Sequential Analysis
Design Methods and Applications

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsqa20

Asymptotic optimality theory for active quickest detection with unknown postchange parameters

Qunzhi Xu & Yajun Mei

To cite this article: Qunzhi Xu & Yajun Mei (2023) Asymptotic optimality theory for active quickest detection with unknown postchange parameters, Sequential Analysis, 42:2, 150-181, DOI: 10.1080/07474946.2023.2187417

To link to this article: https://doi.org/10.1080/07474946.2023.2187417

Published online: 23 May 2023.

SEQUENTIAL ANALYSIS
2023, VOL. 42, NO. 2, 150–181
https://doi.org/10.1080/07474946.2023.2187417

Asymptotic optimality theory for active quickest detection with unknown postchange parameters

Qunzhi Xu and Yajun Mei

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

ABSTRACT
The active quickest detection problem with unknown postchange parameters is studied under the sampling control constraint, where there are p local streams in a system but one is only able to take observations from one and only one of these p local streams at each time instant. The objective is to raise a correct alarm as quickly as possible once the change occurs subject to both false alarm and sampling control constraints. Here we assume that exactly one of the p local streams is affected, and the postchange distribution involves unknown parameters. In this context, we propose an efficient greedy cyclic sampling–based quickest detection algorithm and show that our proposed algorithm is asymptotically optimal in the sense of minimizing the detection delay under both false alarm and sampling control constraints. Numerical studies are conducted to show the effectiveness and applicability of the proposed algorithm.

ARTICLE HISTORY
Received 8 November 2022
Revised 2 February 2023
Accepted 21 February 2023

KEYWORDS
Active sampling; asymptotic optimality; change point detection; CUSUM

1. INTRODUCTION

In the big data age, active quickest detection problems in multistream data have a wide range of applications in quality control, surveillance or security, etc. Under a general setting, there are p data streams available in a system, and at some unknown time an event might occur and affect some local streams in the sense of changing the distribution of its data. Depending on how we access the data, the problems can be divided into two distinct scenarios: the passive change point problem where one passively collects the data and the active change point problem where one is able to actively select the observed data, often with some certain kind of sampling rate constraint.

The passive change point problem and its extension have been well studied in the literature. The classical version of this problem is the case where one monitors $p = 1$ local stream, and many well-known procedures have been developed; see Page (1954), Pollak (1987), Lai (1995), Lorden and Pollak (2008), to name a few. For a review, see books such as Poor and Hadjiliadis (2008) and J. Chen and Gupta (2012). In recent years, research into monitoring $p \geq 2$ local streams in a passive setting has received extensive attention. Mei (2010) used the sum of local cumulative sum (CUSUM) statistics as the global statistic and Y. Xie and Siegmund (2013) suggested a mixture likelihood ratio

CONTACT Qunzhi Xu xuqunzhi@gatech.edu H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 755 Ferst Drive, NW, Atlanta, GA 30332.
This paper was presented in part at the 2021 IEEE International Symposium on Information Theory.

approach. Later, Chan (2017) developed asymptotic optimality theory for large-scale independent Gaussian data streams. For more extensions on the passive monitoring of multiple data streams, see J. Li (2020), who developed nonparametric methods for change detection in high-dimensional data, and Y. Chen and Li (2019), who considered the scenario when each local stream has its own change point. Also see Tartakovsky et al. (2014), Liu et al. (2019) and Wu (2020) for more related contents.

Research is rather limited for active quickest detection problems when monitoring $p \geq 2$ local streams under the sampling control constraint where one needs to decide which local stream is to be observed at each and every time step. This topic was first studied as early as 1963 by Shiryaev with a radar system rotating to observe exactly one out of p possible directions. Though Shiryaev proposed a useful algorithm, there was no asymptotically optimal theorem until Xu, Mei, and Moustakides (2020, 2021) developed the first of its kind under the simplest scenario when there is exactly only one affected local stream and the postchange distribution is completely specified. It is worth mentioning that there are some work which focus on the methodology without any kind of asymptotic optimality, see Liu et al. (2015) and Xie et al. (2021). Recently, Fellouris and Veeravalli (2022) considered a more general setting when the postchange distribution belongs to a finite prespecified set. It remains an open problem to investigate the setting when the postchange distribution involves unknown parameters that might have infinitely many possible values.

In this article, we study the active quickest detection problem with unknown postchange parameters when there is only one affected local stream under the sampling constraint that we are allowed to observe one and only one of the p local streams per time step. We develop an efficient algorithm named greedy cyclic sampling–cumulative sum (GCS-CUSUM) where the postchange distributions are unknown. Conceptually, our proposed algorithm alternates two different sampling policies: one is the greedy sampling policy that observes the local stream that may, and is progressively more likely to, contain the change, and the other is the cyclic sampling policy that switches to the next local stream if no local stream involves a local change. Our main contribution is to prove that even with the sampling rate of $1/p$ at each time step subject to the average run length to false alarm constraint of γ, the proposed GCS-CUSUM algorithm has the remarkable property of having the same detection delay performance up to first order as the oracle procedure that knew which stream is affected when the dimension $p = o(\log \gamma)$ and $\gamma \to \infty$.

We need to point out that this work is a nontrivial extension of our previous work in Xu, Mei, and Moustakides (2021) where the postchange distributions are completely specified. Under our current setting, the postchange distributions involve unknown parameters, and it is highly nontrivial to develop asymptotically optimal theorems with the unknown postchange parameters using existing online or active quickest detection techniques. The main reason is that information might be lost when sequentially estimating the parameters and switching among different streams. To overcome such difficulty, we borrow the tools from Lorden and Pollak (2008) to address the unknown postchange parameters and combine them with those in Xu, Mei, and Moustakides (2021) to incorporate the uncertainty of estimating unknown postchange parameters and the adaptive nature of the sampling policy. We feel that our work is a solid step forward on the active quickest detection problem under sampling control and will shed new light for future research.

It is also useful to point out that the sampling control has been extensively studied in two other well-known problems: the multiarmed bandit problems and sequential hypothesis testing. See Lai (1987), S. Li et al. (2019), Tsopelakos, Fellouris, and Veeravalli (2019), among others. In those contexts, all observations will provide some information for decision making. Here we should emphasize that our setup of sequential change point detection problems under sampling control poses new challenges because observation does not provide information to the quickest detection unless it is taken from the affected stream after the change occurs.

The remainder of the article is organized as follows. In Section 2, we state the mathematical formulation of our problem and review some existing methods. In Section 3, we present our proposed algorithm, and in Section 4 we provide its theoretical properties. Numerical studies are presented in Section 5 to illustrate the performance properties of our proposed algorithm, and we present the proof of the main theorem in Section 6. Finally, we conclude our article in Section 7.

2. PROBLEM FORMULATION AND BACKGROUNDS

For a better presentation, we divide the current section into two parts. In Subsection 2.1, we present the mathematical formulation of our problem, and in Subsection 2.2 we review several existing methods.

2.1. Problem Formulation

Suppose there are p statistically independent local streams in a system, and denote by X_t^i the observation from the ith local stream at time t, for $i = 1, \dots, p$ and $t = 1, 2, \dots$. Let $f_\theta(x) = \exp(\theta x - \psi(\theta))$ be the probability density/mass function of a one-parameter exponential family of distributions. Note that this includes many widely used distributions such as the Gaussian distribution, gamma distribution, and binomial distribution, and it also allows us to investigate the case when there are uncountably many possible values for postchange parameters. Initially, the system is under control and the data $\{X_t^i\}$ from the ith stream are independent and identically distributed (i.i.d.) with the density $f_{\theta_0}(X)$, independent of i. At some unknown time $t = \tau$, a triggering event occurs to the system and affects exactly one of its p streams, say, the ith, in the sense of changing its local distribution to a new unknown postchange density $f_{\theta_i}(X)$. Specifically, if the ith local stream is affected,

$$X_t^i \sim \begin{cases} \exp(\theta_0 x - \psi(\theta_0)), & \text{if } t \le \tau \\ \exp(\theta_i x - \psi(\theta_i)), & \text{if } t > \tau, \end{cases} \tag{2.1}$$

whereas for all other unaffected local streams, $j \neq i$, $X_t^j \sim f_{\theta_0}(x) = \exp(\theta_0 x - \psi(\theta_0))$ all $t > 0$ when $j \neq i$. Here we consider in detail the one-sided change point problem, where it is assumed that θ_0 is known and $\theta_i > \theta_0$ for the given ith affected data stream. In particular, the impact of the change can be different for different streams, and we denote by θ_i the possible postchange parameters if the ith data stream is affected.

Let us now discuss the sampling control constraints. To be rigorous, define a sequence of *sampling indices* $\{R_t\}$ with $R_t \in \{1, \dots, p\}$, where R_t is a random variable and $\{R_t = m\}$ means that we will sample the mth local stream at time instant t. Under our sampling constraint, we are allowed to access only one of these p local streams at each time t, and this can be expressed as

$$1_{\{R_t=1\}} + \cdots + 1_{\{R_t=p\}} = 1 \text{ for all times } t = 1, 2, ..., \tag{2.2}$$

where 1_A denotes the indicator function of the event A.

In the active quickest detection problem under sampling control, an algorithm consists of two components: one is the sampling policy in the sense of dynamically choosing $\{R_t\}$ at each and every time instant t subject to the sampling constraint in (2.2), and the other is the decision policy that is defined as the stopping time T with respect to the observed data sequence $\{X_t^{i=R_t}\}_{t\geq 1}$. Note that the sampling decision R_t depends only on those observed data up to time $t-1$, and the stopping time $\{\mathsf{T} = t\}$ means that we raise an alarm at time t.

Following the classical minimax formulation for quickest detection proposed by Pollak (1985), we are interested in finding a procedure $(\{R_t\}_{t=1,\ldots,\infty}, \mathsf{T})$ that minimizes the worst-case detection delay conditioned on that we stop after the change time t,

$$D_i(\mathsf{T}) = \sup_{t\geq 0} \mathsf{E}_t^i[\mathsf{T} - t | \mathsf{T} > t]. \tag{2.3}$$

for any $i = 1, 2, ..., p$ when the ith local stream is affected by the change, subject to the average run length to false alarm constraint

$$\mathsf{E}_\infty[\mathsf{T}] \geq \gamma > 1. \tag{2.4}$$

Here $\mathsf{P}_t^i(\cdot), \mathsf{E}_t^i[\cdot]$ denote the probability measure and the corresponding expectation induced by the change occurring at the ith local stream at time $\tau = t$ and $\mathsf{P}_\infty(\cdot), \mathsf{E}_\infty[\cdot]$ denote the probability measure and the corresponding expectation induced by the change occurring at ∞.

2.2. Review of Existing Methods

Let us now review some existing research that is related to our problem. First, under the unrealistic scenario where we had the true knowledge on the index i of affected data stream and the postchange parameters f_{θ_i}, it is natural to always sample the ith stream—that is, $R_t \equiv i$ for all t—and utilize the well-known CUSUM procedure to raise an alarm at time

$$\mathsf{T}_{\text{oracle}}(A) = \inf\{t \geq 1 : W_t^i \geq A\}, \tag{2.5}$$

where W_t^i is the CUSUM statistics recursively defined as

$$W_t^i = \max\{W_t^{i-1}, 0\} + \log \frac{f_{\theta_i}(X_t^i)}{f_{\theta_0}(X_t^i)} \quad \text{for } t \geq 1 \tag{2.6}$$

and the initial value $W_0^i = 0$; see Moustakides (1986). Here the threshold A is chosen to satisfy the average run length to false alarm constraint γ in (2.4). We use $\mathsf{T}_{\text{oracle}}$ to emphasize that this CUSUM procedure makes an oracle assumption of known affected local streams and known postchange distribution.

Note that it is highly nontrivial to develop an efficient algorithm under our setup due to two challenges. The first, probably easier, one is that the postchange distributions are unknown. This challenge has been tackled when monitoring $p = 1$ local stream in Lorden and Pollak (2008). Their main idea is to estimate the postchange parameter θ_i by the average of recent observations $\hat{\theta}_{t,i}$ after the candidate change point and update the local statistics as in the classical CUSUM statistic. The local statistics $\tilde{W}_t^i$ can be defined as in the recursion (2.6) with θ_i replaced by $\hat{\theta}_{t,i}$; that is,

$$\tilde{W}_t^i = \max\{\tilde{W}_t^{i-1}, 0\} + \log \frac{f_{\hat{\theta}_{t,i}}(X_t^i)}{f_{\theta_0}(X_t^i)}.$$

This yields Lorden and Pollak's procedure, resulting in

$$\mathsf{T}_{\mathrm{LP}}(A) = \inf\left\{t > 0 : \tilde{W}_t^i \geq A\right\}. \tag{2.7}$$

The second, probably more fundamental, challenge is that the index i of the true affected local stream is unknown, and thus it is unclear how to choose sampling indices $\{R_t\}$ suitably to detect the change quickly. A naive sampling idea is to sample each local process purely cyclically—that is, $R_t = tmodp + 1R_t = tmodp + 1$ for all time instants $t = 1, 2, \ldots$—and each local stream is visited only once during each p time instant. Combing this cyclic sampling policy with Lorden and Pollak's procedure in (2.7) yields the following quickest detection algorithm:

$$\mathsf{T}_{\mathrm{cyclic}}(A) = \inf\left\{t > 0 : \max\left\{\tilde{W}_t^1, \ldots, \tilde{W}_t^p\right\} > A\right\}, \tag{2.8}$$

where $\tilde{W}_t^i (i = 1, \ldots, p)$ is only updated when $R_t = i$. In the sequel we will refer to (2.8) as the cyclic algorithm with the purely cyclic sampling policy.

