

Non-Parametric Quickest Detection of a Change in the Mean of an Observation Sequence

Yuchen Liang

*ECE Department and Coordinated Science Lab
University of Illinois at Urbana-Champaign
Urbana, IL
yliang35@illinois.edu*

Venugopal V. Veeravalli

*ECE Department and Coordinated Science Lab
University of Illinois at Urbana-Champaign
Urbana, IL
vvv@illinois.edu*

Abstract—We study the problem of quickest detection of a change in the mean of an observation sequence, under the assumption that both the pre- and post-change distributions have bounded support. We first study the case where the pre-change distribution is known, and then study the extension where only the mean and variance of the pre-change distribution are known. In both cases, no knowledge of the post-change distribution is assumed other than that it has bounded support. For the case where the pre-change distribution is known, we derive a test that asymptotically minimizes the worst-case detection delay over all post-change distributions, as the false alarm rate goes to zero. We then study the limiting form of the optimal test as the gap between the pre- and post-change means goes to zero, which we call the Mean-Change Test (MCT). We show that the MCT can be designed with only knowledge of the mean and variance of the pre-change distribution. We validate our analysis through numerical results for detecting a change in the mean of a beta distribution. We also demonstrate the use of the MCT for pandemic monitoring.

Index Terms—Quickest change detection (QCD), non-parametric methods, minimax robust detection

I. INTRODUCTION

Quickest Change Detection (QCD) is a fundamental problem in mathematical statistics. Given a stochastic sequence whose distribution changes at some unknown change-point, the goal is to detect the change after it occurs as quickly as possible, subject to false alarm constraints. The QCD framework has had a wide range of applications, including but not limited to line-outage in power systems [1], dim-target manoeuvre detection [2], stochastic process control [3], structural health monitoring [4], and, more recently, in the Multi-Armed Bandit (MAB) problem with piece-wise stationary bandits [5]. Two formulations are used in the classical QCD problem: the Bayesian formulation [6], where the change-point is assumed to follow some prior distribution, and the minimax formulation [7], [8], where the worst-case detection delay is minimized over all possible change-points, subject to false alarm constraints. In both the Bayesian and minimax settings, if the pre- and post-change distributions are known,

This work was supported in part by the National Science Foundation under grant ECCS-2033900, and by the Army Research Laboratory under Cooperative Agreement W911NF-17-2-0196, through the University of Illinois at Urbana-Champaign.

low-complexity efficient solutions to the QCD problem can be found [9].

In many practical situations, we may not know the exact distribution in the pre- or post-change regimes. Although it is reasonable to assume that we can obtain a large amount of data in the pre-change regime, this may not be the case for the post-change regime. In applications such as the epidemic detection problem and the MAB problem, a change in a specific statistic (e.g., the mean) of the distribution is of interest. This is different from the original QCD problem where any distributional change needs to be detected. Furthermore, in many applications, the support of the distribution is bounded. For example, the observations representing the fraction of some specific group in the entire population are bounded between 0 and 1. This is the case, for example, in the pandemic monitoring problem that we discuss in detail in Section IV.

There have been a number of lines of work on the QCD problem when the pre- and/or post-change distributions are not completely known. The most prevalent is the Generalized Likelihood Ratio (GLR) approach, in which the maximum of the test statistics corresponding to each possible change-points is used to construct the test. A GLR approach for a Gaussian mean-change detection problem is proposed in [10]. A GLR test for the case where the pre- and post-change distributions come from one-parameter exponential families is analyzed in [11]. A kernel approach to construct GLR statistics is discussed in [12]. More recently, a GLR mean-change test for sub-Gaussian observations using scan statistics is proposed and analyzed in [13], and this approach is applied to the MAB problem in [5]. In all of these methods, the complexity of computing the test statistic at each time-step grows at least linearly with the number of samples. In practice, a windowed version of the GLR test statistic is used to keep the complexity small, while suffering some loss in performance. An alternative to the GLR approach was given in [14], where a histogram is used to estimate the post-change distribution, and a low complexity test with a recursive structure is constructed.

Another line of work is the one based on a *minimax* robust approach [15], in which it is assumed that the distributions come from mutually exclusive uncertainty classes. Under certain conditions on the uncertainty classes, e.g., joint stochastic boundedness [9], a low-complexity saddle-point solution to

the minimax robust QCD problem can be found. In this paper, we use an asymptotic version of the minimax robust QCD problem formulation to develop algorithms for the non-parametric detection of a change in mean of an observation sequence. Our contributions are as follows:

- 1) We study the problem of quickest detection of a change in the mean of an observation sequence under the assumption that no knowledge of the post-change distribution is available other than that it has bounded support.
- 2) For the case where the pre-change distribution is known, we derive a test that asymptotically minimizes the worst-case detection delay over all possible post-change distributions, as the false alarm rate goes to zero.
- 3) We study the limiting form of the optimal test as the gap between the pre- and post-change means goes to zero, which we call the Mean-Change Test (MCT). We show that the MCT can be designed with only knowledge of the mean and variance of the pre-change distribution. We also study the performance of the MCT when the gap is moderate.
- 4) We validate our analysis through numerical results for detecting a change in the mean of a beta distribution. We also demonstrate the use of the MCT for pandemic monitoring.

