Second-Order Asymptotically Optimal Change-point Detection Algorithm with Sampling Control

Qunzhi Xu
Georgia Institute of Technology
xuqunzhi@gatech.edu

Yajun Mei
Georgia Institute of Technology
ymei@isye.gatech.edu

George V. Moustakides
University of Patras, Greece
moustaki@upatras.gr

***Abstract*—In the sequential change-point detection problem for multi-stream data, it is assumed that there are M processes in a system and at some unknown time, an occurring event impacts one unknown local process in the sense of changing the distribution of observations from that affected local process. In this paper, we consider such problem under the sampling control constraint, in which one is able to take observations from only one of the local processes at each time step. Our objective is to design an adaptive sampling policy and a stopping time policy that is able to raise a correct alarm as quickly as possible subject to the false alarm and sampling control constraint. We develop an efficient sequential change-point detection algorithm under the sampling control that turns out to be second-order asymptotically optimal under the full data scenario. That is, with the sampling rate that is only $1/M$ of the full data scenario, our proposed algorithm has the same performance up to second-order as the optimal procedure under the full data scenario.**

***Index Terms*—Asymptotic optimality, change-point detection, CUSUM, quickest detection.**

I. Introduction

The sequential change-point detection problem for multi-stream data under the sampling constraint has many important real-world applications such as quality control, surveillance or security. Under a general setting, there are M possible data streams or resources available in a system, and at some unknown time, an occurring event impacts one unknown local process in the sense of changing the distribution of observations from that affected local process. However, one is only able to observe one of these M local streams at each time, due to the sampling cost or data process/analysis capacity. The objective is how to take observations adaptively and how to use the observed data to raise a correct alarm as quickly as possible once the change occurs subject to the false alarm constraint.

When the full data is available without any sampling control, this problem has been extensively studied in the sequential change-point detection literature, see [5]. When there is sampling control for monitoring multi-stream data, research has been conducted in the context of sequential hypothesis testing that does not involve change times, and a pioneer work is Chernoff [2] for $M = 2$ processes when the data are Bernoulli distributed. Also see a recent article [1] that extends to a more complicated case with a general M processes when the number of anomaly processes is unknown. For the sequential change-point detection problem for multi-stream data under sampling control, research is rather limited, and the only exception is [4], which proposed an efficient adaptive sampling strategy but did not provide any asymptotic optimality theorems. Finally, in [9] a version of CUSUM is proposed which is based on a time domain data-efficient sampling technique of a single data stream.

In this paper, we develop an efficient sequential change-point detection for multi-stream data with sampling control by taking advantage of the fact that there is only one local process affected by the occurring event. We propose to explore each local process extensively to decide whether or not there is a local change, and then switch to new processes only when we are confident that the existing process does not involve any local changes. It turns out that even with the sampling rate of $1/M$ at each time step, our proposed algorithm has the remarkable property of being second-order asymptotically optimum. To the best of our knowledge, the second-order asymptotic optimality under sampling control is the first of its kind in the sequential change-point detection literature.

The remainder of the paper is organized as follows. In Section II, we state the mathematical formulation of our problem and provide some technical background. In Section III, we present our proposed algorithm for the special case of $M = 2$ processes and then prove its second-order asymptotic optimality properties. In section IV, we extend our results to the general $M \geq 2$ case. Finally, numerical simulations are presented in Section V to illustrate the usefulness of our proposed algorithm. All technical proofs are postponed in the appendix.

II. Problem Formulation and Background

Suppose that there are M processes $\{X_t^i\}, i = 1, \ldots, M$ that are statistically independent. Initially, the system is under control, and in each process the observations $\{X_t^i\}$ are i.i.d. with a density $\mathsf{f}(X)$. At some unknown time τ, the system becomes out of control in the sense that one of its processes switches to a new distribution $\mathsf{g}(X)$. Specifically, if the ith data stream is affected, then

$$X_t^i \sim \begin{cases} \mathsf{f}(X), & \text{if } t \leq \tau; \\ \mathsf{g}(X), & \text{if } t > \tau. \end{cases} \tag{1}$$

The question is then to detect the change-time τ as quickly as possible once the change occurs.

