continuous function, which we will refer to as growth function. Note that the inverse of \( g_\nu \), denoted by \( g_\nu^{-1} \), exists and is also increasing and continuous. We assume that the expected sum of the log-likelihood ratios under \( \mathbb{P}_\nu \), which corresponds to the cumulative KL-divergence for our non-stationary model, matches the value of the growth function at all positive integers, i.e.,

\[
g_\nu (n) = \sum_{i=\nu}^{\nu + n - 1} \mathbb{E}_\nu [Z_{i,\nu}], \forall n \geq 1. \tag{9}
\]

Furthermore, we assume that \( \mathbb{E}_\nu [Z_{i,\nu}] > 0 \) for all \( i \geq \nu \) and that for each \( x > 0 \)

\[
g_\nu^{-1}(x) := \sup_{\nu \geq 1} g_\nu^{-1}(x) \tag{10}
\]

exists. Note that \( g_\nu^{-1} \) is also increasing and continuous. We also assume that \( g_\nu (n) \) diverges for all \( \nu \geq 1 \), and thus \( g_\nu^{-1}(x) \) is properly defined on the entire positive real
line. In the special case where the post-change distribution is invariant to the change-point \( \nu \), i.e., for \( j \geq 0 \), \( p_{1, \nu + j, \nu} \) is not a function of \( \nu \), we have \( g \equiv g_{\nu} \) and \( g^{-1} \equiv g_{\nu}^{-1} \) for all \( \nu \geq 1 \).

In order for change detection procedures to perform well it is necessary for the cumulative KL divergence between post- and pre-change distributions to grow sufficiently fast with \( n \), e.g., faster than \( \log n \). That is, the inverse \( g^{-1}(x) \) cannot grow too fast. This fact was discussed in [9] for the hypothesis testing problem where it was shown that if \( g(n) \sim \log n \), then the performance of Wald’s SPRT is extremely poor. The same is true for change detection problems.

The following key assumption on \( g^{-1}(x) \) guarantees the asymptotic optimality solution to the minimax problem in which we are interested:

\[
\log g^{-1}(x) = o(x) \quad \text{as} \quad x \to \infty. \tag{11}
\]

To better understand this condition, we provide three special cases below:

1) The post-change observations are independent and stationary (as in the classical case). Here \( g^{-1}(x) \) is linear in \( x \). Thus, \( \log g^{-1}(x) = O(\log x) = o(x) \), and condition (11) is always satisfied.

2) The KL divergence between the post- and pre-change distributions always increases (or increases asymptotically). Intuitively, this means that the post-change distributions increasingly drift away from that of the pre-change over time, as in Example 1 below and in the pandemic monitoring example in Section V-D. In this case, \( g^{-1}(x) \) grows at a slower than linear rate, and thus condition (11) is always satisfied.

3) The KL divergence between the post- and pre-change distributions gradually decreases to 0. Intuitively, this means that the post-change distributions gradually recovers that of the pre-change over time. Condition (11) guarantees detection for slow enough recovery, specifically, when the post-change KL divergence satisfies \( \mathbb{E}_\nu [Z_{n, \nu}] \sim n^{-\theta} \) with \( \theta < 1 \) as the decay factor. In this case, \( \log g^{-1}(x) = O(1 - \theta)^{-1} \log x = o(x) \). Obviously, condition (11) fails if \( \mathbb{E}_\nu [Z_{n, \nu}] \sim n^{-1} \), i.e., when \( g(n) \sim \log n \) and hence \( \log g^{-1}(x) \sim x \). We provide an example and some simulations for diminishing KL divergence between the post- and pre-change distributions in Section V-B.

We should note that such a growth function \( g(n) \) has been adopted previously in sequential hypothesis testing with non-stationary observations [1, Sec. 3.4], but not in QCD problem formulations such as the one considered here.

The proof of asymptotic optimality is performed in two steps. First, we derive a first-order asymptotic (as \( \alpha \to 0 \)) lower bound for the maximal expected detection delays \( \text{inf}_{\tau \in \mathcal{C}_\alpha} \text{WADD}^\alpha (\tau) \) and \( \text{inf}_{\tau \in \mathcal{C}_\alpha} \text{SADD}^\alpha (\tau) \). To this end, we need the following right-tail condition for the log-likelihood ratio process:

\[
\sup_{\nu \geq 1} \mathbb{P}_\nu \left\{ \max_{t \leq n} \sum_{i=\nu}^{\nu+t-1} Z_{i, \nu} \geq (1 + \delta) g_{\nu} (n) \right\} \xrightarrow{n \to \infty, \forall \delta > 0} 0, \tag{12}
\]

assuming that for all \( \nu \geq 1 \)

\[
\frac{1}{g_{\nu}(n)} \sum_{i=\nu}^{\nu+n-1} Z_{i, \nu} \xrightarrow{\nu \to \infty} 1, \quad \text{in} \quad \mathbb{P}_{\nu}-\text{probability}.
\]

At the second stage, we show that this lower bound is attained for the WL-CuSum procedure under the following left-tail condition

\[
\max_{\ell \geq \nu \geq 1} \mathbb{P}_\nu \left\{ \sum_{i=\nu}^{\ell+n-1} Z_{i, t} \leq (1 - \delta) g_{\nu} (n) \right\} \xrightarrow{n \to \infty, \forall \delta \in (0, 1)}. \tag{13}
\]

The following lemma provides sufficient conditions under which conditions (12) and (13) hold for the sequence of independent and non-stationary observations. Hereafter we use the notation \( \text{Var}_\nu(Y) = \mathbb{E}_\nu \left[ Y^2 - \mathbb{E}_\nu [Y]^2 \right] \) for variance of the random variable \( Y \) under distribution \( \mathbb{P}_{\nu} \).

**Lemma 1:** Consider the growth function \( g_{\nu} (n) \) defined in (9). Suppose that the sum of variances of the log-likelihood ratios satisfies

\[
\sup_{\ell \geq \nu \geq 1} \frac{1}{g_{\nu}^2(n)} \sum_{i=\nu}^{\ell+n-1} \text{Var}_\nu (Z_{i, t}) \xrightarrow{n \to \infty} 0 \tag{14}
\]

Then condition (12) holds.

If, in addition, for all \( \nu \geq 1 \) and all positive integers \( \Delta \),

\[
\mathbb{E}_\nu [Z_{\nu, \nu}] \leq \mathbb{E}_\nu [Z_{\nu + \Delta, \nu + \Delta}], \tag{15}
\]

then condition (13) holds.

The proof is given in the appendix.

**Remark:** One can generalize condition (15) in a way that either \( \mathbb{E}_\nu [Z_{\nu, \nu}] \leq \mathbb{E}_\nu [Z_{\nu + \Delta, \nu + \Delta}] \) or

\[
\frac{1}{g_{\nu}(n)} \sum_{i=\nu}^{\nu+n-1} \left( \mathbb{E}_\nu [Z_{\nu, \nu}] - \mathbb{E}_\nu [Z_{\nu + \Delta, \nu + \Delta}] \right) = o(1)
\]

holds for all positive integers \( \Delta \).

**Example 1:** Consider the following Gaussian exponential mean-change (GEM) detection problem. Denote by \( \mathcal{N}(\mu_0, \sigma_0^2) \) the Gaussian distribution with mean \( \mu_0 \) and variance \( \sigma_0^2 \). Let \( X_1, \ldots, X_{\nu-1} \) be distributed as \( \mathcal{N}(\mu_0, \sigma_0^2) \), and for all \( n \geq \nu \) let \( X_n \) be distributed as \( \mathcal{N}(\mu_0 e^{(n-\nu)}, \sigma_0^2) \). Here \( \theta \) is some positive fixed parameter. The log-likelihood ratio is given by:

\[
Z_{n, t} = \log \frac{p_{1, n, t}(X_n)}{p_{0}(X_n)} = -\frac{(X_n - \mu_0 e^{(n-t)})^2}{2\sigma_0^2} + \frac{(X_n - \mu_0)^2}{2\sigma_0^2} = \frac{\mu_0}{\sigma_0^2} (e^{(n-\nu)} - 1) X_n - \frac{\mu_0^2}{2\sigma_0^2} (e^{2(n-\nu)} - 1). \tag{16}
\]

Now, the growth function can be calculated as

\[
g_{\nu}(n) = \sum_{i=\nu}^{\nu+n-1} \mathbb{E}_\nu [Z_{\nu, i}] = \sum_{i=\nu}^{\nu+n-1} \frac{\mu_0^2}{2\sigma_0^2} (e^{2i} - 1). \tag{17}
\]

Since the post-change distribution is invariant to the change-point \( \nu \), \( g^{-1}(n) = g^{-1}_1(n) \sim O(\log n) \Rightarrow \log g^{-1}

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
which holds for an arbitrary $\delta$. Also, the sum of variances of the log-likelihood ratios is
\[
\sum_{i=1}^{t+n-1} \mathrm{Var}_\nu(Z_i) = \sum_{i=1}^{t+n-1} \frac{\mu_0^2}{\sigma_0^2} (e^{\nu(i-t)} - 1)^2 \mathrm{Var}_\nu(X_i)
\]
\[
= \frac{2}{\sigma_0^2} g_\nu(n) = o(g_\nu^2(n))
\]
for all $t \geq \nu$, which establishes condition (14). Further, for any $i \geq \nu$ and $\Delta \geq 1$, 
\[
\mathbb{E}_\nu[|Z_i+\Delta|^{\alpha}] = \mathbb{E}_\nu\left[\frac{\mu_0^2}{\sigma_0^2} (e^{\nu(i-\nu)} - 1) \mathbb{E}_\nu[|Z_{i+\Delta}|] - \frac{\mu_0^2}{2\sigma_0^2} (e^{2\nu(i-\nu)} - 1) \right]
\]
\[
\geq \frac{\mu_0}{\sigma_0^2} (e^{\nu(i-\nu)} - 1) \mathbb{E}_\nu[|Z_i|] - \frac{\mu_0^2}{2\sigma_0^2} (e^{2\nu(i-\nu)} - 1) \mathbb{E}_\nu[|Z_{i+\Delta}|]
\]
\[
= \mathbb{E}_\nu[|Z_i|]
\]
which establishes condition (15).