II. PROBLEM STATEMENT

Let $X_1, \dots, X_t, \dots \in [0, 1]$ be independent samples, with $X_1, \dots, X_\nu \sim P_0$, and $X_{\nu+1}, \dots \sim P_1$. Let $\mathbb{P}_\nu^{P_0, P_1} \{\cdot\}$ denote the probability measure on the entire sequence of observations when the pre- and post-change distributions are P_0 and P_1 , respectively, and the change-point is at $t = \nu$, and let $\mathbb{E}_\nu^{P_0, P_1} [\cdot]$ denote the corresponding expectation. We will first consider the case where P_0 is completely known, and then study extensions to the cases where only partial (moment) information about P_0 is available. The change-time ν is unknown but deterministic. The problem is to detect the change quickly while not causing too many false alarms. Let τ be a stopping time [9] defined on the observation sequence associated with the detection rule, i.e. τ is the time at which we stop taking observations and declare that the change has occurred. If both the pre- and post-change distributions are known, Lorden [7] proposed solving the following optimization problem to find the best stopping time τ :

$$\inf_{\tau \in \mathcal{C}_\alpha} \text{WADD}^{P_0, P_1} (\tau) \quad (1)$$

where

$$\begin{aligned} \text{WADD}^{P_0, P_1} (\tau) := \\ \sup_{\nu \geq 1} \text{ess sup} \mathbb{E}_\nu^{P_0, P_1} [(\tau - \nu + 1)^+ | X_1, \dots, X_{\nu-1}] \end{aligned} \quad (2)$$

is a worst-case delay metric and

$$\mathcal{C}_\alpha^{\{P_0\}} := \{\tau : \text{FAR}^{P_0} (\tau) \leq \alpha\} \quad (3)$$

is the feasible set where $\text{FAR}^{P_0} (\tau) := \mathbb{E}_\infty^{P_0, P_1} [\tau]^{-1}$. Note that $\mathbb{E}_\infty^{P_0, P_1} [\cdot]$ is the expectation operator when the change never happens, and $(\cdot)^+ := \max\{0, \cdot\}$.

Lorden also showed that Page's Cumulative Sum (CUSUM) algorithm [16] whose test statistic is given by:

$$\begin{aligned} \Lambda^{P_0, P_1} (t) = \max_{1 \leq k \leq t} \sum_{i=k}^t \ln L^{P_0, P_1} (X_i) \\ = (\Lambda^{P_0, P_1} (t-1) + \ln L^{P_0, P_1} (X_t))^+ \end{aligned} \quad (4)$$

with $\Lambda^{P_0, P_1} (0) = 0$ solves the problem in (1) asymptotically, where $L^{P_0, P_1} (x) = p_1(x)/p_0(x)$ is the likelihood ratio between the densities. The CUSUM stopping rule is given by:

$$\tau (\Lambda^{P_0, P_1}, b_\alpha) := \inf \{t : \Lambda^{P_0, P_1} (t) \geq b_\alpha\} \quad (5)$$

where $b_\alpha := |\ln \alpha|$. It was shown by Moustakides [17] that Page's test is exactly optimal for the problem in (1).

When the pre-change and post-change distributions are unknown but belong to some uncertainty sets, a minimax robust metric can be applied:

$$\inf_{\tau \in \mathcal{C}_\alpha^{\{P_0\}}} \sup_{(P_0, P_1) \in \mathcal{P}_0 \times \mathcal{P}_1} \text{WADD}^{P_0, P_1} (\tau) \quad (6)$$

where the feasible set becomes

$$\mathcal{C}_\alpha^{\mathcal{P}_0} = \left\{ \tau : \sup_{P_0 \in \mathcal{P}_0} \text{FAR}^{P_0} (\tau) \leq \alpha \right\} \quad (7)$$

A pair of uncertainty sets $(\mathcal{P}_0, \mathcal{P}_1)$ is said to be *jointly stochastically (JS) bounded* by $(\bar{P}_0, \bar{P}_1) \in \mathcal{P}_0 \times \mathcal{P}_1$ if, for any $(P_0, P_1) \in \mathcal{P}_0 \times \mathcal{P}_1$ and any $t > 0$,

$$P_0 \{L^{\bar{P}_0, \bar{P}_1} (X) > t\} \leq \bar{P}_0 \{L^{\bar{P}_0, \bar{P}_1} (X) > t\} \quad (8)$$

$$P_1 \{L^{\bar{P}_0, \bar{P}_1} (X) > t\} \geq \bar{P}_1 \{L^{\bar{P}_0, \bar{P}_1} (X) > t\} \quad (9)$$

where $L^{\bar{P}_0, \bar{P}_1}$ is the likelihood ratio between \bar{P}_1 and \bar{P}_0 as defined previously [9]. The distributions \bar{P}_0 and \bar{P}_1 are called least favorable distributions (LFDs) within the classes \mathcal{P}_0 and \mathcal{P}_1 , respectively. If the pair of pre- and post-change uncertainty sets is JS bounded, the test statistic $\Lambda^{\bar{P}_0, \bar{P}_1} (t)$ with the stopping rule $\tau (\Lambda^{\bar{P}_0, \bar{P}_1}, b_\alpha)$ solves (6) exactly [18].