Mathematically, a change-point detection procedure is defined as a stopping time T, where $\{T = t\}$ implies that one raises an alarm at time t based on the observations up to time t. Following the classical minimax formulation for sequential change-point detection proposed by Lorden [3], one is interested in finding a stopping time T that minimizes the worst-case detection delay

$$D(T) = \sup_{t\geq 0} \operatorname{esssup} \mathsf{E}_t[T-t|\mathscr{F}_t, T>t] \tag{2}$$

subject to the average run length to false-alarm constraint

$$\mathsf{E}_\infty[T] \geq \gamma > 1. \tag{3}$$

Here E_t denotes expectation with respect to the measure induced when the change occurs at $\tau = t$ and E_∞ denotes expectation with respect to the nominal measure (namely the change occurs at infinity). In addition, $\mathscr{F}_t$ is the σ-algebra generated by observed data/information up to time t.

If we observe the full data vector $X_t = [X_t^1, \cdots, X_t^M]$ at each time instant t, the problem is completely solved under two different scenarios. The first is when we know exactly which local stream is affected by the change, say, the ith local stream. In this case, the optimal procedure is the well-known CUSUM test

$$\mathsf{T}_i = \mathsf{T}_i(A) = \inf\{t > 0 : W_t^i \geq A\}, \tag{4}$$

where W_t^i is the CUSUM statistic defined recursively by

$$W_t^i = \max\{W_{t-1}^i, 0\} + \log \frac{\mathsf{g}(X_t^i)}{\mathsf{f}(X_t^i)}, \tag{5}$$

for $t \geq 1$ and $W_0^i = 0$ (see [7]). The second scenario with full data is when we do not know which local stream is affected by the change. In this case, an efficient approach is to monitor each local stream individually and raise a global alarm if any local CUSUM procedure T_i raises a local alarm. This yields a full-sampling mechanism that can be denoted as

$$\begin{aligned}\mathsf{T}_{\text{full}} &= \mathsf{T}_{\text{full}}(A) = \min\{\mathsf{T}_1, \cdots, \mathsf{T}_M\} \\ &= \inf\{t > 0 : \max\{W_t^1, \cdots, W_t^M\} \geq A\}.\end{aligned} \tag{6}$$

Note that different values of the threshold A are needed in order for $\mathsf{T}_i(A)$ in (4) and $\mathsf{T}_{\text{full}}(A)$ in (6) to satisfy the same false alarm constraint (3). It turns out that the family of stopping times T_{full} in (6) has the same detection delay (up to second-order) as each optimal procedure T_i in (4) subject to the false alarm constraint (3) and, consequently, it is second-order asymptotically optimal (see [5]).

In this paper, we investigate the above sequential change-point detection problem following a different strategy for the sampling control constraint. To be more specific, at each time step, we are allowed to take observations from exactly *one* of the M processes. Define the index variable $R_t^i = 1$ if the ith process is sampled at time t and $R_t^i = 0$ otherwise. Then the sampling control constraint requires that

$$R_t^1 + \cdots + R_t^M = 1 \qquad \text{for all time steps } t = 1, 2, \cdots \tag{7}$$

Under the sampling control scenario, the observation X_t^i is observable if and only if $R_t^i = 1$. Moreover, a statistical procedure should be defined as a stopping time T with respect to the observed data only. This can be thought of as a sparse quantized version of the full raw data.

The sequential change-point detection problem under the sampling control consists in designing the sampling policies $R_t = [R_t^1, \cdots, R_t^M]$ at each time step t and a corresponding stopping time T that minimizes the detection delay $D(T)$ in (2) subject to two constraints: one is the false alarm constraint in (3) and the other is the sampling constraint in (7). A new challenge arises due to the fact that we need to adaptively select which local process to sample at each time instant based on the previous observations.

