COMPOUND SEQUENTIAL CHANGE-POINT DETECTION IN PARALLEL DATA STREAMS

Yunxiao Chen and Xiaoou Li

London School of Economics and Political Science and University of Minnesota

Abstract: We consider sequential change-point detection in parallel data streams, where each stream has its own change point. Once a change is detected in a data stream, this stream is deactivated permanently. The goal is to maximize the normal operation of the pre-change streams, while controlling the proportion of the post-change streams among the active streams at all time points. Using a Bayesian formulation, we develop a compound decision framework for this problem. A procedure is proposed that is uniformly optimal among all sequential procedures that control the expected proportion of post-change streams at all time points. We also investigate the asymptotic behavior of the proposed method when the number of data streams grows large. Numerical examples are provided to illustrate the use and performance of the proposed method.

Key words and phrases: Change-point detection, compound decision, false non-discovery rate, large-scale inference, sequential analysis.

1. Introduction

Sequential change-point detection, which dates back to the pioneering work of Page (1954, 1955), focuses on the early detection of distributional changes in sequentially observed data. Methods for sequential change-point detection have been applied widely in various fields, including engineering, education, medical diagnostics, and finance, among others, where a change point typically corresponds to a deviation of a data stream from its "normal" state. The classical methods for sequential change-point detection focus on detecting one or multiple changes in a single data stream (Lorden (1971); Page (1954); Roberts (1966); Shewhart (1931); Shiryaev (1963)). With advances in information technology, large-scale streaming data have become more common, and many recent developments tend to focus on change-point detection in multiple data streams (Chan (2017); Chen and Zhang (2015); Chen (2019); Mei (2010); Xie and Siegmund (2013); Fellouris and Sokolov (2016)).

Corresponding author: Xiaoou Li, School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA. E-mail: lixx1766@umn.edu.

We consider sequential change-point detection in multiple parallel data streams, where each stream has its own change point. Once a change is detected in a data stream, this stream is deactivated permanently and its data are no longer collected. The goal is to maximize the normal operation of the pre-change streams, while controlling the proportion of the post-change streams among the active ones at all time points. This problem is commonly encountered in the real world. One such example is the monitoring of an item pool in standardized educational testing (Choe, Zhang and Chang (2018); Cizek and Wollack (2016); van der Linden and Lewis (2015); Veerkamp and Glas (2000)). In this application, each item corresponds to a data stream. Data are collected sequentially from the items' use in test administrations over time. A change point occurs when an item is leaked to future test takers. The goal is to detect and remove changed items in a sequential fashion from an item pool that consists of hundreds or even thousands of items. Once a change point is detected for an item, test administrators would like to remove it from the item pool to ensure test fairness. On the other hand, it is important to maximize the usage of each item before its leakage, owing to the cost of developing new items. There are many other applications, including multichannel spectrum sensing (Chen, Zhang and Poor (2020)) and credit card fraud detection (Dal Pozzolo et al. (2018)).

Despite its wide application, this type of problem is rarely explored in the literature on multi-stream sequential change-point detection. One exception is Chen, Zhang and Poor (2020), who address a similar problem by proposing a sequential version of the Benjamini–Hochberg FDR control procedure (Benjamini and Hochberg (1995)) for detecting and deactivating post-change data streams. However, no optimality theory is provided in Chen, Zhang and Poor (2020). The challenges of developing optimality theory lie in the compound nature of the FDR-type risk measure and the stochastic control component, owing to the deactivation of data streams. In this study, we formulate the problem under a compound decision theory framework and propose an optimal change-point detection procedure. Our contributions are summarized below.

First, we formulate this problem under a Bayesian sequential change-point detection setting, which generalizes the classical Bayesian setting for single-stream change-point detection (Lai (2001)) to parallel streams. Moreover, we introduce new performance metrics, borrowing ideas from compound decision theory for multiple hypothesis testing (Benjamini and Hochberg (1995); Brown and Greenshtein (2009); Cai, Sun and Wang (2019); Efron and Hastie (2016); Efron (2019); Genovese and Wasserman (2002); Sun and Cai (2007); Zhang (2003)). Specifically, we propose controlling a local false non-discovery rate (LFNR) at each

time point, defined as the expected proportion of post-change streams among the active ones under the current posterior measure. This metric adapts the false non-discovery rate for multiple hypothesis testing (Genovese and Wasserman (2002)) to parallel-stream change-point detection. In addition, we introduce a compound stream utilization measure that is closely related to the classical notion of an average run length (Lorden (1971)). A compound sequential detection procedure involves a trade-off between the LFNR and stream utilization at each time point. Our objective is to maximize stream utilization, while controlling the LFNR to be below a prespecified threshold all the time. Compared with classical performance metrics for individual streams, the proposed metrics better evaluate the risk of a sequential decision at an aggregate level, and thus are more suitable for large-scale streaming data.

