

CuSum for sequential change diagnosis

Austin Warner *Student Member, IEEE* and Georgios Fellouris, *Member, IEEE*,

Abstract—The problem of sequential change diagnosis is considered, where a sequence of independent random elements is accessed sequentially, there is an abrupt change in its distribution at some unknown time, and there are two main operational goals: to quickly detect the change and to accurately identify the post-change distribution among a finite set of alternatives. A standard algorithm is considered, which does not explicitly address the isolation task and raises an alarm as soon as the CuSum statistic that corresponds to one of the post-change alternatives exceeds a certain threshold. It is shown that in certain cases, such as the so-called multichannel problem, this algorithm controls the worst-case conditional probability of false isolation and minimizes Lorden's criterion, for every possible post-change distribution, to a first-order asymptotic approximation as the false alarm rate goes to zero sufficiently faster than the worst-case conditional probability of false isolation. These theoretical results are also illustrated with a numerical study.

Index Terms—Sequential Change Detection, Sequential Change Diagnosis, Isolation, Identification.

I. INTRODUCTION

THE problem of quickly detecting a change in sequentially acquired data dates back to the pioneering works of Shewhart [1] and Page [2], and is motivated by a wide range of engineering and scientific applications, such as industrial process quality control [3]–[5], target detection and identification [6], [7], integrity monitoring of navigation systems [8], [9], target tracking [10], network intrusion detection [11], [12], bioterrorism [13], [14], and genomics [15].

In these applications, there are typically many possible post-change distributions, representing different types of change, and it is useful, if not critical, to not only detect the change quickly, but also to correctly identify its type upon stopping. This problem of simultaneously detecting a change in sequentially collected data and identifying the correct post-change distribution among a finite set of alternatives is known as *sequential change diagnosis* and has been studied in various works [16]–[20].

A Bayesian formulation for the sequential change diagnosis problem was proposed in [21], where the change-point has a known geometric prior distribution and the observations are iid before and after the change conditionally on the change-point. A computationally convenient modification of the Bayes rule was proposed and analyzed in [22]. In the same Bayesian context, a two-stage approach was considered in [23]. The special case of the *multichannel* problem was explored in [18],

wherein a number of independent data streams are observed in parallel and a change occurs in one of them. Additionally, general non-iid models were considered for the pre-change and post-change regimes and non-geometric priors were allowed for the change-point.

When the mechanism that generates the change-point is not assumed to be known *a priori*, the sequential change diagnosis problem is arguably significantly more complex than its individual components of sequential change detection and sequential multi-hypothesis testing. Indeed, simple algorithms for sequential change-detection, which do not assume any prior information on the change-point, such as Page's Cumulative Sum (CuSum) [2], were proposed very early on. This algorithm has been studied extensively since its introduction and has been grounded on strong theoretical support. Specifically, in the case that the observations are iid before and after the change with completely specified pre-change and post-change distributions, it has been shown [24] to minimize *Lorden's criterion* [25], i.e., the worst-case conditional expected detection delay with respect to both the history of the observations up to the change and the change-point itself, subject to a user-specified bound on the false alarm rate.

The first scheme that was ever proposed, in [16], for the sequential change diagnosis problem, under the assumption of iid observations before and after the change, was in fact a generalization of the CuSum algorithm, termed “*Generalized CuSum*”. Unlike the original CuSum algorithm, however, this one does not admit a recursive structure and requires a number of operations at each time instance that grows linearly with the number of observations. Since the change can take a very long time to occur, this implies that this procedure needs to be modified to be applicable in practice. A recursive modification was proposed in [26], according to which an alarm is raised as soon as the CuSum statistics of a post-change hypothesis against the pre-change *and all other post-change alternatives* simultaneously exceed a certain threshold. Thus, this algorithm requires the parallel computation of a matrix of CuSum statistics, and for this reason it is referred to as the “*Matrix CuSum*”.

A fundamental drawback of both the “*Generalized CuSum*” and the “*Matrix CuSum*” is that they might fail to isolate the post-change distribution when the change-point is large [27, Section 4.3.1]. Indeed, they have only been shown to control metrics of the false isolation rate *when the change occurs at the beginning of the sampling period*. This can be partially addressed by a window-limited modification of the Generalized CuSum, as the one suggested in [19]. However, this leads to a nontrivial trade-off regarding the size of the window. Indeed, if this deterministic window is selected to be

This research was supported by the US National Science Foundation under grant AMPS 1736454 through the University of Illinois at Urbana-Champaign.