The following theorem gives a lower bound on the worst-case average detection delays as $\alpha \to 0$ in class $C_\alpha$.

**Theorem 1:** For $\delta \in (0,1)$ let
\[
h_\delta(\alpha) := g^{-1}(1/(1-\delta)|\log \alpha|).
\]
Suppose that $g^{-1}(x)$ satisfies (11). Then for all $\delta \in (0,1)$ and some $\nu \geq 1$
\[
\lim_{\alpha \to 0} \sup_{\tau \in C_\alpha} \mathbb{P}_\nu \left\{ \nu \leq \tau < \nu + h_\delta(\alpha) \right\} = 0
\]
and as $\alpha \to 0$
\[
\inf_{\tau \in C_\alpha} \text{WADD}(\tau) \geq \inf_{\tau \in C_\alpha} \text{SADD}(\tau) \geq g^{-1}(|\log \alpha|)(1+o(1)).
\]

**Proof:** Obviously, for any Markov time $\tau$,
\[
\text{WADD}(\tau) \geq \text{SADD}(\tau) \geq \mathbb{E}_\nu \left[ (\tau - \nu)^+ \right].
\]
Therefore, to prove the asymptotic lower bound (20) we have to show that as $\alpha \to 0$
\[
\sup_{\nu \geq 1} \mathbb{E}_\nu \left[ (\tau - \nu)^+ \right] \geq g^{-1}(|\log \alpha|)(1+o(1)),
\]
where the $o(1)$ term on the right-hand side does not depend on $\tau$, i.e., uniform in $\tau \in C_\alpha$.

To begin, let the stopping time $\tau \in C_\alpha$ and note that by Markov’s inequality,
\[
\mathbb{E}_\nu \left[ (\tau - \nu)^+ \right] \geq h_\delta(\alpha) \mathbb{P}_\nu \left\{ (\tau - \nu)^+ \geq h_\delta(\alpha) \right\}.
\]
Hence, if assertion (19) holds, then for some $\nu \geq 1$
\[
\inf_{\tau \in C_\alpha} \mathbb{P}_\nu \left\{ (\tau - \nu)^+ \geq h_\delta(\alpha) \right\} = 1 - o(1)
\]
as $\alpha \to 0$.

This implies the asymptotic inequality
\[
\inf_{\tau \in C_\alpha} \mathbb{E}_\nu \left[ (\tau - \nu)^+ \right] \geq h_\delta(\alpha)(1+o(1)),
\]
which holds for an arbitrary $\delta \in (0,1)$ and some $\nu$. Since by our assumption the function $h_\delta(\alpha)$ is continuous, taking the limit $\delta \to 0$ and maximizing over $\nu \geq 1$ yields inequality (21).

It remains to prove (19). Changing the measure $\mathbb{P}_\infty \to \mathbb{P}_\nu$ and using Wald’s likelihood ratio identity, we obtain
the following chain of equalities and inequalities for any $C > 0$ and $\delta \in (0,1)$:
\[
\mathbb{P}_\infty \left\{ \nu \leq \tau < \nu + h_\delta(\alpha) \right\}
\]
\[
= \mathbb{E}_\nu \left[ \mathbb{1} \{ \nu \leq \tau < \nu + h_\delta(\alpha) \} e^{-\sum_{i=\nu}^{\tau} Z_i} \right]
\]
\[
\geq \mathbb{E}_\nu \left[ \mathbb{1} \{ \nu \leq \tau < \nu + h_\delta(\alpha) \} \sum_{i=\nu}^{\tau} Z_i < C \right] \mathbb{1} \{ \nu \leq \tau < \nu + h_\delta(\alpha) \}
\]
\[
\geq e^{-C} \mathbb{P}_\nu \left[ \nu \leq \tau < \nu + h_\delta(\alpha) \right] \max_{0 \leq n < h_\delta(\alpha)} \sum_{i=\nu}^{n} Z_i < C
\]
\[
\geq e^{-C} \mathbb{P}_\nu \left\{ \nu \leq \tau < \nu + h_\delta(\alpha) \right\}
\]
\[
- \mathbb{P}_\nu \left\{ \max_{0 \leq n < h_\delta(\alpha)} \sum_{i=\nu}^{n} Z_i \geq C \right\}
\]
where the last inequality follows from the fact that $\mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(B^c)$ for any events $A$ and $B$, where $B^c$ is the complement event of $B$. Setting $C = g(h_\delta(\alpha))(1+\delta) = (1-\delta^2)|\log \alpha|$ yields
\[
\mathbb{P}_\nu \left\{ \nu \leq \tau < \nu + h_\delta(\alpha) \right\} \leq \kappa_{\delta,\alpha}^{(\nu)}(\tau) + \max_{\nu \geq 1} \beta_{\delta,\alpha}^{(\nu)}(\tau)
\]
where
\[
\kappa_{\delta,\alpha}^{(\nu)}(\tau) = e^{(1-\delta^2)|\log \alpha|} \mathbb{P}_\infty \left\{ \nu \leq \tau < \nu + h_\delta(\alpha) \right\}
\]
and
\[
\beta_{\delta,\alpha}^{(\nu)}(\tau) = \mathbb{P}_\nu \left\{ \max_{0 \leq n < h_\delta(\alpha)} \sum_{i=\nu}^{n} Z_i \geq (1+\delta) g(h_\delta(\alpha)) \right\}.
\]
Since $g(h_\delta(\alpha)) \to \infty$ as $\alpha \to 0$, by condition (12),
\[
\sup_{\nu \geq 1} \beta_{\delta,\alpha}^{(\nu)}(\tau) \to 0.
\]

Next we turn to the evaluation of the term $\kappa_{\delta,\alpha}^{(\nu)}(\tau)$ for any stopping time $\tau \in C_\alpha$. It follows from Lemma 2.1 in [2, page 72] that for any $M < \alpha^{-1}$, there exists some $\ell \geq 1$ (possibly depending on $\alpha$) such that
\[
\mathbb{P}_\infty \left\{ \ell \leq \tau < \ell + M \right\} \leq \mathbb{P}_\infty \left\{ \tau < \ell + M/\tau \geq \ell \right\} < M \alpha,
\]
so for some $\nu \geq 1$,
\[
\kappa_{\delta,\alpha}^{(\nu)}(\tau) \leq M \alpha e^{(1-\delta^2)|\log \alpha|} = M \alpha^{\delta^2}.
\]
If we choose
\[
M \leq M_\alpha = |h_\delta(\alpha)^2| \left|_{\delta=0} = \left| (g^{-1}(1/(\log \alpha)))^2 \right| \right.,
\]
then for all sufficiently small $\alpha$,
\[
\log M \leq 2 \log g^{-1}(1/(\log \alpha)) = o(|\log \alpha|)
\]
so that the condition (11) is satisfied. Furthermore,
\[
M_\alpha \alpha^p \xrightarrow{\alpha \to 0} 0 \quad \text{as} \quad \alpha \to 0
\]
for any $p > 0$. To see this, assume for purpose of contradiction that there exists some $p_0 > 0$ and $c_0 > 0$ such that $\lim_{\alpha \to 0} M_\alpha \alpha^{p_0} = c_0$. Then, since $\lim_{\alpha \to 0} \alpha^{-p_0} \neq 0$,
\[ \lim_{n \to 0} \log M_n = p_0 \lim_{n \to 0} \log |\log \alpha| + \log c_0 \text{ and thus } \log M_n \not\sim o(|\log \alpha|). \] Hence, it follows that for some \( \nu \geq 1 \), which may depend on \( \alpha \), as \( \alpha \to 0 \)
\[ \inf_{\tau \in C_{3_a}} \kappa_{3_a, \alpha}^{(\nu)}(\tau) \leq M_\alpha c^{\delta \tau} \to 0. \] (26)

Combining (23), (24), and (26) we obtain that for some \( \nu \geq 1 \),
\[ \mathbb{P}_\nu \{ \nu \leq \tau < \nu + h_\delta(\alpha) \} \leq M_\alpha c^{\delta \tau} + \sup_{\nu \geq 1} \beta_{3_a, \alpha}^{(\nu)} = o(1), \]
where the \( o(1) \) term is uniform over all \( \nu \geq 1 \). This yields assertion (19), and the proof is complete. □

C. Asymptotically Optimal Detection for Non-Stationary Post-Change Observations With Known Distributions

Recall that under the classical setting, Page’s CuSum procedure (in (5)) is optimal and has the following structure:
\[ \tau_{Page}(b) = \inf \left\{ n : \max_{1 \leq k \leq n+1} \sum_{i=k}^{n} Z_i \geq b \right\}, \] (27)

where \( Z_i \) is the log-likelihood ratio when the post-change distributions are stationary. When the post-change distributions are potentially non-stationary, the CuSum stopping rule is defined similarly as:
\[ \tau_C(b) := \inf \left\{ n : \max_{1 \leq k \leq n+1} \sum_{i=k}^{n} Z_{i,k} \geq b \right\}, \] (28)

where \( Z_{i,k} \) represents the log-likelihood ratio between densities \( p_{1,i,k} \) and \( p_0 \) for observation \( X_i \) (defined in (8)). Here \( i \) is the time index and \( k \) is the hypothesized change point. Note that if the post-change distributions are indeed stationary, i.e., \( p_{1,i,k} \equiv p_1 \), we would get \( Z_{i,k} \equiv Z_i \) for all \( k \leq i \), and thus \( \tau_C \equiv \tau_{Page} \).