Clearly, the purely cyclic algorithm in (2.8) seems to be inefficient, because it might spend too much time on those $p - 1$ unaffected local streams. To the best of our knowledge, no efficient algorithms have been developed in the quickest detection literature to simultaneously address these two challenges of unknown postchange distribution and unknown index of affected local streams.

3. OUR PROPOSED ALGORITHM

In this section, we present our proposed algorithm, denoted by $\mathsf{T}_{\mathrm{GCS}}$, based on the GCS policy. At a high level, we propose to sample one stream until we are confident in deciding whether a local change has occurred or not. If we detect a local change, then we stop and raise a global alarm. If we decide there is no local change or we have sampled from the same stream for a long time, then we switch to sample from another stream. We repeat these steps until we raise an alarm.

For better presentation, the current section is divided into three subsections: in Subsection 3.1 we define local statistics, which will be the cornerstone of our algorithm. We propose the GCS policy in Subsection 3.2 and the decision policy in Subsection 3.3.

3.1. Local Statistics

For the sake of clarity, we define two sets of local monitoring statistics, $\tilde{W}_t^i$ and $\hat{W}_t^i$, for the ith local stream at time t. The former is used to update the observed data, and the latter also takes into account a possible switch to sampling different data streams.

Let us first define the local statistics $\tilde{W}_t^i$. When the ith local stream is observed, we update its local statistics based on Lorden and Pollak's procedure in (2.7). When the ith local stream is not observed, we treat it as missing data and the corresponding log-likelihood ratio of missing data as 0. Mathematically, at each time instant $t = 1, 2, \ldots,$ let $\hat{\theta}_{t,i}$ be the estimate of the postchange parameter for the ith stream at time t, which will be defined later, and the local statistics $\tilde{W}_t^i$ can be defined recursively as

$$\begin{aligned} \tilde{W}_t^i &= \max\{\tilde{W}_{t-1}^i, 0\} + 1_{\{i=R_t\}} \log \frac{f_{\hat{\theta}_{t,i}}(X_t^i)}{f_{\theta_0}(X_t^i)} \\ &= \begin{cases} \max\{\tilde{W}_{t-1}^i, 0\}, & \text{if } i \neq R_t \\ \max\{\tilde{W}_{t-1}^i, 0\} + \log \dfrac{f_{\hat{\theta}_{t,i}}(X_t^i)}{f_{\theta_0}(X_t^i)}, & \text{if } i = R_t, \end{cases} \end{aligned} \tag{3.1}$$

with the initial values $\tilde{W}_0^i = 0$ for all $i = 1, \ldots, p$.

As for the postchange parameter estimators $\hat{\theta}_{t,i}$, by (3.1), we only need to pay attention to the sampled local stream and thus adopt the same idea as in Lorden and Pollak's procedure (2.7). To be more concrete, at time instant t, assume that we sample at the ith stream. Denote by $M(t)$ the total time instants in which we have consecutively sampled at the ith stream, which can be recursively updated as

$$M(t) = \begin{cases} M(t-1) + 1, & \text{if } R_t = R_{t-1} \\ 1. & \text{otherwise} \end{cases}. \tag{3.2}$$

Here we propose to estimate the postchange parameter based on the observed data from the ith stream during the time period of $t - M(t) + 1$ to $t - 1$, because we save the data at the time instant t for quickest detection, not for parameter estimation. One natural idea is to consider the method of moments estimator of the distribution

$$\hat{\theta}_{t,i}^{\text{MOM}} = \psi_1\left(\frac{\sum_{\ell=t-M(t)+1}^{t-1} X_\ell^i}{M(t) - 1}\right),$$

where we define 0/0=$-\infty$ and $\psi_1(\cdot)$ is the inverse function of $\psi'(\cdot)$. However, it turns out that the proposed method of moments estimator is unable to handle the case when the magnitude of change is extremely small (i.e., θ_i is close to θ_0), and we need to define a constant δ that indicates the smallest magnitude of change—that is, $\theta_i \geq \delta > \theta_0$ for any given ith affected stream. Also, due to technical issues, we need to define an additional constant ζ that indicates the largest magnitude of change; that is, $\theta_i < \zeta$ for all i. It is worth mentioning that for Gaussian mean change cases, the upper bound of change ζ can be removed. Mathematically, our proposed estimator for the ith stream at time t is defined as

$$\hat{\theta}_{t,i} \quad = \min\left\{\zeta, \max\left\{\delta, \psi_1\left(\frac{\sum_{\ell=t-M(t)+1}^{t-1} X_\ell^i}{M(t)-1}\right)\right\}\right\}, \tag{3.3}$$

Next, we define the local statistics $\hat{W}_t^i$ as a modification of $\tilde{W}_t^i$ by taking into account of possible switch of sampling different data streams. At each time instant t, if we propose to switch to sampling a different stream at time $t+1$—that is, $R_{t+1} \neq R_t$,—then we switch all local statistics back to 0. Mathematically,

$$\hat{W}_t^i = \begin{cases} 0, & \text{if } R_{t+1} /= R_t, \\ \tilde{W}_t^i, & \text{if } R_{t+1} = R_t. \end{cases} \tag{3.4}$$

for all $i = 1, \ldots, p$ and time t. In addition, at each time instant, we will further reset $\tilde{W}_t^i = \hat{W}_t^i$ after updating its value from $\tilde{W}_{t-1}^i$ in (3.1) and before updating the values at time $t+1$.

3.2. Greedy Cyclic Sampling Policy

Here we adopt the GCS policy with a twist of avoiding sampling a local stream for too long. On one hand, if the local statistics $\tilde{W}_t^i$ of the streams being sampled are positive, then we should continue to sample the same stream, and if it becomes zero, then we should switch to sampling new local stream. On the other hand, if we sample the same stream for a very long time but the corresponding local statistics are positive but small values, then it might suggest that there is no strong evidence for this stream involving changes, and we might want to explore new local streams.

To start with, we introduce a controlling parameter q, which was first introduced in Lorden and Pollak (2008) and can be thought of as the maximum consecutive time we can tolerate staying in the same stream. In our results below, we set $q = q(A) = Ce^{\epsilon A}$ for some constant $C \in (0, \infty)$ and $\epsilon \in \left(0, \frac{1}{2}\right)$.

Now we are ready to define our sampling policy. If $\tilde{W}_t^{R_t} > 0$ and $M(t) < q(A)$, then we adopt the GCS policy by continuing sampling the same local stream; for example, $R_{t+1} = R_t$. Otherwise, we will switch to sampling the next stream, because we need to avoid sampling the local stream whose local statistic value is zero or staying for too long. Mathematically, we define the sampling index R_{t+1} as

$$R_{t+1} = \begin{cases} R_t & \text{if } \tilde{W}_t^{R_t} > 0 \text{ and } M(t) < q(A), \\ R_t \bmod p + 1 & \text{if } \tilde{W}_t^{R_t} \leq 0 \text{ or } M(t) \geq q(A), \end{cases} \tag{3.5}$$

with the initial value R_1 randomly picked from $\{1, \ldots, p\}$.

3.3. Decision Policy

Our proposed decision policy $\mathsf{T}_{\mathrm{GCS}}$ is inspired by the prior knowledge that there is only one stream that changes, and we thus propose to raise an alarm at

$$\mathsf{T}_{\mathrm{GCS}}(A) = \inf\left\{t > 0 : \max_{1 \leq i \leq p} \hat{W}_t^i \geq A\right\}, \tag{3.6}$$

for some prespecified constant A. Combining the local statistics, sampling policy, and decision policy, our proposed algorithm defined by T_{GCS} can be summarized in Algorithm 1.

Algorithm 1. Our proposed algorithm T_{GCS}.

1. Randomly pick R_1 from $\{1, \ldots, p\}$ and initialize $\tilde{W}_0^i = 0$ for $i = 1, \ldots, p$.
2. **for** each time t **do**
3. Sample the stream R_t.
4. Update the accumulated time $M(t)$ as in (3.2).
5. Update the local statistics $\tilde{W}_t^i$ for $i = 1, \ldots, p$ as in (3.1).
6. Update the sampling index R_{t+1} as in (3.5) and the local statistics $\hat{W}_t^i$ as in (3.4).
7. **if** $\max_{1 \le i \le p} \hat{W}_t^i \ge A$ **then**
8. Raise an alarm at $\mathsf{T}_{\text{GCS}}(A) = t$.
9. **end if**
10. Reset the local statistics $\tilde{W}_t^i = \hat{W}_t^i$ for $i = 1, \ldots, p$.
11. **end for**

4. ASYMPTOTIC OPTIMALITY

In this subsection, we will investigate the theoretical properties of our proposed algorithm T_{GCS} in (3.6). First, let us make some standard assumptions from the quickest detection literature. We assume that Kullback-Leibler information numbers are positive and finite for all $i = 1, 2, \ldots, p$:

$$(A1) : I(\theta_0, \theta_i) = \int \log \frac{f_{\theta_0}(X)}{f_{\theta_i}(X)} f_{\theta_0}(X) dX > 0,$$
$$I(\theta_i, \theta_0) = \int \log \frac{f_{\theta_i}(X)}{f_{\theta_0}(X)} f_{\theta_i}(X) dX > 0,$$

Moreover, we assume that the second-order moments of log-likelihood ratios are bounded away from ∞.

$$(A2) : J(\theta_0, \theta_i) = \int \left(\log \frac{f_{\theta_0}(X)}{f_{\theta_i}(X)} \right)^2 f_{\theta_0}(X) dX > 0,$$
$$J(\theta_i, \theta_0) = \int \left(\log \frac{f_{\theta_i}(X)}{f_{\theta_0}(X)} \right)^2 f_{\theta_i}(X) dX > 0,$$

Now we are ready to present the theoretical properties of our proposed algorithm. The following theorem summarizes the nonasymptotic properties of our algorithm on the average run length to false alarm and detection delay for any threshold $A > 0$.

Theorem 4.1. *For our proposed algorithm T_{GCS} in (3.6), we have*

$$\mathsf{E}_\infty[\mathsf{T}_{\text{GCS}}] \ge e^A. \tag{4.1}$$

Moreover, its detection delay satisfies

$$D_i(\mathsf{T}_{\text{GCS}}) \le \frac{A}{I(\theta_i, \theta_0)} + C_0 \log A + C_1 \sqrt{A} + C_2 p \tag{4.2}$$

as $A \to \infty$ for any $i \in 1, ..., p$. Here C_0, C_1, C_2 are constants depending only on the distributions, not on A and p.

The rigorous proof of Theorem 4.1 will be postponed to Section 6. As one of our reviewers pointed out, the constant C_2 in relationship (4.2) serves as an important role in the detection delay. It can be regarded as time "wasted" looking at each unaffected stream, and once the dimension grows larger, that wasted time can be significant and deteriorate the detection delay performance of our proposed algorithm. More discussions can be found at the end of the section.

By Theorem 4.1, the following corollary establishes the first-order asymptotic optimality properties of our proposed algorithm T_{GCS} in (3.6) in the quickest detection framework when the average run length to false alarm constraint γ in (2.4) goes to ∞.

Corollary 4.1. *Let $A = \log \gamma$; then our proposed algorithm $\mathsf{T}_{\text{GCS}}(A)$ in (3.6) satisfies both the false alarm constraint in (2.4) and the sampling control constraint in (2.2). Moreover, for each $i = 1, ..., p$, its detection delay satisfies*

$$D_i(\mathsf{T}_{\text{GCS}}) - D_i^{\text{orc}} \quad \le C_0 \log \log \gamma + C_1 \sqrt{\log \gamma} + C_2 p, \tag{4.3}$$

where D_i^{orc} is the oracle detection delay achieved by assuming that the index of the affected stream and the postchange parameters are completely specified:

$$D_i^{\text{orc}} = \frac{\log \gamma}{I(\theta_i, \theta_0)} + C_3 \tag{4.4}$$

and C_3 is a constant that only depends on the distributions, not on γ and p.

It is useful to add some remarks.

(1) Note that relationship (4.3) holds for every p and γ. On one hand, our proposed algorithm T_{GCS} has the same detection delay of the oracle or CUSUM procedure up to $O\left(\sqrt{\log \gamma}\right)$ when p is fixed as $\gamma \to \infty$ or when $p = O\left(\sqrt{\log \gamma}\right)$. On the other hand, when p is large but γ is moderately large, the additional term $C_0 \log \log \gamma + C_1 \sqrt{\log \gamma} + C_2 p$ can be comparable to or even larger than D_i^{orc}, and thus the performance of our proposed algorithm will be much worse than the oracle or CUSUM procedure. This is not surprising for a high-dimensional setting, because the sampling control in (2.2) is too restrictive for large p and we should not be able to detect the change quickly if we only sample one out of p local streams at each time instant. In other contexts, we can evaluate the constants $C_0, C_1,$ and C_2 to see the effects of the dimension p on the performance of our proposed algorithm; see also Wang and Mei (2015) for similar contexts. It remains an open problem to develop a general asymptotic optimality theory for high-dimensional streams under the sampling control.

(2) As one of our reviewers correctly pointed out, the nonattainability of the oracle bound under large p can be explained from another viewpoint. To decide which stream is the most likely to have the change, a minimum average number of samples must be taken from each stream before zeroing in on that stream. When the dimension p is large, additional wasted time on the $p - 1$

unaffected stream is significant compared to the oracle bound and thus our proposed algorithm is not asymptotically optimal.

