Optimal Parallel Sequential Change Detection under Generalized Performance Measures

Zexian Lu, *Student Member, IEEE,* Yunxiao Chen, *Member, IEEE,* and Xiaoou Li, *Member, IEEE*

***Abstract*—This paper considers the detection of change points in parallel data streams, a problem widely encountered when analyzing large-scale real-time streaming data. Each stream may have its own change point, at which its data has a distributional change. With sequentially observed data, a decision maker needs to declare whether changes have already occurred to the streams at each time point. Once a stream is declared to have changed, it is deactivated permanently so that its future data will no longer be collected. This is a compound decision problem in the sense that the decision maker may want to optimize certain compound performance metrics that concern all the streams as a whole. Thus, the decisions are not independent for different streams. Our contribution is three-fold. First, we propose a general framework for compound performance metrics that includes the ones considered in the existing works as special cases and introduces new ones that connect closely with the performance metrics for single-stream sequential change detection and large-scale hypothesis testing. Second, data-driven decision procedures are developed under this framework. Finally, optimality results are established for the proposed decision procedures. The proposed methods and theory are evaluated by simulation studies and a case study.**

***Index Terms*—Large-scale inference, multiple change detection, sequential analysis, multiple hypothesis testing**

I. Introduction

Sequential change detection aims to detect distributional changes in sequentially observed data. Classical methods focusing on change detection in a single data stream have received wide applications in various fields, including engineering, education, medical diagnostics and finance [1–4]. Several metrics have been proposed for evaluating their performance, under which optimality theory has been established [5–8]; see [9–11] for a review.

The emergence of large-scale real-time streaming data has motivated multi-stream sequential change detection. Different settings have been considered in the literature, largely motivated by surveillance applications, where each data stream corresponds to a sensor, and the change point is caused by a failure in one or multiple sensors. For example, [12] consider one change point which occurs to one and only one of multiple streams. [13] and [14] consider a setting where all the streams change at the same time. More generally, [15–22] consider settings where a common change occurs to all or a subset of streams. There are also settings under which data streams are gradually affected after some propagation time. For these settings, a change can appear at a predetermined stream [23] or at any stream [24], but eventually all the streams will change. A related problem, which has received much attention recently and will be the focus of the current work, considers a setting where each stream has its own change point [25–28], where the stream-specific change points have independent causes. Real-world applications of such a setting will be discussed in the sequel. A decision maker needs to declare whether a change has already occurred for each stream at each time point. Once a stream is declared to have changed, it is deactivated permanently so that its data is no longer collected. This problem will be referred to as a parallel sequential change detection problem.

The parallel sequential change detection problem is widely encountered in the real world. For example, [25, 29] consider an application to a multichannel dynamic spectrum access problem for cognitive radios. Each cognitive radio channel corresponds to a data stream, and the change corresponds to the time at which the primary user of the channel starts to transmit signals. A false discovery rate (FDR) is proposed to measure the proportion of false discoveries (i.e., unused channels) among the ones detected as occupied by primary users. [26, 27] consider monitoring an item pool for standardized educational testing. In this application, each stream corresponds to a test item that is reused in multiple test administrations, and the change point corresponds to the time at which the item is leaked to the public. A certain false non-discovery rate (FNR) is proposed to measure the proportion of leaked items among the non-detections (i.e., items that are not detected as having leaked). There are many other potential applications, such as the detection of credit card fraud [30], for which each stream corresponds to a credit card account, and the change point corresponds to a fraud event.

We note that it is often not a good idea to run a single-stream change detection procedure independently on individual streams. This is because the decision maker may want to control a certain compound risk that concerns all the streams as a whole, such as the FDR and FNR measures. Consequently, each decision at one time point requires all the up-to-date information from all the streams, making the parallel sequential change detection a challenge.

Several methods have been proposed in [25–27] to control the above compound risk measures in parallel sequential change detection problems. However, these methods, along with their theoretical properties, are established under rel-

Manuscript received June 18, 2023. Yunxiao Chen is partially supported by IEA Research and Development Funds. Xiaoou Li is partially supported by NSF CAREER DMS-2143844.

Zexian Lu and Xiaoou Li are with the School of Statistics, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: lu000276@umn.edu; lixx1766@umn.edu).

Yunxiao Chen is with the Department of Statistics, London School of Economics and Political Science, London, WC2A 2AE UK (e-mail: Y.Chen186@lse.ac.uk).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the authors. The material includes the proofs of all the technical results and an additional simulation study. This material is 400 KB in size.

atively restrictive model assumptions and for specific risk measures. Specifically, [25] proposes a method based on the Benjamini-Hochberg method [31] for FDR control and establishes its asymptotic results. However, no results are given on the method's optimality. Under a Bayesian setting, [26] and [27] propose methods for controlling a certain FNR measure at all time points. As shown in [26], under a geometric change point model and assuming the same pre- and post- change distribution for all the streams, this method maximizes the expected number of remaining streams at all time points while controlling the FNR to be no greater than a pre-specified tolerance level. However, it is unclear whether this optimality theory can be extended to more general models and other sensible risk measures.

The parallel sequential change detection problem is also closely related to the sequential multiple testing problem. The latter can be viewed as a special case when a stream can only change at the beginning of the process or never change. Several methods have been proposed for the sequential multiple testing problem, controlling compound risks. Specifically, [32], [33], and [34] consider controlling a familywise error rate, an FDR/FNR, and a generalized familywise error rate, respectively. While the risk measures may be relevant, their methods and theoretical results can hardly be extended to the current change detection problem.

Statistical methodology for sequential change detection can generally be divided into Bayesian and non-Bayesian methods [10, 11, 13, 35, 36]. Bayesian methods assume a prior distribution for the change point/points, based on which performance metrics are defined, such as average detection delay and probability of false alarm. On the other hand, non-Bayesian methods do not assume any knowledge on the change point distribution, and typically aims to find the best-performing procedure in the worst-case scenario (through a minimax formulation). For these methods, performance metrics are introduced to measure their performance in the worst-case scenario, such as supremum average detection delay, average run length to the false alarm, and worst-case probabilities of missed detection and false alarm; see e.g., [13, 16, 36, 37].

The current work studies parallel sequential change detection under a Bayesian setting. It provides a unified decision theory framework under general classes of change point models and performance measures. A computationally efficient sequential method is developed under the proposed framework. Two optimality criteria are introduced, for which the proposed method is shown to be optimal under suitable conditions.

Our contributions are summarized below:

- We propose a general class of performance metrics to evaluate the sequence procedures. This class of metrics not only includes existing metrics as special cases (e.g., FDR [25] and the local FNR metric [26]) but also introduces new metrics that are closely related to the metrics for single-stream change detection and multiple hypothesis testing. See Section II-E and Section IV-C for more examples.

  Thanks to the generality of these performance metrics, the proposed method can also be used to solve problems considered in [33, 34, 38] for sequential multiple testing. See Section IV-C for a discussion on the connections with several recent works [32–34, 38, 39].
- We propose a sequential procedure (Algorithms 1–4) that is easy-to-implement and is data-driven. It automatically adapts to various model settings when controlling the risk measures to a pre-specified tolerance level, without requiring additional Monte Carlo simulation or bisection search commonly used in sequential problems to determine decision boundaries (see, e.g., [40]).
- We provide two optimality criteria for the parallel sequential change detection problem, including the local and uniform optimalities. The local optimality criterion concerns the maximization of a utility measure in the next step, and the uniform optimality criterion refers to the maximization of the utility measure at all time. We show that the proposed method is locally optimal under very mild conditions and uniformly optimal under stronger conditions (Theorems 1–3).

  We note that the precise characterization of the conditions for uniform optimality requires the analysis of stochastic processes on a special non-Euclidean space. To this end, we develop new analytical tools for comparing vectors and stochastic processes with different dimensions due to early stopping. This analytical tool may be useful in the theoretical analysis of other sequential decision problems.

The remainder of the paper is organized as follows. In Section II, we describe the change point models, the class of parallel sequential change detection methods, a general class of performance metrics, and the optimality criteria. We also provide examples of generalized performance metrics. In Section III, we propose a parallel change detection method (Algorithms 1 and 2) and provide a simplified version of this method under mild conditions on the performance measures (Algorithms 3 and 4). Section IV provides theoretical results for the proposed methods including their optimality properties and the connection with recent works. In Sections V and VI, we evaluate the performance of the proposed method through simulation studies and a case study. Concluding remarks and future directions are given in Section VII. For space reasons, part of the proofs of the theoretical and numerical results are postponed to the Appendix in the supplementary material.

II. Problem Setup

A. Notation

The following notations are used throughout the paper. For two real numbers, $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For two sets A and B, $A \setminus B = \{x : x \in A, x \notin B\}$ denotes the set minus operator. We abbreviate 'almost surely' as 'a.s.' For a set S, $|S|$ denotes its cardinality. $\mathbb{Z}$ and $\mathbb{Z}_+$ denote the set of integers and positive integers, respectively. For a positive integer n, $\langle n \rangle$ denotes the set $\{1, \cdots, n\}$.

B. Model Assumptions

Consider the case where there are $K \geq 2$ data streams, and let $\langle K \rangle$ denote the set $\{1, \cdots, K\}$. At each time epoch $t \in \mathbb{Z}_+$, an observation $X_{k,t}$ is obtained from the kth data stream, for $k \in \langle K \rangle$. Each data stream k is associated with a change

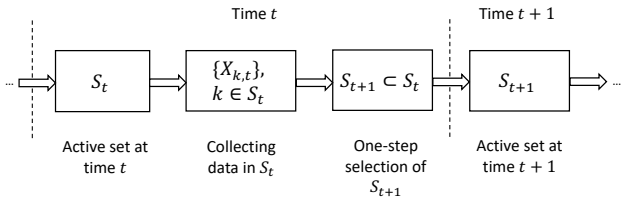

Fig. 1: A flowchart of a sequential decision in $\mathcal{D}$

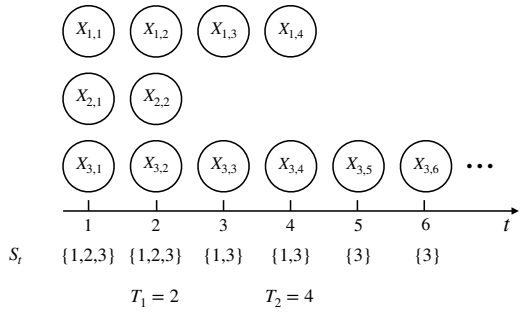

Fig. 2: An example of a parallel sequential change detection procedure where $K = 3$, stream 2 is deactivated at time $t = 2$, stream 3 is deactivated at time $t = 3$, and no more stream is deactivated before $t = 6$. As a result, $S_1 = S_2 = \{1, 2, 3\}$, $S_3 = S_4 = \{1, 2\}$, $S_5 = S_6 = \{1\}$. Correspondingly, $T_1 = 2$, $T_2 = 4$, and $T_3 > 6$.

point $\tau_k \in \{0\} \cup \{\infty\} \cup \mathbb{Z}_+$ for $k \in \langle K \rangle$. Under a Bayesian parallel change point model, the change points $\tau_1, \cdots, \tau_K$ are assumed to be independent and identically distributed (i.i.d.) with

$$\mathbb{P}(\tau_k = s) = \pi_s \tag{1}$$

for $s \in \{0\} \cup \{\infty\} \cup \mathbb{Z}_+$ and $k \in \langle K \rangle$. Given $(\tau_1, \cdots, \tau_K)$, $\{X_{k,t}\}_{t \in \mathbb{Z}_+}$ are independent for $k \in \langle K \rangle$, and have conditional probability density

$$X_{k,t} | \tau_k, \{X_{k,s}\}_{1 \le s \le t-1} \sim \begin{cases} p_{k,t} \text{ if } t \le \tau_k \\ q_{k,t} \text{ if } t \ge \tau_k + 1 \end{cases} \tag{2}$$

with respect to some baseline measure over a measurable space $(\Omega, \mathcal{F})$. That is, $X_{k,t}$ are independent given the change points, and follow pre- and post- change density functions $p_{k,t}$ and $q_{k,t}$, respectively. In particular, $\tau_k = \infty$ corresponds to the case where the change point never occurs to the kth stream. That is, $X_{k,t}$ follows the pre-change density function $p_{k,t}$ for all $t \in \mathbb{Z}_+$. Throughout the paper, all the probabilities and expectations are taken under the model described above.