A pair of uncertainty sets $(\mathcal{P}_0, \mathcal{P}_1)$ is said to be *weakly stochastically (WS) bounded* by $(\bar{P}_0, \bar{P}_1) \in \mathcal{P}_0 \times \mathcal{P}_1$ if

$$D(\tilde{P}_1 || \tilde{P}_0) \leq D(P_1 || \tilde{P}_0) - D(P_1 || \tilde{P}_1) \quad (10)$$

for all $P_1 \in \mathcal{P}_1$, and

$$\mathbb{E}^{P_0} \left[L^{\bar{P}_0, \bar{P}_1} (X) \right] \leq \mathbb{E}^{\bar{P}_0} \left[L^{\bar{P}_0, \bar{P}_1} (X) \right] = 1 \quad (11)$$

for all $P_0 \in \mathcal{P}_0$ [2], where $\mathbb{E}^P [\cdot]$ denotes the expectation operator with respect to distribution P . It is shown in [2] that if a pair of uncertainty sets is JS bounded by (\bar{P}_0, \bar{P}_1) , it is also WS bounded by (\bar{P}_0, \bar{P}_1) .

Theorem II.1. [2] If (\bar{P}_0, \bar{P}_1) solves

$$\inf_{(P_0, P_1) \in \mathcal{P}_0 \times \mathcal{P}_1} D(P_1 || P_0) \quad (12)$$

and if furthermore, \mathcal{P}_1 is convex, and (11) holds for all $P_0 \in \mathcal{P}_0$, then $(\mathcal{P}_0, \mathcal{P}_1)$ is WS bounded by (\bar{P}_0, \bar{P}_1) , and

$\tau(\Lambda^{\tilde{P}_0, \tilde{P}_1}, b_\alpha)$ solves the problem in (6) asymptotically as $\alpha \rightarrow 0$. Also, the worst-case delay is

$$\inf_{\tau \in \mathcal{C}_\alpha^{\mathcal{P}_0}} \sup_{(P_0, P_1) \in \mathcal{P}_0 \times \mathcal{P}_1} WADD^{P_0, P_1}(\tau) = \frac{|\ln \alpha|}{D(\tilde{P}_1 || \tilde{P}_0)} (1 + o(1)) \quad (13)$$

In this paper, we consider the problem of quickest change detection for observations with bounded support. The goal is to construct a test to detect when the mean of the observations exceeds some pre-specified threshold, i.e., $P_1 \in \mathcal{P}_1$ such that

$$\mathcal{P}_1 := \{P : \mathbb{E}^P[X] \geq \eta > \mu_0\} \quad (14)$$

In this expression, X denotes a generic observation in the sequence, η is a pre-designed threshold on the mean, and $\mu_0 := \mathbb{E}^{P_0}[X]$. Define

$$\Delta := \frac{\eta - \mu_0}{2} \quad (15)$$

Also, for $i = 0, 1$, let the density (w.r.t. the Lebesgue measure) of P_i be p_i . Define

$$\kappa_0(\lambda) = \ln \mathbb{E}^{P_0}[e^{\lambda X}] \quad (16)$$

to be the cumulant-generating function (cgf) under measure P_0 . We present our main results below.

III. MAIN RESULTS

A. Known Pre-change Distribution

Throughout we will assume that \mathcal{P}_1 is as defined in (14). In this subsection, we study the case where $\mathcal{P}_0 = \{P_0\}$.

Theorem III.1. For $\mathcal{P}_0 = P_0$, and \mathcal{P}_1 given in (14), define

$$p_1^*(x) = p_0(x)e^{\lambda^*x - \kappa_0(\lambda^*)} \quad (17)$$

where $\kappa_0(\lambda)$ is the cgf under P_0 and λ^* satisfies

$$\kappa'_0(\lambda^*) := \frac{\mathbb{E}^{P_0}[Xe^{\lambda^*X}]}{\mathbb{E}^{P_0}[e^{\lambda^*X}]} = \eta \quad (18)$$

Then, the statistic

$$\Lambda^{P_0, P_1^*}(t) = \max_{1 \leq k \leq t} \sum_{i=k}^t (\lambda^* X_i - \kappa_0(\lambda^*)) \quad (19)$$

and the stopping rule $\tau(\Lambda^{P_0, P_1^*}, b_\alpha)$ with threshold $b_\alpha = |\ln \alpha|$ solves the minimax robust problem in (6) asymptotically as $\alpha \rightarrow 0$, and

$$\inf_{\tau \in \mathcal{C}_\alpha^{\mathcal{P}_0}} \sup_{(P_0, P_1) \in \mathcal{P}_0 \times \mathcal{P}_1} WADD^{P_0, P_1}(\tau) = \frac{|\ln \alpha|}{\lambda^* \eta - \kappa_0(\lambda^*)} (1 + o(1)) \quad (20)$$