Page’s classical CuSum algorithm admits a recursive way to compute its test statistic. Unfortunately, despite having independent observations, the test statistic in (28) cannot be computed recursively, even for the special case where the post-change distribution is invariant to the change-point as in Example 1.

Example 2: Consider the Gaussian Exponential Mean-Change problem defined in Example 1. Suppose \( \mu_0 = \sigma_0^2 = \theta = 1 \). Then, the log-likelihood ratio is given by
\[ Z_{n,t} = (e^{n-t} - 1)X_n - e^{2(n-t)} - 1. \]

Note that \( Z_{n,t} \) is a (linear) function of \( X_n \). Consider the following realization:
\[ X_1 = 1, \quad X_2 = 0, \quad X_3 = 10. \]

It can be verified that
\[ \arg \max_{1 \leq k \leq 3} \sum_{i=k}^{2} Z_{i,k} = 2, \quad \text{and} \quad \arg \max_{1 \leq k \leq 4} \sum_{i=k}^{3} Z_{i,k} = 1. \]

Note that maximizer \( k^* \) goes backward in time in this case, in contrast to what happens when both the pre- and post-change observations follow i.i.d. models. The test statistic at time \( n = 2 \) is a function of only \( X_2 \), and this is insufficient to construct the test statistic at time \( n = 3 \), which is a function of \( X_1 \), in addition to being a function of \( X_2 \), and \( X_3 \).

For computational tractability we therefore consider a window limited version of the CuSum procedure in (28):
\[ \tilde{\tau}_C(b) := \inf \left\{ n : \max_{n-m \leq k \leq n+1} \sum_{i=k}^{n} Z_{i,k} =: W(n) \geq b \right\} \] (29)

where \( m \) is the window size. For \( n < m \) maximization is performed over \( 1 \leq k \leq n \). In the asymptotic setting, \( m = m_\alpha \) depends on \( \alpha \) and should go to infinity as \( \alpha \to 0 \) with certain appropriate rate. Specifically, following a similar condition that Lai [8] used in the asymptotically stationary case, we shall require that \( m_\alpha \to \infty \) as \( \alpha \to 0 \) in such a way that
\[ \lim_{\alpha \to 0} m_\alpha / g^{-1}(|\log \alpha|) > 1. \] (30)

Since the range for the maximum is smaller in \( \tilde{\tau}_C(b) \) than in \( \tau_C(b) \), given any realization of \( X_1, X_2, \ldots \), if the test statistic of \( \tilde{\tau}_C(b) \) crosses the threshold \( b \) at some time \( n \), so does that of \( \tau_C(b) \). Therefore, for any fixed threshold \( b > 0 \),
\[ \tau_C(b) \leq \tilde{\tau}_C(b) \] (31)

almost surely.

In the following, we first control the asymptotic false alarm rate of \( \tilde{\tau}_C(b) \) with an appropriately chosen threshold in Lemma 2. Then we obtain asymptotic approximation of the expected detection delays of \( \tilde{\tau}_C(b) \) in Theorem 2. Finally, we combine these two results and provide an asymptotically optimal solution to the problem in (1) in Theorem 3.

Lemma 2: Suppose that \( b_0 = |\log \alpha| \). Then
\[ \text{FAR} (\tilde{\tau}_C(b_0)) \leq \alpha \quad \text{for all } \alpha \in (0, 1), \] i.e., \( \tilde{\tau}_C(b_0) \in C_\alpha \).

Proof: Define the statistic
\[ R_n = \sum_{k=1}^{n} \exp \left( \sum_{i=k}^{n} Z_{i,k} \right), \quad R_0 = 0 \]

and the corresponding stopping time \( T_b := \inf \{ n : R_n \geq e^b \} \). We now show that \( \mathbb{E}_\infty \left[ T_b \right] \geq e^b \), which implies that \( \mathbb{E}_\infty \left[ \tilde{\tau}_C(b) \right] \geq e^b \) for any \( b > 0 \) since, evidently, \( \tilde{\tau}_C(b) \geq T_b \) for any \( b > 0 \). Recall that \( \mathcal{F}_n = \sigma(X_\ell, 1 \leq \ell \leq n) \) denotes a sigma-algebra generated by \( (X_1, \ldots, X_n) \). Since \( \mathbb{E}_\infty \left[ e^{Z_{n,k}} \mathbb{I}(\mathcal{F}_{n-1}) \right] = 1 \), it is easy to see that
\[ \mathbb{E}_\infty \left[ R_n | \mathcal{F}_{n-1} \right] = 1 + R_{n-1} \quad \text{for } n \geq 1. \]

Consequently, the statistic \( \{ R_n - n \}_{n \geq 1} \) is a zero-mean \( (\mathbb{P}_\infty, \mathcal{F}_n) \)-martingale. It suffices to assume that \( \mathbb{E}_\infty \left[ T_b \right] < \infty \) since otherwise the statement is trivial. Then, \( \mathbb{E}_\infty \left[ R_{T_b} - T_b \right] \) exists and also
\[ \lim_{n \to \infty} \int_{\{T_b > n\}} |R_n - n| d\mathbb{P}_\infty = 0 \]

since \( 0 \leq R_n < e^b \) on the event \( \{T_b > n\} \). Hence, we can apply the optional sampling theorem (see, e.g. [1, Th 2.3.1, page 31]), which yields \( \mathbb{E}_\infty \left[ R_{T_b} \right] = \mathbb{E}_\infty \left[ T_b \right] \). Since \( R_{T_b} \geq e^b \) it follows that \( \mathbb{E}_\infty \left[ \tilde{\tau}_C(b) \right] \geq \mathbb{E}_\infty \left[ T_b \right] \geq e^b \).
Now, setting $b_0 = |\log \alpha|$ implies the inequality
\[
E_{\infty} [\tilde{\tau}_C(b_0)] \geq e^{b_0} = \frac{1}{\alpha} \quad (33)
\]
(for any $m_0 \geq 1$), and therefore (32) follows. \hfill \Box

The following result establishes asymptotic performance of the WL-CuSum procedure given in (29) for large threshold values.

**Theorem 2:** Fix $\delta \in (0, 1)$ and let $N_{b, \delta} := \lfloor g^{-1}(b/(1 - \delta)) \rfloor$. Suppose that in the WL-CuSum procedure the size of the window $m = m_b$ diverges (as $b \to \infty$) in such a way that
\[
m_b \geq N_{b, \delta}(1 + o(1)). \quad (34)
\]
Further, suppose that conditions (12) and (13) hold for $Z_{n,k}$ when $n \geq k \geq 1$. Then, as $b \to \infty$,
\[
\text{SADD} (\tilde{\tau}_C(b)) \sim \text{WADD} (\tilde{\tau}_C(b)) \sim g^{-1}(b). \quad (35)
\]

**Proof:** Since FAR ($\tilde{\tau}_C(b)) \leq e^{-b}$, the WL-CuSum procedure $\tilde{\tau}_C(b)$ belongs to class $C_\nu$ with $\alpha = e^{-b}$. Hence, replacing $n$ by $e^{-b}$ in the asymptotic lower bound (20) in Theorem 1, we obtain that under condition (12) the following asymptotic lower bound holds:
\[
\lim_{b \to \infty} \inf \frac{\text{WADD} (\tilde{\tau}_C(b))}{g^{-1}(b)} \geq \lim_{b \to \infty} \inf \frac{\text{SADD} (\tilde{\tau}_C(b))}{g^{-1}(b)} \geq 1. \quad (36)
\]
Thus, to establish (35) it suffices to show that under condition (13) as $b \to \infty$,
\[
\text{WADD} (\tilde{\tau}_C(b))) \leq g^{-1}(b)(1 + o(1)). \quad (37)
\]
Note that we have the following chain of equalities and inequalities:
\[
\begin{align*}
E_{\nu} \left[ (\tilde{\tau}_C(b) - \nu)^+ | F_{\nu-1} \right] & = \sum_{t=0}^{(\ell+1)N_{b, \delta}} \mathbb{P}_\nu \left\{ \tilde{\tau}_C(b) - \nu > t | F_{\nu-1} \right\} \ dt \\
& \leq N_{b, \delta} + \sum_{t=1}^{(\ell+1)N_{b, \delta}} \mathbb{P}_\nu \left\{ \tilde{\tau}_C(b) - \nu > t | F_{\nu-1} \right\} \ dt \\
& \leq N_{b, \delta} + \sum_{t=1}^{(\ell+1)N_{b, \delta}} \mathbb{P}_\nu \left\{ \tilde{\tau}_C(b) - \nu > t N_{b, \delta} | F_{\nu-1} \right\} \ dt \\
& = N_{b, \delta} \left[ 1 + \sum_{t=1}^{(\ell+1)N_{b, \delta}} \mathbb{P}_\nu \left\{ \tilde{\tau}_C(b) - \nu > t N_{b, \delta} | F_{\nu-1} \right\} \right] . \quad (38)
\end{align*}
\]
Define $\lambda_{n,k} := \sum_{i=k}^{n} Z_{i,k}$ and $K_n := \nu + n N_{b, \delta}$. We have $W(n) = \max_{m - m_b \leq n \leq \lambda_{n,k}} \nu$. Since by condition (34) $m_b > N_{b, \delta}$ (for a sufficiently large $b$), for any $n \geq 1$, $W(n + n N_{b, \delta}) > \lambda_{K_n, K_{n-1}}$, and we have
\[
\begin{align*}
\mathbb{P}_\nu \left\{ \tilde{\tau}_C(b) - \nu > t N_{b, \delta} | F_{\nu-1} \right\} & = \mathbb{P}_\nu \left\{ W(1) \leq b, \ldots, W(\nu + t N_{b, \delta}) < b | F_{\nu-1} \right\} \\
& \leq \mathbb{P}_\nu \left\{ W(\nu + n N_{b, \delta}) < b, \ldots W(\nu + t N_{b, \delta}) < b | F_{\nu-1} \right\} \\
& \leq \mathbb{P}_\nu \left\{ \lambda_{K_1, K_0} < b, \ldots, \lambda_{K_{\ell}, K_{\ell-1}} < b | F_{\nu-1} \right\} \\
& = \prod_{n=1}^{\ell} \mathbb{P}_\nu \left\{ \lambda_{K_n, K_{n-1}} < b \right\} , \quad (39)
\end{align*}
\]
where the last equality follows from independence of the increments of $\{\lambda_{t,n}\}_{n \geq t}$.