C. Parallel Sequential Change Detection Procedures

A decision maker sequentially observes data from the parallel data streams and determines whether change points have already occurred to these data streams at each time. Once a change point is declared, the corresponding data stream is deactivated and its data are no longer collected. This decision process is characterized by an index set process $S_t \subset \langle K \rangle$ for $t \in \mathbb{Z}_+$, where $k \in S_t$ if and only if the decision maker has not declared a change in the kth stream at time t yet (i.e., stream k is active at time t). Specifically, the available information at time t is contained in the historical data $H_t = \{\{X_{k,s}\}_{k \in S_s, 1 \le s \le t}, \{S_s\}_{1 \le s \le t}\}$ and, equivalently, the induced information σ-field $\mathcal{F}_t = \sigma(H_t)$. At each time t, the decision maker selects the index set $S_{t+1} \subset S_t$ based on the current information $\mathcal{F}_t$. That is, S_{t+1} is measurable with respect to $\mathcal{F}_t$. Denote by $\mathcal{D}$ the set of all such compound sequential decisions. A graphical illustration of the decision process is given in Figure 1.

We make a few remarks on the information filtration and the decision process. First, we require $S_1 = \langle K \rangle$, meaning that all the streams are initially active and data from all the streams are collected at time 1. Second, $\{S_s\}_{1 \le s \le t}$ is measurable with respect to $\mathcal{F}_t$, meaning that the decision history is tracked in the current information. Third, $\{X_{k,s}\}_{k \in S_s, 1 \le s \le t}$ is measurable with respect to $\mathcal{F}_t$, indicating that $X_{k,s}$ is observed if and only if stream k is active at time s and $s \le t$ (i.e., $k \in S_s$). Fourth, S_{t+1} is required to be measurable with respect to $\mathcal{F}_t$, meaning that the decision maker selects the active streams for time $t + 1$ based on all the information available at time t. Lastly, S_{t+1} is required to be a subset of S_t for all $t \in \mathbb{Z}_+$, meaning that the deactivation of streams is permanent. That is, no future data will be collected at a stream, once a change is declared at that stream.

Remark 1. *Although described in a different way, the class of sequential decisions defined above is equivalent to that in [25]. In [25], a parallel sequential procedure is defined through a sequence of stopping times $\{T_q\}_{q \ge 1}$ along with a sequence of index sets $\{D_q\}_{q \ge 1}$. At each stopping time T_q, a decision maker declares change points for streams in D_q and exclude those streams from the future decision process. Then, the sequences $\{T_q\}_{q \ge 1}$ and $\{D_q\}_{q \ge 1}$ can be represented using the sequence $\{S_t\}_{t \ge 1}$ as $T_q = \min\{t > T_{q-1} : S_t \backslash S_{t+1} \neq \varnothing\}$ and $D_q = S_{T_q} \setminus S_{T_q+1}$ where $T_0 = 0$, $q = 1$. An example where $K = 3$ is given in Figure 2 for a graphical illustration.*

Another way to understand a compound sequential change detection procedure is to view it as a sequence of mappings $\delta = (d_1, d_2, \cdots, d_t, \cdots)$, where each d_t determines S_{t+1} according to the historical information H_t. That is, d_t is a measurable function with respect to $\mathcal{F}_t$ and $S_{t+1} = d_t(H_t)$ satisfying that $d_t(H_t) \subset S_t$ for all $t \in \mathbb{Z}_+$.

D. Generalized Performance Measures and Optimality Criteria

Ideally, a perfect sequential change detection procedure collects all the pre-change streams in the set S_t at each time point (i.e., $S_t = \{k : \tau_k \ge t\}$). However, this is not achievable by any sequential decision because τ_ks are unobserved. To this end, we consider a general class of performance measures to compare the performance of different sequential decisions. We assume each sequential decision is associated with a risk process, denoted by $\{R_t\}_{t \in \mathbb{Z}_+}$, and a utility process, denoted by $\{U_t\}_{t \in \mathbb{Z}_+}$. The risk process is used to quantify the loss of a sequential decision at time t due to the false detections of pre-change streams and/or the non-detection of post-change

streams, while the utility process is used to reward the correct decisions. Our goal is to find a good sequential decision that has a relatively small R_t and a relatively large U_t at every time point. Below, we first give formal statements of the optimality criteria, and then introduce several examples of R_t and U_t in Section II-E, followed by additional discussions.

Let

$$W_{k,t} = \mathbb{P}(\tau_k < t \mid \mathcal{F}_t) \tag{3}$$

be the posterior probability that the change point τ_k has already occurred at time t for the k-th stream given the information up to time t. Under the Bayesian setting, $W_{k,t}$ is also the best estimator (under the squared error loss) of $\mathbb{1}(\tau_k < t)$, where $\mathbb{1}(\cdot)$ denotes the indicator function. A simple iterative updating rule is derived to calculate $W_{k,t}$ at each time, which will be discussed in Section III.

Throughout the paper, we consider risk and utility processes that are functions of $(\{W_{k,t}\}_{k\in S_t}, S_t, S_{t+1})$. That is, there are pre-specified functions $\{r_t\}_{t\in\mathbb{Z}_+}$ and $\{u_t\}_{t\in\mathbb{Z}_+}$ such that

$$R_t = r_t(\{W_{k,t}\}_{k\in S_t}, S_t, S_{t+1}), \tag{4}$$

and

$$U_t = u_t(\{W_{k,t}\}_{k\in S_t}, S_t, S_{t+1}). \tag{5}$$

From (4) and (5), R_t and U_t can be any processes that are measurable with respect to $\{W_{k,t}\}$, S_t and S_{t+1}. On the other hand, only some choices of R_t and U_t lead to practically meaningful performance measures. A partial list of practically meaningful choices of r_t and u_t are given in Section II-E

Let $\alpha \in \mathbb{R}$ denote a pre-specified tolerance level, and let

$$\mathcal{D}_\alpha = \{\delta \in \mathcal{D} : R_t(\delta) \le \alpha \text{ a.s., for all } t = 1, 2, \cdots\},$$

where $R_t(\delta)$ denotes the risk process associated with the sequential decision δ, and $\mathcal{D}$ denotes the entire set of parallel sequential detection procedures described in Section II-C. The set $\mathcal{D}_\alpha$ collects all sequential decisions that control the risk process to be no greater than the tolerance level α at all time points.

We note that risk process $\{R_t\}_{t\in\mathbb{Z}_+}$ is an adaptive stochastic process with respect to the information filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}_+}$. It is easy to verify that $\mathbb{E}[R_t(\delta)] \le \alpha$ for $\delta \in \mathcal{D}_\alpha$. That is, the expected risk is also controlled below or equal to the same tolerance level. In addition, any weighted average of $R_t(\delta)$ across different time points are also controlled. We provide additional discussion and theoretical results regarding this point in Section IV-C.

The following regularity assumptions over the risk and utility functions are imposed throughout the paper.

Assumption 1. *For any $\{W_{k,t}\}_{k\in S_t}$ and S_t, $\min_{S\in\{\emptyset,S_t\}} r_t(\{W_{k,t}\}_{k\in S_t}, S_t, S) \le \alpha$. In addition, the utility function u_t is bounded at each time t.*

The assumption on r_t guarantees that the class of sequential decisions controlling the risk process at a pre-specified level is non-empty, i.e., $\mathcal{D}_\alpha \neq \emptyset$. The boundedness assumption on u_t is a mild condition to ensure the integrability of the utility process.

Given a pre-specified tolerance level α and sequences of functions $\{r_t\}_{t\in\mathbb{Z}_+}$ and $\{u_t\}_{t\in\mathbb{Z}_+}$, we define two optimality criteria for sequential decisions in $\mathcal{D}_\alpha$.

Definition 1 (Uniform Optimality)**.** *A sequential decision $\delta^* \in \mathcal{D}_\alpha$ is called uniformly optimal if*

$$\mathbb{E}\left(U_t\left(\delta^*\right)\right) = \sup_{\delta\in\mathcal{D}_\alpha} \mathbb{E}\left(U_t(\delta)\right),$$

for all $t \in \mathbb{Z}_+$, where $U_t(\delta^*)$ and $U_t(\delta)$ denote the utility process associated with sequential decisions δ^* and δ, respectively.

Definition 2 (Local Optimality)**.** *A sequential decision $\delta^* = (d_1^*, d_2^*, \cdots, d_t^*, \cdots) \in \mathcal{D}_\alpha$ is called locally optimal at time t, if*

$$\mathbb{E}(U_t(\delta^*)) \ge \mathbb{E}(U_t(\delta))$$

for any $\delta = (d_1, d_2, \cdots, d_t, \cdots) \in \mathcal{D}_\alpha$ satisfying $d_s = d_s^*$, for $s = 1, \ldots, t-1$.

We make a few remarks on the above optimality criteria. First, in most applications, there is a trade-off between minimizing the risk and maximizing the utility. That is, a sequential decision that has relatively small risk tends to have relatively small utility at the same time. Thus, we define both uniform and local optimality through constrained optimization problems, where the overall goal is to find a sequential decision so that its corresponding risk process is controlled to be no greater than the tolerance level while the expected utility is no less than any other sequential decisions that control the risk process at the same level. Second, a uniformly optimal sequential decision has the largest expected utility among all decisions in $\mathcal{D}_\alpha$ at *every* time point. In contrast, a locally optimal sequential decision only has the largest expected utility at *a given time point* t given the decisions at previous time points. Thus, uniform optimality is a stronger notion than local optimality. A sequential decision that is locally optimal at every time point does not necessarily imply that it is also uniformly optimal. See [26, Example 3] for a counterexample where a locally optimal decision exists for $R_t = \text{LFNR}_t$ in (12) and $U_t = |S_{t+1}|$ but there is no uniformly optimal decision. In later sections, we show that locally optimal sequential decisions exist under very weak assumptions on the risk and utility measures, while uniformly optimal sequential decisions only exist under stronger assumptions of the change point model and the performance measures. Third, we assume the same tolerance level α for every time t for ease of presentation. Our methods and theory can be easily extended to the class of sequential decisions whose risk is controlled at different levels at different time points. That is, $\{\delta \in \mathcal{D} : R_t(\delta) \le \alpha_t \text{ a.s., for all } t\}$ for a sequence of constants α_t. We can see this by redefining the risk process as $R_t - \alpha_t$ and replacing α_t by 0.

E. Examples of Generalized Performance Measures

We start with several examples of performance measures in the forms of (4) and (5), which are motivated by common risk measures in the literature of multiple hypotheses testing [31, 41–43].

For the consistency of notation, the sum over an empty set is defined to be 0 (i.e., $\sum_{i\in\emptyset} a_i = 0$), and the product over an empty set is defined to be 1 (i.e., $\prod_{i\in\emptyset} a_i = 1$).

Example 1 (Local family-wise error rate (LFWER))**.** *Consider the event*

$$E_{1,t} = \{\textit{There exists } k \in \langle K\rangle \textit{ such that } \tau_k < t, k \in S_{t+1}\}, \tag{6}$$

which happens when at least one false non-detection occurs at time t. Because $E_{1,t}$ is not directly observed, we consider the its posterior probability given the information up to time t,

$$\textit{LFWER}_t := \mathbb{P}(E_{1,t}|\mathcal{F}_t) = 1 - \prod_{k\in S_{t+1}} (1 - W_{k,t}). \tag{7}$$

Example 2 (Generalized local family-wise error rate (GLFWER))**.** *Given $m \geq 1$, we consider the event*

$$E_{m,t} = \{|\{k \in \langle K\rangle \textit{ such that } \tau_k < t, k \in S_{t+1}\}| \geq m\}. \tag{8}$$

This event happens when false non-detections occur in at least m data streams. Its posterior probability given information up to time t is

$$\textit{GLFWER}_{m,t} := \mathbb{P}(E_{m,t}|\mathcal{F}_t) \tag{9}$$

$$=1 - \sum_{j=0}^{m-1} \sum_{\substack{I\subset S_{t+1}\\ |I|=j}} \left(\prod_{i\in I} W_{i,t}\right) \prod_{k\in S_{t+1}\setminus I} (1 - W_{k,t}). \tag{10}$$

In addition, $\textit{GLFWER}_{m,t} = 0$ if $S_{t+1} = \varnothing$.