A. Warner and G. Fellouris are with Department of Statistics of the University of Illinois at Urbana-Champaign. Email: {awarner5,fellouri}@illinois.edu.

too small, it will degrade the ability of the algorithm to detect a change of weak signal. On the other hand, if it is chosen to be too large, it will fail to ameliorate the issue stated above.

A recursive, CuSum-based algorithm that, under certain conditions can control the *worst-case* conditional probability of false isolation given that there was no false alarm with respect to the change-point, was proposed in [20]. This algorithm requires the computation of only the original K CuSum statistics, which are used for the pure change detection problem, and for this reason it is referred to as the “*Vector CuSum*”. The worst-case conditional expected detection delay of this scheme, however, was shown to be bounded only with respect to the change-point, not the data up to the change. That is, its delay analysis was based on Pollak’s criterion [28], not Lorden’s.

The goal of the present work is to study a very simple algorithm for the change diagnosis problem, which does not explicitly address the isolation task and incidentally admits a convenient decentralized implementation (see, e.g., [29]). Specifically, this procedure raises an alarm when the value of a CuSum statistic that corresponds to a post-change distribution exceeds a certain threshold. This algorithm is shown to control the worst-case conditional probability of false isolation in certain problems where the post-change distributions are more similar, in a certain sense, to the pre-change distribution than to each other. This is the case, for example, in the *multichannel* problem mentioned above. In such a setup, this algorithm is also shown to minimize Lorden’s criterion, *under every possible post-change alternative*, to a first order asymptotic approximation as the prescribed false alarm rate goes to 0, sufficiently faster than the prescribed worst-case probability of false isolation.

II. PROBLEM STATEMENT

Let $X \equiv \{X_n, n \in \mathbb{N}\}$ be a sequence of independent $\mathbb{S}$ -valued random elements, where $\mathbb{S}$ is an arbitrary Polish space. We denote by $\{\mathcal{F}_n : n \in \mathbb{N}\}$ the natural filtration of X , i.e., $\mathcal{F}_n \equiv \sigma(X_1, \dots, X_n)$, $n \in \mathbb{N}$, and by $\mathcal{F}_0$ the trivial σ -algebra. We assume that each term of this sequence has a density with respect to a σ -finite measure λ , which is f up to and including some time instance $\nu \in \mathbb{N} \cup \{0\}$ and g after ν .

We assume that the *change-point*, ν , is deterministic and completely unknown, and that the pre-change density, f , is completely specified. On the other hand, we allow for uncertainty regarding the distribution after the change. Specifically, we postulate K alternatives for the post-change density, i.e., $g \in \{g_1, \dots, g_K\}$, and assume that the Kullback-Leibler information number

$$I_i \equiv \int \log(g_i/f) g_i d\lambda$$

is positive and finite for every $i \in [K] \equiv \{1, \dots, K\}$.

We denote by $\mathbb{P}_\infty$ the distribution of X , and by $\mathbb{E}_\infty$ the corresponding expectation, when the change never occurs, i.e., when X is a sequence of independent random elements with common density f . Moreover, we denote by $\mathbb{P}_{\nu,i}$ the

distribution of X , and by $\mathbb{E}_{\nu,i}$ the corresponding expectation, when the change occurs at time ν and the post-change density is g_i . For simplicity, when the change occurs from the very beginning, i.e., when $\nu = 0$, we suppress the dependence on the change-point by setting $\mathbb{P}_i \equiv \mathbb{P}_{0,i}$ and $\mathbb{E}_i \equiv \mathbb{E}_{0,i}$.

We assume that the terms of the sequence X are observed sequentially and that the problem is to quickly detect the change, as well as to identify the correct post-change distribution upon stopping. Thus, we need to specify a random time, T , at which we declare that the change has occurred, as well as a $[K]$ -valued random variable, D , that represents our decision regarding the post-change density at the time of stopping. On the event $\{T = n, D = i\}$ the alarm is raised and g_i is declared as the correct post-change density after having taken n observations, where $n \in \mathbb{N}$ and $i \in [K]$. We refer to such a pair (T, D) as a *diagnosis procedure* if T is an $\{\mathcal{F}_n\}$ -stopping time and D is $\mathcal{F}_T$ -measurable, or in other words if

$$\{T = n\}, \{T = n, D = i\} \in \mathcal{F}_n \quad \forall n \in \mathbb{N}, i \in [K].$$

We denote by $\mathcal{C}$ the family of all diagnosis procedures and, without loss of generality, we restrict ourselves to stopping times that are not almost surely bounded under any $\mathbb{P}_{\nu,i}$.