By condition (13), for a sufficiently large $b$ there exists a small $\varepsilon_b$ such that
\[
\mathbb{P}_\nu \left\{ \lambda_{K_n, K_{n-1}} < b \right\} \leq \varepsilon_b, \quad \forall n \geq 1.
\]
Therefore, for any $\ell \geq 1$,
\[
\mathbb{P}_\nu \left\{ \tilde{\tau}_C(b) - \nu > \ell N_{b, \delta} | F_{\nu-1} \right\} \leq \varepsilon_b.
\]
Combining this inequality with (38) and using the fact that $\sum_{t=1}^{\ell} \varepsilon_b = \varepsilon_b(1 - \varepsilon_b)^{-1}$, we obtain
\[
\begin{align*}
E_{\nu} \left[ (\tilde{\tau}_C(b) - \nu)^+ | F_{\nu-1} \right] & \leq N_{b, \delta} \left[ 1 + \frac{\varepsilon_b}{1 - \varepsilon_b} \right] \\
& = \frac{g^{-1}(b/(1 - \delta))}{1 - \varepsilon_b}. \quad (40)
\end{align*}
\]
Since the right-hand side of this inequality does not depend on $\nu$, $g^{-1}(b/(1 - \delta)) \to \infty$ as $b \to \infty$ and $\varepsilon_b$ and $\delta$ can be arbitrarily small numbers, this implies the upper bound (37). The proof is complete. \hfill \Box

Using Lemma 2 and Theorem 2, we obtain the following asymptotic result which establishes asymptotic optimality of the WL-CuSum procedure and its asymptotic operating characteristics.

**Theorem 3:** Suppose that threshold $b_0$ is so selected that $b_0 \sim |\log \alpha|$ as $\alpha \to 0$, in particular as $b_0 \sim |\log \alpha|$. Further, suppose that left-tail (12) and right-tail (13) conditions hold for $Z_{n,k}$ when $n \geq k \geq 1$. Then, the WL-CuSum procedure in (29) with the window size $m_\alpha$ that satisfies the condition
\[
m_\alpha \geq g^{-1}(|\log \alpha|)(1 + o(1)) \text{ as } \alpha \to 0 \quad (41)
\]
solves the problems (1) and (7) asymptotically to first order as $\alpha \to 0$, i.e.,
\[
\begin{align*}
\inf_{\tau \in C_{\alpha}} \text{WADD} (\tau) & \sim \text{WADD} (\tilde{\tau}_C(b_0)), \\
\inf_{\tau \in C_{\alpha}} \text{SADD} (\tau) & \sim \text{SADD} (\tilde{\tau}_C(b_0)), \quad (42)
\end{align*}
\]
and
\[
\begin{align*}
\text{SADD} (\tilde{\tau}_C(b_0)) & \sim \text{WADD} (\tilde{\tau}_C(b_0)) \sim g^{-1}(|\log \alpha|). \quad (43)
\end{align*}
\]

**Proof:** Let $b_0$ be so selected that FAR ($\tilde{\tau}_C(b_0)) \leq \alpha$ and $b_0 \sim |\log \alpha|$ as $\alpha \to 0$. Then by Theorem 2, as $\alpha \to 0$.

SADD ($\tilde{\tau}_C(b_0)) \sim \text{WADD} (\tilde{\tau}_C(b_0)) \sim g^{-1}(|\log \alpha|)$, i.e., asymptotic approximations (43) hold. Comparing these asymptotic equalities with the asymptotic lower bound (20) in Theorem 1 immediately yields asymptotics (42). In particular, if $b_0 = |\log \alpha|$, then by Lemma 2 FAR ($\tilde{\tau}_C(b_0)) \leq \alpha$, and therefore the assertions hold. \hfill \Box

Remark: Clearly, the asymptotic optimality result still holds in the case where no window is applied, i.e., $m_\alpha = n - 1$.

**Example 3:** Consider the same setting as in Example 1. We have shown that conditions (14) and (15) hold in this setting, and thus (12) and (13) also hold by Lemma 1. Considering the growth function $g_\nu(n)$ given in (17), as $n \to \infty$, we obtain
\[
g_\nu(n) = \sum_{i=0}^{n-1} \frac{\mu^2}{2\sigma_0^2} (e^{\theta i} - 1)^2 = \frac{\mu^2}{2\sigma_0^2} e^{2\theta(n-1)}(1 + o(1)).
\]
Thus, as $y \to \infty$,
\[ g^{-1}(y) = \frac{1}{2\theta} \log \left( \frac{2\sigma^2}{\mu_0} y \right) (1 + o(1)) \]
and if $b_\alpha = |\log \alpha|$ or more generally $b_\alpha \sim |\log \alpha|$ as $\alpha \to 0$ we obtain
\[
\text{WADD} (\tau_C(b_\alpha)) = \frac{1}{2\theta} \log \left( \frac{2\sigma^2}{\mu_0} |\log \alpha| \right) (1 + o(1)) = O \left( \frac{1}{2\theta} \log(|\log \alpha|) \right). \tag{44}
\]

III. ASYMPTOTICALLY OPTIMAL PROCEDURE FOR NON-STATIONARY POST-CHANGE OBSERVATIONS WITH PARAMETRIC UNCERTAINTY

We now study the case where the evolution of the post-change distribution is parametrized by an unknown but deterministic parameter $\theta \in \mathbb{R}^d$. Let $X_0, X_{i+1}, \ldots$ have density $p_{i,\nu,\nu}, p_{i,\nu,\nu+1}, \ldots$, respectively, with respect to the common non-degenerate measure $\mu$, when post-change parameter is $\theta$. Let $P_{\theta_i, \nu}$ and $E_{\theta, \nu}$ denote, respectively, the probability measure on the entire sequence of observations and expectation when the change point is $\nu = k < \infty$ and the post-change parameter is $\theta$. Let $\Theta \subset \mathbb{R}^d$ be an open and bounded set of parameter values. For any $n \geq k$ and $\theta \in \Theta$ the log-likelihood ratio process is given by
\[
Z_{n,k}^\theta = \frac{p_{i,\nu,n}(X_n)}{p_0(X_n)}. \tag{45}
\]

Also, the growth function in (9) is redefined as
\[
g_{\nu,\theta}(n) = \sum_{\nu=\nu+1}^{\nu+n-1} E_{\nu,\theta}[Z_{\nu,\nu}', \ \forall n \geq 1, \tag{46}\]
and it is assumed that $g_{\nu}^{-1}(x) = \sup_{\nu \geq 1} g_{\nu,\theta}(x)$ exists. It is also assumed that
\[
\log g_{\nu}^{-1}(x) = o(x) \quad \text{as } x \to \infty. \tag{47}
\]

The goal in this section is to solve the optimization problems (1) and (7) asymptotically as $\alpha \to 0$ under parametric uncertainty. More specifically, for $\theta \in \Theta$, define Lorden’s and Pollak’s worst-case expected detection delay measures
\[
\text{WADD}_\theta (\tau) := \text{ess sup} \sup_{\nu \geq 1} E_{\nu,\theta} [(\tau - \nu + 1)^+ | F_{\nu-1}] \tag{48}
\]
and
\[
\text{SADD}_\theta (\tau) := \sup_{\nu \geq 1} E_{\nu,\theta} [\tau - \nu + 1 | \tau \geq \nu]
\]
and the corresponding asymptotic optimization problems: find a change detection procedure $\tau^*$ that minimizes these measures to first order in class $C_\alpha$, i.e., for all $\theta \in \Theta$
\[
\lim_{\alpha \to 0} \inf_{\tau \in C_\alpha} \frac{\text{WADD}_\theta (\tau)}{\text{WADD}_\theta (\tau^*)} = 1, \quad \lim_{\alpha \to 0} \inf_{\tau \in C_\alpha} \frac{\text{SADD}_\theta (\tau)}{\text{SADD}_\theta (\tau^*)} = 1. \tag{49}
\]

Consider the following WL-GLR-CuSum change detection procedure
\[
\tilde{\tau}_G (b) := \inf \left\{ n : \max_{n-m_b \leq k \leq n} \sup_{\theta \in \Theta} \sum_{i=k}^{n} Z_{i,k}^\theta \geq b \right\}, \tag{49}
\]
where $\Theta_b \not\supset \Theta$ as $b \to \infty$. For $n < m_b$ maximization is performed over $1 \leq k \leq n$. Therefore, it is guaranteed that $\theta \in \Theta_b$ for all large enough $b$. Since we are interested in class $C_\alpha = \{ \tau : \text{FAR} (\tau) \leq \alpha \}$, in which case both threshold $b = b_\alpha$ and window size $m_b = m_\alpha$ are the functions of $\alpha$, we will write $\Theta_b = \Theta_\alpha$ and suppose that $\Theta_\alpha \subset \mathbb{R}^d$ is compact for each $\alpha$. Hereafter we omit the dependency of $\hat{\theta}_n, k$ on $\alpha$ for brevity. In this paper, we focus on the case where $\Theta_\alpha$ is continuous for all $\alpha$’s. The discrete case is simpler and will be considered elsewhere.