Comparing (7) with (9), we can see that GLFWER extends LFWER by allowing for more false non-detections. Under a large-scale setting with many data streams, it may be more sensible to use GLFWER with its m value chosen based on the total number of streams K to achieve a balance between false detections and false non-detections. Similar risk measures have been proposed for sequential multiple testing [34].

Example 3 (Local false non-discovery rate (LFNR))**.** *Local false non-discovery rate (LFNR) is defined in [26], which extends the concept of LFNR in multiple testing to parallel sequential change detection. It is defined as follows. First, the false non-discovery proportion (FNP) is defined as*

$$\textit{FNP}_t := \frac{\sum_{k\in S_{t+1}} \mathbb{1}(\tau_k < t)}{|S_{t+1}| \vee 1}. \tag{11}$$

FNP describes the proportion of post-change streams among the active ones. Then, the local false non-discovery rate (LFNR) at time t is defined as the Bayes estimator (i.e., posterior mean) of $\textit{FNP}_t$ given information up to time t. That is,

$$\textit{LFNR}_t := \mathbb{E}(\textit{FNP}_t \mid \mathcal{F}_t) = \frac{\sum_{k\in S_{t+1}} W_{k,t}}{|S_{t+1}| \vee 1}. \tag{12}$$

Compared with LFWER and GLFWER, LFNR depends on $W_{k,t}$ in a linear rather than multivariate polynomial form. In addition, LFNR is scalable under a large-scale setting in the sense that the same tolerance level $\alpha \in (0, 1)$ can be used as K grows large.

Example 4 (Local False Discovery Rate (LFDR))**.** *False discovery proportion (FDP) and local false discovery rate (LFDR) are defined by replacing $\tau_k < t$ and S_{t+1} with $\tau_k \geq t$ and $S_t \setminus S_{t+1}$ respectively in* (11) *and* (12)*. That is,*

$$\textit{FDP}_t := \frac{\sum_{k\in S_t\setminus S_{t+1}} \mathbb{1}(\tau_k \geq t)}{|S_t \setminus S_{t+1}| \vee 1}, \tag{13}$$

and

$$\textit{LFDR}_t := \mathbb{E}(\textit{FDP}_t \mid \mathcal{F}_t) = \frac{\sum_{k\in S_t\setminus S_{t+1}} (1 - W_{k,t})}{|S_t \setminus S_{t+1}| \vee 1}. \tag{14}$$

Similar to LFNR, LFDR also has the appealing feature of scalability for large K. The difference between LFNR and LFDR lies in whether focusing on false detections or false non-detections.

In [25], an aggregated version of false discovery rate (AFDR)[1] is considered, which can be viewed as the expectation of a weighted average of LFDR at different time points. More discussions on the connection between LFDR and AFDR will be provided in Section IV.

Next, we provide two examples of performance measures motivated by single-stream sequential change detection. Denote by N_k the detection time of the kth stream,

$$N_k = \sup\{t : k \in S_t\}. \tag{15}$$

Note that N_k plays a similar role as the stopping time in the standard single-stream sequential change detection problem. Indeed, N_k is a stopping time with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{Z}_+}$ for all $k \in \langle K\rangle$.

Example 5 (Incremental Average Run Length (IARL))**.** *We define the incremental run length (IRL) aggregated over different streams as*

$$\begin{aligned}\textit{IRL}_t :&= \sum_{k=1}^{K}\{\tau_k \wedge N_k \wedge (t+1)\} - \sum_{k=1}^{K}\{\tau_k \wedge N_k \wedge t\}\\ &= \sum_{k\in S_{t+1}}^{K} \mathbb{1}(\tau_k > t)\end{aligned} \tag{16}$$

IRL indicates the total number of pre-change streams being used at a given time. We refer to its posterior mean as the incremental average run length (IARL), defined as

$$\textit{IARL}_t := \mathbb{E}(\textit{IRL}_t \mid \mathcal{F}_t) = \sum_{k\in S_{t+1}} \{1 - g(W_{k,t})\}, \tag{17}$$

where

$$g(W_{k,t}) = \mathbb{P}(\tau_k \leq t|\mathcal{F}_t) = \bar{\pi}_t^{-1}\pi_t + \big(1 - \bar{\pi}_t^{-1}\pi_t\big)W_{k,t}, \tag{18}$$

$\bar{\pi}_s = \mathbb{P}\big(\tau_k \geq s\big) = \pi_\infty + \sum_{l=s}^{\infty} \pi_l$, and the proof for equation (18) *is given in Appendix D.*

IRL and IARL are closely related to the average run length to false alarm (ARL2FA) that is commonly used to measure the propensity for making a false detection in a single-stream

[1]In [25], this risk measure is referred to as 'false discovery rate (FDR)'. Here, we name it as AFDR to distinguish it from LFDR.

sequential change detection problem. Specifically, taking summation of IRL_t over t, we obtain

$$\sum_{s=0}^{t-1}\text{IRL}_s=\sum_{k=1}^{K}\big(\tau_k\wedge N_k\wedge t\big), \tag{19}$$

which is the total run length from different data streams up to the change point by time t. Moreover, we have

$$\mathbb{E}(\sum_{s=0}^{t-1}\text{IARL}_s)=\mathbb{E}(\sum_{s=0}^{t-1}\text{IRL}_s)=\sum_{k=1}^{K}\mathbb{E}\big(\tau_k\wedge N_k\wedge t\big). \tag{20}$$

Thus, the sum of the expected value of IARL across time leads to the total averaged run length up to the change point.

Example 6 (Incremental Average Detection Delay (IADD))**.** *We define the incremental detection delay (IDD) aggregated over all the streams as*

$$\begin{aligned}\textit{IDD}_t:&=\sum_{k=1}^{K}\{(N_k\wedge(t+1)-\tau_k-1)_+-(N_k\wedge t-\tau_k-1)_+\}\\&=\sum_{k\in S_{t+1}}\mathbb{1}(\tau_k<t).\end{aligned} \tag{21}$$

IDD counts the total number of post-change streams that are active at a given time. We refer to its posterior mean as the incremental average detection delay (IADD), defined as

$$\textit{IADD}_t:=\mathbb{E}(\textit{IDD}_t\mid\mathcal{F}_t)=\sum_{k\in S_{t+1}}W_{k,t}. \tag{22}$$

By taking summation over t, we have

$$\sum_{s=0}^{t-1}\text{IDD}_s=\sum_{k=1}^{K}(N_k\wedge t-\tau_k-1)_+ \tag{23}$$

and

$$\mathbb{E}\big(\sum_{s=0}^{t-1}\text{IADD}_s\big)=\mathbb{E}\big(\sum_{k=1}^{K}(N_k\wedge t-\tau_k-1)_+\big). \tag{24}$$

Remark 2. *IDD and IADD are closely related to the concept of average detection delay (ADD), which is commonly used to measure false non-detection (i.e., the change point has occurred but the sequential decision fails to detect it) in single stream sequential change detection [44]. We clarify that IDD and IADD are random, and ADD for a single data stream is a non-random number.*

Next, we give a precise characterization of the relationship between IDD, IADD, and ADD for parallel change detection. In [44], for a sequential detection rule with a stopping time N and a change point τ following some prior distribution, its ADD is defined as

$$\textit{ADD}(N,\tau)=\mathbb{E}\big(N-\tau|N>\tau\big).$$

It can be shown that IDD, IADD, and ADD have the following relationship

$$\mathbb{E}\big(\sum_{s=0}^{\infty}\textit{IADD}_s\big)=\sum_{k=1}^{K}\{\textit{ADD}(N_k,\tau_k)-1\}\mathbb{P}(N_k>\tau_k) \tag{25}$$

given that $N_1,\cdots,N_K<\infty$ almost surely. In other words, if we aggregate the expected value of IADD_t over all the time t, then it is the same as a weighted sum of ADD -1 across different data streams, where the weight is determined by the probability of the stopping time to be greater than the corresponding change point. The proof of (25) *is given in Appendix D.*

Remark 3. *Note that if $S_{t+1}=\emptyset$, then LFWER_t in Example 1, GLFWER_t in Example 2, LFNR_t in Example 3, IARL_t in Example 5 and IADD_t in Example 6 become zero, regardless of the value of S_t and $\{W_{k,t}\}_{k\in S_t}$. Similarly, if we take $S_{t+1}=S_t$, then LFDR_t in Example 4 becomes zero, regardless of the value of S_t and $\{W_{k,t}\}_{k\in S_t}$. Thus, $\min_{S\in\{\emptyset,S_t\}}r_t(\{W_{k,t}\}_{k\in S_t},S_t,S)=0$ and Assumption 1 is satisfied for all $\alpha\geq 0$. Thus, all of the risk measures discussed in this section satisfy Assumption 1 for $\alpha\geq 0$.*

Among the above examples, LFWER, GLFWER, and LFNR are error rates for false non-detections, LFDR is an error rate for false detections, IARL estimates the number of pre-change streams that are active, and IADD estimates the number of post-change and active streams. Because a small value of LFWER (or GLFWER/LFNR/IADD) and a large value of IARL (or minus LFDR) is desired, we could choose the risk process $R_t\in\{\text{LFWER}_t,\text{GLFWER}_t,\text{LFNR}_t,\text{IADD}_t\}$ and the utility process $U_t\in\{\text{IARL}_t,-\text{LFDR}_t\}$, or $R_t\in\{\text{LFDR}_t,-\text{IARL}_t\}$ and $U_t\in\{-\text{LFWER}_t,-\text{GLFWER}_t,-\text{LFNR}_t,-\text{IADD}_t\}$. Note that in the above examples, there is a trade-off between R_t and U_t. That is, if one declares detection at more data streams, then the corresponding LFWER, GLFWER, LFNR, and IADD tend to be smaller and IARL and minus LFDR tend to be smaller as well. Thus, the optimality criteria (Definitions 1 and 2) formulated through constrained optimization are reasonable.

The choices of R_t and U_t should be application-driven. In practice, we suggest to choose the risk process R_t with a known range so that the tolerance level is easy to specify. For example, LFWER, GLFWER, LFNR, and LFDR represent certain probability/expected proportions that are known to be between $[0,1]$. Thus, they are sensible choices of R_t, for which setting the tolerance level $\alpha\in[0,1]$ is relatively straightforward.

III. Proposed Sequential Decision Procedures

In this section, we first provide a formula for computing the posterior probability $W_{k,t}=\mathbb{P}(\tau_k<t|\mathcal{F}_t)$, which is a key quantity in computing the risk and utility measures. Then, we present our proposed sequential decisions for controlling the risk process at a given level, followed by a simplified version of the algorithm to reduce the computational complexity.

A. Recursive Formula for $W_{k,t}$

Recall $\pi_s=\mathbb{P}(\tau_k=s)$ and $\bar{\pi}_s=\mathbb{P}\big(\tau_k\geq s\big)$. Let

$$Q_{k,t}=\bar{\pi}_t^{-1}\sum_{s=0}^{t-1}\pi_s\prod_{r=s+1}^{t}\frac{q_{k,r}\left(X_{k,r}\right)}{p_{k,r}\left(X_{k,r}\right)}\text{ with }Q_{k,0}=0. \tag{26}$$

Given $Q_{k,t}$ and $X_{k,t+1}$, $Q_{k,t+1}$ can be computed using the recursive formula

$$Q_{k,t+1} = \bar{\pi}_{t+1}^{-1}\big(\bar{\pi}_t Q_{k,t} + \pi_t\big) L_{k,t+1}, \tag{27}$$

where we define $L_{k,t+1} = q_{k,t+1}\left(X_{k,t+1}\right)/p_{k,t+1}(X_{k,t+1})$. Then, we obtain

$$W_{k,0} = 0 \text{ and } W_{k,t} = \frac{Q_{k,t}}{Q_{k,t}+1}. \tag{28}$$

We note that Q_{kt} calculates the odds of stream k having changed given the up-to-date information $\mathcal{F}_t$. It is introduced to obtain a recursive formula for calculating W_{kt}. The above recursive equations (27) and (28) are extensions of classic results in single-stream Bayesian sequential change detection problems [45]. Their rigorous justifications are given in Appendix D.