Note that it is highly non-trivial to develop an efficient sequential change-point detection procedure under the sampling control in (7). For instance, a naive idea would be to adopt a block sampling policy where one splits the time domain into blocks with each block consisting of M time steps, and then one samples each local process from the first process to the last process within each block. That is, the naive sampling policy can be defined as $R_t^i = 1$ if and only if $t \bmod M = i$ for all $i = 1, \cdots, M$ and for all time steps $t = 1, 2, \cdots$. It is not difficult to show that the detection delay of this naive block sampling mechanism will be $MD(\mathsf{T}_{\text{full}})$, i.e., a factor of M times larger than the detection delay of the full sampling policy.

While one might expect that the sampling control constraint in (7) will lead to a larger detection delay, there is no reason to be by a factor of M as in the naive block sampling policy. Next we propose an efficient sequential change-point detection procedure under the sampling control in (7) that has the *same* asymptotic properties up to second-order as the optimal procedure T_i in (4) or the full-sampling procedure T_{full} in (6), subject to the false alarm constraint in (3). Our results show that the sampling control constraint in (7) has little impact on information bounds of the detection delay whenever only one local process can provide information to the occurring event.

III. Main results when $M = 2$

To highlight our main ideas, in this section, we investigate the simple scenario of $M = 2$ processes. In the next section these results are extended to cover the general case of M processes. Our current section is divided into two parts: in the first we present our candidate algorithm under the sampling control in (7) for $M = 2$, and in the second we establish its second-order asymptotic optimality properties.

A. Proposed algorithm

At a high level, our algorithm consists in exploring the fact that there is only one local process affected by the occurring event. Thus we propose to sample a local process extensively until we are confident to decide whether or not there is a local change. If there is a local change, then we stop and raise a global alarm. If there is no local change, then we switch to sample the next local process until we are confident to decide whether or not there is a local change. We repeat these steps until we raise a global alarm.

Below we present two equivalent ways to define our proposed algorithm and its corresponding stopping time $\mathsf{T} = \mathsf{T}(A)$. The first form follows the SPRT (Sequential Probability Ratio Test) representation of the optimal CUSUM procedure for T_i in (4), and allows us to derive the statistical properties of our proposed algorithm. The second is inspired by the full-sampling procedure T_{full} in (6), which allows us to easily implement our proposed algorithm.

Let us begin with the SPRT-based definition of our algorithm for monitoring $M = 2$ processes. Without loss of generality, at time $t = 1$, we start sampling from the first process. Then we keep on sampling the first process until the log-likelihood ratio statistic $S_t^1 = \sum_{l=1}^t \log \frac{\mathsf{g}(X_l^1)}{\mathsf{f}(X_l^1)} \notin (0, A)$. In other words, we sample the first process until the SPRT stops at time

$$\tau_1 = \inf\{t > 0 : S_t^1 \notin (0, A)\}. \tag{8}$$

When $S_{\tau_1}^1 \geq A$ we declare that the first process has a local alarm, and thus we raise a global alarm at time $\mathsf{T} = \tau_1$. When $S_{\tau_1}^1 \leq 0$ we declare that the first process is under control, and we switch to the second process and start taking samples at time $\tau_1 + 1$. That is, we define another SPRT from the second process

$$\tau_2 = \inf\{t > \tau_1 : S_t^2 = \sum_{l=\tau_1+1}^{t} \log \frac{\mathsf{g}(X_l^2)}{\mathsf{f}(X_l^2)} \notin (0, A)\}. \tag{9}$$

If $S_{\tau_2}^2 \geq A$ we declare that the second process has a local alarm and we raise a global alarm at time $\mathsf{T} = \tau_2$. When $S_{\tau_2} \leq 0$ we decide that the second process is under control and we switch back to sampling the first process. We apply again the SPRT to the new observations from the first process, starting from time $\tau_2 + 1$. This leads to a new SPRT from the first process

$$\tau_3 = \inf\{t > \tau_2 : \sum_{l=\tau_2+1}^{t} \log \frac{\mathsf{g}(X_l^1)}{\mathsf{f}(X_l^1)} \notin (0, A)\}. \tag{10}$$

If the upper bound A is crossed at time τ_3 we raise the global alarm at time $\mathsf{T} = \tau_3$. If the lower bound 0 is crossed then we switch to sampling the second process. We repeat this procedure until we stop.