The following assumption is made to guarantee the existence of an upper bound on FAR.

Assumption 1: There exists $\varepsilon > 0$ such that for any large enough $b > 0$,
\[
\inf \left\{ \sum_{k \leq n \leq k+m_b} \| \hat{\theta}_n, k \|_2 \leq \varepsilon b \right\} \leq 2 \varepsilon b \tag{50}
\]
where $\lambda_{\max} (A)$ represents the maximum absolute eigenvalue of a symmetric matrix $A$ and $\varepsilon_b \leq 0$ as $b \to \infty$.

Example 4: Consider again the Gaussian exponential mean-change detection problem in Example 1. Now we consider the case where the exact value of the post-change exponent coefficient $\theta$ is unknown and belongs to $\Theta = \Theta_{\min} \cup \Theta_{\max}$. Note that $\theta$ characterizes the entire post-change evolution rather than a single post-change distribution.

We shall verify Assumption 1 below.

Recalling the definition of log-likelihood ratio given in (16), for any $\theta \in \Theta$ and $k \leq i \leq n$ where $n \geq m_b$, we have
\[
-\frac{\partial^2}{\partial \theta^2} Z_{i,k}^\theta = -\frac{\partial^2}{\partial \theta^2} \left( \frac{\mu_0}{\sigma_0^2} (e^{\theta(i-k)} - 1) X_i - \mu_0^2 e^{2\theta(i-k)} \right)
= -\frac{\mu_0^2}{\sigma_0^2} (i-k)^2 e^{2\theta(i-k)} X_i + 2(i-k)^2 \mu_0^2 e^{2\theta(i-k)}
= \frac{\mu_0^2}{\sigma_0^2} (i-k)^2 e^{\theta(i-k)} (2\mu_0 e^{\theta(i-k)} - X_i). \tag{51}
\]

Therefore,
\[
\max_{(k,n) : k \leq n \leq k+m_b} \text{ess sup}_{\theta \in \Theta} \left| \frac{\partial^2}{\partial \theta^2} \sum_{i=k}^{n} Z_{i,k}^\theta \right|
= \max_{(k,n) : k \leq n \leq k+m_b} \text{ess sup}_{\theta \in \Theta} \left| \frac{\mu_0^2}{\sigma_0^2} \sum_{i=k}^{n} (i-k)^2 e^{2\theta(i-k)} (2\mu_0 e^{\theta(i-k)} - X_i) \right|
\leq \max_{(k,n) : k \leq n \leq k+m_b} \text{ess sup}_{\theta \in \Theta} \left| \frac{\mu_0^2}{\sigma_0^2} m_b^2 e^{2\theta m_b} + \max_{(k,n) : k \leq n \leq k+m_b} \left| \sum_{i=k}^{n} X_i \right| \right|
\leq \frac{4\mu_0^2}{\sigma_0^2} m_b^2 e^{2\theta m_b} \leq \frac{4\mu_0^2}{\sigma_0^2} m_b^2 e^{2\theta m_b}, \tag{52}
\]
where (*) is true provided that

$$\max_{(k,n):k \leq n \leq k + m_\sigma} \left\{ \sum_{i=k}^{n} X_i \right\} < 2 \mu_0 m_b e^{\Theta_{\max}}.$$

Since \( X_i \)'s are i.i.d. under \( P_\infty \), \( \sum_{i=k}^{n} X_i \) has a Gaussian distribution with mean \( (m_n + 1) \mu_0 \) and variance \( (m_b + 1) \sigma_0^2 \). Therefore, for any \( \theta \in \Theta \),

$$P_\infty \left\{ \max_{(k,n):k \leq n \leq k + m_\sigma} \left\{ \sum_{i=k}^{n} X_i \right\} > 2 \mu_0 m_b e^{\Theta_{\max}} \right\} \leq P_\infty \left\{ \sum_{i=1}^{m_b} X_i > 2 \mu_0 m_b e^{\Theta_{\max}} \right\} = 2 Q \left( \frac{2 \mu_0 m_b e^{\Theta_{\max}} - m_b \mu_0}{\sigma_0 \sqrt{m_b + 1}} \right) \leq 2 \exp \left( -\frac{2 \mu_0^2 m_b^2 e^{2 \Theta_{\max}}}{\sigma_0^2 (m_b + 1)^2} \right) \to 0 \text{ as } b \to \infty$$

where \( Q(x) = \left( \frac{1}{\sqrt{2\pi}} \right) \int_x^{\infty} e^{-t^2/2} dt \) is the standard Q-function.

Recalling the condition in (34) on the window size and using the formula (44) for the worst-case expected delay, we obtain that if we set

$$m_b = \frac{1}{2\Theta_{\min}} \log b$$

then

$$\frac{4 \mu_0^2}{\sigma_0^2} m_b^3 e^{2 \Theta_{\max} m_b} \sim (\log b)^3 \Theta_{\max}/\Theta_{\min}.$$ 

Then Assumption 1 holds when \( \varepsilon = (1 + \delta) \Theta_{\max}/\Theta_{\min} \) with arbitrary \( \delta > 0 \).

Note that WADD\( \theta \left( \tau_G(b) \right) \leq \text{WADD}\( \theta \left( \tau_C(b) \right) \) for any threshold \( b > 0 \). In order to establish asymptotic optimality of the WL-GLR-CuSum procedure we need the following lemma that allows us to select threshold \( b = b_\alpha \) in such a way that the FAR of \( \tau_G(b) \) is controlled at least asymptotically.

**Lemma 3:** Suppose that the log-likelihood ratio \( \left\{ Z_{n,k} \right\}_{n \geq k} \) satisfies (50). Then, as \( b \to \infty \),

$$\text{FAR} \left( \tau_G(b) \right) \leq |\Theta_\alpha| C_d^{-1} b^2 \alpha^{-1-b} (1 + o(1)),$$

where \( C_d = \frac{\pi^{d/2}}{\Gamma(1+d/2)} \) is a constant that does not depend on \( \alpha \). Consequently, if \( b = b_\alpha \) satisfies equation

$$|\Theta_\alpha| C_d^{-1} b_\alpha^2 \alpha^{-1-b_\alpha} = \alpha,$$

then \( \text{FAR} \left( \tau_G(b_\alpha) \right) \leq \alpha(1 + o(1)) \) as \( \alpha \to 0 \).

**Remark:** Since \( |\Theta_\alpha| \leq |\Theta| < \infty \), it follows from (54) that \( b_\alpha \sim |\log \alpha| \) as \( \alpha \to 0 \).

The proof of Lemma 3 is given in the appendix. The following theorem establishes asymptotic optimality properties of the WL-GLR-CuSum detection procedure.

**Theorem 4:** Suppose that threshold \( b = b_\alpha \) is so selected that \( \text{FAR} \left( \tau_C(b_\alpha) \right) \leq \alpha \) or at least so that \( \text{FAR} \left( \tau_C(b_\alpha) \right) \leq \alpha(1 + o(1)) \) and \( b_\alpha \sim |\log \alpha| \) as \( \alpha \to 0 \), in particular from equation (54) in Lemma 3. Further, suppose that conditions (12), (13) and (50) hold for \( \left\{ Z_{n,k} \right\}_{n \geq k} \). Then, the WL-GLR-CuSum procedure \( \tau_G(b_\alpha) \) defined by (49) with the window size \( m_\alpha \) that satisfies the condition (41) solves first-order asymptotic optimization problems (48) uniformly for all parameter values \( \theta \in \Theta \), and

$$\text{SADD}_\theta \left( \tau_G(b_\alpha) \right) \sim \text{WADD}_\theta \left( \tau_G(b_\alpha) \right) \sim g_\theta^{-1}(|\log \alpha|), \forall \theta \in \Theta.$$

as \( \alpha \to 0 \).

**Proof:** Evidently, for any \( \theta \in \Theta \) and any threshold \( b > 0 \),

$$\text{WADD}_\theta \left( \tau_G(b) \right) \leq \text{WADD}_\theta \left( \tau_C(b) \right),$$

$$\text{SADD}_\theta \left( \tau_G(b) \right) \leq \text{SADD}_\theta \left( \tau_C(b) \right).$$

Let \( b = b_\alpha \) be so selected that \( \text{FAR} \left( \tau_G(b_\alpha) \right) \leq \alpha \) and \( b_\alpha \sim |\log \alpha| \) as \( \alpha \to 0 \). Then it follows from the asymptotic approximations (43) in Theorem 3 that, as \( \alpha \to 0 \),

$$\text{SADD}_\theta \left( \tau_G(b_\alpha) \right) \leq \text{WADD}_\theta \left( \tau_G(b_\alpha) \right) \leq g_\theta^{-1}(|\log \alpha|)(1 + o(1)).$$

Comparing these asymptotic inequalities with the asymptotic lower bound (20) in Theorem 1, immediately yields (55), which is asymptotically the best one can do to first order according to Theorem 1.

In particular, if \( b_\alpha \) is found from equation (54), then \( b_\alpha \sim |\log \alpha| \) as \( \alpha \to 0 \) and by Lemma 3 \( \text{FAR} \left( \tau_G(b_\alpha) \right) \leq \alpha(1 + o(1)) \), and therefore the assertions hold.