Remark 4. *In many applications, the prior distribution and pre- and post-change distributions may be unknown. To apply the proposed method, one may combine it with a plug-in estimator for these unknown distributions. In particular, assuming that the prior distribution and the pre- and post-distributions follow a parametric model (i.e., the distributions only depend on a finite number of parameters), we could take an empirical Bayes approach to estimate the unknown parameters and use an Expectation-maximization (EM) algorithm [46] for its computation. Then, we apply the proposed method by replacing π_t, $\bar{\pi}_t$, $q_{k,t}$, $p_{k,t}$ in* (26) – (28) *with their estimates. According to a simulation study in Appendix A, this approach seems to perform well in controlling a compound risk. The theoretical properties of this approach are left for future investigation.*

B. Proposed Sequential Decision for Unstructured Risk and Utility

We first propose a one-step selection rule to select S_{t+1}, given S_t and $\{W_{k,t}\}_{k\in S_t}$ so that the risk R_t is controlled to be no greater than α. This one-step selection rule goes over all $2^{|S_t|}$ possible subsets of S_t, and then select the one which attains the highest utility U_t. Algorithm 1 implements this idea. According to Assumption 1, $\{S : \gamma_S \le \alpha \text{ and } S \subset S_t\} \neq \emptyset$.

Algorithm 1 One-step selection rule at time t.

1: **Input:** Tolerance level α, the current index set S_t, and posterior probabilities $\{W_{k,t}\}_{k\in S_t}$, where $W_{k,t} = \mathbb{P}\left(\tau_k < t \mid \mathcal{F}_t\right)$ is computed according to (28).
2: For all $S \subset S_t$, compute $\gamma_S = r_t(\{W_{k,t}\}_{k\in S_t}, S_t, S)$ and $\mu_S = u_t(\{W_{k,t}\}_{k\in S_t} S_t, S)$.
3: **Output:**
$S_{t+1} = \arg\max_S \mu_S$ subject to $\gamma_S \le \alpha$ and $S \subset S_t$.[1]

Thus, S_{t+1} in line 3 of the above algorithm is well-defined. The next proposition states that the above one-step selection rule can control the risk process at any given level.

Proposition 1. *Under Assumption 1, the index set S_{t+1} selected by Algorithm 1 satisfies $R_t \le \alpha$ a.s.*

[1]If the solution is not unique, S_{t+1} can be any one of the solutions.

Proof. According to the second and third lines of Algorithm 1, S_{t+1} output by Algorithm 1 belongs to the set $\{S : \gamma_S \le \alpha \text{ and } S \subset S_t\}$. Thus, $R_t = r_t(\{W_{k,t}\}_{k\in S_t}, S_t, S_{t+1}) = \gamma_{S_{t+1}} \le \alpha$. □

Note that Proposition 1 does not require any assumptions on R_t and U_t except for Assumption 1, which ensures the existence of the set S_{t+1} in the last line of Algorithm 1.

Next, we combine Algorithm 1 at different time points to obtain a sequential decision in $\mathcal{D}_\alpha$. At each time t, this sequential decision selects S_{t+1} using Algorithm 1 and deactivates data streams that are not in the index set. Algorithm 2 below implements this idea.

Algorithm 2 Proposed sequential decision δ_{P}.

1: **Input:** Tolerance level α.
2: Initialize: set $t = 1$, $S_t = \langle K\rangle$ and compute $W_{k,t}$ for $k \in S_t$ using equations (27) and (28).
3: Select: input α, S_t and $(W_{k,t})_{k\in S_t}$ to Algorithm 1, and obtain S_{t+1}.
4: Update: deactivate streams in $S_t \setminus S_{t+1}$. If $S_{t+1} = \varnothing$, stop; otherwise, update $\{W_{k,t+1}\}_{k\in S_{t+1}}$ using equations (27) and (28).
5: Iterate: set $t = t+1$ and return to line 2.
6: **Output:** $\{S_t\}_{t\ge 1}$.

Proposition 2. *Under Assumption 1, $\delta_{\mathrm{P}} \in \mathcal{D}_\alpha$. That is, the proposed sequential decision given by Algorithm 2 controls the risk process at level α at every time point.*

Proof. For each t, S_{t+1} is obtained through Algorithm 1. Thus, $R_t(\delta_{\mathrm{P}}) \le \alpha$ a.s. for all $t \in \mathbb{Z}_+$, according to Proposition 1. This implies $\delta_{\mathrm{P}} \in \mathcal{D}_\alpha$. □

C. Simplified Sequential Decision for 'Monotone' Risk

At each time t, directly applying Algorithm 1 requires evaluating and comparing the risk and utility associated with $2^{|S_t|}$ subsets, which is computationally intensive when $|S_t|$ is large. In many cases where the risk and utility satisfy additional monotonicity assumptions, this algorithm can be simplified, reducing the computational complexity significantly. In this section, we provide one such assumption, under which the proposed sequential decision only requires evaluating and comparing the risks associated with $|S_t|+1$ subsets.

Assumption 2. *For all non-empty $S_0 \subset \langle K\rangle$, $\mathbf{w} = (w_1, \cdots, w_{|S_0|}) \in [0,1]^{|S_0|}$, $S \subset S_0$, $i \in S, j \in S_0 \setminus S$, and $w_i \ge w_j$, we have $r_t(\mathbf{w}, S_0, S) \ge r_t(\mathbf{w}, S_0, (S\setminus\{i\})\cup\{j\})$ and $u_t(\mathbf{w}, S_0, S) \le u_t(\mathbf{w}, S_0, (S\setminus\{i\})\cup\{j\})$.*

Under Assumption 2, R_t tends to become larger and U_t tends to become smaller if we keep streams with relatively smaller posterior probability active. Under this assumption, Algorithm 1 can be simplified to the following Algorithm 3, and it also controls R_t to be below a pre-specified level α. As will be discussed in Corollary 1, all the risk and utility measures presented in Examples 1 – 6 satisfy this assumption.

The following Algorithm 3 selects S_{t+1} so that streams with relatively large posterior probabilities are detected and those

with relatively small posterior probabilities are kept active. The cut-off point for the detection is decided by maximizing the utility while controlling the risk at time t. Because Algorithm 3 restricts S_{t+1} to be a subset of streams with relatively small posterior probability, it only involves evaluating and comparing the risk and utility functions associated with $|S_t|+1$ subsets, and, thus, reduces the computational complexity from the order $O(2^{|S_t|})$ to the order $O(|S_t|\log(|S_t|))$.

Algorithm 3 Simplified one-step selection rule.

1: **Input:** Tolerance level α, the current index set S_t, and posterior probabilities $\{W_{k,t}\}_{k\in S_t}$.
2: Arrange posterior probabilities in an ascending order[2], i.e.

$$W_{k_1,t} \le W_{k_2,t} \le \cdots \le W_{k_{|S_t|},t}.$$

3: For $n = 1, \ldots, |S_t|$, compute

$$\gamma_n = r_t(\{W_{k,t}\}_{k\in S_t}, S_t, \{k_i\}_{i=1}^n)$$

and

$$\mu_n = u_t(\{W_{k,t}\}_{k\in S_t}, S_t, \{k_i\}_{i=1}^n).$$

For $n = 0$, compute $\gamma_0 = r_t(\{W_{k,t}\}_{k\in S_t}, S_t, \varnothing)$ and $\mu_0 = u_t(\{W_{k,t}\}_{k\in S_t}, S_t, \varnothing)$.
4: Set $n^* \in \{0, \cdots, |S_t|\}$ as the solution to the problem[3]

$$n^* = \arg\max_n \mu_n \text{ subject to } \gamma_n \le \alpha.$$

5: **Output:** $S_{t+1} = \{k_1, \ldots, k_{n^*}\}$ if $n^* \ge 1$ and $S_{t+1} = \varnothing$ if $n^* = 0$.

Note that under Assumption 1, $\{n : \gamma_n \le \alpha\} \neq \emptyset$. Thus, the fourth line of the above Algorithm 3 is well-defined. The following Algorithm 4 gives an overall sequential decision rule δ_S by adopting Algorithm 3 at every time point.

Algorithm 4 Simplified decision procedure δ_S.

1: **Input:** Tolerance level α.
2: Initialize: set $t = 1$. $S_t = \langle K \rangle$ and compute $W_{k,t}$ for $k \in S_t$ using equations (27) and (28).
3: Select: input α, S_t and $(W_{k,t})_{k\in S_t}$ to Algorithm 3, and obtain S_{t+1}.
4: Update: deactivate streams in $S_t \setminus S_{t+1}$. If $S_{t+1} = \varnothing$, stop; otherwise, update $\{W_{k,t+1}\}_{k\in S_{t+1}}$ using equations (27) and (28).
5: Iterate: set $t = t + 1$ and return to Step 3.
6: **Output:** $\{S_t\}_{t\ge 1}$.

Proposition 3. *Under Assumptions 1, the sequential decision δ_S given by Algorithm 4 satisfies $\delta_S \in \mathcal{D}_\alpha$.*

Proof. Under Assumption 1, $R_t = r_t(\{W_{k,t}\}_{k\in S_t}, S_t, S_{t+1})$, where S_{t+1} is obtained by Algorithm 3. According to the third and fourth lines of Algorithm 3, it satisfies $R_t = r_t(\{W_{k,t}\}_{k\in S_t}, S_t, S_{t+1}) \le \alpha$. Thus, $\delta_S \in \mathcal{D}_\alpha$. □

[2]If $W_{k_i,t} = W_{k_j,t}$ for $1 \le i < j \le n$, we choose $k_i < k_j$ to avoid additional randomness because of ties.

[3]If the solution is not unique, we choose n^* to be the largest solution.

IV. Theoretical Properties of Proposed Methods

In this section, we first show that the proposed sequential decision δ_P is locally optimal under very weak assumptions in Section IV-A. Then, we show that the simplified sequential decision δ_S is uniformly optimal under stronger model assumptions and additional monotonicity assumptions on risk and utility measures in Section IV-B. We also provide theoretical results on aggregated risk and utility measures of the proposed methods in Section IV-C.

A. Local Optimality Results

The following two theorems show that the proposed sequential decision δ_P is locally optimal under Assumption 1 while δ_S is locally optimal under Assumptions 1 and 2. That is, they satisfy Definition 2.

Theorem 1. *Under Assumption 1, the sequential decision δ_P described in Algorithm 2 is locally optimal.*

Proof. First, we know that $\delta_P = (d_1^*, d_2^*, \cdots, d_t^*, \cdots) \in \mathcal{D}_\alpha$ according to Proposition 2. We compare it with an arbitrary sequential decision $\delta = (d_1, d_2, \cdots, d_t, \cdots) \in \mathcal{D}_\alpha$ satisfying $d_s = d_s^*$, for $s = 1, \cdots, t-1$. Let $\{S_t^*\}_{t\in\mathbb{Z}_+}$ be the index set of active streams following δ_P at all time points, and S_{t+1} be the set selected by δ at time $t+1$. Note that both δ_P and δ select $S_1^*, \cdots, S_t^*$ as the index sets at time $1, \cdots, t$, according to the assumption that $d_s = d_s^*$ for $s = 1, \cdots, t-1$.

According to the second and third line of Algorithm 1, S_{t+1}^* satisfies $u_t(\{W_{k,t}\}_{k\in S_t^*}, S_t^*, S_{t+1}^*) = \max_{S\subset S_t^*} u_t(\{W_{k,t}\}_{k\in S_t^*}, S_t^*, S)$ subject to $\gamma_S \le \alpha$. Because $\delta \in \mathcal{D}_\alpha$, and the index set selected by δ and δ_P at time t are both S_t^*, we have $R_t(\delta) = \gamma_{S_{t+1}} \le \alpha$ a.s. This further implies $u_t(\{W_{k,t}\}_{k\in S_t^*}, S_t^*, S_{t+1}^*) \ge u_t(\{W_{k,t}\}_{k\in S_t^*}, S_t^*, S_{t+1})$. That is, $U_t(\delta_P) \ge U_t(\delta)$. The proof is completed by taking expectation on both sides. □

Theorem 2. *Under Assumptions 1 and 2, the sequential decision δ_S described in Algorithm 4 is locally optimal.*

Proof. The proof is given in Appendix B. □

The next corollary applies the above results to examples given in Section II-E.