Proof. We follow the procedure outlined in [19, Sec. 6.4.1]. We want to minimize $D(P_1 || P_0) = \mathbb{E}^{P_1}[\ln(p_1(x)/p_0(x))]$ subject to $\mathbb{E}^{P_1}[X] \geq \eta$. We consider the Lagrangian

$$\begin{aligned} L(p_1, \lambda, \mu) &= \mathbb{E}^{P_1}[\ln(p_1(x)/p_0(x))] \\ &\quad + \lambda(\eta - \mathbb{E}^{P_1}[X]) + \mu \left(1 - \int_{[0,1]} p_1(x) dx \right) \\ &= \int_{[0,1]} \left(\ln \frac{p_1(x)}{p_0(x)} - \lambda x - \mu \right) p_1(x) dx \\ &\quad + \lambda \eta + \mu \end{aligned} \quad (21)$$

where the Lagrange multiplier $\lambda \geq 0$ corresponds to the constraint that the post-change mean is greater than η , and μ corresponds to the constraint that $p_1(x)$ is a probability measure. For an arbitrary direction z , since p_1 is continuous by assumption, we take the Gateaux derivative with respect to p_1 :

$$\begin{aligned} \nabla_{p_1, z} L(p_1, \lambda, \mu) &:= \lim_{h \rightarrow 0} \frac{L(p_1 + hz, \lambda, \mu) - L(p_1, \lambda, \mu)}{h} \\ &= \int_{[0,1]} \left(\ln \frac{p_1(x)}{p_0(x)} - \lambda x - \mu' \right) z dx \end{aligned} \quad (22)$$

where $\mu' = \mu - 1$, and since z is arbitrary, we arrive at

$$\ln \frac{p_1(x)}{p_0(x)} - \lambda x - \mu' = 0 \quad (23)$$

By the Generalized Kuhn–Tucker Theorem [20], if $p_0(x)$ is bounded, $p_1(x) = p_0(x)e^{\lambda x + \mu}$ is a necessary condition for optimality. Furthermore, since $L(p_1, \lambda, \mu)$ is convex in p_1 , this is also a global optimum. To satisfy the constraints,

$$\mu' = -\ln \int_{[0,1]} p_0 e^{\lambda x} dx = -\kappa_0(\lambda) \quad (24)$$

and λ^* satisfies

$$\eta = \mathbb{E}^{P_1}[X] = \frac{\mathbb{E}^{P_0}[Xe^{\lambda^*X}]}{\mathbb{E}^{P_0}[e^{\lambda^*X}]} = \kappa'_0(\lambda^*) \quad (25)$$

Thus, P_1^* in (17) minimizes the KL divergence when P_0 is known. By Theorem II.1, since (P_0, P_1^*) minimizes the KL divergence, \mathcal{P}_1 is convex, and \mathcal{P}_0 is a singleton, the uncertainty classes $\mathcal{P}_0 \times \mathcal{P}_1$ are WS bounded by (P_0, P_1^*) . Therefore, an asymptotically optimal test is Λ^{P_0, P_1^*} as in (19).

Furthermore, the minimum KL-divergence is

$$\begin{aligned} D(P_1^* || P_0) &= \int_{[0,1]} (\lambda^* x - \kappa_0(\lambda^*)) p_1^*(x) dx \\ &= \lambda^* \eta - \kappa_0(\lambda^*) \end{aligned} \quad (26)$$

Hence, the worst-case delay satisfies

$$\begin{aligned} \inf_{\tau \in \mathcal{C}_\alpha^{\mathcal{P}_0}} \sup_{P_1 \in \mathcal{P}_1} WADD^{P_0, P_1}(\tau) &= \frac{|\ln \alpha|}{D(P_1^* || P_0)} (1 + o(1)) \\ &= \frac{|\ln \alpha|}{\lambda^* \eta - \kappa_0(\lambda^*)} (1 + o(1)) \end{aligned} \quad (27)$$

as $\alpha \rightarrow 0$. \square

Since p_0 is a density on $[0, 1]$, p^* is also a density on $[0, 1]$. Indeed, p^* is an exponentially-tilted version (or the Esscher transform) of p_0 .

B. Approximation for Small Δ

Even though we have an expression for the test statistic when P_0 is known, as given in (19), the exact solution of λ^* is not available in closed-form. Fortunately, if the mean-change gap Δ is small, we obtain a low-complexity test in terms of only the pre-change mean and variance that closely approximates the performance of the asymptotically minimax optimal test in the previous section.

As the gap is asymptotically small, the worst-case post-change mean $\eta \rightarrow \mu_0$. Hence, $\lambda^* \rightarrow 0$. From the Taylor expansion on κ_0 around 0, we obtain

$$\begin{aligned}\kappa_0(\lambda^*) &= \kappa_0(0) + \kappa'_0(0)\lambda^* + \frac{\kappa''_0(0)}{2}(\lambda^*)^2 + o((\lambda^*)^2) \\ &= \mu_0\lambda^* + \frac{\sigma_0^2}{2}(\lambda^*)^2 + o((\lambda^*)^2)\end{aligned}\quad (28)$$