(3) We use Pollak's criterion for measuring the worst-case delay, and it is well known that one alternative criterion is Lorden's criterion, proposed in Lorden (1971):

$$D(T) = \sup_{t \geq 0} \operatorname{esssup} \ \mathsf{E}_t \left[(T-t)^+ | \mathcal{F}_t \right], \tag{4.5}$$

where $\mathcal{F}_t$ is the filtration of the information up to time t. Though Pollak's criterion and Lorden's criterion are asymptotically equivalent when monitoring i.i.d. data streams with known pre- and postchange distributions, they are very different under our context with sampling policies. In particular, Lorden's criterion involves the" ess sup" over all possible sampling policies, and the main technical difficulty occurs when we are sampling on the unaffected data streams but somehow with large local statistics $\tilde{W}_t^i$, in which case it might take a long time to switch from this unaffected local stream to the other streams. Though we are able to establish the asymptotic optimality theories under Pollak's criterion, it remains an open problem to establish asymptotic optimality theories under Lorden's criterion.

(4) Our algorithm can be implemented much more efficiently from a computational point of view. We employ two sets of statistics: $\tilde{W}_t^i$ in (3.1) and $\hat{W}_t^i$ in (3.4), where the former statistics in (3.1) are updated based on the samples and the latter statistics in (3.4) are to reset those in (3.1) to zero when the stream is switched. Moreover, at most one of these p statistics is nonzero, and thus we can compress our algorithm and reduce the number of registers from p to 1 by focusing only on this nonzero statistics.

(5) The key issue in sampling policy is how to break ties. Under our context of only one affected data stream, any reasonable algorithm would like to sample from the stream with the highest local statistics. This applies to our proposed GCS-CUSUM algorithm, but the main issue is what to sample next if the local statistics $\tilde{W}_t^i$ in (3.1) are negative for $i = R_t$ and all other $p-1$ local statistics are zero. We need to decide how to break ties. Cyclic sampling is one way to break ties, so that we have opportunities to explore all local streams. The other sensible way to break ties is to randomly select the index among all remaining $p-1$ local streams with a zero value. The corresponding procedure has similar theoretical properties as our proposed algorithm, although the proofs become more complicated. Thus, we adopt the cyclic sampling and leave random sampling to break ties as a remark.

5. NUMERICAL STUDIES

In this section, we conduct simulation studies to demonstrate the performance properties of our proposed algorithm $\mathsf{T}_{\mathrm{GCS}}$. Below we consider two types of numerical examples: in Subsection 5.1 we perform Monte Carlo simulations to compare the performance of $\mathsf{T}_{\mathrm{GCS}}$ against some benchmarking algorithms, and in Subsection 5.2 we study its application on a hot-forming process.

5.1. Comparison of T_{GCS} against $\mathsf{T}_{\text{cyclic}}$

In our first simulation, we consider two choices on the number p of local streams: $p = 2$ or $p = 10$. For each choice of p streams, we consider two different distributions f; one is a normal distribution and the other is an exponential distribution. Because of space limitations, we only present the homogeneous setting (i.e., prechange $f_{\theta_0} = f$ and post-change $f_{\theta_i} = g$ for any $i = 1, \ldots, p$).

- Mean shift in normal distribution from 0 to $\mu \geq 0.5$.
- Mean shift in exponential distribution from 1 to $\lambda \geq 2$.

In each case, we set the false alarm constraint $\gamma = 50,000$. For our proposed algorithm $\mathsf{T}_{\text{GCS}}(A)$ and the purely cyclic method $\mathsf{T}_{\text{cyclic}}(A)$ in (2.8), we first use the bisection method to find a suitable threshold A to attain the false alarm constraint and then simulate the worst-case detection delay under different postchange scenarios where the change occurs to the pth stream (because our algorithm starts to sample at the first stream). After obtaining the detection delay for $p = 2, 10,$ we estimate the parameter C_2 in relationship (4.2) by calculating the difference of these two delays and dividing it by the difference of dimension.

Table 1 reports the detection delay of our proposed algorithm T_{GCS} and the cyclic algorithm $\mathsf{T}_{\text{cyclic}}$ in (2.8). In addition, we report the oracle detection delay of the CUSUM procedure in (2.5) and the ratio C_2/oracle delay, which denotes the impact of increasing dimension p. All numerical results are based on 50, 000 Monte Carlo runs. From the tables, it is clear that our proposed algorithm T_{GCS} is much better than the naive method $\mathsf{T}_{\text{cyclic}}$ and can reduce the detection delay by at least 25% when $p = 2$ and 50% when $p = 10$. In other words, compared to naive purely cyclic sampling, our proposed GCS policy can lead to a significant improvement on the detection delay performance.

Moreover, our results show that as the dimension p increases from $p = 2$ to $p = 10$, the detection delays of both our proposed algorithm T_{GCS} and the purely cyclic method $\mathsf{T}_{\text{naive}}$ in (2.8) increase significantly. The detection delay will increase by 20% to 40% of the oracle delay when the dimension p increases by 1. We conjecture that the oracle

Table 1. Comparison of detection delays.

$\gamma = 50000$			$p = 2$		$p = 10$	
μ	Oracle	C_2/Oracle (%)	Cyclic	T_{GCS}	Cyclic	T_{GCS}
0.5	61.87	29.00	144.01	90.56	701.23	234.10
0.75	29.62	25.73	64.13	39.07	308.52	100.06
1.0	17.20	27.76	36.45	22.65	174.67	60.85
1.25	11.35	30.69	23.40	15.46	112.12	43.33
1.5	7.93	37.54	16.60	11.21	80.28	35.03

$\gamma = 50000$			$p = 2$		$p = 10$	
λ	Oracle	C_2/Oracle (%)	Cyclic	T_{GCS}	Cyclic	T_{GCS}
2.0	26.78	28.93	57.50	39.62	286.09	101.60
2.25	19.39	30.77	41.58	28.78	206.86	76.52
2.5	15.18	33.21	32.17	22.49	159.72	62.83
2.75	12.06	38.15	25.59	17.40	126.66	54.21
3.0	9.84	42.28	21.49	14.76	105.10	48.05

Top table: normal distribution. Bottom table: exponential distribution.

bound of the CUSUM procedure is unattainable for high-dimensional monitoring under sampling control, but we are unable to provide a rigorous proof.

5.2. Hot-Forming Process

In this subsection, we will evaluate the performance of GCS-CUSUM algorithm based on a hot-forming process. A Bayesian network (BN) for the hot-forming process was identified by J. Li and Jin (2010) and a physical illustration is shown in Figure 1.

The linear Gaussian parameterization of a BN is assumed to be known:

$$\mu(X_i) = \sum_{k=1}^{\mathrm{card}(\mathrm{PA}(i))} p(\mathrm{PA}_k(i), i)\mu(X_{\mathrm{PA}\ _k(i)}) + V_i, \tag{5.1}$$

where $\mathrm{PA}(i)$ denotes the parents of node i, $p(\mathrm{PA}_k(i), i)$ is called the path coefficient (the number annotated on the arc of the BN), and $V_i \sim N(0, 1)$ represents the random noise that cannot be described by the linear model and is assumed to be independent of $X_{\mathrm{PA}_k(i)}$ and $V_i(j \neq i)$.

For any stopping time T in the hot-forming process, we consider two performance metrics: one is the detection performance $D_i(\mathsf{T})$ when the ith node is affected by the change and the other is the sampling ratio (SR) of the node where the change is detected. Mathematically, if a change is detected on the jth node at time T, then the SR of the stopping time T is defined as

$$\mathrm{SR}(\mathrm{T}) = \frac{\sum_{\ell=1}^{T} 1_{\{R_\ell = j\}}}{T}. \tag{5.2}$$

where R_ℓ is the index of the observed node at time ℓ.

In this example, we consider the situation when there is only a single mean shift occurring at some variable in the hot-forming process. The mean shift will propagate and dilute along the BN (e.g., a mean shift in X_2 with $u_2 = 1$ will result in a mean shift in X_1 with $u_1 = 0.522$). The detailed settings are summarized as follows:

- In the in-control state, $X_i \sim N(\mu(X_i), 1)$ and the mean $\mu(X_i)$ satisfies the BN structure in Figure 1 and relationship (5.1) for $i = 1, 2, 3, 4, 5$.

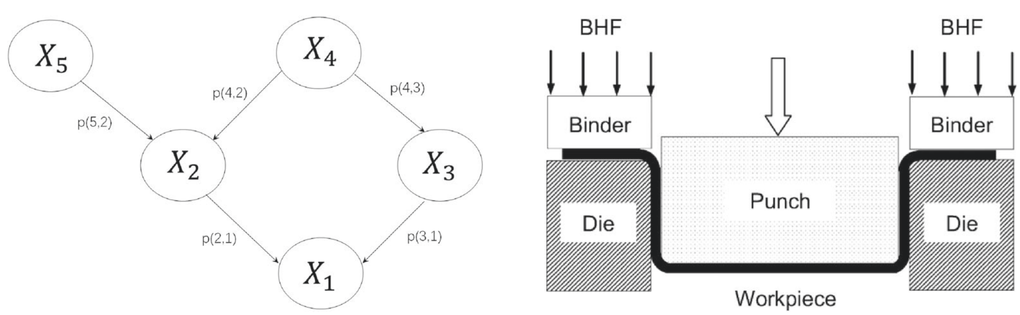

Figure 1. BN structure and physical illustration of a hot-forming process.

- The postchange mean is set to $\mu = 0.5$ and the potential affected nodes are X_1, X_2, X_3, X_4, X_5. For each case, we calculate the detection delay and SR in (5.2) of $\mathsf{T}_{\mathrm{GCS}}$ and T_{RD}.
- The path coefficients are set to $p(2,1) = 0.574, p(3,1) = 0.335, p(4,2) = 0.493, p(4.3) = 0.688, p(5,2) = 0.325$.
- The false alarm control constraint $\gamma = 5000$.

Table 2 summarizes the detection delay and the sampling ratio of our proposed algorithm $\mathsf{T}_{\mathrm{GCS}}$ and the random sampling algorithm T_{RS}. All numerical results are based on 20,000 Monte Carlo runs. From the tables, it is clear that our proposed GCS-CUSUM algorithm is more efficient than the random sampling algorithm and can reduce the detection delay by at least 50%. Moreover, Table 2 shows that the GCS-CUSUM algorithm is focused and always spends more than 50% of the time on one node. Compared to the random sampling algorithm, our proposed algorithm gains more information from one affected node and thus can detect the change more quickly.

6. PROOF OF THE MAIN THEOREM

In this section, we provide the rigorous proof for Theorem 4.1. The proof is divided into two parts: in the first part, we study the false alarm relationship (4.1), and in the second part we study the detection delay relationship (4.2).

6.1. False Alarm Relationship

Let us begin with the proof of false alarm relationship in (4.1), which is the easier one. The key idea in proving (4.1) is to construct a new stopping time $\mathsf{T}_*(A)$ such that $\mathsf{E}_\infty[\mathsf{T}_*(A)] \le \mathsf{E}_\infty[\mathsf{T}_{\mathrm{GCS}}(A)]$ for all thresholds A and prove that $\mathsf{T}_*(A)$ satisfies the false alarm constraint in (2.4). Mathematically, $\mathsf{T}_*(A)$ is defined as

$$\mathsf{T}_*(A) = \inf\left\{ t \ge 1 : S_t^* = \sum_{\ell=1}^{t} \prod_{m=l}^{t} \frac{f_{\hat{\theta}_{m,R_m}}(X_m^{R_m})}{f_0(X_m^{R_m})} \ge e^A \right\}. \tag{6.1}$$

It is easy to verify that $\log S_t^* \ge \max\{\tilde{W}_t^1, \ldots, \tilde{W}_t^p\}$ for all times t and thus $\mathsf{E}_\infty[\mathsf{T}_*(A)] \le \mathsf{E}_\infty[\mathsf{T}_{\mathrm{GCS}}(A)]$ holds for any threshold A. Note that $\{S_t^* - t\}$ is a P_∞- martingale with zero expectation, and applying the optional sampling theorem obtains that

$$\mathsf{E}_\infty[\mathsf{T}_{\mathrm{GCS}}] \ge \mathsf{E}_\infty[\mathsf{T}_*] = \mathsf{E}_\infty\left[S_{\mathsf{T}_*}^*\right] \ge e^A. \tag{6.2}$$

Table 2. Comparison of detection delay and sampling ratio.