Corollary 1. *If $\alpha > 0$ and $R_t \in \{LFWER_t, GLFWER_t, LFNR_t, IADD_t\}$, $U_t \in \{IARL_t, -LFDR_t\}$, or $R_t = LFDR_t$ and $U_t \in \{-LFWER_t, -GLFWER_t, -LFNR_t, -IADD_t\}$, then the simplified sequential decision δ_S is locally optimal.*

If $\alpha < 0$ and $R_t = -IARL_t$ and $U_t \in \{-LFWER_t, -GLFWER_t, -LFNR_t, -IADD_t\}$, then the simplified sequential decision δ_S is locally optimal.

Proof. The proof is given in Appendix B. □

B. Uniform Optimality Results

In this section, we show that the proposed sequential decision rule δ_S defined in Algorithm 4 is uniformly optimal under stronger assumptions. We note that the uniform optimality results developed in the current work are non-trivial

extensions of those in [26]. In particular, we consider a general class of risk and utility measures while [26] only allows the risk measure to be LFNR. Moreover, time-heterogeneous pre/post- change distributions and non-geometric priors for the change points are allowed in the current work. These extensions require a delicate analysis of a special class of monotone functions and stochastic processes defined over a non-Euclidean space.

The assumptions for establishing the uniform optimality results include monotonicity assumptions on the risk and utility processes and assumptions on the pre- and post- change distributions. We point out that the monotonicity assumptions are made on functions over a special non-Euclidean space

$$\mathcal{S}_o = \bigcup_{k=1}^{K} \{(v_1, \cdots, v_k) : 0 \le v_1 \le \cdots v_k \le 1\} \cup \{\varnothing\}, \quad (29)$$

which contains ordered vectors of different dimensions. Thus, the definition of monotonicity is non-standard.

Specifically, for functions maps $\mathcal{S}_o$ to $\mathbb{R}$, we define two types of monotonicity.

Definition 3 (Entrywise increasing functions)**.** *A function $f : \mathcal{S}_o \to \mathbb{R}$ is* "entrywise increasing"*, if $f(\mathbf{u}) \le f(\mathbf{v})$ for all $m \in \langle K\rangle$, $\mathbf{u} = (u_1, \cdots, u_m), \mathbf{v} = (v_1, \cdots, v_m) \in \mathcal{S}_o$, satisfying $u_j \le v_j$ for $1 \le j \le m$. In addition, a function f is* "entrywise decreasing" *if $-f$ is "entrywise increasing".*

Definition 4 (Appending increasing functions)**.** *A function $f : \mathcal{S}_o \to \mathbb{R}$ is* "appending increasing"*, if for all $m \in \langle K\rangle$, $\mathbf{u} = (u_1, \cdots, u_m) \in \mathcal{S}_o$, $f(u_1, \cdots, u_k) \le f(\mathbf{u})$, for all $k \le m$. In addition, $f(\varnothing) \le f(u_1)$ for $u_1 \in [0, 1]$.*

For each vector $\mathbf{v} = (v_1, \cdots, v_m)$, denote its order statistic by $[\mathbf{v}] = (v_{(1)}, \cdots, v_{(m)})$. That is, $[\mathbf{v}]$ is a permutation of $\mathbf{v}$ satisfying $v_{(1)} \le \cdots \le v_{(m)}$. We can see that if $v_k \in [0, 1]$ for all $k \in \langle K\rangle$, then $[\mathbf{v}] \in \mathcal{S}_o$.

Assumption 3. *There exists a measurable function $\widetilde{r}_t : \mathcal{S}_o \to \mathbb{R}$ such that $r_t(\{W_{k,t}\}_{k\in S_t}, S_t, S_{t+1}) = \widetilde{r}_t\big([\{W_{k,t}\}_{k\in S_{t+1}}]\big)$. In addition, $\widetilde{r}_t$ is entrywise increasing and appending increasing.*

Assumption 4. *There exists a measurable function $\widetilde{u}_t : \mathcal{S}_o \to \mathbb{R}$ such that $u_t(\{W_{k,t}\}_{k\in S_t}, S_t, S_{t+1}) = \widetilde{u}_t\big([\{W_{k,t}\}_{k\in S_{t+1}}]\big)$. In addition, $\widetilde{u}_t$ is entrywise decreasing and appending increasing.*

Assumption 5. *The pre- and post-change distributions $\{p_{k,t}\}_{t\ge1}$ and $\{q_{k,t}\}_{t\ge1}$ are the same for different $k \in \langle K\rangle$. That is, $p_{1,t} = \ldots = p_{K,t}$ and $q_{1,t} = \ldots = q_{K,t}$ for all t.*

Theorem 3. *Under Assumptions 1, 3, 4 and 5, the sequential decision δ_S described in Algorithm 4 is uniformly optimal.*

Proof. The proof is involved that requires monotone coupling for stochastic processes living on the space $\mathcal{S}_o$. It is given in Appendix C. □

We make several remarks on the above theorem. First, under Assumptions 3 and 4, risk and utility measures are symmetric functions $(W_{k,t})_{k\in S_{t+1}}$. These assumptions rule out the cases (e.g., LFDR defined in Example 4) where the risk also depends on $W_{k,t}$ for $k \notin S_{t+1}$, without which the uniform optimal solution may not exist (see Counterexample 1 below). Second, under the monotonicity assumptions that $\widetilde{r}_t$s are entrywise increasing, the risk process tends to be larger if the posterior probability of the change points associated with the selected streams is larger. It is also appending increasing, meaning that the risk tends to be larger if more streams are kept active. Similarly, the utility process tends to be larger if fewer streams are kept active and the posterior probabilities associated with the selected streams are smaller. Third, we require the pre- and post- stream distributions $p_{k,t}$ and $q_{k,t}$ to be identical for different streams. In this case, the process $\{W_{k,t}\}_{t\in\mathbb{Z}_+}$ has identical distribution for different k and contributed in a symmetric way to the risk and utility processes.

For most of applications, it is easy to check Assumptions 1 and 5. In some cases, additional efforts are needed to verify monotonicity assumptions described in Assumptions 3 and 4. Below we provide sufficient conditions for the monotonicity conditions. Note that the risk and utility measures described in Examples 1, 2, 3, 5, and 6 are all symmetric multivariate polynomials of the posterior probabilities. Thus, we restrict the analysis to the polynomial case in the next proposition.

Proposition 4 (Polynomial case)**.** *Let $\widetilde{r} : \mathcal{S}_o \to \mathbb{R}$ be a function in the following form*

$$\widetilde{r}(\mathbf{u}) = \sum_{p=0}^{\infty} \mathbb{1}(\dim(\mathbf{u}) = p) \sum_{k=1}^{p} C_{p,k} \sum_{i_1<i_2<\cdots<i_k} \prod_{j=1}^{k} u_{i_j}. \quad (30)$$

Note that $\widetilde{r}(\varnothing) = 0$. If $\widetilde{r}(\cdot)$ satisfies

$$\widetilde{r}(\mathbf{u}^{i,p-i}) \le \widetilde{r}(\mathbf{u}^{i-1,p-i+1}), \quad \text{for all } i \in \langle p\rangle \text{ and } p \ge 1, \quad (31)$$

where $\langle p\rangle = \{1, \cdots, p\}$ and $\mathbf{u}^{i,p-i}$ denotes the p dimensional vector whose first i elements are 0 and last $p-i$ elements are all 1, then $\widetilde{r}$ is entrywise increasing.

Moreover, if $\widetilde{r}$ satisfies (31) *and*

$$\widetilde{r}(\mathbf{u}^{i,p-i}) \le \widetilde{r}(\mathbf{u}^{i+1,p-i}), \quad \text{for all } i \in \{0, \cdots, p\} \text{ and } p \ge 0, \quad (32)$$

then $\widetilde{r}$ is also appending increasing.

Proof. The proof is given in Appendix C. □

Remark 5. *The inequalities* (31) *and* (32) *are equivalent to*

$$\sum_{k=1}^{p-i} C_{p,k} \binom{p-i-1}{k-1} \ge 0, \quad \text{for } i = 0, \cdots, p-1, \quad (33)$$

$$\sum_{k=1}^{p-i} (C_{p+1,k} - C_{p,k}) \binom{p-i}{k} \ge 0, \text{ for } i = 0, \cdots, p, \quad (34)$$

and all $p \ge 0$. We leave the rigorous justification for the above statements in Appendix C.

Proof. The proof is given in Appendix C. □

Now we apply the uniform optimality result in Theorem 3 to performance measures described in Examples 1, 2, 3 and 5.

Corollary 2. *If $\alpha > 0$, $R_t \in \{LFNR_t, LFWER_t, GLFWER_t, IADD_t\}$, and $U_t = IARL_t$, then under Assumption 5, δ_S is uniformly optimal.*

Proof. The proof is given in Appendix C. □

We point out that LFDR_t described in Example 4 does not satisfy the assumptions made in Theorem 3. Thus, we do not have uniform optimality results for it. Indeed, if $R_t = \text{LFDR}_t$, then the uniformly optimal sequential decision may not exist. A counterexample is given below.

Counterexample 1. *Let $K = 3$, $p_{k,t}(x) = (0.01)^x(0.99)^{1-x}$, $q_{k,t}(x) = (0.99)^x(0.01)^{1-x}$, and $\mathbb{P}(\tau_k = l) = 1/3$ for $k = 1, 2, 3$, $x = 0, 1$, $l = 0, 1, 2$, and $t \geq 1$. That is, the pre- and post- change distributions are Bernoulli distributions with parameters 0.01 and 0.99, respectively, and τ_ks are uniformly distributed over $\{0, 1, 2\}$. We further assume that the tolerance level $\alpha = 0.51$, the risk process $R_t = LFDR_t$ (see Example 4) and the utility process $U_t = -IADD_t$ (see Example 6).*

In this setting, there does not exist a sequential decision achieving the maximum of the expected utility at both times 1 and 2. This implies that there is no uniformly optimal sequential decision. We leave detailed calculation in Appendix C.

We make a connection between the current notion of optimality and that under a non-Bayesian setting.

Remark 6. *We remark that our utility and risk processes are defined under a Bayesian setting, and so do the local and uniform optimalities. Alternatively, one may study the current problem under a non-Bayesian setting, without assuming a prior distribution for the change points. The performance metrics defined in Section II-D can be naturally extended to the non-Bayesian setting. For example, one may replace $LFDR_t$ by a supremum false discovery rate*

$$\sup_{\tau_1,\cdots,\tau_K \in \mathbb{Z}_+ \cup \{0\} \cup \{\infty\}} \mathbb{E}_{\tau_1,\cdots,\tau_K}(FDP_t)$$

which measures the false discovery rate in the worst-case scenario with respect to the change points τ_k. The other performance metrics can be extended similarly. Taking one worst-case metric as the risk, it is possible to come up with sequential procedures that control the risk under a pre-specified threshold at all time points. However, it is a challenge to extend the current notion of optimality to the non-Bayesian setting. A more sensible optimal criterion may be needed to establish optimality results under the non-Bayesian setting. We leave it for future investigation.

C. Implications on Aggregated Risk

Let $\{a_t\}_{t\geq 1}$ be a sequence of non-negative random variables satisfying $\sum_{t=1}^{\infty} a_t = 1$, and $\{b_t\}_{t\geq 1}$ be a sequence of non-negative constants. Consider the following aggregated risk (AR) and aggregated utility (AU),

$$\text{AR} = \mathbb{E}\big(\sum_{t=1}^{\infty} a_t R_t\big) \text{ and } \text{AU} = \mathbb{E}\big(\sum_{t=1}^{\infty} b_t U_t\big). \tag{35}$$

The aggregated risk and utility metrics defined above provide a summary of the performance across time. These types of risk and utility measures are considered in many recent works on multi-stream sequential change detection and hypothesis testing, including [25, 33, 34, 38].