As in the change detection problem, we measure the ability of $(T, D) \in \mathcal{C}$ to avoid false alarms using the average number of observations until stopping under $\mathbb{P}_\infty$, i.e., $\mathbb{E}_\infty[T]$. Similarly, to measure the ability of $(T, D) \in \mathcal{C}$ to quickly detect the change when the post-change density is g_i for some $i \in [K]$, we adopt Lorden’s criterion [25], which utilizes the worst-case conditional expected detection delay with respect to both the change-point and the data up to the change:

$$\mathcal{J}_i(T) \equiv \sup_{\nu \geq 0} \text{esssup}_{\mathbb{P}_{\nu,i}} [T - \nu | \mathcal{F}_\nu, T > \nu]. \quad (1)$$

Finally, to measure the ability of $(T, D) \in \mathcal{C}$ to isolate the post-change density when the change occurs at some time ν , we use the conditional probability of selecting the wrong post-change density at the time of stopping given that there was no false alarm:

$$\max_{j \in [K]} \mathbb{P}_{\nu,j}(D \neq j | T > \nu).$$

We denote by $\mathcal{C}(\alpha)$ the family of diagnosis procedures that control the expected time until stopping under $\mathbb{P}_\infty$ by at least $1/\alpha$, i.e.,

$$\mathcal{C}(\alpha) \equiv \{(T, D) \in \mathcal{C} : \mathbb{E}_\infty[T] \geq 1/\alpha\},$$

where $\alpha \in (0, 1)$ is a user-specified value that represent tolerance to false alarms. Moreover, we denote by $\mathcal{C}(\alpha, \beta, \mathcal{N})$ the subfamily of diagnosis procedures that also control the conditional probability of false isolation below β when the change point belongs to a set $\mathcal{N} \subseteq \{0\} \cup \mathbb{N}$, i.e.,

$$\mathcal{C}(\alpha, \beta, \mathcal{N}) \equiv \{(T, D) \in \mathcal{C}(\alpha) : \max_{j \in [K]} \sup_{\nu \in \mathcal{N}} \mathbb{P}_{\nu,j}(D \neq j | T > \nu) \leq \beta\}, \quad (2)$$

where $\beta \in (0, 1)$ is a user-specified value that represent tolerance to false isolations. Clearly, we would like to have a diagnosis scheme that can be designed to belong to $\mathcal{C}(\alpha, \beta, \mathcal{N})$,

for $\alpha, \beta \in (0, 1)$ arbitrarily small and $\mathcal{N}$ as large as possible, ideally $\{0\} \cup \mathbb{N}$, while achieving

$$\inf_{(T,D) \in \mathcal{C}(\alpha, \beta, \mathcal{N})} \mathcal{J}_i(T), \quad (3)$$

at least approximately, for every possible post-change alternative, i.e., for every $i \in [K]$. As we mentioned in the Introduction, existing algorithms in the literature have been shown to enjoy such an asymptotic optimality property (only) when the change occurs from the beginning, i.e., $\mathcal{N} = \{0\}$. Our main goal in this work is to show that, in certain problems, a simple algorithm that does not explicitly address the isolation task can actually control the *worst-case* probability of false isolation and enjoy the above asymptotic optimality property with $\mathcal{N} = \{0\} \cup \mathbb{N}$.