$\gamma = 5000$	Detection delay		Sampling ratio	
Affected node	$\mathsf{T}_{\mathrm{cyclic}}$	$\mathsf{T}_{\mathrm{GCS}}$	$\mathsf{T}_{\mathrm{cyclic}}$ (%)	$\mathsf{T}_{\mathrm{GCS}}$ (%)
X_1	315.75	138.60	20.00	56.17
X_2	315.02	138.94	20.00	56.87
X_3	318.09	152.97	20.00	52.41
X_4	314.61	133.97	20.00	57.90
X_5	312.71	159.19	20.00	50.15

6.2. Detection Delay Relationship

Consider now the detection delay relationship in (4.2). Below we write the log likelihood ratio as

$$\log\left(\frac{f_{\theta_j}(X)}{f_{\theta_0}(X)}\right) = (\theta_j - \theta_0)X - \psi(\theta_j) + \psi(\theta_0). \tag{6.3}$$

The main idea in proving (4.2) is to present an equivalent definition of our proposed algorithm $\mathsf{T}_{\mathrm{GCS}}(A)$ by the sequential tests. For each $j(j = 1, \ldots, p)$, a prototype sequential test applied to the jth local stream is defined as

$$\mathsf{T}^j = \inf\left\{ t \geq 1 : t = q(A) \text{ or } S_t^j = \sum_{\ell=1}^{t}((\hat{\theta}_\ell - \theta_0)X_\ell^j - \psi(\hat{\theta}_\ell) + \psi(\theta_0)) \notin (0, A)\right\}. \tag{6.4}$$

with the estimator $\hat{\theta}_\ell$

$$\hat{\theta}_\ell = \min\left\{\zeta, \max\left\{\delta, \psi_1\left(\frac{X_1^j + \cdots + X_{\ell-1}^j}{\ell - 1}\right)\right\}\right\}. \tag{6.5}$$

For notational simplicity, here we use the same notation $\hat{\theta}_\ell$ for different sequential tests T^j and omit the subscript j.

Define the sequence $\{\mathsf{T}_m\}, m = 1, 2, \ldots,$ of sequential tests applied to each local stream. Then it is clear that each T_m has the same distribution as a particular prototype sequential probability ratio test T^j. In addition, if we define (a stopping time) K to be the first time the sequential test T_m of the local stream being tested crosses the upper boundary A, then our proposed stopping time $\mathsf{T}_{\mathrm{GCS}}$ can be written as the sum

$$\mathsf{T}_{\mathrm{GCS}} = \mathsf{T}_1 + \mathsf{T}_2 + \cdots + \mathsf{T}_K = \sum_{m=1}^{K} \mathsf{T}_m. \tag{6.6}$$

Without loss of generality, we assume that the change occurs to the pth local stream at change time $\tau = k$, and we are monitoring the ith (i.e., $R_k = i$) local stream when the change occurs. To better characterize the sequential test on the ith stream when the change occurs at time $\tau = k$, we define

- $M(k)$: the total time instants in which we have consecutively sampled at the stream when the change occurs (the ith stream) up to time k.
- $S(k) = \sum_{\ell=k-M(k)+1}^{k} X_\ell^i$ is the sum of observations on the ith stream within the window $[k - M(k) + 1, k]$.
- $G(k) = \sum_{\ell=k-M(k)+1}^{k}((\hat{\theta}_\ell - \theta_0)X_\ell^i - \psi(\hat{\theta}_\ell) + \psi(\theta_0))$ is the local statistic on the ith stream at time k.
- $\mathsf{T}_k^i(A)$: the sequential test on the ith stream after the change occurs.

$$\mathsf{T}_k^i(A) = \inf\{t - k : t \geq k + 1, t - k + M(k) = q(A) \text{ or } S_{t,k}^i \notin (-G(k), A - G(k))\}, \tag{6.7}$$

where $S^i_{t,k} = \sum_{\ell=k+1}^{t}((\tilde{\theta}_\ell - \theta_0)X^i_\ell - \psi(\tilde{\theta}_\ell) + \psi(\theta_0))$ and $\tilde{\theta}_\ell$ is defined as

$$\tilde{\theta}_\ell = \min\left\{\zeta, \max\left\{\delta, \psi_1\left(\frac{S(k) + X^i_{k+1} + \cdots + X^i_{\ell-1}}{M(k) + \ell - 1 - k}\right)\right\}\right\}. \tag{6.8}$$

Here we use the notation $\tilde{\theta}_\ell$ to distinguish it from the estimator $\hat{\theta}_\ell$, which is not affected by $M(k), S(k)$.

With the definition of $\mathsf{T}^i_k(A)$, we are able to write the detection delay of our proposed algorithm as the sum of sequential tests. We divide the discussions into two parts:

- We first consider the simplest scenario when the change occurs to the pth stream at time $\tau = 0$ and R_1 is randomly set to 1; that is, we start sampling from the stream. This clearly generates a worse detection delay, because the change occurs to the pth stream and we need to go over $p-1$ unaffected stream.

$$\begin{aligned}
E^p_0\left[T_{\mathrm{GCS}}(A)|R_1 = 1\right] &= E^p_0\left[\sum\nolimits_{m=1}^{K} T_m\right] = \sum\nolimits_{m=1}^{\infty} E^p_0[T_m]P^p_0(K \geq m) \\
&= \Omega_p + \omega_p\Omega_p + \omega_p^2\Omega_p + \cdots = \frac{\Omega_p}{1-\omega_p}, E^p_0\left[T_{\mathrm{GCS}}(A)|R_1 = 1\right] \\
&= E^p_0\left[\sum\nolimits_{m=1}^{K} T_m\right] = \sum\nolimits_{m=1}^{\infty} E^p_0[T_m]P^p_0(K \geq m) \\
&= \Omega_p + \omega_p\Omega_p + \omega_p^2\Omega_p + \cdots = \frac{\Omega_p}{1-\omega_p},
\end{aligned}$$

where

$$\begin{aligned}
\omega_p &= \beta_p \prod\nolimits_{j=1}^{p-1}(1-\alpha_i) \\
\Omega_p &= E^p_0[T^1] + E^p_0[T^2]\Big(1-\alpha_1\Big) + \cdots + \cdots + E^p_0[T^p]\prod_{j=1}^{p-1}(1-\alpha_i) \\
&= E_\infty[T^1] + E_\infty[T^2]\Big(1-\alpha_1\Big) + \cdots + \cdots + E^p_0[T^p]\prod_{j=1}^{p-1}(1-\alpha_i)
\end{aligned}$$

where $\beta_p = \mathsf{P}^p_0(S^p_{\mathsf{T}^p} < A)$ denotes the type II error probability of the sequential test T^p and $\alpha_i = \mathsf{P}_\infty(S^i_{\mathsf{T}^i} \geq A)$ denotes the type I error probability of the sequential test T^i $(i = 1, \ldots, p-1)$. Because $0 \leq \alpha_i, \beta_i \leq 1$, we have the following upper bound:

$$\mathsf{E}^p_0[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1] \leq \frac{\mathsf{E}^p_0[\mathsf{T}^p]}{1-\beta_p} + \frac{1}{1-\beta_p}\sum_{j=1}^{p-1}\mathsf{E}_\infty[\mathsf{T}^i]. \tag{6.9}$$

- Then we consider the general scenario when the change occurs to the pth stream at time $\tau = k$ and we are monitoring the ith stream ($R_k = i$) at the same time. Note that if we do not detect a change on the ith stream, we will switch to the next stream, and this actually returns to the simplest scenario when the change occurs at time $\tau = 0$. This can be further divided into two cases:

- If we are not monitoring the affected stream when the change occurs ($i \neq p$), the detection delay can be written as

$$\begin{aligned} D_p(\mathsf{T}_{\mathrm{GCS}}(A)|\ i \neq p) \quad &\leq\ \mathsf{E}_k^p\left[\mathsf{T}_k^i(A)|\ i \neq p, \mathsf{T}_{\mathrm{GCS}} > k\right] \\ &+\ \mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1] \end{aligned}.$$

- If we are just monitoring the affected stream when the change occurs ($i = p$), the detection delay can be written as

$$\begin{aligned} D_p(\mathsf{T}_{\mathrm{GCS}}(A)|\ i = p) \quad &\leq\ \mathsf{E}_k^p\left[\mathsf{T}_k^p(A)|\ i = p, \mathsf{T}_{\mathrm{GCS}} > k\right] \\ &+\ \beta_{k,p}\mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1] \end{aligned},$$

where $\beta_{k,p} = \mathsf{P}_k^p(S^p_{\mathsf{T}_k^p+k,k} \geq A - G(k))$ is the type II error probability of the sequential test $\mathsf{T}_k^p(A)$.

If we are able to show that both $D_p(\mathsf{T}_{\mathrm{GCS}}(A)|\ i \neq p), D_p(\mathsf{T}_{\mathrm{GCS}}(A)|\ i = p)$ satisfy the relationship (4.2), the problem is solved. It suffices to bound the detection delay $\mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1]$ and the sequential tests $\mathsf{T}_k^i(A)$ for $i \neq p, i = p$. For better presentation, below we divide the proof into two subsubsections. In the first subsubsection, we bound the detection delay $\mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1]$, and in the second one we focus on the sequential tests $\mathsf{T}_k^i(A)$.

6.2.1. Detection Delay When Change Occurs to the pth Stream at Time $\tau = 0$

We first consider the detection delay $\mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1]$, which is the easier one. Below we divide the proofs into a series of lemmas. The notation of constants is as follows: C_{ij} refers to the constant j in Lemma i, and C_{ijk} refers to the constant k in Lemma ij. (Thus, C_{121} refers to a constant in Lemma 6.12.) Constants of the form $C_{(i)}$ refer to a constant that does not appear in the lemma. When constants add up, we freely use another constant to bound their sum.

Lemma 6.1. *For any $j = 1, \ldots, p$, there exist constants C_{11}, C_{12}, C_{13} such that for some $\eta > 0$ and all times t,*

$$\mathsf{P}_0^j\left(\left|\frac{\sum_{\ell=1}^t X_\ell^j}{t} - \psi'(\theta_j)\right| > \eta\right) < C_{11}e^{-C_{12}\eta^2 t}, \tag{6.10}$$

$$\mathsf{E}_0^j\left[\left(\frac{\sum_{\ell=1}^t X_\ell^j}{t} - \psi'(\theta_j)\right)^4\right] < \frac{C_{13}}{t^2}. \tag{6.11}$$

The same relationship holds when $\psi'(\theta_j)$ is replaced by $\psi'(\theta_0)$ and $\mathsf{P}_0^j, \mathsf{E}_0^j$ is replaced by $\mathsf{P}_\infty, \mathsf{E}_\infty$.

Proof. Standard large deviations arguments.

Lemma 6.2. *For the sequential test T^p in (6.4), let $\beta_p = \mathsf{P}_0^p(S^p_{\mathsf{T}^p} < A)$, and we have*

$$\mathsf{E}_0^p[\mathsf{T}^p] \leq \frac{(1-\beta_p)A}{I(\theta_p,\theta_0)} + x_1 + x_2 + x_3, \tag{6.12}$$

where

$$x_1 \quad \leq C_{21}\left(\mathsf{E}_0^p\left[\sum_{\ell=1}^{\mathsf{T}^p}(\hat{\theta}_\ell - \theta_p)^2\right]\right)^{1/2} + C_{22}(\mathsf{E}_0^p[\mathsf{T}^p])^{1/2} \tag{6.13}$$

$$x_2 \leq C_{23}\mathsf{E}_0^p\left[\sum_{\ell=1}^{\mathsf{T}^p}(\hat{\theta}_\ell - \theta_p)^2\right] + C_{24}\mathsf{E}_0^p\left[\sum_{\ell=1}^{\mathsf{T}^p}(\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^2\right] \tag{6.14}$$

$$x_3 \quad \leq C_{25}A\mathsf{P}_0^p(\mathsf{T}^p = q(A)) \tag{6.15}$$

for some constants $C_{21}, C_{22}, C_{23}, C_{24}$ that only depend on the parameter θ_p but not on the threshold A.

Proof. Applying Wald's equation and optional stopping theorem obtains that

$$\begin{aligned}
E_0^p[T^p] &= \frac{E_0^p[\sum_{\ell=1}^{T^p}(\theta_p - \theta_0)X_\ell^p - \psi(\theta_p) + \psi(\theta_0))]}{I(\theta_p,\theta_0)} \\
&= \frac{E_0^p[\sum_{\ell=1}^{T^p}((\hat{\theta}_\ell - \theta_0)X_\ell^p - \psi(\hat{\theta}_\ell) + \psi(\theta_0))]}{I(\theta_p,\theta_0)} \\
&\quad + \frac{E_0^p[\sum_{\ell=1}^{T^p}((\theta_p - \hat{\theta}_\ell)X_\ell^p - \psi(\theta_p) + \psi(\hat{\theta}_\ell))]}{I(\theta_p,\theta_0)} \\
&= \frac{E_0^p[S_{T^p}^p]}{I(\theta_p,\theta_0)} + \frac{E_0^p[\sum_{\ell=1}^{T^p}((\theta_p - \hat{\theta}_\ell)\psi'(\theta_p) - \psi(\theta_p) + \psi(\hat{\theta}_\ell))]}{I(\theta_p,\theta_0)} \\
&\leq \frac{P_0^p(S_{T^p}^p \geq A)A}{I(\theta_p,\theta_0)} + \frac{P_0^p(S_{T^p}^p \geq A)E_0^p[S_{T^p}^p - A | S_{T^p}^p \geq A]}{I(\theta_p,\theta_0)} \\
&\quad + \frac{E_0^p[\sum_{\ell=1}^{T^p}((\theta_p - \hat{\theta}_\ell)\psi'(\theta_p) - \psi(\theta_p) + \psi(\hat{\theta}_\ell))]}{I(\theta_p,\theta_0)} + \frac{P_0^p(T^p = q(A))E_0^p[S_{T^p}^p | T^p = q(A)]}{I(\theta_p,\theta_0)} \\
&= \frac{P_0^p(S_{T^p}^p \geq A)A}{I(\theta_p,\theta_0)} + x_1 + x_2 + x_3.
\end{aligned}$$

where $S^p_{\mathsf{T}^p}$ is the local statistics defined in (6.4) and x_1, x_2, x_3, are the three corresponding components in the second to last inequality.

$$
\begin{aligned}
x_1 &\le \frac{E_0^p[|(\hat{\theta}_{T^p}-\theta_0)X_{T^p}^p-\psi(\hat{\theta}_{T^p})+\psi(\theta_0)|]}{I(\theta_p,\theta_0)}\\
&\le \frac{E_0^p[|(\hat{\theta}_{T^p}-\theta_p)X_{T^p}^p|]}{I(\theta_p,\theta_0)}+\frac{E_0^p[|\theta_0-\theta_p|X_{T^p}^p]}{I(\theta_p,\theta_0)}+C_{26}\\
&\le \frac{\left(E_0^p\left[\sum_{\ell=1}^{T^p}(\hat{\theta}_\ell-\theta_p)^2(X_\ell^p)^2\right]\right)^{1/2}}{I(\theta_p,\theta_0)}+\frac{\left(E_0^p\left[\sum_{\ell=1}^{T^p}(\theta_0-\theta_p)^2(X_\ell^p)^2\right]\right)^{1/2}}{I(\theta_p,\theta_0)}+C_{26}\\
&\le C_{21}\left(E_0^p\left[\sum_{\ell=1}^{T^p}(\hat{\theta}_\ell-\theta_p)^2\right]\right)^{1/2}+C_{22}(E_0^p[T^p])^{1/2}.
\end{aligned}
$$

Note that $\hat{\theta}_\ell$ is universally bounded by δ,ζ defined in (3.3) and thus $\mathsf{E}_0^p[|\psi(\theta_p)-\psi(\hat{\theta}_\ell)|]$ is bounded by a constant C_{26}. We omit this constant in the following proof.