In summary, when monitoring $M = 2$ processes, the sampling policy of our proposed algorithm is to define $R_t^1 = 1$ if $t \in [1, \tau_1] \cup [\tau_2 + 1, \tau_3] \cup \cdots$, whereas $R_t^2 = 1$ if $t \in [\tau_1 + 1, \tau_2] \cup [\tau_3 + 1, \tau_4] \cup \cdots$. Moreover, our proposed algorithm raises a global alarm at time

$$\mathsf{T} = \min\Big\{\inf\{\tau_{2k+1} : S_{\tau_{2k+1}}^1 \geq A\}, \; \inf\{\tau_{2k} : S_{\tau_{2k}}^2 \geq A\}\Big\}. \tag{11}$$

The definition of our proposed algorithm in the form of (11) allows us to easily derive the theoretical properties, but it is computationally inefficient. Inspired by the full-sampling procedure T_{full} in (6), we propose an equivalent way to define T that is computationally simple.

To define the alternative form we initialize the CUSUM-type statistics $W_0^i = 0, i = 1, 2$. At each time instant, we update W_t^i as in the classical CUSUM statistic in (5) if the ith process is sampled and re-initialize the other process that is not being sampled. Specifically, for $i = 1, 2$ we have

$$W_t^i = \begin{cases} 0, & \text{if } R_t^i = 0 \\ \max\{W_{t-1}^i, 0\} + \log \frac{\mathsf{g}(X_t^i)}{\mathsf{f}(X_t^i)}, & \text{if } R_t^i = 1. \end{cases} \tag{12}$$

We note that the main difference between the previous form of W_t^i and the one used in the classical CUSUM statistics in (5) is that W_t^i is kept equal to 0 whenever $R_t^i = 0$, e.g., when the ith process is not sampled.

With the previous definition our proposed stopping time T defined in (11) can be equivalently written as

$$\mathsf{T} = \{t > 0 : \max\{W_t^1, W_t^2\} \geq A\}. \tag{13}$$

Moreover, the sampling policy of our proposed method can be summarized by the following algorithm:

Step 1: Sample Process 1 until $W_t^1 \notin (0, A)$. If we do not stop sampling according to the stopping rule in (13), switch to Process 2.

Step 2: Sample Process 2 until $W_t^2 \notin (0, A)$. If we do not stop sampling according to the stopping rule in (13), switch to Process 1.

Step 3: Go back to Step 1.

B. Asymptotic optimality properties

Before presenting the asymtotic optimality properties, it is useful to mention a standard assumption in the classical sequential analysis literature: we assume that the following Kullback-Leibler information numbers are well-defined:

$$\begin{aligned} I(\mathsf{f}, \mathsf{g}) &= \int \log \frac{\mathsf{f}(X)}{\mathsf{g}(X)} \mathsf{f}(X) dX, \\ I(\mathsf{g}, \mathsf{f}) &= \int \log \frac{\mathsf{g}(X)}{\mathsf{f}(X)} \mathsf{g}(X) dX. \end{aligned} \tag{14}$$

Here $I(\mathsf{f}, \mathsf{g})$ and $I(\mathsf{g}, \mathsf{f})$ will be used to estimate the expected sample size of the SPRTs.

Clearly our proposed algorithm satisfies the sampling control constraint in (7). Thus it suffices to analyze the statistical properties of the proposed stopping time T in (11) or (13) in terms of the average run length to false alarms and detection delay. The following theorem summarizes the main results.

Theorem 1 *For our proposed stopping time* T *in (11) or (13), we have*

$$\mathsf{E}_\infty[\mathsf{T}] \geq e^A. \tag{15}$$

Moreover, its detection delay satisfies

$$D(\mathsf{T}) \leq \frac{A}{I(\mathsf{g}, \mathsf{f})} + O(1), \tag{16}$$

as $A \to \infty$.