III. THE MIN-CUSUM ALGORITHM

In this section we introduce the algorithm that we consider in this work for the sequential change diagnosis problem, and review some of its well known properties. To this end, for each $i \in [K]$ and $n \in \mathbb{N}$, we introduce the following notation:

$$\ell_i(n) \equiv \log(g_i(X_n)/f(X_n)). \quad (4)$$

A. An algorithm for change-detection

Page's CuSum algorithm [2] for detecting a change from f to g_i raises an alarm at the first time $n \in \mathbb{N}$ the statistic

$$Y_i(n) \equiv \max_{0 \leq m \leq n} \sum_{t=m+1}^n \ell_i(t) \quad (5)$$

exceeds a threshold $b_i > 0$, i.e., at

$$\sigma_i(b_i) \equiv \inf\{n \geq 1 : Y_i(n) \geq b_i\}. \quad (6)$$

An important property of this algorithm concerning its implementation in practice is that its statistic can be evaluated via the following recursion:

$$Y_i(n) = (Y_i(n-1) + \ell_i(n))^+, \quad n \in \mathbb{N}, \quad (7)$$

with $Y_i(0) = 0$. Moreover, it is well known [24] that it minimizes $\mathcal{J}_i$ in $\mathcal{C}(\alpha)$ when the threshold b_i is selected so that the false alarm constraint is satisfied with equality. In the presence of uncertainty regarding the post-change distribution, which is the case in our setup whenever $K > 1$, a standard approach for detecting the change is to run in parallel all these CuSum algorithms, $\sigma_1(b_1), \dots, \sigma_K(b_K)$, and stop as soon as one of them does, i.e., at

$$\min_{i \in [K]} \sigma_i(b_i).$$

In what follows, we set $b_1 = \dots = b_K = b$ (and refer to [30, Section 4] for a discussion of how to select unequal thresholds). Moreover, we denote the resulting stopping time by $\sigma(b)$, i.e.,

$$\sigma(b) \equiv \min_{i \in [K]} \sigma_i(b),$$

and refer to it as the *min-CuSum* stopping time. It is well known (see, e.g., [17, Theorem 3.1]) that, for any $b > 0$,

$$\mathbb{E}_\infty[\sigma(b)] \geq e^b/K, \quad (8)$$

which implies that the false alarm constraint is satisfied with $b = b_\alpha$, where

$$b_\alpha \equiv |\log \alpha| + \log K.$$

Moreover, it is well known that, for every $i \in [K]$ and $b > 0$,

$$\mathcal{J}_i[\sigma(b)] = \mathbb{E}_i[\sigma(b)], \quad (9)$$

and that $\sigma(b_\alpha)$ achieves the minimum value of $\mathcal{J}_i$ in $\mathcal{C}(\alpha)$ for every $i \in [K]$, to a first-order asymptotic approximation as $\alpha \rightarrow 0$. In particular, from [31, Section 9.2] it follows that

$$\mathcal{J}_i[\sigma(b_\alpha)] \sim \frac{|\log \alpha|}{I_i} \sim \inf_{(T,D) \in \mathcal{C}(\alpha)} \mathcal{J}_i[T] \quad (10)$$

for every $i \in [K]$ as $\alpha \rightarrow 0$, i.e.,

$$\lim_{\alpha \rightarrow 0} \frac{\mathcal{J}_i[\sigma(b_\alpha)]}{\inf_{(T,D) \in \mathcal{C}(\alpha)} \mathcal{J}_i[T]} = 1.$$

B. An algorithm for diagnosis

The stopping time $\sigma(b)$ is associated with a natural decision rule, which is to select the post-change hypothesis that corresponds to the CuSum stopping time that stops first. We denote this decision rule by $\hat{\sigma}(b)$, i.e.,

$$\hat{\sigma}(b) \in \arg \min_{i \in [K]} \sigma_i(b).$$

When $\hat{\sigma}(b)$ is not uniquely defined, we assume that a hypothesis in $\hat{\sigma}(b)$ is selected in some arbitrary way.

In the next section we show that, in certain problems, the diagnosis scheme $(\sigma(b), \hat{\sigma}(b))$ controls the *worst-case* conditional probability of false isolation and achieves the infimum in (3), with $\mathcal{N} = \{0\} \cup \mathbb{N}$, as α goes to 0 sufficiently faster than β .

IV. MAIN RESULTS

We start by stating the assumptions that we need for the results of this work.