For the x_2 term, as $\psi''(\theta)>0$ for all θ, we have

$$
\begin{aligned}
x_2 &= \frac{E_0^p[\sum_{\ell=1}^{T^p}((\theta_p-\hat{\theta}_\ell)\psi'(\theta_p)-\psi(\theta_p)+\psi(\hat{\theta}_\ell))]}{I(\theta_p,\theta_0)}\\
&\le \frac{E_0^p[\sum_{\ell=1}^{T^p}((\theta_p-\hat{\theta}_\ell)(\psi'(\theta_p)-\psi'(\theta_*))]}{I(\theta_p,\theta_0)}]\\
&\le \frac{E_0^p[\sum_{\ell=1}^{T^p}((\theta_p-\hat{\theta}_\ell)(\psi'(\theta_p)-\psi'(\theta_\ell))]}{I(\theta_p,\theta_0)}\\
&\le C_{23}E_0^p\left[\sum_{\ell=1}^{T^p}(\hat{\theta}_\ell-\theta_p)^2\right]+C_{24}E_0^p\left[\sum_{\ell=1}^{T^p}(\psi'(\hat{\theta}_\ell)-\psi'(\theta_p))^2\right]
\end{aligned}
$$

for some constants $C_{23}=C_{24}=\frac{1}{2}I(\theta_p,\theta_0)$. For the x_3 term, note that $\mathsf{E}_0^p[S_{\mathsf{T}^p}^p|\mathsf{T}^p=q(A)]\le A$ and thus $x_3\le C_{25}A\mathsf{P}_0^p(\mathsf{T}^p=q(A))$ for some constants $C_{25}=\frac{1}{I}(\theta_p,\theta_0)$.

Lemma 6.3. *There exist constants C_{31},C_{32} such that*

$$
\mathsf{P}_\infty\left[\mathsf{T}^j\ge\ell\right]\le C_{31}e^{-C32\ell} \tag{6.16}
$$

for all $j=1,\dots,p$ and $\ell=1,2,\dots$.

Proof. Let $0<u<1$, and we have

$$
\mathsf{E}_\infty\left[e^{u((\hat{\theta}_\ell-\theta_0)X_\ell^j-\psi(\hat{\theta}_\ell)+\psi(\theta_0))}|X_1^j,\dots,X_{\ell-1}^j\right]=e^{\psi(u\hat{\theta}_\ell+(1-u)\theta_0)-u\psi(\hat{\theta}_\ell)-(1-u)\psi(\theta_0)}. \tag{6.17}
$$

Note that

$$
\frac{\partial}{\partial\theta}(\psi(u\theta+(1-u)\theta_0)-u\psi(\theta)-(1-u)\psi(\theta_0))=u(\psi'(u\theta+(1-u)\theta_0)-\psi'(\theta))<-a. \tag{6.18}
$$

where $a=\min_{\theta\in[\delta,\zeta]}(-u(\psi'(u\theta+(1-u)\theta_0)-\psi'(\theta)))>0$. This is because $0<u<1$ and $\psi'(\theta)$ is strictly increasing. Thus,

$$
e^{\psi(u\hat{\theta}_\ell+(1-u)\theta_0)-u\psi(\hat{\theta}_\ell)-(1-u)\psi(\theta_0)}\le e^{-bu}. \tag{6.19}
$$

for some constant $b=a(\delta-\theta_0)$.

It follows that $e^{uS_\ell^j+bu\ell}$ is a supermartingale under P_∞ and we have

$$\begin{aligned}\mathsf{P}_\infty\left[\mathsf{T}^j \geq \ell\right] &\leq \mathsf{P}_\infty\left[uS_\ell^j > 0\right] \\ &= \mathsf{P}_\infty\left[e^{uS_\ell^j} > 1\right] \\ &\leq \mathsf{E}_\infty\left[e^{uS_\ell^j}\right] \\ &\leq e^{-bu\ell}\mathsf{E}_\infty\left[e^{uS_\ell^j+bu\ell}\right] \\ &\leq e^{-bu\ell}\mathsf{E}_\infty\left[e^{uS_1^j+bu}\right].\end{aligned}$$

Relationship (6.16) is then $C_{31} = \mathsf{E}_\infty[e^{uS_1^j+bu}]$ and $C_{32} = bu$. □

Lemma 6.4. *There exists a constant C_{41} such that*

$$\mathsf{E}_\infty[\mathsf{T}^j] < C_{41}, \tag{6.20}$$

for all $j = 1, ..., p$.

Proof. Applying Lemma 6.3 obtains that

$$\mathsf{E}_\infty[\mathsf{T}^j] = \sum_{\ell=1}^{\infty}\mathsf{P}_\infty(\mathsf{T}^j \geq \ell) \leq C_{41}. \tag{6.21}$$

for some constant C_{41}. □

Lemma 6.5. *For any sufficiently large threshold $A \geq A_0$, $\beta_p = \mathsf{P}_0^p(S_{\mathsf{T}^p}^p < A)$, which is the type II error probability of the sequential test $\mathsf{T}^p(A)$, satisfies the lower bound*

$$\beta_p \leq C_{51}, \tag{6.22}$$

where $0 < C_{51} < 1$ is a constant that does not depend on the threshold A.

Proof. We consider the following event:

$$D = \left\{\left|\frac{\sum_{\ell=1}^t X_\ell^p}{t} - \psi'(\theta_p)\right| \leq \eta \quad \text{for all } t \geq 1 \text{ and some } \eta > 0.\right\}$$

From relationship (6.10) in Lemma 6.1, we can verify that $\mathsf{P}_0^p(D) > C_{52}$ for some constant $C_{52} > 0$. Under the event D, the estimator $\hat{\theta}_\ell$ satisfies $|\hat{\theta}_\ell - \theta_p| < \lambda$ for all times ℓ and some constant $\lambda = \lambda(\eta) > 0$. It follows that the log-likelihood ratio satisfies

$$(\hat{\theta}_\ell - \theta_0)X_\ell^p - \psi(\hat{\theta}_\ell) + \psi(\theta_0) \geq Y_\ell = (\theta_p - \theta_0)X_\ell^p - \lambda|X_\ell^p| - C_{53}, \tag{6.23}$$

where $C_{53} = \max_{|\theta-\theta_p|<\lambda}(\psi(\theta) - \psi(\theta_0))$. We choose proper η, λ such that $\mathsf{E}_0^p[Y_\ell] > 0$ and define the stopping time and the event as

$$N(A) = \inf\left\{t > 0 : \sum_{\ell=1}^t Y_\ell \leq 0\right\}.$$

$$E = \{N(A) = \infty\}$$

If event E occurs, the relationship (6.23) implies that our summary statistics also never return to 0. We can further prove that $\mathsf{P}_0^p(E) > C_{54}$ for some constant $C_{54} > 0$ in the renewal theory. Now consider the third event with sufficiently large A such that $A/q(A) < \mathsf{E}_0^p[Y_\ell]/2$:

$$F = \left\{ \sum_{\ell=1}^{q(A)} Y_l \leq A \right\}.$$

We have

$$\begin{aligned}
\mathsf{P}_0^p(F) &= \mathsf{P}_0^p\left(\frac{\sum_{\ell=1}^{q(A)} Y_l}{q(A)} \leq \frac{A}{q(A)} \right) \\
&\leq \mathsf{P}_0^p\left(\left| \frac{\sum_{\ell=1}^{q(A)} Y_l}{q(A)} - \mathsf{E}_0^p[Y_\ell] \right| \geq \frac{\mathsf{E}_0^p[Y_\ell]}{2} \right) \\
&\leq C_{55}
\end{aligned}$$

for some constant $0 < C_{55} < 1$ which can be derived from Chebyshev's inequality. Under the event D, E and F^C, we have $S_{\mathsf{T}^p}^p \geq A$ and thus

$$\beta_p \leq 1 - \mathsf{P}_0^p(D)\mathsf{P}_0^p(E)\mathsf{P}_0^p(F^C) \leq C_{51}. \tag{6.24}$$

for some constant $0 < C_{51} < 1$.

Lemma 6.6. *There exist constants C_{61}, C_{62} such that for sufficiently large $A \geq A_0$ and every $r = 1, 2, \ldots,$*

$$\mathsf{P}_0^p\left(\mathsf{T}^p > \frac{10A}{I(\theta_p, \theta_0)} + r \right) \leq C_{61} e^{-C_{62} r}. \tag{6.25}$$

Proof. Let $s = [\frac{5A}{I(\theta_p, \theta_0)}]$, we define the following event for some constant η,

$$E = \left\{ |\hat{\theta}_\ell - \theta_p| < \lambda \quad \text{for all } \ell \geq s + \frac{r}{2} \right\}.$$

From relationship (6.10) we have $\mathsf{P}_0^p(E^C) \leq C_{63} e^{-C_{64} r}$ for some constant C_{63}, C_{64}. Similar to Lemma 6.5, we consider a new stopping time

$$N_*(A) = \inf\left\{ t > 0 : \sum_{\ell=1}^{t} Y_\ell \geq A \right\}, \tag{6.26}$$

where Y_ℓ is defined in (6.23). It is easy to verify that if $\mathsf{T}^p(A) > 10A/I(\theta_p, \theta_0) + r,$ then $N_*(A) > 10A/I(\theta_p, \theta_0) + r$. Moreover, it is straightforward to see that $\mathsf{P}_0^p(N_*(A) > 10A/I(\theta_p, \theta_0) + r)$ is bounded exponentially in r and we can prove that there exist constants C_{61}, C_{62} such that

$$
\begin{aligned}
\mathsf{P}_0^p\left(\mathsf{T}^p > \frac{10A}{I(\theta_p,\theta_0)} + r\right) &= \mathsf{P}_0^p(\mathsf{T}^p > \frac{10A}{I(\theta_p,\theta_0)} + r|E)\mathsf{P}_0^p(E) \\
&\quad + \mathsf{P}_0^p(\mathsf{T}^p > \frac{10A}{I(\theta_p,\theta_0)} + r|E^C)\mathsf{P}_0^p(E^C) \\
&\le \mathsf{P}_0^p(\mathsf{T}^p > \frac{10A}{I(\theta_p,\theta_0)} + r|E) + \mathsf{P}_0^p(E^C) \\
&\le \mathsf{P}_0^p(N_* > \frac{10A}{I(\theta_p,\theta_0)} + r) + \mathsf{P}_0^p(E^C) \\
&\le C_{61}e^{-C_{62}r}.
\end{aligned}
\tag{6.27}
$$

Lemma 6.7. *There exist constants C_{71}, C_{72} such that*

$$
\mathsf{E}_0^p\left[\sum_{\ell=1}^{\mathsf{T}^p}(\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^2\right] \le C_{71}\log A + C_{72}. \tag{6.28}
$$

Proof. From relationship (6.11), we obtain that there exists constant C_{73} such that

$$
\begin{aligned}
\mathsf{E}_0^p\left[(\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^4\right]^{1/2} &= \mathsf{E}_0^p\left[\left(\tfrac{X_1^p+\cdots+X_{\ell-1}^p}{\ell-1} - \psi'(\theta_p)\right)^4\right]^{1/2} \\
&= \frac{C_{73}}{\ell-1}.
\end{aligned}
$$

Moreover, we have

$$
\begin{aligned}
E_0^p\left[\sum_{\ell=1}^{T^p}(\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^2\right] &= E_0^p\left[\sum_{\ell=1}^{\infty}(\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^2 \mathbb{1}_{\{T^p\ge\ell\}}\right] \\
&\le \sum_{\ell=1}^{\infty}(E_0^p[(\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^4])^{1/2}(P_0^p(T^p\ge\ell))^{1/2} \\
&= \sum_{\ell=1}^{10A/I(\theta_p,\theta_0)} (E_0^p[((\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^4])^{1/2}(P_0^p(T^p\ge\ell))^{1/2} \\
&\quad + \sum_{\ell=10A/I(\theta_p,\theta_0)+1}^{\infty} (E_0^p[(\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^4])^{1/2}(P_0^p(T^p\ge\ell))^{1/2} \\
&\le \sum_{\ell=2}^{10A/I(\theta_p,\theta_0)} \frac{C_{73}}{\ell} + \sum\nolimits_{\ell=10A/I(\theta_p,\theta_0)+1}^{\infty} \frac{C_{73}}{\ell-1}\left(C_{61}e^{-C_{62}(\ell-10A/I(\theta_p,\theta_0))}\right)^{1/2} \\
&\le C_{71}\log A + C_{72}.
\end{aligned}
$$

for some constants $C_{71} = C_{73}$ and $C_{72} = C_{73}\log(\frac{10}{I}(\theta_p,\theta_0)) + 2C_{73}\frac{\sqrt{C_{61}}}{C_{62}}$.