In this same regime, by continuity of $\kappa'_0(\cdot)$,

$$\begin{aligned}\lambda^* &= \frac{\kappa'_0(\lambda^*) - \kappa'_0(0)}{\kappa''_0(0)} + o(\Delta) \\ &= \frac{\eta - \mu_0}{\sigma_0^2} + o(\Delta) \\ &= \frac{2\Delta}{\sigma_0^2} + o(\Delta)\end{aligned}\quad (29)$$

where we have used $\kappa'_0(\lambda^*) = \eta$. Hence, the approximate test statistic is

$$\begin{aligned}\lambda^* X_t - \kappa_0(\lambda^*) &= \lambda^* X_t - (\mu_0\lambda^* + \frac{\sigma_0^2}{2}(\lambda^*)^2) + o((\lambda^*)^2) \\ &= \frac{2\Delta}{\sigma_0^2}(X_t - \mu_0) - \frac{\sigma_0^2}{2} \left(\frac{2\Delta}{\sigma_0^2} \right)^2 + o(\Delta^2) \\ &= \frac{2\Delta}{\sigma_0^2} \left(X_t - \frac{\mu_0 + \eta}{2} \right) + o(\Delta^2)\end{aligned}\quad (30)$$

and the corresponding minimum KL-divergence is approximated as:

$$D(P_1^* || P_0) = \frac{2\Delta^2}{\sigma_0^2} + o(\Delta^2) \quad (31)$$

Let σ_0^2 be the variance of X under P_0 . Since

$$\frac{2\Delta}{\sigma_0^2} \left(X_t - \frac{\mu_0 + \eta}{2} \right) > b_\alpha \iff X_t - \frac{\mu_0 + \eta}{2} > \tilde{b}_\alpha \quad (32)$$

where

$$\tilde{b}_\alpha := \frac{|\ln \alpha| \sigma_0^2}{2\Delta} = \frac{|\ln \alpha| \sigma_0^2}{\eta - \mu_0}, \quad (33)$$

the stopping rule $\tau(\Lambda^{P_0, P_1^*}, b_\alpha)$ can be approximated by the stopping rule $\tau(\tilde{\Lambda}^{\mu_0, \eta}, \tilde{b}_\alpha)$, where

$$\begin{aligned}\tilde{\Lambda}^{\mu_0, \eta}(t) &= \max_{1 \leq k \leq t} \sum_{i=k}^t \left(X_i - \frac{\mu_0 + \eta}{2} \right) \\ &= \left(\tilde{\Lambda}^{\mu_0, \eta}(t-1) + \left(X_t - \frac{\mu_0 + \eta}{2} \right) \right)^+\end{aligned}\quad (34)$$

with $\tilde{\Lambda}^{\mu_0, \eta}(0) = 0$. We call $\tau(\tilde{\Lambda}^{\mu_0, \eta}, \tilde{b}_\alpha)$ the Mean-Change Test (MCT), and $\tilde{\Lambda}^{\mu_0, \eta}$ the MCT statistic.

Note that the worst-case delay satisfies

$$\begin{aligned}\inf_{\tau \in \mathcal{C}_\alpha^{\{P_0\}}} \sup_{P_1 \in \mathcal{P}_1} \text{WADD}^{P_0, P_1}(\tau) &= \frac{|\ln \alpha|}{D(P_1^* || P_0)} (1 + o(1)) \\ &\approx \frac{|\ln \alpha| \sigma_0^2}{2\Delta^2} (1 + o(1))\end{aligned}\quad (35)$$

where the approximation becomes more accurate as $\Delta \rightarrow 0$.

Therefore, if the gap is small, it is sufficient to know only the mean and variance to construct a good approximation to the asymptotically minimax robust test. Furthermore, only the mean of the pre-change distribution is needed to construct the MCT statistic. From the simulation results in Section IV, we see that the performance of the MCT can be very close to that of the asymptotically minimax robust test even for moderate values of Δ . Since the mean and variance of a distribution are much easier and more accurately estimated than the entire density, this test can be useful and accurate when only a moderate number of observations in the pre-change regime is available.

C. Performance Analysis of MCT for moderate Δ

We now study the asymptotic performance of the MCT for fixed Δ , as $\alpha \rightarrow 0$.

Lemma III.2. Fix $P_0 \in \mathcal{P}_0$. For simplicity, denote $\tau := \tau(\tilde{\Lambda}^{\mu_0, \eta}, b)$, $Z_i := X_i - (\mu_0 + \eta)/2$, and $S_t = \sum_{i=1}^t Z_i$. For any threshold $b > 1$,

$$\begin{aligned}P_0 \{ S_\tau \geq b \} &\leq 2R_0 \sqrt{\frac{b^2}{\Delta^2}} K_1 \left(\frac{R_0^2 b \Delta}{\sigma_0^2} \right) \exp \left(-\frac{R_0^2 \Delta}{\sigma_0^2} b \right) \\ &\sim \sqrt{\frac{2\pi \sigma_0^2 b}{\Delta^3}} \exp \left(-\frac{2R_0^2 \Delta}{\sigma_0^2} b \right), \text{ as } b \rightarrow \infty\end{aligned}\quad (36)$$

where

$$R_0 = \sigma_0^2 / (\sigma_0^2 + \Delta \cdot \max\{\mu_0, 1 - \mu_0\}/3) \quad (37)$$

and $K_\beta(z)$ is the modified Bessel function of the second kind with order β .