The next proposition shows that if the risk process is controlled at the desired tolerance level at every time point, then the aggregated risk is also controlled at the same level.

Proposition 5. *Let $\delta \in \mathcal{D}_\alpha$ and AR(δ) be the corresponding aggregated risk defined in* (35)*. Then, AR(δ) $\leq \alpha$.*

Proof. According to the definition of $\mathcal{D}_\alpha$, $\delta \in \mathcal{D}_\alpha$ implies $R_t(\delta) \leq \alpha$ a.s. for all $t \in \mathbb{Z}_+$. Thus, $\text{AR}(\delta) = \mathbb{E}\big(\sum_{t=1}^{\infty} a_t R_t\big) \leq \alpha \mathbb{E}\big(\sum_{t=1}^{\infty} a_t\big) = \alpha$, where the last equation is due to $\sum_{t=1}^{\infty} a_t = 1$. □

Note that the reverse statement does not hold. That is, the aggregated risk being controlled does not imply the risk at each time t being controlled.

The next proposition shows that a uniformly optimal sequential decision also maximizes the aggregated utility.

Proposition 6. *Suppose that δ is uniformly optimal in $\mathcal{D}_\alpha$. Then, for the aggregated utility defined in* (35)*,*

$$AU(\delta) = \sup_{\delta' \in \mathcal{D}_\alpha} AU(\delta'),$$

where AU(δ) and AU(δ') denote the aggregated utility associated with δ and δ', respectively.

Proof. For any $\delta' \in \mathcal{D}_\alpha$, $\text{AU}(\delta') = \mathbb{E}\big(\sum_{t=1}^{\infty} b_t U_t(\delta')\big) = \sum_{t=1}^{\infty} b_t \mathbb{E}\big(U_t(\delta')\big) \leq \sum_{t=1}^{\infty} b_t \mathbb{E}\big(U_t(\delta)\big) = \text{AU}(\delta)$, where the second last inequality is due to the assumption that δ is uniformly optimal, and the last equation is obtained based on the definition of aggregated risk. □

Next, we use Propositions 5 and 6 to make a connection between the current results and recent works on the sequential multiple testing and parallel sequential change detection [25, 26, 38].

1) Controlling generalized error rates in multi-stream sequential hypothesis testing: Note that if $\pi_0 + \pi_\infty = 1$ (i.e., change points either occur at the beginning or never occur), the sequential change point detection problem reduces to a sequential multiple hypotheses testing problem, where the goal is to choose between H_0^k and H_1^k for $k = 1, \cdots K$,

$$H_0^k : X_{k,t} \sim p_{k,t} \text{ for all } t \text{ against } H_1^k : X_{k,t} \sim q_{k,t} \text{ for all } t,$$

under a Bayesian setting, where $\mathbb{P}(H_0^k \text{ holds}) = \pi_\infty$ and $\mathbb{P}(H_1^k \text{ holds}) = \pi_0$. In addition, we assume that $X_{k,t}$ are jointly independent.

Let $m \geq 1$. We define the generalized family-wise error rate (GFWER) as

$$\text{GFWER}_m := \mathbb{P}(E_{m,T}), \tag{36}$$

where T is a stopping time and the event $E_{m,t}$ is defined in (8). GFWER_m can be viewed as a generalized family-wise error rate measuring type-II errors in sequential multiple hypotheses testing, which takes a similar form as the generalized type-II

error rate in [32, 34, 38, 39]. Specifically, if we reject H_0^k at time t if and only if $k \in S_{t+1}$. Then,

$$E_{m,t} = \Big\{ \sum_{k=1}^{K} \mathbb{1}(H_1^k \text{ holds and } H_0^k \text{ is chosen at time } t) \geq m \Big\}$$

in the context of multiple hypotheses testing.

The next corollary of Proposition 5 shows that the proposed method δ_{S} controls the GFWER in the perspective of the hypothesis testing problem.

Corollary 3. *Given the tolerance level α, consider the sequential decision δ_S in Algorithm 3 with the risk process $R_t = GLFWER_{m,t}$ defined in Example 2 and any utility process. Then, the generalized family-wise error rate $GFWER_m$ defined in* (36) *is also controlled to be no greater than α for any stopping time T.*

Proof. By comparing the definition of GFWER_m with that of $\mathrm{GLFWER}_{m,t}$ defined in (9), we have $\mathrm{GFWER}_m = \mathbb{E}(R_T) = \mathbb{E}\big(\sum_{t=1}^{\infty} \mathbb{1}(T=t) R_t\big)$. The proof is completed by applying Proposition 5 with $a_t = \mathbb{1}(T=t)$. □

Note that the above Corollary 3 holds for any stopping time T. In particular, if we let T grow to infinity, then Corollary 3 states that the GFWER accumulated over all the time points is controlled to be no greater than α. If we let $T = T_1$ as defined in Remark 1, then different data streams are stopped at the same detection time T_1. In this case, the proposed sequential procedure belongs to the class of sequential multiple testing procedures described in [34].

2) Controlling aggregated false discovery rate: The aggregated false discovery rate (AFDR) is considered in [25],

$$\mathrm{AFDR} = \mathbb{E}\left(\frac{\sum_{t=1}^{\bar{N}-1} \sum_{k=1}^{K} \mathbb{1}\left(\tau_k \geq t, N_k = t\right)}{\Big(\sum_{t=1}^{\bar{N}-1} \sum_{k=1}^{K} \mathbb{1}\left(N_k = t\right)\Big) \vee 1} \right), \quad (37)$$

where $\bar{N}$ is a positive integer that is referred to as a 'deadline'. The next proposition states that any decision that controls LFDR_t at every time also controls AFDR asymptotically.

Proposition 7. *Let $R_t = LFDR_t$ and $\delta \in \mathcal{D}_\alpha$. Assume that there exist a sequence of constants $\{C_t\}_{t\geq 1}$ and a sequence of random variables $\{A_t\}_{t\geq 1}$ such that $K^{-1}\sum_{k=1}^{K} \mathbb{1}\left(N_k = t\right)$ converges to C_t in probability and FDP_t converges to A_t in probability for all $t \geq 1$ as K grows to infinity. Then, $\lim_{K\to\infty} AFDR(\delta) \leq \alpha$. That is, AFDR is controlled to be no greater than α asymptotically.*

Proof. The proof is given in Appendix C. □

3) Maximizing total average run length: Let the total average run length (TARL) be

$$\mathrm{TARL} = \sum_{k=1}^{K} (\tau_k \wedge N_k), \quad (38)$$

where N_k is defined in (15). TARL aggregates IRL_t across different time points, and can be viewed as an extension of the classic ARL2FA to multi-stream problems. The next corollary of Proposition 6 shows that the proposed method also maximizes TARL under certain conditions.

Corollary 4. *Under Assumptions 5, and $R_t \in \{LFNR_t, LFWER_t, GLFWER_t, IADD_t\}$,*

$$\mathbb{E}(TARL(\delta_S)) = \sup_{\delta\in\mathcal{D}_\alpha} \mathbb{E}(TARL(\delta)),$$

where δ_S is obtained from Algorithm 4 by letting $U_t = IARL_t$, and $TARL(\delta_S)$ and $TARL(\delta)$ denote the the total average run length (TARL) of the decision δ_S and δ, respectively.

Proof. According to the definition of IRL_t in Example 5, $\mathrm{TARL} = \sum_{s=0}^{\infty} \mathrm{IRL}_s$. This implies $\mathbb{E}(\mathrm{TARL}) = \mathbb{E}(\sum_{s=0}^{\infty} \mathbb{E}(\mathrm{IRL}_s|\mathcal{F}_s)) = \mathbb{E}(\sum_{s=0}^{\infty} \mathrm{IARL}_s)$. According to Corollary 2, δ_{S} is uniformly optimal when $U_t = \mathrm{IARL}_t$. We complete the proof by letting $b_t = 1$ for all t, $U_t = \mathrm{IARL}_t$ and $\mathrm{AU}(\delta) = \mathrm{TARL}(\delta)$ in Proposition 6. □

V. A Simulation Study

In this section, we study the performance of the proposed sequential decision δ_{S} defined in Algorithm 4 through three simulation studies. Throughout the simulation studies, we let the tolerance level $\alpha = 0.1$.

A. Study I: Controlling LFDR

In this simulation study, we choose $R_t = \mathrm{LFDR}_t$ and $U_t = -\mathrm{IADD}_t$ in Algorithm 4. We compare the performance of the proposed method with the MD-FDR method proposed in [25].

Let $\pi_t = (1-\pi_\infty)\theta(1-\theta)^t$ for $t \geq 0$, $p_{k,t}(x) = (2\pi)^{-1/2} e^{-\frac{(x-\mu_0)^2}{2}}$ and $q_{k,t}(x) = (2\pi)^{-1/2} e^{-\frac{(x-\mu_1)^2}{2}}$ for all k and t. We set $\pi_\infty = 0.2$, $\theta = 0.1$, $\mu_0 = 0$ and $\mu_1 = 1$. That is, we set the pre- and post-change probability distributions to be the Gaussian distributions $\mathcal{N}(0,1)$ and $\mathcal{N}(1,1)$, respectively, and set the prior distribution for the change point τ_k to be a mixture of a point mass at infinity and a geometric distribution.

We assess and compare the performance of two sequential decisions. The first sequential decision is the the proposed method δ_{S} described in Algorithm 4 with $R_t = \mathrm{LFDR}_t$ (defined in Example 4) and $U_t = -\mathrm{IADD}_t$ (defined in Example 6). With this choice of R_t and U_t, line 4 in Algorithm 3 can be simplified as

$$n^* = \arg \min_{n=0,1,\ldots,|S_t|} \{n : \gamma_n \leq \alpha\}.$$

The other sequential decision is the MD-FDR method developed in [25]. Following the MD-FDR method, the risk measure AFDR defined in (37) is guaranteed to be no greater than the tolerance level α.

We first compare the proposed method with the MD-FDR method in terms of their FDP_t (defined in (13)) and IDD_t (defined in (21)) for fixed $K = 500$ with 1000 independent Monte Carlo simulations. The averaged FDP_t and IDD_t across the 1000 replications are plotted in Figures 3 and 4 as functions of t. According to Figure 3, the averaged FDP_t of both methods are below 0.1 for all t with a trend of first increasing and then decreasing as t increases. The FDP_t of the proposed method has a plateau near 0.1 for $t \in [5, 20]$. In addition, the FDP_t of the proposed method is larger than that of the MD-FDR method, which suggests that the proposed method is less conservative while still controlled under the target tolerance

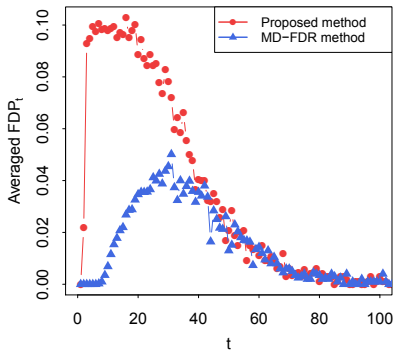

Fig. 3: FDP$_t$ averaged over 1000 Monte Carlo simulations at $K = 500$

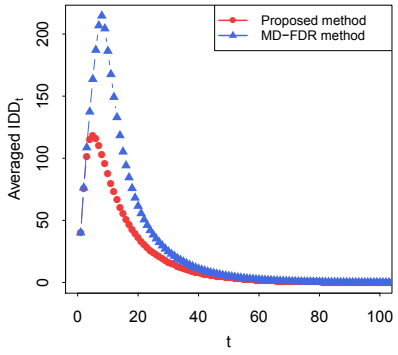

Fig. 4: IDD$_t$ averaged over 1000 Monte Carlo simulations at $K = 500$

level. Figure 4 compares the averaged IDD$_t$ of the proposed method and the MD-FDR method for different t. It displays that, for both methods, IDD$_t$ first increases and then decreases as t increases. The proposed method has a lower averaged IDD$_t$ than the MD-FDR method for all t, indicating a smaller detection delay.