A. Assumptions

Our first assumption is that each ℓ_i has a finite exponential moment under $\mathbb{P}_\infty$. Specifically, we denote by ψ_i the cumulant generating function of $\ell_i(n)$ under $\mathbb{P}_\infty$, i.e.,

$$\psi_i(\theta) \equiv \log(\mathbb{E}_\infty[\exp\{\theta \ell_i(n)\}]), \quad \theta \in \mathbb{R},$$

and assume that

$$\psi_i \text{ is finite around 0 for every } i \in [K]. \quad (A1)$$

Our second assumption essentially requires that any two post-change distributions are more similar to the pre-change distribution than to each other, in the sense that

$$\int (g_i/f) g_j d\lambda \leq 1 \quad \text{for every } i, j \in [K] : j \neq i. \quad (A2)$$

Indeed, these two assumptions imply that, for any $i, j \in [K]$ with $i \neq j$, the process

$$\{\exp\{\ell_j(n)\}, n \in \mathbb{N}\}$$

is a $\mathbb{P}_i$ -supermartingale, and consequently that the CuSum statistic for detecting a change from f to g_j , namely Y_j , will not grow after the change when the true post-change distribution is g_i .

Remark: Assumption (A1) is satisfied, for example, whenever the post-change densities belong to the same exponential family as the pre-change density. Specifically, let

$$\Theta \equiv \{\theta \in \mathbb{R} : \zeta(\theta) < \infty\},$$

$$\text{where } \zeta(\theta) \equiv \int \exp\{\theta x\} f(x) d\lambda,$$

and for any such $\theta \in \Theta$ consider the following density with respect to λ :

$$f(x; \theta) \equiv f(x) \exp\{\theta x - \zeta(\theta)\}.$$

If Θ contains a neighborhood of 0 and $g_i = f(\cdot; \theta_i)$ for every $i \in [K]$, then (A2) holds. If also $K = 2$ and $\theta_1 < 0 < \theta_2$, then (A2) also holds.

Remark: Assumptions (A1) and (A2) are also satisfied in the multichannel setup of Subsection IV-C, e.g., when X is a sequence of Gaussian random vectors of independent components and g_i represents a mean-shift in the i -th component of this vector.

B. Main results

We next state an upper bound for the conditional probability of false isolation of the Min-CuSum, which is uniform for a certain family of change points. Due to the space constraints, the proof is omitted and will be provided elsewhere.

Lemma 1: Suppose that (A1)-(A2) hold and let $\nu \geq 0$ be such that

$$\mathbb{P}_\infty(Y_i(\nu) \geq x) \geq \mathbb{P}_\infty(Y_i(\nu) \geq x \mid \sigma(b) > \nu) \quad (11)$$

$$\text{for all } x \geq 0, b > 0, i \in [K].$$

Then, for every $j \in [K]$ we have:

$$\mathbb{P}_{\nu,j}(\hat{\sigma}(b) \neq j \mid \sigma(b) > \nu) \leq C_K b e^{-b}(1 + \phi(b)),$$

where ϕ is a real function, which does not depend on ν or j , such that $\phi(b) \rightarrow 0$ as $b \rightarrow \infty$, and

$$C_K \equiv (K-1) \left(1 + \max_{i \in [K]} (1/I_i)\right).$$

The previous lemma implies that if $\mathcal{N}$ is the set of change-points that satisfy (11), which is nonempty because $\{0\} \in \mathcal{N}$, then

- (i) for any $\alpha, \beta \in (0, 1)$, we can select b large enough so that

$$(\sigma(b), \hat{\sigma}(b)) \in \mathcal{C}(\alpha, \beta, \mathcal{N})$$

- (ii) the infimum in (3) is achieved, to a first-order asymptotic approximation as $\alpha, \beta \rightarrow 0$, by

$$(\sigma(b_\alpha), \hat{\sigma}(b_\alpha)),$$

as long as β is sufficiently larger than α .

These two implications are formalized and stated in the following theorem.

Theorem 2: Suppose that (A1)-(A2) hold and let $\mathcal{N}$ denote the set of ν 's that satisfy (11).

- (i) For any $\alpha, \beta \in (0, 1)$, there is a $b > 0$ such that

$$(\sigma(b), \hat{\sigma}(b)) \in \mathcal{C}(\alpha, \beta, \mathcal{N}).$$

- (ii) For any $\alpha \in (0, 1)$ let β be such that

$$C_K b_\alpha e^{-b_\alpha} (1 + \phi(b_\alpha)) \leq \beta. \quad (12)$$

Then: $\sigma(b_\alpha) \in \mathcal{C}(\alpha, \beta, \mathcal{N})$ and, as $\alpha \rightarrow 0$,

$$\mathcal{J}_i[\sigma(b_\alpha)] \sim \frac{|\log \alpha|}{I_i} \sim \inf_{(T,D) \in \mathcal{C}(\alpha, \beta, \mathcal{N})} \mathcal{J}_i[T]$$

for all $i \in [K]$.