Proof. Let $M = \max\{\mu_0/3, (1 - \mu_0)/3\}$. Note that $\mathbb{E}^{P_0}[Z_i] = (\mu_0 - \eta)/2 = -\Delta$. Thus, we have

$$\begin{aligned}P_0 \{ S_\tau \geq b \} &= P_0 \left\{ \sum_{i=1}^\tau Z_i \geq b \right\} \\ &= \sum_{t=1}^\infty P_0 \left\{ \sum_{i=1}^t Z_i \geq b, t = \tau \right\} \\ &\leq \sum_{t=1}^\infty P_0 \left\{ \sum_{i=1}^t Z_i \geq b \right\} \\ &= \sum_{t=1}^\infty P_0 \left\{ \sum_{i=1}^t (Z_i + \Delta) \geq b + t\Delta \right\}\end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{\leq} \sum_{t=1}^{\infty} \exp \left(-\frac{(b+t\Delta)^2}{2(t\sigma_0^2 + M(b+t\Delta))} \right) \\
&\stackrel{(b)}{\leq} \int_0^{\infty} \exp \left(-\frac{(b+x\Delta)^2}{2(x\sigma_0^2 + M(b+x\Delta))} \right) dx \\
&= a \int_0^{\infty} \exp \left(-\frac{(a\Delta y + C)^2}{2y} \right) dy \\
&= ae^{-a\Delta C} \int_0^{\infty} e^{-((a^2\Delta^2/2)y + (C^2/2)y^{-1})} dy \\
&= \frac{2C}{\Delta} e^{-a\Delta C} K_1(a\Delta C)
\end{aligned} \tag{38}$$

where $a := (\sigma_0^2 + M\Delta)^{-1}$ and $C := \sigma_0^2 b / (\sigma_0^2 + M\Delta)$. In the series of inequalities above, (a) is by Bernstein's inequality [21, p. 9], and (b) is from bounding the sum with an integral. Since $K_1(z) \sim \sqrt{\frac{\pi}{2z}} e^{-z}$ as $|z| \rightarrow \infty$, the asymptotic result follows. \square

Theorem III.3. *Let*

$$\sqrt{\frac{2\pi\sigma_0^2\tilde{b}'_\alpha}{\Delta^3}} \exp \left(-\frac{2R_0^2\Delta}{\sigma_0^2} \tilde{b}'_\alpha \right) = \alpha \tag{39}$$

then (7) is satisfied asymptotically as $\alpha \rightarrow 0$. Furthermore,

$$\tilde{b}'_\alpha = \frac{\tilde{b}_\alpha}{R_0^2} (1 + o(1)) \tag{40}$$

where R_0 is defined in (37).

Proof. As $\alpha \rightarrow 0$, $\tilde{b}'_\alpha \rightarrow \infty$. As a result, for any $P_1 \in \mathcal{P}_1$, $P_0 \{S_\tau \geq b\} \leq \alpha(1 + o(1))$. From [22, Sec. 2.6], it can be shown that

$$\begin{aligned}
\mathbb{E}_\infty^{P_0, P_1} [\tau] &= \frac{\mathbb{E}^{P_0} [S_\tau]}{P_0 \{S_\tau \geq b\}} \\
&\geq \frac{1}{P_0 \{S_\tau \geq b\}} \\
&\geq \alpha^{-1} (1 + o(1))
\end{aligned} \tag{41}$$

since $\mathbb{E}^{P_0} [S_\tau] \geq 1$. Thus, the false alarm constraint is satisfied asymptotically.

For the second result, it is sufficient to show that $(\tilde{b}'_\alpha - \tilde{b}_\alpha)/\tilde{b}_\alpha = R_0^{-2} - 1 + o(1)$. Let

$$D := \frac{2\Delta}{\sigma_0^2} \tilde{b}'_\alpha - |\ln \alpha| \tag{42}$$

Note that the desired ratio $(\tilde{b}'_\alpha - \tilde{b}_\alpha)/\tilde{b}_\alpha = D/|\ln \alpha|$. Plugging into (36), we have

$$\sqrt{\frac{\sigma_0^4 \pi}{\Delta^4} (D + |\ln \alpha|)} e^{-R_0^2(D + |\ln \alpha|)} = \alpha \tag{43}$$

$$-\frac{1}{2} \ln \left(\frac{\sigma_0^4 \pi}{\Delta^4} (D + |\ln \alpha|) \right) + R_0^2(D + |\ln \alpha|) = |\ln \alpha| \tag{44}$$

Now, we hypothesize that $D = D_1 |\ln \alpha| + o(|\ln \alpha|)$. The first term then vanishes because

$$\begin{aligned}
\ln \left(\frac{\sigma_0^4 \pi}{\Delta^4} (D + |\ln \alpha|) \right) &= \ln \left(\frac{\sigma_0^4 \pi}{\Delta^4} ((D_1 + 1) |\ln \alpha| + o(|\ln \alpha|)) \right) \\
&= O(\ln(|\ln \alpha|)) \\
&= o(|\ln \alpha|)
\end{aligned} \tag{45}$$

From the second term, we validate our hypothesis that

$$D = (R_0^{-2} - 1) |\ln \alpha| + o(|\ln \alpha|) \tag{46}$$

and the second result follows. \square

Remark. In practice, the threshold can be set to be $\sigma_0^2/(2R_0^2\Delta)$ using equation (40).