Next, we compare the two methods in terms of aggregated performance measures. In particular, we consider the aggregated risk AFDR defined in (37), where we set the deadline parameter $\bar{N} = 500$. For aggregated utility, we consider the the total average detection delay (TADD), defined as

$$\text{TADD} = \mathbb{E}\Big(\sum_{s=0}^{\bar{N}-1} \text{IDD}_s\Big) = \mathbb{E}\Big(\sum_{s=0}^{\bar{N}-1} \text{IADD}_s\Big), \tag{39}$$

where IDD$_s$ is defined in (23). Then, we let the aggregated utility be AU $= -$TADD. A higher utility, which corresponds to a lower TADD, reflects a quicker detection of the changes.

Tables I and II compare the two methods in terms of their aggregated risk AFDR and the aggregated utility TADD, respectively, which are estimated based on a Monte Carlo simulation with 1000 replications. From Table I, we can see that both the proposed method and MD-FDR method control AFDR below the tolerance level $\alpha = 0.1$, while the MD-FDR method is more conservative. We also note that as K grows larger, AFDR of the proposed method is approaching $\alpha = 0.1$. From Table II, we can see that the TADD of the proposed method is significantly less than that of the MD-FDR method, indicating that the proposed method detects changes faster than the MD-FDR method, when the AFDR of both methods are controlled at the same level. An interesting observation is that TADD of both methods scale with K as K grows. That is, TADD$/K$ seems to converge to a constant as K grows large. Specifically, for the proposed method, TADD$/K$ is around 3.9. For the MD-FDR method, TADD$/K$ is around 6.5 for large K.

Overall, these results suggests that the proposed method is less conservative and adapts better to the tolerance level than the MD-FDR method.

K	Proposed method	MD-FDR method
10	7.0×10^{-2} (3×10^{-3})	2.8×10^{-2} (2×10^{-3})
100	8.6×10^{-2} (9×10^{-4})	2.4×10^{-2} (5×10^{-4})
200	9.2×10^{-2} (7×10^{-4})	2.4×10^{-2} (4×10^{-4})
500	9.6×10^{-2} (5×10^{-4})	2.3×10^{-2} (2×10^{-4})
1000	9.8×10^{-2} (3×10^{-4})	2.4×10^{-2} (2×10^{-4})

TABLE I: Estimated AFDR (standard deviation in parenthesis)

K	Proposed method	MD-FDR method
10	45.8 (0.5)	61.4 (0.5)
100	413.8 (1.3)	650 (1.7)
200	799.8 (1.9)	1304.1 (2.4)
500	1964.9 (3.0)	3264 (3.7)
1000	3891.4 (4.0)	6535.3 (5.2)

TABLE II: Estimated TADD (standard deviation in parenthesis)

B. Study II: Time-heterogeneous distributions

We choose $R_t = \text{LFDR}_t$ and $U_t = -\text{IADD}_t$ and let $p_{k,t}(x) = (2\pi\sigma_t^2)^{-1/2} e^{-\frac{x^2}{2\sigma_t^2}}$ and $q_{k,t}(x) = (2\pi\sigma_t^2)^{-1/2} e^{-\frac{(x-1)^2}{2\sigma_t^2}}$ for all k and t. That is, the pre- and post- change distributions at the kth stream for time t are $N(0, \sigma_t^2)$ and $N(1, \sigma_t^2)$, respectively. This simulation study has a similar setting as Study I in Section V-A except that the pre-/post- change distribution have time-heterogeneous standard deviation. We generate $\{\sigma_t\}_{t\geq 1}$ independently from a uniform distribution over $\{2/\sqrt{1}, \cdots, 2/\sqrt{10}\}$.

Figure 5 and 6 present the averaged FDP$_t$ and IDD$_t$ over 1000 Monte Carlo replications, respectively. As we can see in Figure 5, the averaged FDP$_t$ is controlled at the level $\alpha = 0.1$ and fluctuates for $t < 50$. Figure 6 displays a first increasing then decreasing trend as t increases, which is similar to the trend in Figure 4.

C. Study III: Controlling LFNR

In this simulation study, we let $p_{k,t}(x) = (2\pi)^{-1/2} e^{-\frac{x^2}{2}}$ and $q_{k,t}(x) = (2\pi)^{-1/2} e^{-\frac{(x-1)^2}{2}}$ for all k and t and $\pi_\infty = 0.2$, and $\pi_t = (t+2)!/(2\,t!) \cdot 0.8 \cdot (0.1)^3 (0.9)^t$ for $t \geq 0$. In addition, we set $K = 100$ and consider the risk process $R_t = \text{LFNR}_t$ (see Example 3) and the utility process $U_t = \text{IRL}_t$ (see Example 5).

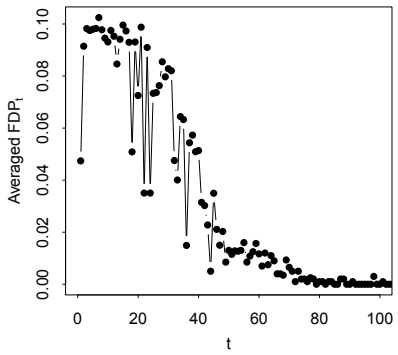

Fig. 5: FDP$_t$ averaged over 1000 Monte Carlo simulations at $K = 500$

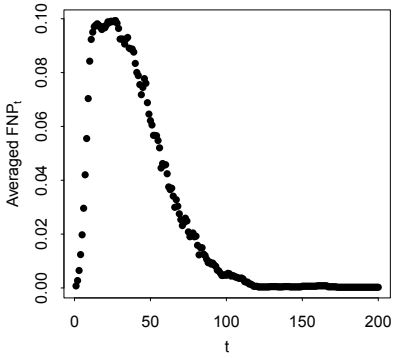

Fig. 7: FNP$_t$ averaged over 1000 Monte Carlo simulations at $K = 100$

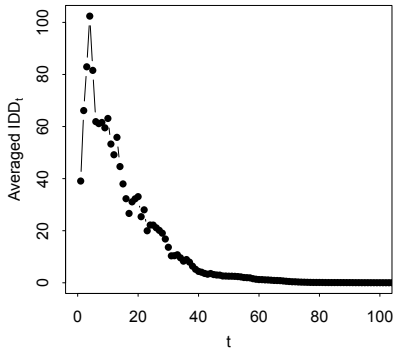

Fig. 6: IDD$_t$ averaged over 1000 Monte Carlo simulations at $K = 500$

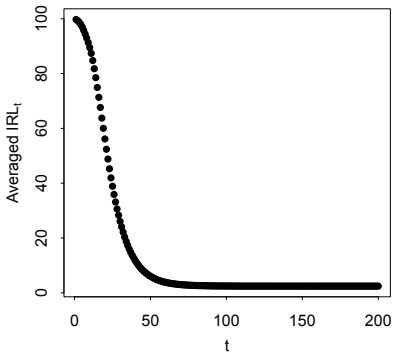

Fig. 8: IRL$_t$ averaged over 1000 Monte Carlo simulations at $K = 100$

We plot the averaged risk measure FNP$_t$ and the averaged utility measure of the proposed method δ_S defined in Algorithm 4 in Figures 7 and 8 based on a Monte Carlo simulation with 1000 replications. From Figure 7, we can see that the averaged FNP$_t$ is below 0.1 with a peak at around $t = 27$, which is consistent with Proposition 3. From Figure 8, we can see that IRL$_t$ gradually decreases to 0 as t increases.

VI. A Case Study: Multi-Channel Spectrum Sensing in Cognitive Radios

In this section, we conduct a case study on a multi-channel spectrum sensing problem for cognitive radios, following the settings described in [25]. Cognitive radios are radios that can dynamically and automatically adjust their operational parameters according to the environment so that the spectrum is utilized more efficiently [47, 48]. To make the most out of a spectrum, a cognitive user is allowed to use the idle spectrum band when the primary user is not transmitting. However, when the primary user starts transmission, the cognitive user should detect the change and vacate the spectrum band as soon as possible. The detection of the transmission of the primary user can be formulated as a sequential change detection problem, where the transmission time corresponds to the change point [25, 29].

We consider a multi-channel spectrum sensing problem for cognitive radios, where there are K independent frequency channels assigned to K independent primary users. The cognitive users monitor the spectrum bands and collect signal samples sequentially. The distribution of the signals will change when a primary user starts transmission. As soon as the change is detected, the cognitive user vacates the spectrum band, so that the primary user can use it without interference. Here, each channel corresponds to a data stream, and the time that a primary user starts transmission corresponds to a change point in that data stream. Our goal is to have a sequential decision that can detect the transmission of the primary user at each channel quickly to reduce the interference, while controlling the false discovery rate, which corresponds to the expected proportion of unoccupied channels among the detected ones.

Specifically, we assume that $X_{k,t}$ is the signal collected from the kth cognitive user at time t, τ_k is the time when the k-th primary user starts transmission, and $X_{k,t}$s and τ_ks follow the change point model described in (1) and (2). For the change point τ_k, we further assume that

$$\pi_t = (1 - \pi_\infty)\theta(1 - \theta)^t$$

with $\pi_\infty = 0.1$ and $\theta = 0.05$. That is, τ_k follows a mixture distribution of a point mass at infinity and a geometric distribution.

For the pre- and post- change distributions, we assume

$$X_{k,t} = \begin{cases} Y_{k,t}, & \text{if } t \le \tau_k \\ Y_{k,t} + Z_{k,t}, & \text{if } t > \tau_k \end{cases},$$

where $Y_{k,t}$ denotes a Gaussian white noise and $Z_{k,t}$ denotes the faded received primary radio signal at the cognitive user's end. We further assume that $Y_{k,t} \sim \mathcal{CN}(0, \sigma^2)$, $Z_{k,t} \sim \mathcal{CN}(0, \lambda_k)$, and $Y_{k,t}$s and $Z_{k,t}$s are independent, where $\mathcal{CN}(0, \sigma^2)$ and $\mathcal{CN}(0, \lambda_k)$ denote the circularly-symmetric complex Gaussian distributions with mean 0 and the complex variance σ^2 and λ_k, respectively. Note that a complex random variable has a circularly-symmetric complex Gaussian distribution with a variance σ^2 if its real and imaginary parts are independent and identically distributed univariate Gaussian random variables with the mean zero and the variance $\sigma^2/2$.

Under this model, $X_{k,t}$ has the distribution

$$X_{k,t} \sim \begin{cases} \mathcal{CN}(0, \sigma^2), & \text{if } t \le \tau_k \\ \mathcal{CN}(0, \sigma^2 + \lambda_k), & \text{if } t > \tau_k \end{cases}.$$

Notice that in this setting, the streams share the same pre-change distribution, but have different post-change distributions characterized by their different variances. The above distribution assumptions are commonly adopted in the literature [25, 29].

In this case study, we assume $\sigma^2 = 2$ and sample independent λ_ks from a uniform distribution over $[1, 2]$. We then treat λ_k as known parameters. Here, we sample λ_k from an interval to mimic the practical situation where the signals sent by the primary users may experience channel attenuation at the cognitive user's end, which results in a range of variance-distinct post-change signals.

Let the tolerance level $\alpha = 0.1$. We compare the performance of the proposed sequential decision following Algorithm 3 (with $R_t = \text{LFDR}_t$ and $U_t = -\text{IADD}_t$) and the MD-FDR method proposed in [25]. We also consider the aggregated risks AFDR (defined in (37)) and the aggregated utility TADD (defined in (39)).

Figures 9 and 10 show the averaged FDP_t and IDD_t for different t based on a Monte Carlo simulation with 1000 replications. We see that FNP_t of both methods are below 0.1 with a peak at around $t = 12$ and $t = 42$, respectively. According to Figure 9, the averaged FNP_t of the MD-FDR method appears to be smaller than that of the proposed method for time $t < 50$, while both of them decline at a similar rate after time $t = 50$. For larger t, the averaged FNP_t of both methods are close to zero. According to Figure 10, the averaged IDD_t of the MD-FDR method is larger than that of the proposed method for all t, which suggests that the proposed method detects changes more quickly than that of the MD-FDR method.