Proof When the system is under control, both processes have the same pre-change distribution, and applying SPRT between the processes is equivalent to applying the SPRT to the same process, which implies that our proposed stopping time T is stochastically equivalent to the classical CUSUM procedure T_i depicted in (4). Hence, relation (15) follows directly from this and the well-known property of CUSUM that $\mathsf{E}_\infty[\mathsf{T}_i] \geq e^A$, see Theorem 2 in [3].

The proof of (16) is based on the relationship between our proposed stopping time T and the classical CUSUM procedure T_i in (4). However, they are no longer stochastically equivalent under the alternative hypothesis when a change occurs. This is because only one of the two processes is under the post-change distribution and the other is still under the pre-change. Switching to sample from the process under the pre-change distribution will increase the detection delay of our proposed stopping time T as compared to the classical CUSUM procedure T_i in (4). The good news is that such increase on the detection delay is of order $O(1)$, since the SPRTs from the process under the pre-change distribution can easily cross the lower bound 0 and have little impact on the false alarm or detection delays. More technical details are presented in the appendix for completeness.

The following corollary shows that our proposed algorithm is asymptotically optimal up to $O(1)$, as the false alarm constraint $\gamma \to \infty$. In the sequential change-point detection literature, the order $o(\log\gamma)$ is often referred as the first-order, and the order $O(1)$ is referred as the second-order. Thus our proposed algorithm is asymptotically optimal up to second-order.

Corollary 1 *Let $A = \log\gamma$. Then our proposed algorithm T in (11) or (13) satisfies the average run length to false alarm constraint in (3). Moreover, its detection delay satisfies the bounds*

$$D_{\text{opt}} \leq D(\mathsf{T}) \leq D_{\text{opt}} + O(1), \tag{17}$$

where D_{opt} is the smallest detection delay of any stopping time under the full-sampling scenario that satisfies the average run length to false alarm constraint in (3):

$$D_{\text{opt}} = \frac{\log\gamma}{I(\mathsf{g},\mathsf{f})} - C_0 \tag{18}$$

for some constant C_0 that only depends on f and g.

Proof The smallest bound D_{opt} is from the optimality properties of the CUSUM procedure T_i in (4) when it is known that the ith process is affected by the occurring event, also see equation (2.14) of [6] for the same bound. All other statements in the corollary follow directly from Theorem 1.

IV. Extension to the General M Processes

Our proposed algorithm for $M = 2$ processes can be easily extended to the general case of $M \geq 2$ processes, by switching among all M processes. To be more precise, let us consider the CUSUM-type definition of our proposed algorithm, and define the CUSUM-type statistics $W_t^i, i = 1, \dots, M$. Here at each time step t, $M-1$ values of the W_t^i will be set to 0 corresponding to the processes not being sampled (i.e. $R_t^i = 0$) and only one statistic will be properly updated corresponding to the process being sampled (i.e. $R_t^i = 1$). The updating equation is exactly the same as in (12) only now $i = 1, \dots, M$. Also the stopping rule in (13) becomes

$$\mathsf{T} = \{t > 0 : \max\{W_t^1, \dots, W_t^M\} \geq A\}. \tag{19}$$

For sampling we apply a cyclical sampling rule with respect to the available processes:

Step 1: Sample Process 1 until $W_t^1 \notin (0, A)$. If we do not stop sampling according to the stopping rule in (19), switch to Process 2.

Step 2: Sample Process 2 until $W_t^2 \notin (0, A)$. If we do not stop sampling according to the stopping rule in (19), switch to Process 3.

⋮

Step M: Sample Process M until $W_t^M \notin (0, A)$. If we do not stop sampling according to the stopping rule in (19), switch to Process 1.

Step $M+1$: Go back to Step 1.

Theorem 1 and Corollary 1 still hold for T in the general M case in the classical asymptotic setting when M is fixed as $\gamma \to \infty$. In such a setting, the constant $O(1)$ will be proportional to the number M of processes. Because of this fact, even if we allow M to grow to infinity with γ, our test is still *first-order* asymptotically optimum as long as $M = o(\log\gamma)$. It remains an open problem to derive an efficient algorithm when $M \gg \log\gamma$.