**IV. EXTENSIONS TO POINTWISE OPTIMALITY AND DEPENDENT NON-HOMOGENEOUS MODELS**

The measure of FAR that we have used in this paper (see (4)) is the inverse of the MTFA. However, the MTFA is a good measure of the FAR if, and only if, the pre-change distributions of the WL-CuSum stopping time \( \tilde{\tau}_G(b) \) and the WL-GLR-CuSum stopping time \( \tilde{\tau}_C(b) \) are approximately geometric. While this geometric property can be established for i.i.d. data models (see, e.g., Pollak and Tartakovsky [20] and Yakir [21]), it is not necessarily true for non-homogeneous and dependent data, as discussed in Mei [22] and Tartakovsky [23]. Therefore, in general, the MTFA is not appropriate for measuring the FAR. In fact, large values of MTFA may not necessarily guarantee small values of the probability of false alarm as discussed in detail in [1] and [23]. When the post-change model is Gaussian non-stationary as defined in Example 1, the MTFA may still be an appropriate measure for false alarm rate, as shown in the simulation study in Section V-C. Based on this result we conjecture that the MTFA-based FAR constraint may be suitable for other independent and non-stationary data models as well. However, in general, this may not be the case, and a more appropriate measure of the FAR in the general case may be the maximal (local) conditional probability of false alarm in the time interval \( [k, k + m] \) defined as [1]:

$$\text{SPFA}_m(\tau) = \sup_{k \geq 0} \mathbb{P}_\infty \left\{ \tau \leq k + m \mid |\tau > k \right\}.$$
asymptotic pointwise optimality as $\beta \to 0$ (i.e., for all change points $\nu \geq 1$) of the Shiryaev-Roberts (SR) procedure for the simple post-change hypothesis, and the mixture SR for the composite post-change hypothesis in class $C_{\beta,m}$, when $m = m_{\beta}$ depends on $\beta$ and goes to infinity as $\beta \to 0$ at such a rate that $\log m_{\beta} = o(\log \beta)$.

The results of [14] and [24] can be readily extended to the asymptotically non-homogeneous case where the function $g(n)$ increases with $n$ faster than $\log n$. In particular, using the developed in [14] and [24] techniques based on embedding class $C_{\beta,m}$ in the Bayesian class with a geometric prior distribution for the change point and the upper-bounded weighted PFA, it can be shown that the WL-CuSum procedure (29) with $m_{\alpha}$ replaced by $m_{\beta}$ is first-order pointwise asymptotically optimal in class $C_{\beta,m_{\beta}} = C_{\beta}$ as long as the uniform complete version of the strong law of large numbers for the log-likelihood ratio holds, i.e., for all $\delta > 0$

$$\sum_{n=1}^{\infty} \sup_{\nu \geq 1} \left\{ \frac{1}{g_\nu(n)} \sum_{i=\nu}^{\nu+n-1} Z_{i,\nu} - 1 \right\} > \delta \} < \infty,$$

where in the general non-i.i.d. case the partial LLR $Z_{i,\nu}$ is

$$Z_{i,\nu} = \log \frac{p_{0,i,\nu}(X_i|X_1,\ldots,X_{i-1})}{p_{0,i}(X_i|X_1,\ldots,X_{i-1})}.$$

Specifically, it can be established that for all fixed $\nu \geq 1$, as $\beta \to 0$,

$$\inf_{\tau \in C_{\beta}} \text{ADD}_{\nu}(\tau) \sim \text{ADD}_{\nu}(\tau_{C_{\beta}(h_0)}) \sim g^{-1}(\log \beta),$$

where we used the notation $\text{ADD}_{\nu}(\tau) = E_{\nu}[\tau - \nu|\tau \geq \nu]$ for the conditional average delay to detection. Similar results also hold for the maximal average detection delays WADD $(\tau)$ and SADD $(\tau) = \sup_{\nu \geq 1} \text{ADD}_{\nu}(\tau)$.

It is worth noting that it follows from the proof of Theorem 1 that under condition (12) the following asymptotic lower bound holds for the average detection delay $\text{ADD}_{\nu}(\tau)$ uniformly for all values of the change point in class $C_{\beta}$:

$$\inf_{\tau \in C_{\beta}} \text{ADD}_{\nu}(\tau) \geq g^{-1}(\log \beta)(1 + o(1)), \quad \forall \nu \geq 1$$

as $\beta \to 0$.

In the case where the post-change observations have parameter uncertainty, sufficient conditions for the optimality of the WL-GLR-CuSum procedure are more sophisticated – a probability in the vicinity of the true post-change parameter should be involved [14].

Further details and the proofs are omitted and will be given elsewhere.

V. Numerical Results

A. Performance Analysis for GEM Problem

In Fig. 1, we study the performance of the proposed WL-CuSum procedure in (29) through Monte Carlo (MC) simulations for the Gaussian exponential mean-change detection problem (see Example 1), with known post-change parameter $\theta$. The change-point is taken to be $\nu = 1.3$.

We consider the performance of the WL-CuSum procedure (29) tuned to the true post-change parameter and the WL-GLR-CuSum procedure (49) where only the set of post-change parameter values is known. It is seen that the operating characteristic of the WL-GLR-CuSum procedure is close to that of the WL-CuSum procedure for a sufficiently large window-size. We also observe that procedures with slightly insufficiently large window-sizes perform similarly to those with sufficiently large window sizes.

Fig. 1. Performances of the WL-CuSum with different window-sizes for the Gaussian exponential mean-change detection problem with $\mu_0 = 0.1$, $\sigma_\nu^2 = 10000$, and $\theta = 0.4$. The change-point is $\nu = 1$. The window-sizes are considered, with the window size of 12 being smaller than the range expected delay values in the plot, and therefore not large enough to satisfy condition (30). The window size of 25 is sufficiently large, and the window size of 100 essentially corresponds to having no window at all. It is seen that the performance is nearly identical for all window sizes considered. We also observe that the expected delay is $O(\log(\log \alpha))$, which matches our theoretical analysis in (44).

In Fig. 2, we compare, also through MC simulations for the problem of Example 1, the performance of the WL-CuSum procedure (29) tuned to the true post-change parameter and the WL-GLR-CuSum procedure (49) where only the set of post-change parameter values is known. It is seen that the operating characteristic of the WL-GLR-CuSum procedure is close to that of the WL-CuSum procedure for a sufficiently large window-size. We also observe that procedures with slightly insufficiently large window-sizes perform similarly to those with sufficiently large window sizes.
B. Performance Analysis for Gaussian Observations With Decaying Post-Change Mean

In this subsection, we apply the WL-CuSum and WL-GLR-CuSum procedures for the QCD problem with Gaussian observations, where the post-change mean gradually decays to the pre-change mean. Specifically,

\[ X_n \sim N(0, \sigma^2), \forall n < \nu \]
\[ X_n \sim N(\mu_1(n - \nu + 1)^{-\theta}, \sigma^2), \forall n \geq \nu \]  

(56)

for some decay parameter \( \theta \in (0, 1/2) \). The growth function for this model is

\[ g_{\nu, \theta}(n) = \sum_{i=\nu}^{n+1} E_{\nu, \theta} \left[ z_{i,\nu}^\theta \right] = \sum_{i=1}^{n} \frac{\mu_1^2}{2\sigma^2} \cdot i^{-2\theta} \]
\[ = \frac{\mu_1^2}{2\sigma^2(1 - 2\theta)} n^{1-2\theta}(1 + o(1)), \text{ as } n \to \infty, \]

and thus,

\[ g^{-1}(x) = (2\sigma^2 \mu_1^{-2}(1 - 2\theta)x) \cdot \left(1 + o(1)\right) \]  

(57)

Therefore, \( \log g^{-1}(x) = O(\log x) = o(x) \) and condition (11) is satisfied. Also note that since \((1 - 2\theta)^{-1} > 1\), the optimal WADD and CADD are asymptotically super-linear with \(|\log \alpha|\).

In Fig. 3, we study the performance of the proposed WL-CuSum procedure (29) through MC simulations for the model (56), with known decay parameter \( \theta = 0.35 \). The change-point is taken to be \( \nu = 1 \). Four window-sizes are considered, with the window sizes of 15 and 25 being smaller than the range of expected delay values in the plot, and therefore, not large enough to satisfy condition (30). It is seen that the performance improves significantly with an initial increase of window-size, with diminishing returns when the window-size becomes large enough. We also observe that the expected delay is super-linear with \(|\log \alpha|\), which matches our theoretical analysis in (57).

In Fig. 4, we compare, also through MC simulations for the model in (56), the performance of the WL-CuSum procedure (29) tuned to the true post-change parameter and the WL-GLR-CuSum procedure (49) where only the set of post-change parameter values is known. We observe that the operating characteristic of WL-GLR-CuSum procedure is nearly identical to that of the WL-CuSum procedure for large enough window-size.

C. Analysis of MTFA as False Alarm Measure

In Fig. 5, we study the distribution of the WL-CuSum stopping times under \( P_{\infty} \) using simulation results from the Gaussian exponential mean-change detection problem with \( \mu_0 = 0.1, \sigma_0^2 = 10000, \) and \( \theta = 0.4 \). In all subplots, the x-axis shows the theoretical quantiles of the best-fit geometric distribution and the y-axis shows the experimental quantiles of distributions of the stopping times. The first row corresponds to WL-CuSum procedure (29) and the second row corresponds to the full-history CuSum procedure (28).
appropriate false alarm performance measure, and our measure of FAR as the reciprocal of the MTFA is justified.