Lemma 6.8. *There exist constants C_{81}, C_{82} such that*

$$
\mathsf{E}_0^p\left[\sum_{\ell=1}^{\mathsf{T}^p}(\hat{\theta}_\ell - \theta_p)^2\right] \le C_{81}\log A + C_{82}. \tag{6.29}
$$

Proof. Because $\hat{\theta}_\ell$ is universally bounded, we define $\rho = \max_{\delta \le \theta \le \zeta} \psi''(\theta)$ and obtain that

$$\begin{aligned} \mathsf{E}_0^p\left[\sum_{\ell=1}^{\mathsf{T}^p}(\hat{\theta}_\ell - \theta_p)^2\right] &\le \frac{\mathsf{E}_0^p\left[\sum_{\ell=1}^{\mathsf{T}^p}(\psi'(\hat{\theta}_\ell) - \psi'(\theta_p))^2\right]}{\rho^2} \\ &\le C_{81}\log A + C_{82}. \end{aligned}$$

Lemma 6.9. *There exists a constant C_{91}, such that*

$$x_3 \le C_{25} A \mathsf{P}_0^p(\mathsf{T}^p = q(A)) < C_{91}, \tag{6.30}$$

for all thresholds A.

Proof. This results from relationship Lemma 6.3 directly.

Combining the results above, we have

$$\begin{aligned} E_0^p[T_{\mathrm{GCS}}(A)|R_1 = 1] &\le \frac{E_0^p[T^p]}{1-\beta_p} + \frac{1}{1-\beta_p}\sum\nolimits_{j=1}^{p-1} E_\infty[T^i]. \\ &\le \frac{\frac{(1-\beta_p)A}{I(\theta_p,\theta_0)} + x_1 + x_2 + x_3}{1-\beta_p} + \frac{C_{41}}{1-C_{51}}(p-1) \\ &\le \frac{A}{I(\theta_p,\theta_0)} + \frac{x_1+x_2+x_3}{1-C_{51}} + \frac{C_{41}}{1-C_{51}}(p-1) \\ &\le \frac{A}{I(\theta_p,\theta_0)} + C_{(1)}\log A + C_{(2)}\sqrt{E_0^p[T_{\mathrm{GCS}}(A)|R_1=1]} + C_{(3)}p, \end{aligned}$$

for some suitable constants $C_{(1)} = C_{81}(C_{21} + C_{23}) + C_{71}C_{24}$, $C_{(2)} = C_{22}$, $C_{(3)} = \frac{C_{41}}{1-C_{51}}$.

6.2.2. Analysis of the Sequential Tests $\mathsf{T}_k^i(A)$

We then consider the sequential tests $\mathsf{T}_k^i(A)$ on the stream being monitored (the ith stream) when the change occurs to the pth stream at time $\tau = k$. We need to bound two kinds of sequential tests:

- $\mathsf{E}_k^p[\mathsf{T}_k^i(A)|\ i \ne p, \mathsf{T}_{\mathrm{GCS}} > k]$, where we are monitoring another stream $i(i \ne p)$ when the change occurs to the pth stream. Observations under this scenario are under a prechange distribution.
- $\mathsf{E}_k^p[\mathsf{T}_k^p(A)|\ i = p, \mathsf{T}_{\mathrm{GCS}} > k]$, where we are just monitoring the affected stream p when the change occurs to the pth stream. Observations under this scenario are under a postchange distribution.

We note that $\mathsf{T}_k^i(A), \mathsf{T}_k^p(A)$ are completely characterized by the statistics $M(k), S(k), G(k)$ defined at the beginning of Subsection 6.2. The following lemmas give the bounds on these statistics.

Lemma 6.10. *For any set A in the domain of $M(k)$, we have*

$$\mathsf{P}_k^p(M(k) \in A | \mathsf{T}_{\mathrm{GCS}} > k, i \neq p) = \mathsf{P}_k^p(M(k) \in A | \mathsf{T}_{\mathrm{GCS}} > k, i = p). \tag{6.31}$$

Similar relationship holds for $S(k), G(k)$.. Thus, in the following contexts on these three statistics, we will omit the conditions $i \neq p$ and $i = p$.

Proof. Due to the random starting value of R_1, $\mathsf{P}_k^p(M(k) \in A | \mathsf{T}_{\mathrm{GCS}} > k, i = j)$ should be identical for any $j = 1, ..., p$. Relationship (6.31) is then proved.

Lemma 6.11. *There exists a constant C_{111} such that*

$$\mathsf{P}_k^p(\mathsf{T}_{\mathrm{GCS}} > k | \mathsf{T}_{\mathrm{GCS}} > k - M(k)) \geq C_{111} \tag{6.32}$$

for some constant C_{111} that is not related to k, A.

Proof.

$$\begin{aligned}
& \mathsf{P}_k^p(\mathsf{T}_{\mathrm{GCS}} \leq k | \mathsf{T}_{\mathrm{GCS}} > k - M(k)) \\
= \; & \mathsf{P}_\infty(\mathsf{T}_{\mathrm{GCS}} < M(k)) \\
\leq \; & \mathsf{P}_\infty(\mathsf{T}_* < M(k)) \\
\leq \; & \sum_{\ell=1}^{M(k)-1} \mathsf{P}_\infty(S_\ell^* \geq e^A) \\
\leq \; & \sum_{\ell=1}^{M(k)-1} \frac{\ell}{e^A} \\
\leq \; & \frac{q^2(A)}{A} \leq C_{112}
\end{aligned} \tag{6.33}$$

for some constant $C_{112} < 1$. Here T_* and S_t^* are the sequential test and statistics defined in (6.1). Relationship (6.32) is then proved with $C_{111} = 1 - C_{112}$.

Lemma 6.12. *There exist constants C_{121} and C_{122} such that*

$$\mathsf{P}_k^p(M(k) \geq \ell) \leq C_{121} e^{-C_{122}\ell}. \tag{6.34}$$

Proof. Applying Lemma 6.3 obtains that

$$\begin{aligned}
\mathsf{P}_k^p(M(k) \geq \ell) \; &= \mathsf{P}_\infty(M(k) \geq \ell) \\
&= \sum_{m=1}^{k-\ell} \mathsf{P}_\infty(k - M(k) = m, \tilde{W}_t^p > 0 \quad \text{for all} \;\; m+1 < t < k) \\
&\leq \sum_{m=1}^{k-\ell} \mathsf{P}_\infty(\mathsf{T}^i > k - m) \\
&\leq \sum_{m=1}^{k-\ell} C_{31} e^{-C32(k-m)} \\
&\leq C_{121} e^{-C_{122}\ell}
\end{aligned}$$

for some constants C_{121}, C_{122}.

Lemma 6.13. *There exist constants C_{131} and C_{132} such that*

$$\mathsf{P}_k^p(M(k) \geq \ell | \mathsf{T}_{\mathrm{GCS}} > k) \leq C_{131} e^{-C_{132}\ell}. \tag{6.35}$$

Proof.

$$\begin{aligned}
\mathsf{P}_k^p(M(k) \geq \ell | \mathsf{T}_{\mathrm{GCS}} > k) &= \mathsf{P}_\infty(M(k) \geq \ell | \mathsf{T}_{\mathrm{GCS}} > k) \\
&= \mathsf{P}_\infty(M(k) \geq \ell | \mathsf{T}_{\mathrm{GCS}} > k, \mathsf{T}_{\mathrm{GCS}} > k - M(k)) \\
&= \frac{\mathsf{P}_\infty(M(k) \geq \ell | \mathsf{T}_{\mathrm{GCS}} > k - M(k))}{\mathsf{P}_\infty(\mathsf{T}_{\mathrm{GCS}} > k | \mathsf{T}_{\mathrm{GCS}} > k - M(k))} \\
&\leq \frac{C_{121}}{C_{111}} e^{-C_{122}\ell}.
\end{aligned}$$

Thus, relationship (6.35) is proved with $C_{131} = C_{121}/C_{111}$ and $C_{132} = C_{122}$.

Lemma 6.14. *There exists constant C_{141} such that for $r = 1, 2, \ldots$,*

$$\mathsf{E}_k^p\left[S(k)^4 | \mathsf{T}_{\mathrm{GCS}} > k\right] \leq C_{141}, \tag{6.36}$$

$$\mathsf{E}_k^p\left[G(k)^4 | \mathsf{T}_{\mathrm{GCS}} > k\right] \leq C_{142}, \tag{6.37}$$

$$\mathsf{E}_k^p\left[\left(\frac{S(k)}{M(k)+r}\right)^4 | \mathsf{T}_{\mathrm{GCS}} > k\right] \leq \frac{C_{141}}{r^4}, \tag{6.38}$$

Proof.

$$\begin{aligned}
E_k^p[S(k)^4 | T_{\mathrm{GCS}} > k] &= E_k^p[S(k)^4 | T_{\mathrm{GCS}} > k, T_{\mathrm{GCS}} > k - M(k)] \\
&= \frac{E_k^p[S(k)^4 | T_{\mathrm{GCS}} > k - M(k)]}{P_k^p(T_{\mathrm{GCS}} > k | T_{\mathrm{GCS}} > k - M(k))} \\
&= \frac{\sum_{u=1}^{q(A)-1} E_k^p[S(k)^4 \mathbb{1}_{\{M(k)=u\}} | T_{\mathrm{GCS}} > k - M(k)]}{P_k^p(T_{\mathrm{GCS}} > k | T_{\mathrm{GCS}} > k - M(k))} \\
&= \frac{\sum_{u=0}^{q(A)-1} E_k^p[(\sum_{\ell=k-u}^{k} X_\ell^i)^4 \mathbb{1}_{\{M(k)=u\}} | T_{\mathrm{GCS}} > k - M(k)]}{P_k^p(T_{\mathrm{GCS}} > k | T_{\mathrm{GCS}} > k - M(k))} \\
&\leq \frac{\sum_{u=0}^{q(A)-1} E_k^p[(\sum_{\ell=k-u}^{k} X_\ell^i)^8]^{1/2} P_k^p(T^i \geq u)^{1/2}}{P_k^p(T_{\mathrm{GCS}} > k | T_{\mathrm{GCS}} > k - M(k))} \\
&\leq \frac{\sum_{u=0}^{q(A)-1} C_{143} u^2 (C_{31} e^{-C_{32}u})^{1/2}}{C_{111}} \\
&\leq C_{141}.
\end{aligned}$$

for some constants C_{141} and C_{143} that can be derived from the property of exponential family and the fact that the summation $\sum_{u=1}^{\infty} u^2 e^{-u}$ is bounded. The proof of relationship (6.37) follows the same path as the proof of (6.36) and relationship (6.38) follows directly from (6.36).

Now we have bounded $M(k), S(k), G(k)$ in Lemma 6.10 to Lemma 6.14, and here we first consider the sequential test $\mathsf{T}_k^i(A)$ when we are not monitoring the affected stream at the change time ($i \neq p$):

$$\mathsf{T}_k^i(A) = \inf\{t-k : t \geq k+1, t-k+M(k) = q(A) \text{ or } S_{t,k}^i \notin (-G(k), A-G(k))\}, \tag{6.39}$$

where $S_{t,k}^i = \sum_{\ell=k+1}^t ((\tilde{\theta}_\ell - \theta_0)X_\ell^i - \psi(\tilde{\theta}_\ell) + \psi(\theta_0))$ and $\tilde{\theta}_\ell$ is defined as

$$\tilde{\theta}_\ell = \min\left\{\zeta, \max\left\{\delta, \psi_1\left(\frac{S(k) + X_{k+1}^i + \cdots + X_{\ell-1}^i}{M(k) + \ell - 1 - k}\right)\right\}\right\}. \tag{6.40}$$

The following lemma gives the bounds on $\mathsf{T}_k^i(A)$ when $i \neq p$.

Lemma 6.15. *For the estimator $\tilde{\theta}_\ell$ defined in (6.40), there exists a constant C_{151} such that*

$$\mathsf{P}_k^p(|\tilde{\theta}_r - \delta| > \lambda | \mathsf{T}_{\mathrm{GCS}} > k) < \frac{C_{151}}{(r-k)^4} \tag{6.41}$$

for all $r = k+1, k+2, \ldots$.