Proof: (i) Let $\alpha, \beta \in (0, 1)$ be arbitrary. From (8) it follows that for $b \geq b_\alpha$ we have

$$\mathbb{E}_\infty[\sigma(b)] \geq 1/\alpha.$$

Moreover, from Lemma 1 it follows that there is a b large enough, which depends on β , so that

$$\max_{j \in [K]} \sup_{\nu \in \mathcal{N}} \mathbb{P}_{\nu,j}(\hat{\sigma}(b) \neq j \mid \sigma(b) > \nu) \leq \beta.$$

Thus, we can select b large enough so that both inequalities are satisfied and consequently $(\sigma(b), \hat{\sigma}(b)) \in \mathcal{C}(\alpha, \beta, \mathcal{N})$.

(ii) For any $\alpha, \beta \in (0, 1)$, we have $\mathcal{C}(\alpha, \beta, \mathcal{N}) \subseteq \mathcal{C}(\alpha)$, therefore

$$\inf_{(T,D) \in \mathcal{C}(\alpha, \beta, \mathcal{N})} \mathcal{J}_i[T] \geq \inf_{T \in \mathcal{C}(\alpha)} \mathcal{J}_i[T]. \quad (13)$$

In view of (10), it suffices to show that $\sigma(b_\alpha) \in \mathcal{C}(\alpha, \beta, \mathcal{N})$ when β satisfies (12), which follows directly by Lemma 1. ■

Condition (11) is always satisfied (trivially) when $\nu = 0$. Therefore, the previous theorem shows that the min-CuSum controls the false isolation rate *at least* when the change occurs from the very beginning. We next show that the previous theorem holds with $\mathcal{N} = \{0\} \cup \mathbb{N}$ in the special but important case of the multichannel problem. That is, in that case, the min-CuSum controls the *worst-case* conditional probability of false isolation, with respect to all possible change points.

C. The multichannel problem

We conclude these theoretical developments by focusing on the special case of the *multichannel problem*, in which d independent channels are monitored, i.e.,

$$X_n = (X_{1,n}, X_{2,n}, \dots, X_{d,n}), \quad n \in \mathbb{N} \quad (14)$$

and each channel i takes values in some Polish space $\mathbb{S}_i$, so that

$$\mathbb{S} = \mathbb{S}_1 \times \dots \times \mathbb{S}_d. \quad (15)$$

Moreover, we assume that each $X_{i,n}$ has a density with respect to a σ -finite measure λ_i that is p_i for every $n \in \mathbb{N}$ if the channel does not undergo any change, whereas it is p_i for every $n \leq \nu$ and q_i for every $n > \nu$ if the change occurs in channel i . We assume that the change occurs in exactly one channel, therefore $K = d$ and

$$\begin{aligned} f(x_1, \dots, x_d) &= \prod_{j=1}^d p_j(x_j) \\ g_i(x_1, \dots, x_d) &= q_i(x_i) \prod_{j \neq i} p_j(x_j) \end{aligned} \quad (16)$$

for $(x_1, \dots, x_d) \in \mathbb{S}$. We also assume that, for each $i \in [K]$, the Kullback-Leibler divergences between p_i and q_i are positive and finite, i.e.,

$$\begin{aligned} \int \log(q_i/p_i) q_i d\lambda_i &\in (0, \infty) \\ \int \log(p_i/q_i) p_i d\lambda_i &\in (0, \infty), \end{aligned} \quad (17)$$

as well as that

$$\theta \in \mathbb{R} \rightarrow \log \left(\int q_i^\theta p_i^{1-\theta} d\lambda_i \right) \text{ is finite around } 0. \quad (18)$$

Proposition 1: Suppose that (14)-(18) hold. Then, conditions (A1)-(A2) are satisfied and (11) holds for every $\nu \in \{0\} \cup \mathbb{N}$.