Theorem III.4. *Fix $P_0 \in \mathcal{P}_0$. The worst-case delay satisfies*

$$\sup_{P_1 \in \mathcal{P}_1} \text{WADD}^{P_0, P_1} \left(\tau(\tilde{\Lambda}^{\mu_0, \eta}, \tilde{b}'_\alpha) \right) = \frac{|\ln \alpha| \sigma_0^2}{2\Delta^2 R_0^2} (1 + o(1)) \tag{47}$$

as $\alpha \rightarrow 0$.

Proof. For any $P_1 \in \mathcal{P}_1$, as $\alpha \rightarrow 0$,

$$\begin{aligned}
\text{WADD}^{P_0, P_1} \left(\tau(\tilde{\Lambda}^{\mu_0, \eta}, \tilde{b}'_\alpha) \right) &= \frac{\tilde{b}'_\alpha}{\Delta} (1 + o(1)) \\
&= \frac{\tilde{b}_\alpha}{\Delta R_0^2} (1 + o(1)) \\
&= \frac{|\ln \alpha| \sigma_0^2}{2\Delta^2 R_0^2} (1 + o(1))
\end{aligned} \tag{48}$$

where the first line is by renewal theory [23]. \square

Remark. As $\Delta \rightarrow 0$, $R_0 \rightarrow 1$. Thus, the result above becomes

$$\sup_{P_1 \in \mathcal{P}_1} \text{WADD}^{P_0, P_1} \left(\tau(\tilde{\Lambda}^{\mu_0, \eta}, \tilde{b}'_\alpha) \right) \sim \frac{|\ln \alpha| \sigma_0^2}{2\Delta^2} (1 + o(1)) \tag{49}$$

where $o(1)$ goes zero as α and Δ go to zero, which coincides with the minimax robust worst-case delay in (35).

Remark. The WADD analysis above can be extended to the case where in the post-change regime the observations are independent but not necessarily identically distributed, as long as all of these distributions belong to the uncertainty class \mathcal{P}_1 .

IV. NUMERICAL RESULTS AND DISCUSSION

We study the performance of the proposed tests through simulations for the case where the pre- and post-change distributions are beta(4,16) ($\mu_0 = 0.2$) and beta(4.5,16) ($\mu_1 = 0.2195$), respectively. The mean-threshold η is set to be 0.21. In particular, we compare the performances for the following three statistics:

- 1) The CUSUM statistic that knows both the pre- and post-change distributions, defined in (4).
- 2) The statistic when only the pre-change distribution is known, defined in (19).
- 3) The MCT statistic defined in (34).

For all three statistics, based on their recursive structure, it is easy to show that the worst-case value of the change-point for computing WADD in (1) is when $\nu = 0$. Therefore we can estimate the worst-case delay by simulating the post-change distribution from time 0.

We see in Fig. 1 that the performance of MCT is very close to that of the asymptotically minimax robust optimal test that uses the full knowledge of the pre-change distribution. Note that the MCT statistic only uses the pre-change mean;

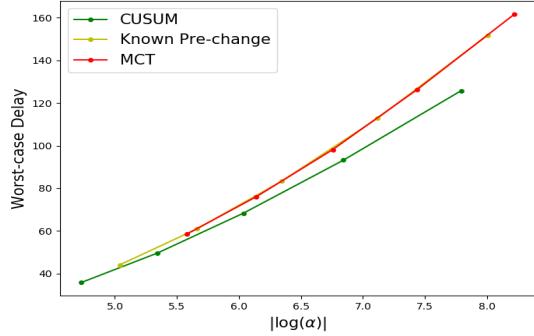


Fig. 1. Performances of different statistics. The pre- and post-change distribution are Beta(4,16) ($\mu_0 = 0.2$) and Beta(4.5,16) ($\mu_1 = 0.2195$), respectively. The mean-threshold $\eta = 0.21$.

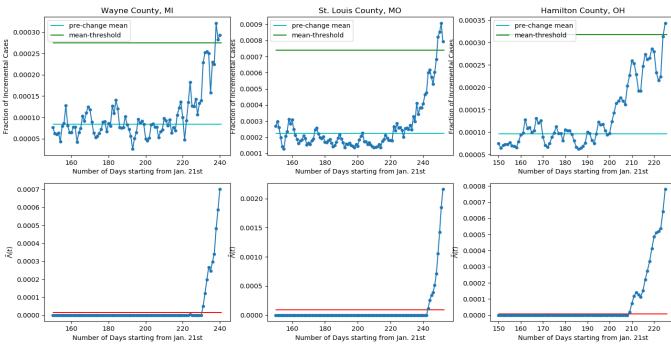


Fig. 2. COVID-19 monitoring example. The upper subplot is the three-day moving average of the new cases of COVID-19 as a fraction of the population in Wayne County, MI (left), St. Louis County, MO (middle), and Hamilton County, OH (right). The x-axis is the number days elapsed after January 21, 2020. The pre-change mean and variance are estimated using data from days 120 to 150. The FAR threshold α is set to 0.01. For each county, the mean-threshold η (in green) is set to be 3.3 times of the estimated pre-change mean (in cyan). The lower subplot shows the evolution of the statistic \bar{A}_t in the corresponding county. The Λ -threshold \bar{b}_α (in red) is calculated using equation (33).

the variance is only needed for setting the threshold to meet a given FAR constraint. In Fig. 2, we apply the MCT to monitoring the spread of COVID-19 using new case data from various counties in the US [24]. The incremental cases from day to day can be assumed to be roughly independent. The goal is to detect the onset of a new wave of the pandemic based on the incremental cases as a fraction of the county population exceeding some pre-specified level. The pre-change mean and variance are estimated using observations for periods in which the increments remain low and roughly constant. We set the mean-threshold η to be a multiple of the pre-change mean, with understanding that such a threshold might be indicative of a new wave. With this choice, we observe that the MCT statistic significantly and persistently crosses the test-threshold around late November 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 MCT 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 nearly vertically after the onset.