Tables III and IV show the AFDR and TADD for both methods for $K \in \{10, 100, 200, 500, 1000\}$. According to the tables, the AFDR of both methods are controlled to be less than $\alpha = 0.1$, with the AFDR of the MD-FDR method smaller than that of the proposed method for all K. This indicates that the proposed method is less conservative in controlling FDR-type of risks, when compared with the MD-FDR method.

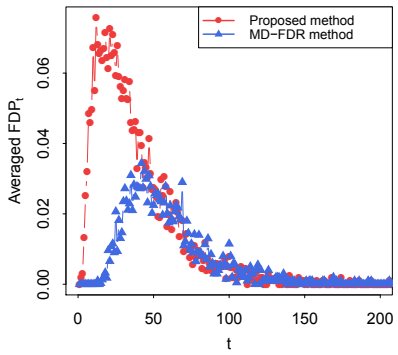

Fig. 9: FDP_t averaged over 1000 Monte Carlo simulations at $K = 100$ in Case Study

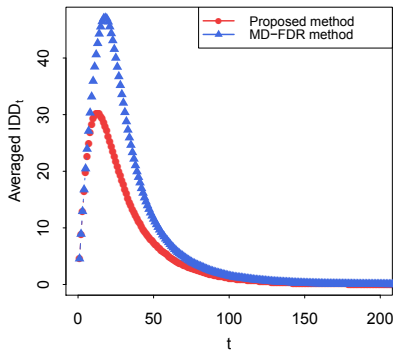

Fig. 10: IDD_t averaged over 1000 Monte Carlo simulations at $K = 100$ in Case Study

Moreover, the proposed method has a much smaller TADD that that of the MD-FDR method for all K, indicating that the proposed method has a smaller detection delay.

VII. Conclusions

The parallel sequential change detection problem is widely encountered in the analysis of large-scale real-time streaming

K	Proposed method	MD-FDR method
10	6.7×10^{-2} (3×10^{-3})	3.2×10^{-2} (2×10^{-3})
100	8.5×10^{-2} (9×10^{-4})	2.9×10^{-2} (6×10^{-4})
200	9.0×10^{-2} (7×10^{-4})	2.9×10^{-2} (4×10^{-4})
500	9.5×10^{-2} (4×10^{-4})	2.8×10^{-2} (2×10^{-4})
1000	9.7×10^{-2} (3×10^{-4})	2.8×10^{-2} (2×10^{-4})

TABLE III: Estimated AFDR in case study (standard deviation in parenthesis)

K	Proposed method	MD-FDR method
10	122.1 (1.2)	162 (1.4)
100	1115.8 (3.7)	1708.5 (4.6)
200	2178.2 (5.1)	3434.8 (6.6)
500	5293.4 (8.1)	8609.4 (10.4)
1000	10460.1 (11.3)	17246.7 (14.8)

TABLE IV: Estimated TADD in case study (standard deviation in parenthesis)

data. This study introduces a general decision theory framework for this problem, covering many compound performance metrics. It further proposes a sequential procedure under this general framework and proves its optimal properties under reasonable conditions. Simulation and case studies evaluate the performance of the proposed method and compare it with the method proposed in [25]. The results support the theoretical developments and also show that the proposed method outperforms in our simulation studies and case study.

The current study can be extended in several directions. First, the current parallel sequential change detection framework may be extended to account for multiple types of decisions, including alerting the changes without stopping the streams and diagnosis of the post-change distribution upon stopping, which is also known as the sequential change diagnosis [49, 50]. We may also consider transient changes (i.e., changes that can appear and then disappear; see [16, 36, 37]) and allow stream reactivation to be one possible decision. Second, in many applications, the post-change distribution of data is challenging to obtain. Also, it is sometimes difficult to specify a prior distribution for the change points. In these cases, it is desirable to formulate the problem in a non-Bayesian decision theory framework, and develop a flexible parallel sequential change detection method that is robust for unknown post-change distributions under this framework. One possible direction is to study the worst-case scenario, by developing a minimax formulation under a non-Bayesian setting. Third, we assume that the change points are independent for different data streams. For some applications, it is reasonable to extend the methods to the case where the change points are dependent. For example, the change points may be driven by the same event [19] or propagated by each other [51].

REFERENCES

[1] W. A. Shewhart, *Economic control of quality of manufactured product*. Oxford, England: Van Nostrand, 1931.

[2] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, 1954.

[3] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory of Probability & Its Applications*, vol. 8, pp. 22–46, 1963.

[4] S. Roberts, "A comparison of some control chart procedures," *Technometrics*, vol. 8, pp. 411–430, 1966.

[5] G. Lorden, "Procedures for reacting to a change in distribution," *The Annals of Mathematical Statistics*, vol. 42, pp. 1897–1908, 1971.

[6] M. Pollak, "Optimal detection of a change in distribution," *The Annals of Statistics*, vol. 13, pp. 206–227, 1985.

[7] ——, "Average run lengths of an optimal method of detecting a change in distribution," *The Annals of Statistics*, vol. 15, pp. 749–779, 1987.

[8] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *The Annals of Statistics*, vol. 14, pp. 1379–1387, 1986.

[9] T. L. Lai, "Sequential analysis: Some classical problems and new challenges," *Statistica Sinica*, vol. 11, pp. 303–351, 2001.

[10] H. V. Poor and O. Hadjiliadis, *Quickest detection*. Cambridge, England: Cambridge University Press, 2008.

[11] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: Theory and application*. Upper Saddle River, NJ: Prentice Hall, 1993.

[12] A. G. Tartakovsky, B. L. Rozovskii, R. B. Blažek, and H. Kim, "Detection of intrusions in information systems by sequential change-point methods," *Statistical methodology*, vol. 3, no. 3, pp. 252–293, 2006.

[13] A. G. Tartakovsky and V. V. Veeravalli, "Asymptotically optimal quickest change detection in distributed sensor systems," *Sequential Analysis*, vol. 27, pp. 441–475, 2008.

[14] S. Zarrin and T. J. Lim, "Cooperative quickest spectrum sensing in cognitive radios with unknown parameters," in *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*. IEEE, 2009, pp. 1–6.

[15] Y. Mei, "Quickest detection in censoring sensor networks," in *2011 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2011, pp. 2148–2152.

[16] D. Egea-Roca, G. Seco-Granados, J. A. López-Salcedo, and S. Kim, "Space-time cusum for distributed quickest detection using randomly spaced sensors along a path," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 2443–2447.

[17] H. P. Chan, "Optimal sequential detection in multi-stream data," *The Annals of Statistics*, vol. 45, pp. 2736–2763, 2017.

[18] Y. Mei, "Efficient scalable schemes for monitoring a large number of data streams," *Biometrika*, vol. 97, pp. 419–433, 2010.

[19] Y. Xie and D. Siegmund, "Sequential multi-sensor change-point detection," *The Annals of Statistics*, vol. 41, pp. 670–692, 2013.

[20] G. Fellouris and G. Sokolov, "Second-order asymptotic optimality in multisensor sequential change detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3662–3675, 2016.

[21] H. Chen and N. Zhang, "Graph-based change-point detection," *The Annals of Statistics*, vol. 43, pp. 139–176, 2015.

[22] H. Chen, "Sequential change-point detection based on nearest neighbors," *The Annals of Statistics*, vol. 47, pp. 1381–1407, 2019.

[23] V. Raghavan and V. V. Veeravalli, "Quickest detection of a change process across a sensor array," in *2008 11th International Conference on Information Fusion*. IEEE, 2008, pp. 1–8.

[24] D. Li, L. Lai, and S. Cui, "Quickest change detection and identification across a sensor array," in *2013 IEEE Global Conference on Signal and Information Processing*. IEEE, 2013, pp. 145–148.

[25] J. Chen, W. Zhang, and H. V. Poor, "A false discovery rate oriented approach to parallel sequential change detection problems," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1823–1836, 2020.

[26] Y. Chen and X. Li, "Compound sequential change-point

detection in parallel data streams," *Statistica Sinica*, vol. 33, no. 1, 2023.

[27] Y. Chen, Y.-H. Lee, and X. Li, "Item pool quality control in educational testing: Change point model, compound risk, and sequential detection," *Journal of Educational and Behavioral Statistics*, vol. 47, no. 3, pp. 322–352, 2022.

[28] J. Chen, W. Zhang, and H. V. Poor, "Non-Bayesian multiple change-point detection controlling false discovery rate," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 31–35.

[29] L. Lai, Y. Fan, and H. V. Poor, "Quickest detection in cognitive radio: A sequential change detection framework," in *IEEE GLOBECOM 2008-2008 IEEE Global Telecommunications Conference*. IEEE, 2008, pp. 1–5.

[30] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: a realistic modeling and a novel learning strategy," *IEEE transactions on neural networks and learning systems*, vol. 29, pp. 3784–3797, 2017.

[31] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, pp. 289–300, 1995.

[32] J. Bartroff and J. Song, "Sequential tests of multiple hypotheses controlling type I and II familywise error rates," *Journal of statistical planning and inference*, vol. 153, pp. 100–114, 2014.

[33] ——, "Sequential tests of multiple hypotheses controlling false discovery and nondiscovery rates," *Sequential analysis*, vol. 39, no. 1, pp. 65–91, 2020.

[34] Y. Song and G. Fellouris, "Sequential multiple testing with generalized error control: An asymptotic optimality theory," *The Annals of Statistics*, vol. 47, pp. 1776–1803, 2019.

[35] D. Egea-Roca, J. A. López-Salcedo, G. Seco-Granados, and H. V. Poor, "Performance bounds for finite moving average tests in transient change detection," *IEEE Transactions on Signal Processing*, vol. 66, no. 6, pp. 1594–1606, 2018.

[36] D. Egea-Roca, B. K. Guépié, J. A. López-Salcedo, G. Seco-Granados, and I. V. Nikiforov, "Two strategies in transient change detection," *IEEE Transactions on Signal Processing*, vol. 70, pp. 1418–1433, 2022.

[37] B. K. Guépié, L. Fillatre, and I. Nikiforov, "Sequential detection of transient changes," *Sequential Analysis*, vol. 31, no. 4, pp. 528–547, 2012.

[38] Y. Song and G. Fellouris, "Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals," *Electronic Journal of Statistics*, vol. 11, no. 1, pp. 338–363, 2017.

[39] J. Bartroff, "Multiple hypothesis tests controlling generalized error rates for sequential data," *Statistica Sinica*, vol. 28, pp. 363–398, 2018.

[40] J. Bartroff, M. Finkelman, and T. L. Lai, "Modern sequential analysis and its applications to computerized adaptive testing," *Psychometrika*, vol. 73, no. 3, pp. 473–486, 2008.

[41] B. Efron, "Large-scale simultaneous hypothesis testing: The choice of a null hypothesis," *Journal of the American Statistical Association*, vol. 99, pp. 96–104, 2004.

[42] B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher, "Empirical Bayes analysis of a microarray experiment," *Journal of the American Statistical Association*, vol. 96, pp. 1151–1160, 2001.

[43] B. Efron, *Large-scale inference: Empirical Bayes methods for estimation, testing, and prediction*. Cambridge, England: Cambridge University Press, 2012.

[44] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential analysis: Hypothesis testing and changepoint detection*. CRC Press, 2014.

[45] A. S. Polunchenko and A. G. Tartakovsky, "State-of-the-art in sequential change-point detection," *Methodology and Computing in Applied Probability*, vol. 14, pp. 649–684, 2012.

[46] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[47] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE journal on selected areas in communications*, vol. 23, no. 2, pp. 201–220, 2005.

[48] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE personal communications*, vol. 6, no. 4, pp. 13–18, 1999.

[49] X. Ma, L. Lai, and S. Cui, "Two-stage Bayesian sequential change diagnosis," *IEEE Transactions on Information Theory*, 2020.

[50] S. Dayanik, C. Goulding, and H. V. Poor, "Bayesian sequential change diagnosis," *Mathematics of Operations Research*, vol. 33, no. 2, pp. 475–496, 2008.

[51] R. Zhang, Y. Xie, R. Yao, and F. Qiu, "Online detection of cascading change-points," *arXiv preprint arXiv:1911.05610*, 2021.