D. Application: Monitoring COVID-19 Second Wave

Next, we apply the developed WL-GLR-CuSum algorithm to monitoring the spread of COVID-19 using new case data from various counties in the US [25]. The goal is to detect the onset of a new wave of the pandemic based on the incremental daily cases. The problem is modeled as one of detecting a change in the mean of a Beta distribution as in [26]. Let $B(x; a, b)$ denote the density of the Beta distribution with shape parameters $a$ and $b$, i.e.,

$$B(x; a, b) = \frac{x^{a-1}(1-x)^{b-1} \Gamma(a+b)}{\Gamma(a)\Gamma(b)}, \forall x \in [0, 1],$$

where $\Gamma$ represents the gamma function. Note that the mean of an observation under density $B(x; a, b)$ is $a/(a+b)$. Let

$$p_0(x) = B(x; a_0, b_0),$$

$$p_1(x, n, k) = B(x; a_0 h_0(n-k), b_0), \quad \forall n \geq k. \quad (58)$$

Here, $h_0$ is a function such that $h_0(x) \geq 1, \forall x > 0$. Note that if $a_0 < b_0$ and $h_0(n-\nu)$ is not too large,

$$E_{\nu, \theta}|X_n| = \frac{a_0 h_0(n-\nu)}{a_0 h_0(n-\nu) + b_0} \approx \frac{a_0}{b_0} h_0(n-\nu) \quad (59)$$

for all $n \geq \nu$. We design $h_\theta$ to capture the behavior of the average fraction of daily incremental cases. In particular, we model $h_\theta$ as

$$h_\theta(x) = 1 + \frac{10^{\theta_0}}{\theta_2} \exp \left( -\frac{(x - \theta_1)^2}{2\theta_2^2} \right), \quad (60)$$

where $\theta_0, \theta_1, \theta_2 \geq 0$ are the model parameters and $\theta = (\theta_0, \theta_1, \theta_2) \in \Theta$. When $n - \nu$ is small, $h_\theta(n-\nu)$ grows like the left tail of a Gaussian density, which matches the exponential growth in the average fraction of daily incremental cases seen at the beginning of a new wave of the pandemic. Also, $h_\theta(n-\nu) \to 0$ as $n \to \infty$, which corresponds to the daily incremental cases eventually vanishing at the end of the pandemic. In Fig. 6, we validate the choice of distribution model defined in (58) using data from COVID-19 wave of Fall 2020. In the simulation, $a_0$ and $b_0$ are estimated using observations from previous periods in which the increments remain low and roughly constant. It is observed that the mean of the daily fraction of incremental cases matches well with the mean of the fitted Beta distribution with $h_\theta$ in (60).

Note that the growth condition (47) required for our asymptotic analysis is not satisfied for the observation model (58) with $h_\theta$ given in (60). Nevertheless, we expect the WL-GLR-CuSum procedure to perform as predicted by our analysis if the procedure stops during a time interval where $h_\theta$ is still increasing, which is what we would require of a useful procedure for detecting the onset of a new wave of the pandemic anyway.

In Fig. 7, we illustrate the use the WL-GLR-CuSum procedure with the distribution model (58) for the detection of the onset of a new wave of COVID-19. We assumed a start date of June 15th, 2021 for the monitoring, at which time the pandemic appeared to be in a steady state with incremental cases staying relatively flat. We observe that the WL-GLR-CuSum statistic significantly and persistently crosses the detection threshold around late July in all counties, which is strong indication of a new wave of the pandemic. More importantly, unlike the raw observations which are highly varying, the WL-GLR-CuSum statistic shows a clear dichotomy between the pre- and post-change settings, with the statistic staying near zero before the purported onset of the new wave, and taking off very rapidly (nearly vertically) after the onset.

VI. CONCLUSION

We considered the problem of the quickest detection of a change in the distribution of a sequence of independent observations, assuming that the pre-change observation are
stationary with known distribution, while the post-change observations are non-stationary with possible parametric uncertainty. Specifically, we assumed that the cumulative KL divergence between the post-change and the pre-change distributions grows at least logarithmically after the change point. We derived a universal asymptotic lower bound on the worst-case expected detection delay under a constraint on the false alarm rate in this non-stationary setting, which had been previously derived only in the asymptotically stationary setting. We showed that the developed WL-CuSum procedure for known post-change distribution, as well as the developed WL-GLR-CuSum procedure for the unknown post-change parameters, asymptotically achieve the lower bound on the worst-case expected detection delay, as the false alarm rate goes to zero. We validated these theoretical results through numerical Monte-Carlo simulations. We also demonstrated that the proposed WL-GLR-CuSum procedure can be effectively used in monitoring pandemics. We provided in Section IV some possible avenues for future research, in particular, those allowing for dependent observations and more general false alarm constraints.

APPENDIX

Proof Proof of Lemma 1: For the first inequality, fix \( \nu \geq 1 \) and \( \delta > 0 \). Note that
\[
g_{\nu}(n) = \sum_{i=\nu}^{\nu+n-1} \mathbb{E}_{\nu}[Z_{i,\nu}].
\]
Since \( \mathbb{E}_{\nu}[Z_{i,\nu}] > 0 \) for all \( i \geq \nu \), for any \( t \leq n \),
\[
g_{\nu}(n) = \sum_{t=\nu}^{\nu+n-1} \mathbb{E}_{\nu}[Z_{t,\nu}] \geq g_{\nu}(t)
\]
and, by definition,
\[
\mathbb{E}_{\nu}\left[\sum_{i=\nu}^{\nu+n-1} Z_{i,\nu} - g_{\nu}(t)\right] = 0.
\]
Thus, for an arbitrary \( \nu > 1 \) we have
\[
\mathbb{P}_{\nu}\left\{ \max_{t \leq n} \sum_{i=\nu}^{\nu+n-1} Z_{i,\nu} \geq (1 + \delta)g_{\nu}(n) \right\} \leq \mathbb{P}_{\nu}\left\{ \max_{t \leq n} \sum_{i=\nu}^{\nu+n-1} Z_{i,\nu} - g_{\nu}(t) \geq \delta g_{\nu}(n) \right\} \leq \mathbb{P}_{\nu}\left\{ \max_{t \leq n} \left| \sum_{i=\nu}^{\nu+n-1} Z_{i,\nu} - g_{\nu}(t) \right| \geq \delta g_{\nu}(n) \right\} \leq 1 \sup_{t \geq \nu, t \geq 1} \frac{1}{\delta^2 g_{\nu}^2(n)} \sum_{i=\nu}^{\nu+n-1} \mathbb{V}_{\nu}(Z_{i,\nu}) = 1 \sup_{t \geq \nu, t \geq 1} \frac{1}{\delta^2 g_{\nu}^2(n)} \sum_{i=\nu}^{\nu+n-1} \mathbb{V}_{\nu}(Z_{i,\nu}) (62)
\]
where \((*)\) follows from Kolmogorov’s inequality and the last line follows by independence. Hence,
\[
\sup_{\nu \geq 1} \mathbb{P}_{\nu}\left\{ \max_{t \leq n} \sum_{i=\nu}^{\nu+n-1} Z_{i,\nu} \geq (1 + \delta)g_{\nu}(n) \right\} \leq \frac{1}{\delta^2} \sup_{t \geq \nu, t \geq 1} \frac{1}{g_{\nu}^2(n)} \sum_{i=\nu}^{\nu+n-1} \mathbb{V}_{\nu}(Z_{i,\nu}) n^{n-\infty} \rightarrow 0
\]
where the limit follows from condition (14).
For the second inequality, fix \( \nu \) and \( t \) such that \( t \geq \nu \geq 1 \). For any \( \delta \in (0, 1) \), we have
\[
\mathbb{P}_{\nu}\left\{ \sum_{i=t}^{t+n-1} Z_{i,\nu} \leq (1 - \delta)g_{\nu}(n) \right\} = \mathbb{P}_{\nu}\left\{ \sum_{i=t}^{t+n-1} (Z_{i,\nu} - \mathbb{E}_{\nu}[Z_{i,\nu}]) \leq -\delta g_{\nu}(n) \right\} + \sum_{i=\nu}^{\nu+n-1} \mathbb{E}_{\nu}(Z_{i,\nu}) - \sum_{i=t}^{t+n-1} \mathbb{E}_{\nu}(Z_{i,\nu}) \leq 0 \text{ by (15)}
\]
\[
\leq \mathbb{P}_{\nu}\left\{ \sum_{i=t}^{t+n-1} Z_{i,\nu} - \mathbb{E}_{\nu}[Z_{i,\nu}] \right\} \leq -\delta g_{\nu}(n) \right\} \leq \mathbb{P}_{\nu}\left\{ \sum_{i=t}^{t+n-1} Z_{i,\nu} - \mathbb{E}_{\nu}[Z_{i,\nu}] \right\} \geq \delta g_{\nu}(n) \right\} \leq \mathbb{P}_{\nu}\left\{ \frac{1}{\delta^2 g_{\nu}^2(n)} \sum_{i=t}^{t+n-1} \mathbb{V}_{\nu}(Z_{i,\nu}) \right\}
\]
(62)

The proof is complete. \( \square \)

Proof Proof of Lemma 3: Let \( \lambda_{n,k}^\theta = \sum_{i=k}^n Z_{i,k}^\theta \) denote the log-likelihood ratio between the hypotheses that \( \nu = k \) with the parameter \( \theta \) against \( \nu = \infty \) in the sample \( (X_1, \ldots, X_n) \). We rewrite the definition of \( \tilde{\tau}_G(b) \) in (49) as:
\[
\tilde{\tau}_G(b) := \inf \left\{ n : \max_{n-m \leq k \leq n} \lambda_{n,k}^\theta \geq b \right\}
\]
where \( \hat{\theta}_{n,k} \in \Theta_b \) solves
\[
\sup_{\theta \in \Theta_b} \sum_{i=k}^n Z_{i,k}^\theta
\]
(64)
for a given pair \( (k, n) \) where \( k \leq n \). Note that now instead of \( \Theta_\alpha \) we use the notation \( \Theta_b \), which is a compact subset of \( \Theta \).
Let \( \Pi \) be a probability measure on \( \Theta_b \). Recall that \( \mathcal{F}_n = \sigma(X_1, \ldots, X_n) \) and \( \mathcal{F}_\infty = \sigma(X_1, X_1, \ldots) \). Given this mixing distribution over \( \Theta_b \), define \( Q_k \) by
\[
Q_k(A) = \int_{\Theta_b} P_{\theta, k}(A) \Pi(d\theta), \quad A \in \mathcal{F}_\infty.
\]
Then \( Q_k \) is easily seen to be a probability measure. Moreover, letting \( P_{\infty,n}^k \) and \( Q_k^n \) denote the restrictions of \( P_{\infty,n} \) and \( Q_k \) to the sigma-algebra \( \mathcal{F}_n \), introduce the likelihood ratio
\[
L_{n,k} := \frac{dQ_k^n}{dP_{\infty,n}^k} = \int_{\Theta_b} \exp(\lambda_{n,k}^\theta) \Pi(d\theta), \quad n \geq k.
\]
Further, let

\[ R_n := \sum_{k=1}^n L_{n,k}. \]