Proof. Due to the continuity of ψ' and Lemma 6.14, we can select suitable $\eta = \eta(\lambda)$ such that

$$\begin{aligned}
&P_k^p(|\tilde{\theta}_r - \delta| > \lambda | T_{\mathrm{GCS}} > k) \\
&\leq P_k^p\left(\left|\frac{S(k) + X_{k+1}^i + \cdots + X_{r-1}^i}{M(k) + r - 1 - k} - \psi'(\theta_0)\right| > \eta | T_{\mathrm{GCS}} > k\right) \\
&\leq P_k^p\left(\left|\frac{S(k) + X_{k+1}^i + \cdots + X_{r-1}^i}{M(k) + r - 1 - k} - \psi'(\theta_0)\right| > \eta, \frac{|S(k)|}{M(k) + r - 1 - k} < \frac{\eta}{4} | T_{\mathrm{GCS}} > k\right) \\
&+ P_k^p\left(\frac{|S(k)|}{M(k) + r - 1 - k} \geq \frac{\eta}{4} || T_{\mathrm{GCS}} > k\right) \\
&\leq P_k^p\left(\left|\frac{X_{k+1}^i + \cdots + X_{r-1}^i}{M(k) + r - 1 - k} - \psi'(\theta_0)\right| > \frac{\eta}{2} | T_{\mathrm{GCS}} > k\right) + \frac{C_{152}}{(r-k)^4} \\
&\leq P_k^p\left(\left|\frac{X_{k+1}^i + \cdots + X_{r-1}^i}{r - 1 - k} - \frac{M(k) + r - 1 - k}{r - 1 - k}\psi'(\theta_0)\right| > \frac{\eta}{2} | T_{\mathrm{GCS}} > k\right) + \frac{C_{152}}{(r-k)^4} \\
&\leq P_k^p\left(\left|\frac{X_{k+1}^i + \cdots + X_{r-1}^i}{r - 1 - k} - \psi'(\theta_0)\right| > \frac{\eta}{2} - \frac{M(k)}{r - 1 - k}|\psi'(\theta_0)|| T_{\mathrm{GCS}} > k\right) + \frac{C_{152}}{(r-k)^4} \\
&\leq P_k^p\left(\left|\frac{X_{k+1}^i + \cdots + X_{r-1}^i}{r - 1 - k} - \psi'(\theta_0)\right| > \frac{\eta}{4} | T_{\mathrm{GCS}} > k\right) \\
&+ P_k^p\left(\frac{M(k)}{r - 1 - k}|\psi'(\theta_0)| \geq \frac{\eta}{4} | T_{\mathrm{GCS}} \geq k\right) + \frac{C_{152}}{(r-k)^4} \\
&\leq \frac{C_{151}}{(r-k)^4}
\end{aligned}$$

for some constants $\eta = \eta(\lambda), C_{151}, C_{152}$ that can be derived from Lemma 6.1 and Lemma 6.14 and all $r = k+1, k+2, \ldots$. □

Lemma 6.16. *There exists a constant C_{161} such that*

$$\mathsf{E}_k^p\left[\mathsf{T}_k^i(A)|\mathsf{T}_{\mathrm{GCS}} \geq k\right] \leq C_{161}. \tag{6.42}$$

Proof. We consider a new sequential test:

$$\mathsf{T}_k^{i,*}(A) = \inf\{t-k : t \geq k+1, \sum_{\ell=k+1}^{t}((\theta_0-\tilde{\theta}_\ell)X_\ell^i + \psi(\tilde{\theta}_\ell) - \psi(\theta_0)) \notin (-\infty, G(k))\}, \tag{6.43}$$

and it is clear that $\mathsf{T}_k^{i,*}(A) \geq \mathsf{T}_k^i(A)$. We claim that

$$\mathsf{P}_k^p(\mathsf{T}_k^{i,*}(A) \geq r|\mathsf{T}_{\mathrm{GCS}} > k) \leq \frac{C_{162}}{r^2} \tag{6.44}$$

for some constant C_{162} and all $r \geq 1$. To prove relationship (6.44), consider the event

$$H = \{|\tilde{\theta}_\ell - \delta| \leq \lambda \text{ for all } \ell \geq k+r\}.$$

From Lemma 6.15, we obtain that $\mathsf{P}_k^p(H^C|\mathsf{T}_{\mathrm{GCS}} > k) \leq \frac{C_{163}}{r^3}$ for some constant C_{163}. Under the event H, we have

$$(\theta_0-\tilde{\theta}_\ell)X_\ell^i + \psi(\tilde{\theta}_\ell) - \psi(\theta_0) \geq Z_\ell = (\theta_0-\delta)X_\ell^i + \psi(\delta) - \psi(\theta_0) - \lambda|X_\ell^i| - \max_{|\theta-\delta|\leq\lambda}(\psi(\delta)-\psi(\theta)). \tag{6.45}$$

We select suitable λ so that $\mathsf{E}_k^p[Z_\ell] > 0$ and

$$\begin{aligned}
P_k^p(T_k^{i,*}(A) \geq r|T_{\mathrm{GCS}} > k) &= P_k^p\left(\sum\nolimits_{\ell=k+1}^{k+r}((\theta_0-\tilde{\theta}_\ell)X_\ell^i + \psi(\tilde{\theta}_\ell) - \psi(\theta_0)) \leq G(k)\right) \\
&\leq P_k^p\left(\sum\nolimits_{\ell=k+1}^{k+r} Z_\ell \leq G(k)) + P_k^p(H^C|T_{\mathrm{GCS}} > k\right) \\
&\leq P_k^p\left(\sum\nolimits_{\ell=k+1}^{k+r} Z_\ell \leq \frac{rE_k^p[Z_\ell]}{2}\right) + P_k^p\left(G(k) \geq \frac{rE_k^p[Z_\ell]}{2}|T_{\mathrm{GCS}} > k\right) + \frac{C_{153}}{r^3} \\
&\leq P_k^p\left(\left|\frac{\sum_{\ell=k+1}^{k+r} Z_\ell}{r} - E_k^p[Z_\ell]\right| \geq \frac{E_k^p[Z_\ell]}{2}\right) + \frac{C_{164}}{r^4} + \frac{C_{163}}{r^3}
\end{aligned}$$

for some constant C_{164} that can be derived from Lemma 6.14. Below we write Z_ℓ as

$$Z_\ell = X_\ell + Y_\ell + C_{165},$$

where $X_\ell = (\theta_0-\delta)X_\ell^i$, $Y_\ell = \lambda|X_\ell^i|$ and $C_{165} = \psi(\delta) - \psi(\theta_0) - \max_{|\theta-\delta|\leq\lambda}(\psi(\delta) - \psi(\theta))$. We have

$$
\begin{aligned}
&P_k^p\left(\left|\frac{\sum_{\ell=k+1}^{k+r} Z_\ell}{r} - E_k^p[Z_\ell]\right| \geq \frac{E_k^p[Z_\ell]}{2}\right)\\
&= P_k^p\left(\left|\frac{\sum_{\ell=k+1}^{k+r} X_\ell}{r} + \frac{\sum_{\ell=k+1}^{k+r} Y_\ell}{r} - E_k^p[X_\ell] - E_k^p[Y_\ell]\right| \geq \frac{E_k^p[Z_\ell]}{2}\right)\\
&\leq P_k^p\left(\left|\frac{\sum_{\ell=k+1}^{k+r} X_\ell}{r} - E_k^p[X_\ell]\right| + \left|\frac{\sum_{\ell=k+1}^{k+r} Y_\ell}{r} - E_k^p[Y_\ell]\right| \geq \frac{E_k^p[Z_\ell]}{2}\right)\\
&\leq \min\left\{P_k^p\left(\left|\frac{\sum_{\ell=k+1}^{k+r} X_\ell}{r} - E_k^p[X_\ell]\right| \geq \frac{E_k^p[Z_\ell]}{4}\right), P_k^p\left(\left|\frac{\sum_{\ell=k+1}^{k+r} Y_\ell}{r} - E_k^p[Y_\ell]\right| \geq \frac{E_k^p[Z_\ell]}{4}\right)\right\}\\
&\leq C_{166} e^{-C_{167} r}
\end{aligned}
$$

for some constants C_{166}, C_{167} that can be derived from Lemma 6.1. Relationship (6.44) is then proved and relationship (6.42) follows directly from (6.44). □

With Lemma 6.16, we are able to write the detection delay when we are not monitoring the affected stream at the change time as

$$
\begin{aligned}
D_p(\mathsf{T}_{\mathrm{GCS}}(A)|\ i \neq p) &\leq \mathsf{E}_k^p\left[\mathsf{T}_k^i(A)|\ i \neq p, \mathsf{T}_{\mathrm{GCS}} > k\right] + \mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1]\\
&\leq C_{161} + \frac{A}{I(\theta_p, \theta_0)} + C_{(1)} \log A + C_{(2)} \sqrt{\mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1]} + C_{(3)}\ .\\
&\leq \frac{A}{I(\theta_p, \theta_0)} + C_{(1)} \log A + C_{(2)} \sqrt{\mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1 = 1]} + C_{(4)}
\end{aligned} \tag{6.46}
$$

Consider now the second scenario when we are just monitoring the affected stream (the pth stream) when the change occurs. The sequential test $\mathsf{T}_k^p(A)$ is defined as

$$\mathsf{T}_k^p(A) = \inf\{t - k : t \geq k, t - k + M(k) = q(A) \text{ or } S_{t,k}^p \notin (-G(k), A - G(k))\}, \tag{6.47}$$

where $S_{t,k}^p = \sum_{\ell=k+1}^{t}((\tilde{\theta}_\ell - \theta_0)X_\ell^p - \psi(\tilde{\theta}_\ell) + \psi(\theta_0))$ and $\tilde{\theta}_\ell$ is defined by

$$\tilde{\theta}_\ell = \min\left\{\zeta, \max\left\{\delta, \psi^{(1)}\left(\frac{S(k) + X_{k+1}^p + \cdots + X_{\ell-1}^p}{M(k) + \ell - 1 - k}\right)\right\}\right\}. \tag{6.48}$$

Lemma 6.17. *For the estimator $\tilde{\theta}_\ell$ defined in (6.48), there exists constant C_{171} such that*

$$\mathsf{P}_k^p(|\tilde{\theta}_\ell - \theta_p| > \lambda|\mathsf{T}_{\mathrm{GCS}} > k) < \frac{C_{171}}{(\ell - k)^4} \tag{6.49}$$

for all $\ell = k + 1, k + 2, \ldots$.

Proof. The proof is the same as the proof of Lemma 6.15.

Lemma 6.18. *Let $\beta_{p,k} = \mathsf{P}_k^p(S_{\mathsf{T}_k^p+k,k}^p \geq A - G(k))$, and we can write $\mathsf{E}_k^p[\mathsf{T}_k^p|\mathsf{T}_{\mathrm{GCS}} > k]$ as*

$$\mathsf{E}_k^p\left[\mathsf{T}_k^p|\mathsf{T}_{\mathrm{GCS}} > k\right] \leq \frac{(1 - \beta_{p,k})A}{I(\theta_p, \theta_0)} + x_1 + x_2 + x_3, \tag{6.50}$$

where

$$x_1 \le C_{181}\Big(E_k^p\Big[\sum\nolimits_{\ell=k+1}^{k+T_k^p}(\tilde{\theta}_\ell-\theta_p)^2|T_{\mathrm{GCS}}>k\Big]\Big)^{1/2}+C_{182}(E_k^p[T_k^p]\Big)^{1/2}$$
$$x_2 \le C_{183}E_k^p\Big[\sum\nolimits_{\ell=k+1}^{k+T_k^p}(\tilde{\theta}_\ell-\theta_p)^2|T_{\mathrm{GCS}}>k\Big]+C_{184}E_k^p\Big[\sum\nolimits_{\ell=k+1}^{k+T_k^p}(\psi'(\tilde{\theta}_\ell)-\psi'(\theta_p))^2|T_{\mathrm{GCS}}>k\Big]$$
$$x_3 \le C_{185}AP_k^p(T_k^p=q(A)-M(k)|T_{\mathrm{GCS}}>k)$$

for some constants $C_{181}, C_{182}, C_{183}, C_{184}, C_{185}$.

Proof. The proof is the same as the proof of Lemma 6.2.