V. Simulation

We conduct the Monte Carlo simulation studies to illustrate the usefulness of our theoretical results. Assume $\mathsf{f} \sim N(0,1)$ and $\mathsf{g} \sim N(1,1)$. In our simulations, we consider two cases

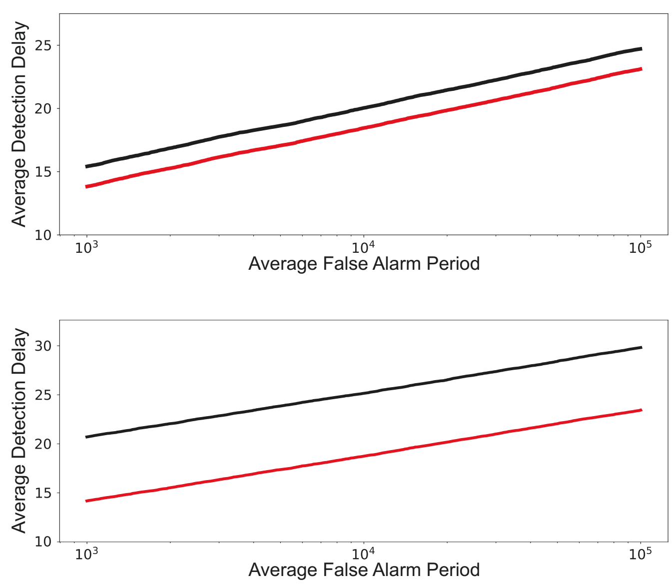

Fig. 1. Top: $M = 2$ processes. Bottom: $M = 5$ processes. In each plot, the x-axis is the false alarm constraint γ value (in a base 10 logarithmic scale) in (3) varying from 10^3 to 10^5, and the y-axis is the detection delay. There are two curves in each plot: the red curve represents the optimal CUSUM procedure, whereas the black curve represents our proposed algorithm.

for the number of available processes, namely $M = 2, 5$. In each case, we compute the detection delay of our proposed algorithm $D(\mathsf{T})$ and compare it against the classical CUSUM procedure T_i when the ith local process is known to be affected by the occurring event. Since our detection delay is defined in the worst-case sense, we simulate $D(\mathsf{T})$ by assuming that the change occurs to the Mth process at time $\tau = 0$ which suggests that we need to test all non-affected processes before testing the one that has changed. This clearly generates the worst-case detection delay. We compare the two tests for the same run length to false alarm (average false alarm period). To construct our graphs we performed 2500 Monte Carlo independent runs for different values of the threshold A.

From Figure 1, it is clear that the gap between the two curves remains constant as $\gamma \to \infty$. This is consistent with our theoretical result that our proposed algorithm is asymptotically optimal up to $O(1)$. Also the gap increases with the number M of sensors, which is also consistent with our theoretical results.

Appendix: proof of delay (16) in Theorem 1

To prove the detection delay relation in (16), without loss of generality, we assume that the change occurs to the second process at time $\tau = 0$. Recall that our proposed algorithm T in (11) can be defined as the sum of stopping time differences $\mathsf{T} = \tau_k = (\tau_k - \tau_{k-1}) + \cdots + (\tau_2 - \tau_1) + (\tau_1 - 0)$ where the difference $\tau_i - \tau_{i-1}$ corresponds to an SPRT decision delay. For the detection delay of T, about half of these k differences follow the pre-change distribution f, and the remaining ones the post-change g.

For notational convenience, let δ be the SPRT in the problem of testing $H_0 : \mathsf{f}$ against $H_1 : \mathsf{g}$ based on the observations $X_1, \cdots, X_t$:

$$\delta = \inf\{t > 0 : S_t = \sum_{l=1}^{t} \log \frac{\mathsf{g}(X_l)}{\mathsf{f}(X_l)} \notin (0, A)\}.$$