Obviously, \( E_\infty [L_{n,k} | F_{n-1}] = L_{n-1,k} \) and \( E_\infty [L_{n,k}] = 1 \), so \( \{L_{n,k}\}_{n,k \geq 1} \) is a \((\mathbb{P}_\infty, F_n)\)-martingale with unit expectation. Hence, \( E_\infty [R_n | F_{n-1}] = 1 + R_{n-1} \) and the statistic \( R_n - \sum_{k=1}^{n-1} L_{n,k} \) is a zero-mean \((\mathbb{P}_\infty, F_n)\)-martingale. By the optional sampling theorem (see, e.g., [1, Th 2.3.1, page 31]), for any proper stopping time \( \tau \), \( E_\infty [R_\tau] = E_\infty [\tau] \), and in particular, \( E_\infty [R_{\hat{T}_C(b)}] = E_\infty [\hat{T}_C(b)] \).

Now, set \( \Pi(d\theta) = |\Theta_b|^{-1} d\theta \) (uniform on \( \Theta_b \)). At the next step we show that as \( b \to \infty \)

\[ E_\infty [R_{\hat{T}_C(b)}] \geq |\Theta_b|^{-1} e^{-1} \frac{\pi^{d/2}}{(1 + d/2)} b^{-d/2} e^b (1 + o(1)), \]

which along with the previous argument implies that

\[ E_\infty [\hat{T}_C(b)] \geq |\Theta_b|^{-1} e^{-1} \frac{\pi^{d/2}}{(1 + d/2)} b^{-d/2} e^b (1 + o(1)). \]

This inequality implies inequality (53). Thus, it remains to prove the asymptotic inequality (65).

By assumption, \( \tilde{\theta}_{n,k} \) lies in the interior of \( \Theta_b \) (for sufficiently large \( b \)). Using Taylor’s expansion, for any \( k \leq n \) and \( \theta \in \Theta_b \),

\[
\begin{align*}
\lambda_{n,k}^\theta &= \lambda_{n,k}^{\tilde{\theta}_{n,k}} + \nabla_\theta \lambda_{n,k}^{\tilde{\theta}_{n,k}} |_{\theta = \tilde{\theta}_{n,k}} (\theta - \tilde{\theta}_{n,k}) + \frac{1}{2} (\theta - \tilde{\theta}_{n,k})^T \nabla^2_{\theta} \lambda_{n,k}^{\tilde{\theta}_{n,k}} |_{\theta = \tilde{\theta}_{n,k}} (\theta - \tilde{\theta}_{n,k}) \\
&= \lambda_{n,k}^{\theta_{n,k}} + \frac{1}{2} (\theta - \tilde{\theta}_{n,k})^T \nabla^2_{\theta} \lambda_{n,k}^{\tilde{\theta}_{n,k}} |_{\theta = \theta^*_{n,k}} (\theta - \tilde{\theta}_{n,k})
\end{align*}
\]

where \( \theta^*_{n,k} \) is an intermediate point between \( \theta \) and \( \tilde{\theta}_{n,k} \), i.e., \( \theta^*_{n,k} = \rho \theta + (1 - \rho) \tilde{\theta}_{n,k} \) for some \( \rho \in (0, 1) \). The last equality follows from (64). This further implies that

\[
\lambda_{n,k}^\theta - \lambda_{n,k}^{\tilde{\theta}_{n,k}} \geq -\frac{1}{2} \lambda_{n,k}^{\theta_{n,k}} \left( -\nabla^2_{\theta} \lambda_{n,k}^{\tilde{\theta}_{n,k}} |_{\theta = \theta^*_{n,k}} \right) (\theta - \tilde{\theta}_{n,k}, \tilde{\theta}_{n,k}) \geq -\frac{1}{2} \Lambda_{n,k} \| \theta - \tilde{\theta}_{n,k} \|^2,
\]

where

\[ \Lambda_{n,k} := \sup_{\theta : \| \theta - \tilde{\theta}_{n,k} \| < b^{-\frac{d}{2}}} \lambda_{n,k}^{\theta_{n,k}} \left( -\nabla^2_{\theta} \sum_{i=1}^{n} Z_i^{\theta_{n,k}} \right). \]

Fix \( \varepsilon > 0 \) such that Assumption 1 is satisfied. Write \( S(b) := \{ \theta : \| \theta - \tilde{\theta}_{n,k} \| < b^{-\frac{d}{2}} \} \). Since \( S(b) \) \( \subset \) \( \Theta_b \) \( \setminus \) while \( \Theta_b \) \( \setminus \) \( \Theta \) as \( b \to \infty \), it follows that \( S(b) \subset \Theta_b \) for all sufficiently large \( b \).

Denote \( b_0 := \inf \{ b > 0 : S(b) \subset \Theta_b \} \). For any \( b > b_0 \), we have

\[
L_{n,k} e^{-\lambda_{n,k}}^{\tilde{\theta}_{n,k}} \geq |\Theta_b|^{-1} \int_{S(b)} \exp \left( \lambda_{n,k}^{\theta} - \lambda_{n,k}^{\tilde{\theta}_{n,k}} \right) d\theta \]

\[
\geq |\Theta_b|^{-1} \int_{S(b)} \exp \left( \lambda_{n,k}^{\theta} - \lambda_{n,k}^{\tilde{\theta}_{n,k}} \right) d\theta \]

\[
\geq |\Theta_b|^{-1} \int_{S(b)} \exp \left( -\frac{1}{2} \Lambda_{n,k} \| \theta - \tilde{\theta}_{n,k} \|^2 \right) d\theta \]

\[
\geq |\Theta_b|^{-1} \int_{S(b)} \exp \left( -\frac{1}{2} \Lambda_{n,k} b^{-\varepsilon} \right) d\theta \]

\[
\geq |\Theta_b|^{-1} \exp \left( -\frac{1}{2} \Lambda_{n,k} b^{-\varepsilon} \right) C_d b^{-\frac{d}{2}}, \]

where \( C_d := \frac{\pi^{d/2}}{(1 + d/2)} b^{-d/2} \). The last inequality follows because the volume of a \( d \)-dimensional ball with radius \( b^{-\frac{d}{2}} \) is given by \( \frac{\pi^{d/2}}{(1 + d/2)} b^{-d/2} \), where \( \Gamma(\cdot) \) is the gamma function.

Therefore,

\[ L_{n,k} \geq \exp \left( \lambda_{n,k}^{\tilde{\theta}_{n,k}} \right) |\Theta_b|^{-1} \exp \left( -\frac{1}{2} \Lambda_{n,k} b^{-\varepsilon} \right) C_d b^{-\frac{d}{2}}. \]

Write

\[ V_n := \max_{n-m_b \leq k \leq n} \exp \left( \lambda_{n,k}^{\tilde{\theta}_{n,k}} \right) \]

and note that \( V_{\hat{T}_C(b)} \geq e^b \) on \{ \hat{T}_C(b) < \infty \} (by the definition of the stopping time \( \hat{T}_C(b) \) in (63)). It follows that

\[ E_\infty [R_{\hat{T}_C(b)}] = E_\infty \left[ \sum_{k=1}^{\hat{T}_C(b)} L_{n,k} \right] \]

\[ \geq |\Theta_b|^{-1} C_d b^{-\frac{d}{2}} \]

\[ E_\infty \left[ \max_{1 \leq k \leq \hat{T}_C(b)} \exp \left( \lambda_{n,k}^{\tilde{\theta}_{n,k}} \right) \exp \left( -\frac{1}{2} \Lambda_{n,k} b^{-\varepsilon} \right) \right] \]

\[ \geq |\Theta_b|^{-1} C_d b^{-\frac{d}{2}} \]

\[ E_\infty \left[ \exp \left( -\frac{1}{2} \Lambda_{n,k} b^{-\varepsilon} \right) \max_{m_b \leq k \leq \hat{T}_C(b)} L_{n,k} \right] \]

\[ \geq \left[ \exp \left( -\frac{1}{2} \Lambda_{n,k} b^{-\varepsilon} \right) \right] e^b. \]

By Assumption 1,

\[ \lim_{b \to \infty} \mathbb{P} \left\{ \max_{n-m_b \leq k \leq n} \Lambda_{n,k} \leq 2 b^\varepsilon \right\} = 1, \]

and therefore, as \( b \to \infty \),

\[ \mathbb{E}_\infty \left[ \exp \left( -\frac{1}{2} \Lambda_{n,k} b^{-\varepsilon} \right) \right] \]

which along with inequality (69) yields inequality (65).

ACKNOWLEDGMENT

The authors would like to thank Don Towsley for suggesting the example in (56). They are also grateful to the anonymous referees for comments that improved the article.

REFERENCES

[1] A. G. Tartakovsky, I. Nikiforov, and M. Basseville, Sequential Change Detection and Hypothesis Testing General Non-I.I.D. Stochastic Models and Asymptotically Optimal Rules (Monographs on Statistics and Applied Probability), vol. 165. Boca Raton, FL, USA: Chapman & Hall, 2015.

[2] A. G. Tartakovsky, Sequential Change Detection and Hypothesis Testing (Monographs on Statistics and Applied Probability), vol. 136. Boca Raton, FL, USA: Chapman & Hall, 2015.