Lemma 6.19. *There exists constant* C_{191} *such that for sufficient large* $A \ge A_0$ *and every* $r = 1, 2, \ldots,$

$$\mathsf{P}_k^p\left(\mathsf{T}_k^p > \frac{10A}{I(\theta_p,\theta_0)}+r|\mathsf{T}_{\mathrm{GCS}}>k\right)\le\frac{C_{191}}{r^3}. \tag{6.51}$$

Proof. Let $s=[\frac{5A}{I(\theta_p,\theta_0)}]$; then we define the following event for some constant η :

$$E=\left\{|\tilde{\theta}_\ell-\theta_p|<\lambda \quad \text{for all } \ell \ge k+\frac{r}{2}\right\}.$$

From relationship (6.49) we have $\mathsf{P}_k^p(E^C|\mathsf{T}_{\mathrm{GCS}}>k)\le\frac{C_{192}}{r^3}$ for some constant C_{192}. It follows that the log-likelihood ratio satisfies

$$(\tilde{\theta}_\ell-\theta_0)X_\ell^p-\psi(\tilde{\theta}_\ell)+\psi(\theta_0)\ge Y_\ell=(\theta_p-\theta_0)X_\ell^p-\lambda|X_\ell^p|-C_{112}, \tag{6.52}$$

where $C_{193}=\max_{|\theta-\theta_p|<\lambda}(\psi(\theta)-\psi(\theta_0))$. We choose proper η,λ such that $\mathsf{E}_k^p[Y_\ell]>0$ and define the stopping time

$$N_*(A)=\inf\left\{t>0:\sum_{\ell=1}^{t}Y_\ell\ge A\right\}. \tag{6.53}$$

It is easy to verify that if $\mathsf{T}_k^p(A)>10A/I(\theta_p,\theta_0)+r$, then $N_*(A)>10A/I(\theta_p,\theta_0)+r$. Moreover, it is straightforward to see that $\mathsf{P}_k^p(N_*(A)>10A/I(\theta_p,\theta_0)+r|\mathsf{T}_{\mathrm{GCS}}>k)$ is bounded exponentially in r, and we can prove that there exist constants C_{194}, C_{195} such that

$$\begin{aligned}\mathsf{P}_k^p\left(\mathsf{T}_k^p>\frac{10A}{I(\theta_p,\theta_0)}+r|\mathsf{T}_{\mathrm{GCS}}>k\right)&=\mathsf{P}_k^p\left(\mathsf{T}_k^p>\frac{10A}{I(\theta_p,\theta_0)}+r|\mathsf{T}_{\mathrm{GCS}}>k,E\right)\mathsf{P}_k^p(E|\mathsf{T}_{\mathrm{GCS}}>k)\\&\quad+\mathsf{P}_k^p(\mathsf{T}_k^p>\frac{10A}{I(\theta_p,\theta_0)}+r|\mathsf{T}_{\mathrm{GCS}}>k,E^C)\mathsf{P}_k^p(E^C|\mathsf{T}_{\mathrm{GCS}}>k)\\&\le\mathsf{P}_k^p\left(\mathsf{T}_k^p>\frac{10A}{I(\theta_p,\theta_0)}+r||\mathsf{T}_{\mathrm{GCS}}>k,E\right)+\mathsf{P}_k^p(E^C|\mathsf{T}_{\mathrm{GCS}}>k)\\&\le\mathsf{P}_k^p\left(N_*>\frac{10A}{I(\theta_p,\theta_0)}+r|\mathsf{T}_{\mathrm{GCS}}>k\right)+\mathsf{P}_k^p(E^C|\mathsf{T}_{\mathrm{GCS}}>k)\\&\le\frac{C_{191}}{r^3}.\end{aligned} \tag{6.54}$$

for some constant C_{191}. □

Lemma 6.20. *There exist constants C_{201}, C_{202} such that*

$$\mathsf{E}_k^p\left[\sum_{\ell=k+1}^{k+\mathsf{T}_k^p}(\psi'(\tilde{\theta}_\ell)-\psi'(\theta_p))^2|\mathsf{T}_{\mathrm{GCS}}\geq k\right]\leq C_{201}\log A+C_{202}. \tag{6.55}$$

Proof.

$$\begin{aligned}
&E_k^p\left[\sum_{\ell=k+1}^{k+T_k^p}(\psi'(\tilde{\theta}_\ell)-\psi'(\theta_p))^2|T_{\mathrm{GCS}}>k\right]\\
&\leq E_k^p\left[\sum_{\ell=k+1}^{k+q(A)}(\psi'(\tilde{\theta}_\ell)-\psi'(\theta_p))^2|T_{\mathrm{GCS}}>k\right]\\
&\leq E_k^p\left[\sum_{\ell=k+1}^{k+q(A)}(\psi'(\tilde{\theta}_\ell)-\psi'(\theta_p))^2|T_{\mathrm{GCS}}>k\right]\\
&\leq E_k^p\left[\sum_{\ell=k+1}^{k+q(A)}\left(\frac{S(k)+X_{k+1}^p+\cdots+X_{\ell-1}^p}{M(k)+\ell-1-k}-\psi'(\theta_p)\right)^2|T_{\mathrm{GCS}}>k\right]\\
&\leq 2\left\{E_k^p\left[\sum_{\ell=k+1}^{k+q(A)}\left(\frac{S(k)-M(k)\psi'(\theta_p)}{M(k)+\ell-1-k}\right)^2+\left(\frac{\sum_{r=k+1}^{l-1}X_r^p-(\ell-1-k)\psi'(\theta_p)}{M(k)+\ell-1-k}\right)^2\middle|T_{\mathrm{GCS}}>k\right]\right\}\\
&\leq C_{201}\log A+C_{202}
\end{aligned}$$

for some constants C_{201}, C_{202} that can be derived from Lemma 6.1 and Lemma 6.14.

Lemma 6.21. *There exists constant C_{211} such that*

$$\mathsf{P}_k^p(\mathsf{T}_k^p=q(A)-M(k)|\mathsf{T}_{\mathrm{GCS}}>k)\leq\frac{211}{q(A)^3} \tag{6.56}$$

for all sufficiently large threshold $A\geq A_0$.

Proof.

$$\begin{aligned}
\mathsf{P}_k^p(\mathsf{T}_k^p=q(A)-M(k)|\mathsf{T}_{\mathrm{GCS}}>k)&\leq\mathsf{P}_k^p\left(\mathsf{T}_k^p=\frac{q(A)}{2}|\mathsf{T}_{\mathrm{GCS}}>k\right)+\mathsf{P}_k^p\left(M(k)=\frac{q(A)}{2}|\mathsf{T}_{\mathrm{GCS}}>k\right)\\
&\leq\frac{C_{212}}{q(A)^3}+\frac{C_{213}}{q(A)^4}\\
&\leq\frac{C_{211}}{q(A)^3}
\end{aligned}$$

for some constant $C_{211}, C_{212}, C_{213}$. □

With these lemmas, we can write the detection delay when we are just monitoring the affected stream at the change time as

$$\begin{aligned}
D_p(\mathsf{T}_{\mathrm{GCS}}|i=p)&=\mathsf{E}_k^p\left[\mathsf{T}_k^p|\mathsf{T}_{\mathrm{GCS}}\geq k\right]+\beta_{p,k}\mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1=1]\\
&\leq\frac{(1-\beta_{p,k})A}{I(\theta_p,\theta_0)}+x_1+x_2+x_3+\beta_{p,k}\mathsf{E}_0^p[\mathsf{T}_{\mathrm{GCS}}(A)|R_1=1]\\
&\leq\frac{A}{I(\theta_p,\theta_0)}+C_{(5)}\log A+C_{(6)}\sqrt{D_p(\mathsf{T}_{\mathrm{GCS}})}+C_{(7)}p,
\end{aligned} \tag{6.57}$$

for some suitable constants $C_{(5)}, C_{(6)}, C_{(7)}$. Relationship (4.1) follows directly by combining relationships (6.46) and (6.57).

7. CONCLUSIONS

In this article, we consider the active quickest detection problem with unknown post-change parameters, where there are p local streams in a system but one is only able to take observations from one of these p local streams at each time instant. We propose an efficient GCS-based quickest detection algorithm under sampling control constraint, whose main idea is to keep monitoring one stream until we are confident to switch or raise an alarm. Our proposed algorithm is shown to have first-order asymptotic optimality in the sense of minimizing detection delay when the false alarm constraint γ goes to infinity and $p = o(\log \gamma)$. Numerical studies are conducted to show the effectiveness and applicability of the proposed algorithm.

There are a number of interesting problems that have not been addressed here. In practice, one may be interested in the more general scenario when there are $s > 1$ affected data streams and we are allowed to sample from $q > 1$ streams per time step. Our conjecture is that we are still able to achieve asymptotic optimality results when $q \geq s$, but are unable to do so if $q < s$. Another research direction would be to consider monitoring the system based on a linear projection of the complete data. The sparse structure of the changes might help us to detect the change in high-dimensional data with only a low-dimensional projection. Therefore, this article is just the beginning of further investigation.

DISCLOSURE STATEMENT

The authors have no conflicts of interest to report.

FUNDING

This work was partially supported by NSF grant DMS-2015405 and the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002378. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ORCID

Qunzhi Xu http://orcid.org/0000-0003-2524-2260
Yajun Mei http://orcid.org/0000-0002-1015-990X

REFERENCES

Chan, H. P. 2017. "Optimal Sequential Detection in Multi-Stream Data." *The Annals of Statistics* 45 (6):2736–63. doi:10.1214/17-AOS1546

Chen, J., and A. Gupta. 2012. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. 2nd ed. New York: Birkhauser.

Chen, Y., and X. Li. 2019. “Compound Sequential Change Point Detection in Multiple Data Streams.” arXiv preprint arXiv:1909.05903.

Fellouris, G., and V. V. Veeravalli. 2022. “Quickest Change Detection with Controlled Sensing.” In *2022 IEEE International Symposium on Information Theory (ISIT)*, 1921–1926, doi:10.1109/ISIT50566.2022.9834351

Lai, T. L. 1987. “Adaptive Treatment Allocation and the Multi-Armed Bandit Problem.” *The Annals of Statistics* 15 (3):1114. doi:10.1214/aos/1176350495

Lai, T. L. 1995. “Sequential Changepoint Detection in Quality Control and Dynamical Systems.” *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (4):613–44. doi:10.1111/j.2517-6161.1995.tb02052.x

Li, J. 2020. “Asymptotic Distribution-Free Change-Point Detection Based on Interpoint Distances for High-Dimensional Data.” *Journal of Nonparametric Statistics* 32 (1):157–84. doi:10.1080/10485252.2019.1710505

Li, J., and J. Jin. 2010. “Optimal Sensor Allocation by Integrating Causal Models and Set-Covering Algorithms.” *IIE Transaction* 42 (8):564–76. doi:10.1080/07408170903232597

Li, S., X. Li, X. Wang, and J. Liu. 2019. “Sequential Hypothesis Test with Online Usage-Constrained Sensor Selection.” *IEEE Transactions on Information Theory* 65 (7):4392–410. doi:10.1109/TIT.2019.2910730

Liu, K., R. Zhang, and Y. Mei. 2019. “Scalable SUM-Shrinkage Schemes for Distributed Monitoring Large-Scale Data Streams.” *Statistica Sinica* 29:1–22.

Liu, K., Y. Mei, and J. Shi. 2015. “An Adaptive Sampling Strategy for Online High-Dimensional Process Monitoring.” *Technometrics* 57 (3):305–19. doi:10.1080/00401706.2014.947005

Lorden, G. 1971. “Procedures for Reacting to a Change in Distribution.” *The Annals of Mathematical Statistics* 42 (6):1897–908. doi:10.1214/aoms/1177693055

Lorden, G., and M. Pollak. 2008. “Sequential Change-Point Detection Procedures That Are Nearly Optimal and Computationally Simple.” *Sequential Analysis* 27 (4):476–512. doi:10.1080/07474940802446244

Mei, Y. 2010. “Efficient Scalable Schemes for Monitoring a Large Number of Data Streams.” *Biometrika* 97 (2):419–33. doi:10.1093/biomet/asq010

Moustakides, G. V. 1986. “Optimal Stopping Times for Detecting Changes in Distributions.” *The Annals of Statistics* 14 (4):1379–87. doi:10.1214/aos/1176350164

Page, E. S. 1954. “Continuous Inspection Schemes.” *Biometrika* 41 (1–2):100–15. no doi:10.2307/2333009

Pollak, M. 1985. “Optimal Detection of a Change in Distribution.” *The Annals of Statistics* 13 (1):206–77. doi:10.1214/aos/1176346587

Pollak, M. 1987. “Average Run Lengths of an Optimal Method of Detecting a Change in Distribution.” *The Annals of Statistics* 15 (2):749–79. doi:10.1214/aos/1176350373

Poor, H. V., and O. Hadjiliadis. 2008. *Quickest Detection*. New York: Cambridge University Press.

Shiryaev, A. N. 1963. “On the Detection of Disorder in a Manufacturing Process. I.” *Theory of Probability and Its Applications* 8 (3):247–65.

Tartakovsky, A. G., I. Nikiforov, and M. Basseville. 2014. *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. Boca Raton: Chapman and Hall/CRC.

Tsopelakos, A., G. Fellouris, and V. V. Veeravalli. 2019. “Sequential Anomaly Detection with Observation Control.” In *2019 IEEE International Symposium on Information Theory (ISIT)*, 2389–2393. doi:10.1109/ISIT.2019.8849555

Wang, Y., and Y. Mei. 2015. “Large-Scale Multi-Stream Quickest Change Detection via Shrinkage Post-Change Estimation.” *IEEE Transactions on Information Theory* 61 (12):6926–38. doi:10.1109/TIT.2015.2495361

Wu, Y. 2020. “Estimation of Common Change Point and Isolation of Changed Panels after Sequential Detection.” *Sequential Analysis* 39 (1):52–64. doi:10.1080/07474946.2020.1726685

Xie, L., S. Zou, Y. Xie, and V. V. Veeravalli. 2021. “Sequential (Quickest) Change Detection: classical Results and New Directions.” *IEEE Journal on Selected Areas in Information Theory* 2 (2):494–514. doi:10.1109/JSAIT.2021.3072962.

Xie, Y., and D. Siegmund. 2013. "Sequential Multi-Sensor Change-Point Detection." *The Annals of Statistics* 41 (2):670–92. doi:10.1214/13-AOS1094

Xu, Q., Y. Mei, and G. V. Moustakides. 2020. "Second-Order Asymptotically Optimal Change-Point Detection Algorithm with Sampling Control." In *2020 IEEE International Symposium on Information Theory (ISIT)*, 1136–1140, doi:10.1109/ISIT44484.2020.9174114

Xu, Q., Y. Mei, and G. V. Moustakides. 2021. "Optimum Multi-Stream Sequential Change-Point Detection with Sampling Control." *IEEE Transactions on Information Theory* 67 (11):7627–36. (Accepted). doi:10.1109/TIT.2021.3074961.