Proof: Let $i, j \in [K]$ with $i \neq j$. First of all, we observe that by (14) and (16) we have

$$\ell_i(n) = \log \left(\frac{q_i(X_{i,n})}{p_i(X_{i,n})} \right).$$

As a result, for every $\theta \in \mathbb{R}$ we have

$$\psi_i(\theta) = \log \int q_i^\theta p_i^{1-\theta} d\lambda_i$$

which proves that (A1) holds in view of (18). Moreover,

$$\int (g_i/f) g_j d\lambda = \int (q_i/p_i) p_i d\lambda_i = 1,$$

which proves that (A2) holds. We know that $\{0\} \in \mathcal{N}$, so it remains to show that (11) holds for every $\nu \in \mathbb{N}$. Indeed, for any $\nu \in \mathbb{N}$ and $i \in [K]$, $b > 0$, $x \geq 0$ we have

$$\begin{aligned} \mathbb{P}_\infty(Y_i(\nu) > x \mid \sigma(b) > \nu) &= \mathbb{P}_\infty(Y_i(\nu) > x \mid \sigma_j(b) > \nu \text{ for every } j \in [K]) \\ &= \mathbb{P}_\infty(Y_i(\nu) > x \mid \sigma_i(b) > \nu) \\ &\leq \mathbb{P}_\infty(Y_i(\nu) > x), \end{aligned}$$

where the first equality follows by the definition of $\sigma(b)$, the second by the independence of the channels, and the inequality follows from [32, Theorem 1] and the fact that $\{Y_i(n), n \in \mathbb{N}\}$ is a nonnegative stochastically monotone Markov process. ■

V. NUMERICAL ILLUSTRATION

Consider the multichannel problem of Subsection IV-C with $d = 2$ channels, where both generate iid standard Gaussian observations before the change, and after the change the mean changes to 1 in one of the two sequences, i.e., $p_1 = p_2 = \text{Gaussian}(0, 1)$ and $q_1 = q_2 = \text{Gaussian}(1, 1)$.

Without loss of generality, due to symmetry, we consider the case that the change happens in the first channel. We consider different change-points, and for each of them we estimate the conditional probability of false isolation

$$\mathbb{P}_{\nu,1}(\hat{\sigma}(b) \neq 1 \mid \sigma(b) > \nu)$$

for different values of the threshold, b , using Monte Carlo simulation, with $N = 5 \times 10^4$ runs. In Figure 1, for each of these change-points we plot the conditional probability of false isolation as a function of b , and we see that it indeed decays exponentially in b . Moreover, we see that the curve stabilizes as the change-point increases. In particular, while the conditional probability of false isolation is point-wise smaller when $\nu = 0$ compared to $\nu > 0$, there is not a perceptible difference between the cases of $\nu = 50$ and $\nu = 100$. Standard errors are included as vertical bars in the figure.

VI. CONCLUSION

We have shown that in certain cases, such as the multichannel problem, a simple algorithm can control the worst-case probability of false isolation and minimize Lorden's delay criterion in an asymptotic sense, for every possible post-change distribution. Future work on this problem will address the case where the underlying problem does not have such a special structure, and it is necessary to design algorithms that explicitly address the isolation task.

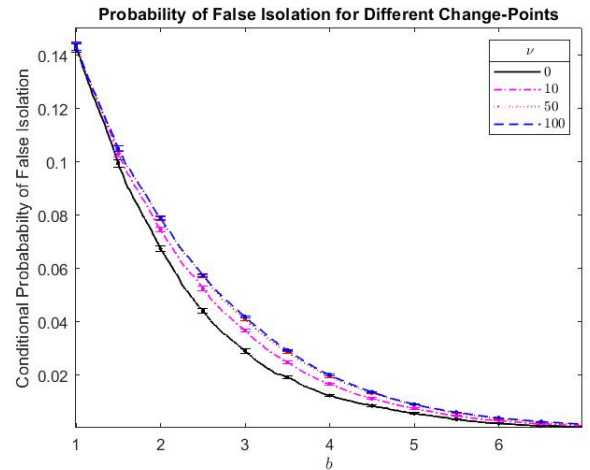


Fig. 1. In the multichannel problem, the probabilities of false isolation are stable as the change-point increases and decrease exponentially in the threshold b .