Denote the SPRTs under the pre-change as $\delta_1^{(0)}, \delta_2^{(0)}, \cdots$, whereas denote the SPRTs under the post-change as $\delta_1^{(1)}, \delta_2^{(1)}, \cdots$. Then it is not difficult to show that under P_0, our proposed stopping time T satisfies

$$\mathsf{T} \le \sum_{j=1}^{k} \delta_j^{(0)} + \sum_{j=1}^{k} \delta_j^{(1)},$$

where k is the first time when the SPRTs $\delta_i^{(1)}$ crosses the upper bound A. Note that the procedure in the right-hand side never stops sampling at the first-process, and only stops sampling and raises a global alarm when the second process raises an alarm. Moreover, under P_0, the classical CUSUM procedure T_i in (4) for process $i = 2$ can be written as

$$\mathsf{T}_2 = \sum_{j=1}^{k} \delta_j^{(1)},$$

see equation (2.50) in [8]. In addition, equation (2.52) in [8] shows that

$$\begin{aligned} \mathsf{E}_0[\mathsf{T}_2] &= \mathsf{E}_0\left[\sum_{j=1}^{k} \delta_j^{(1)}\right] = \mathsf{E}_0\left[\delta_1^{(1)}\right] \mathsf{E}_0[k] \\ &= \frac{\mathsf{E}_{\mathsf{g}}[\delta]}{\mathsf{P}_{\mathsf{g}}(S_\delta \ge A)} = \frac{A}{I(\mathsf{g}, \mathsf{f})} + O(1), \end{aligned}$$

as $A \to \infty$.

As compared to the classical CUSUM procedure T_i in (4), our proposed algorithm has an extra term

$$\mathsf{E}\left[\sum_{j=1}^{k} \delta_j^{(0)}\right] = \mathsf{E}[\delta_1^{(0)}]\mathsf{E}[k] = \frac{\mathsf{E}_{\mathsf{f}}[\delta]}{\mathsf{P}_{\mathsf{g}}(S_\delta \ge A)}.$$

Here it is important to note that k depends on the second process under the post-change hypothesis, whereas $\delta_j^{(0)}$ are the SPRTs from the first process under the pre-change hypothesis.

When $I(\mathsf{f}, \mathsf{g}) > 0$, it can be shown that $\mathsf{E}_{\mathsf{f}}(\delta) = O(1)$. Also when $I(\mathsf{g}, \mathsf{f}) > 0$, we have $\inf_{A>0} \mathsf{P}_{\mathsf{g}}(S_\delta \ge A) > \beta_0$, for some $\beta_0 > 0$ bounded away from 0, e.g. the SPRT will cross the upper bound A with non-zero probability under the alternative hypothesis when g is true. Combining these results yields that $\mathsf{E}[\sum_{j=1}^{k} \delta_j^{(0)}] = O(1)$. Hence, the detection delay relation (16) in Theorem 1 is valid.

Acknowledgement

This work was supported in part by the US National Science Foundation under Grant CIF 1513373 through Rutgers University and under Grant DMS 1830344 through Georgia Institute of Technology.

References

[1] A. Tsopelakos, G. Fellouris, and V. V. Veeravalli, “Sequential anomaly detection with observation control,” 2019 IEEE International Symposium on Information Theory (ISIT), 2019.

[2] H. Chernoff, “Sequential design of experiments,” *The Annals of Mathematical Statistics* , vol. 30, no. 3, pp. 755–770, 1959.

[3] G. Lorden, “Procedures for reacting to a change in distribution,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.

[4] K. Liu, Y. Mei, J. Shi, “An adaptive sampling strategy for online high-dimensional process monitoring,” *Technometrics*, vol. 57, no. 3, pp. 305–319, 2015.

[5] A. Tartakovsky, I. Nikiforov, M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*, Chapman and Hall/CRC, 2014.

[6] Y. Mei, “Sequential change-point detection when unknown parameters are present in the pre-change distribution,” *The Annals of Statistics*, vol. 34, no. 1, pp. 92–122, 2006.

[7] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions,” *The Annals of Statistics*, vol. 14, no. 4, pp. 1379–1387, 1986.

[8] D. Siegmund, *Sequential analysis: Tests and confidence intervals*, Springer, 1985.

[9] T. Banerjee and V. V. Veeravalli, “Data-efficient quickest change detection in minimax settings,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6917-6931, 2013.