REFERENCES

- [1] Y. C. Chen, T. Banerjee, A. D. Dominguez-Garcia, and V. V. Veeravalli, “Quickest line outage detection and identification,” *IEEE Transactions on Power Systems*, vol. 1, no. 31, pp. 749–758, 2016.
- [2] T. L. Molloy and J. J. Ford, “Misspecified and asymptotically minimax robust quickest change detection,” *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5730–5742, 2017.
- [3] G. Tagaras, “A survey of recent developments in the design of adaptive control charts,” *Journal of Quality Technology*, vol. 30, no. 3, pp. 212–231, 1998.
- [4] J. Czarnecki and C. Farrar, “Structural health monitoring using statistical process control,” *Journal of Structural Engineering*, vol. 126, no. 11, pp. 1356–1363, Nov. 2000.
- [5] L. Besson and E. Kaufmann, “The generalized likelihood ratio test meets KL-UCB: An improved algorithm for piece-wise non-stationary bandits,” *arXiv preprint arXiv:1902.01575*, 2019.
- [6] A. Tartakovsky and V. Veeravalli, “General asymptotic bayesian theory of quickest change detection,” *SIAM Theory of Probability and Its Applications*, vol. 49, Jan. 2005.
- [7] G. Lorden, “Procedures for reacting to a change in distribution,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, Dec. 1971.
- [8] M. Pollak, “Average run lengths of an optimal method of detecting a change in distribution,” *The Annals of Statistics*, vol. 15, no. 2, pp. 749–779, Jun. 1987.
- [9] P. Moulin and V. V. Veeravalli, *Statistical Inference for Engineers and Data Scientists*. Cambridge, UK: Cambridge University Press, 2018.
- [10] D. Siegmund and E. S. Venkatraman, “Using the generalized likelihood ratio statistic for sequential detection of a change-point,” *The Annals of Statistics*, vol. 23, no. 1, pp. 255–271, Feb. 1995.
- [11] T. Lai and H. Xing, “Sequential change-point detection when the pre- and post-change parameters are unknown,” *Sequential Analysis*, vol. 29, no. 2, pp. 162–175, May 2010.
- [12] S. Li, Y. Xie, H. Dai, and L. Song, “M-statistic for kernel change-point detection,” in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015, pp. 3366–3374.
- [13] O.-A. Maillard, “Sequential change-point detection: Laplace concentration of scan statistics and non-asymptotic delay bounds,” in *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, ser. Proceedings of Machine Learning Research, A. Garivier and S. Kale, Eds., vol. 98. Chicago, Illinois: PMLR, Mar. 2019, pp. 610–632.
- [14] T. Lau, W. P. Tay, and V. Veeravalli, “A binning approach to quickest change detection with unknown post-change distribution,” *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 609–621, Nov. 2018.
- [15] P. J. Huber, “A robust version of the probability ratio test,” *The Annals of Mathematical Statistics*, vol. 36, no. 6, pp. 1753–1758, Dec. 1965.
- [16] E. S. PAGE, “Continuous Inspection Schemes,” *Biometrika*, vol. 41, no. 1/2, pp. 100–115, Jun. 1954.
- [17] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, Dec. 1986.
- [18] J. Unnikrishnan, V. V. Veeravalli, and S. P. Meyn, “Minimax robust quickest change detection,” *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1604–1614, 2011.
- [19] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. New York, NY: Springer, 2008.
- [20] D. G. Luenberger, *Optimization by Vector Space Methods*, 1st ed. 605 Third Ave, New York, NY, US: John Wiley & Sons, Inc., Jan. 1997.
- [21] L. Wasserman, *All of Nonparametric Statistics*. 233 Spring Street, New York, NY 10013, USA: Springer, 2006.
- [22] D. Siegmund, *Sequential analysis: Tests and confidence intervals*. Springer, 1985.
- [23] 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.
- [24] Rearc. AWS Marketplace: Coronavirus (COVID-19) Data in the United States — The New York Times. [Online]. Available: [https://urldefense.com/v3/_https://aws.amazon.com/marketplace/pp/Coronavirus-COVID-19-Data-in-the-United-States-The/prodview-jmb464qw2yg74_!!IDZ3fjgloSxTS0VgkopT_Gqp0XiRZGO24wcouQ5xRXh389-5cY88mGOynh9goKlh8x9_ogUY\\$](https://urldefense.com/v3/_https://aws.amazon.com/marketplace/pp/Coronavirus-COVID-19-Data-in-the-United-States-The/prodview-jmb464qw2yg74_!!IDZ3fjgloSxTS0VgkopT_Gqp0XiRZGO24wcouQ5xRXh389-5cY88mGOynh9goKlh8x9_ogUY$)