REFERENCES

- [1] W. A. Shewhart, *Economic control of quality of manufactured product*. Macmillan And Co Ltd, London, 1931.
- [2] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, no. 1/2, pp. 100–115, 1954.
- [3] A. Bissell, "Cusum techniques for quality control," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 18, no. 1, pp. 1–25, 1969.
- [4] D. M. Hawkins, P. Qiu, and C. W. Kang, "The changepoint model for statistical process control," *Journal of quality technology*, vol. 35, no. 4, pp. 355–366, 2003.
- [5] S. Joe Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 17, no. 8-9, pp. 480–502, 2003.
- [6] J. Ru, V. P. Jilkov, X. R. Li, and A. Bashi, "Detection of target maneuver onset," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 2, pp. 536–554, 2009.
- [7] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.
- [8] I. Nikiforov, V. Varavva, and V. Kireichikov, "Application of statistical fault detection algorithms to navigation systems monitoring," *Automatica*, vol. 29, no. 5, pp. 1275–1290, 1993.
- [9] B. Bakhache and I. Nikiforov, "Reliable detection of faults in navigation systems," in *Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No. 99CH36304)*, vol. 5. IEEE, 1999, pp. 4976–4981.
- [10] A. G. Tartakovsky, X. R. Li, and G. Yaralov, "Sequential detection of targets in multichannel systems," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 425–445, 2003.
- [11] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim, "Detection of intrusions in information systems by sequential change-point methods," *Statistical methodology*, vol. 3, no. 3, pp. 252–293, 2006.
- [12] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods," *IEEE transactions on signal processing*, vol. 54, no. 9, pp. 3372–3382, 2006.
- [13] H. Rolka, H. Burkom, G. F. Cooper, M. Kulldorff, D. Madigan, and W.-K. Wong, "Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: research needs," *Statistics in Medicine*, vol. 26, no. 8, pp. 1834–1856, 2007.
- [14] S. E. Fienberg and G. Shmueli, "Statistical issues and challenges associated with rapid detection of bio-terrorist attacks," *Statistics in Medicine*, vol. 24, no. 4, pp. 513–529, 2005.
- [15] D. Siegmund, "Change-points: from sequential detection to biology and back," *Sequential analysis*, vol. 32, no. 1, pp. 2–14, 2013.
- [16] I. V. Nikiforov, "A generalized change detection problem," *IEEE Transactions on Information theory*, vol. 41, no. 1, pp. 171–187, 1995.
- [17] A. G. Tartakovsky, "Multidecision quickest change-point detection: Previous achievements and open problems," *Sequential Analysis*, vol. 27, no. 2, pp. 201–231, 2008.
- [18] —, "An asymptotic theory of joint sequential changepoint detection and identification for general stochastic models," *IEEE Transactions on Information Theory*, 2021.
- [19] T. L. Lai, "Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 595–608, 2000.
- [20] I. V. Nikiforov, "A simple recursive algorithm for diagnosis of abrupt changes in random signals," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2740–2746, 2000.
- [21] S. Dayanik, C. Goulding, and H. V. Poor, "Bayesian sequential change diagnosis," *Mathematics of Operations Research*, vol. 33, no. 2, pp. 475–496, 2008.
- [22] S. Dayanik, W. B. Powell, and K. Yamazaki, "Asymptotically optimal bayesian sequential change detection and identification rules," *Annals of Operations Research*, vol. 208, no. 1, pp. 337–370, 2013.
- [23] X. Ma, L. Lai, and S. Cui, "Two-stage bayesian sequential change diagnosis," *IEEE Transactions on Signal Processing*, vol. 69, pp. 6131–6147, 2021.
- [24] G. V. Moustakides *et al.*, "Optimal stopping times for detecting changes in distributions," *the Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.
- [25] G. Lorden *et al.*, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.
- [26] T. Oskiper and H. V. Poor, "Online activity detection in a multiuser environment using the matrix cusum algorithm," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 477–493, 2002.
- [27] I. V. Nikiforov, "Sequential detection/isolation of abrupt changes," *Sequential Analysis*, vol. 35, no. 3, pp. 268–301, 2016.
- [28] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, pp. 206–227, 1985.
- [29] O. Hadjiliadis, H. Zhang, and H. V. Poor, "One shot schemes for decentralized quickest change detection," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3346–3359, 2009.
- [30] G. Fellouris and G. Sokolov, "Second-order asymptotic optimality in multisensor sequential change detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3662–3675, 2016.
- [31] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.
- [32] M. Pollak and D. Siegmund, "Convergence of quasi-stationary to stationary distributions for stochastically monotone markov processes," *Journal of applied probability*, vol. 23, no. 1, pp. 215–220, 1986.