Second, we propose a sequential decision procedure that can control the LFNR under any prespecified threshold. Under a class of Bayesian change-point models, we show that this procedure is uniformly optimal among all sequential detection procedures under the same LFNR constraint, in the sense that the proposed procedure has the highest stream utilization at any time. Note that this is a non-asymptotic result that applies to any finite number of data streams. This implies that this compound change-point detection problem is special, in the sense that a myopic decision rule that maximizes the next-step stream utilization under the LFNR constraint is also uniformly optimal throughout time. This phenomenon does not hold in general for stochastic control problems (Howard (1960)). The proof of the uniform optimality result is nontrivial. As such, we develop new mathematical tools, including a monotone coupling over a partially ordered space (Thorisson (2000)) for comparing stochastic processes with different dimensions, owing to the deactivation step. In addition to the nonasymptotic optimality, we establish asymptotic theory to characterize the performance of the proposed method when the number of data streams grows large.

The current setting is substantially different from most of existing works on multi-stream sequential change-point detection, including Mei (2010), Xie and Siegmund (2013), Chen and Zhang (2015), Chan (2017), Chen (2019), and Chen, Wang and Samworth (2022). These works consider the detection of a *single* change point, after which all (or part) of the data streams deviate from their initial states. On the other hand, the current work detects multiple change points in parallel streams. Because the dimension of the action space at each time point grows exponentially with the number of data streams, the current problem tends to be computationally and theoretically more challenging.

2. Compound Sequential Change-point Detection

2.1. Bayesian change-point model for parallel streams

Consider, in total, K parallel data streams. For each $k = 1, \ldots, K$, the observations from the kth stream are $X_{k,t}$, $t = 1, 2, \ldots$. Each stream k is associated with a change point, denoted by τ_k, which takes a value in $\{0\} \cup \{\infty\} \cup \mathbb{Z}_+$. The random vector $(\tau_1, \ldots, \tau_K)$ is assumed to follow a known prior distribution. Given the change points, the data points $X_{k,t}$ from the kth stream at time t are independent for different t and k. It is further assumed that the pre- and post-change distributions of $X_{k,t}$ have the density functions $p_{k,t}(\cdot)$ and $q_{k,t}(\cdot)$, respectively, with respect to some baseline measure μ. That is, $X_{k,t}$ has the following conditional density functions:

$$X_{k,t} \mid \tau_1, \ldots, \tau_K, \{X_{l,s}; 1 \le l \le K, 1 \le s \le t-1\} \sim \begin{cases} p_{k,t} & \text{if } t \le \tau_k, \\ q_{k,t} & \text{if } t > \tau_k. \end{cases} \quad (2.1)$$

Remark 1. We assume that the prior distribution for the change points and the pre- and post-change distributions are known, which is a standard assumption in single-stream Bayesian sequential change detection (e.g., (Shiryaev (1963)). Similar assumptions are adopted in recent developments on multi-stream sequential multiple testing (Song and Fellouris (2019)) and multi-stream sequential change detection (Chen, Zhang and Poor (2020)).

When these distributions are unknown, the current results provide the oracle procedure and theoretical guidance for the development and analysis of sequential change detection procedures. In addition, the proposed procedure can be extended to handle the unknown distribution scenario by using an empirical Bayes approach (see e.g., Efron (2008); Jiang and Zhang (2009); Robbins (1956); Zhang (2003)). Alternatively, we can run the proposed procedure under the worst case model, if such a model can be specified using domain knowledge. This procedure will preserve some properties of the oracle procedure, when the change point model enjoys certain stochastic ordering properties.

Equation (2.1) provides a general model for change points in parallel data streams. It contains some commonly used models as special cases. We provide two examples below.

Example 1 (A partially dependent model). Let τ_0 be a nonnegative random variable, and let $\tau_1, \ldots, \tau_K$ be independent and identically distributed (i.i.d.) conditional on τ_0, with conditional distribution $\Pr(\tau_k = m | \tau_0 = m) = \eta$ and $\Pr(\tau_k = \infty | \tau_0 = m) = 1 - \eta$, for $m = 0, 1, \ldots$ and some parameter $\eta \in [0, 1]$.

This model describes the situation where there is a single change point for all of the data streams. After the change point, all or part of the data streams have a distributional change. If we further let $p_{k,t}$ be the density function of the standard normal distribution $N(0,1)$, and let $q_{k,t}$ be the density function of $N(\mu, 1)$, for some $\mu > 0$, then this model becomes a Bayesian formulation of the change-point models studied in Mei (2010), Xie and Siegmund (2013), and Chan (2017). An interesting boundary case is $\eta = 1$, where all the change points $\tau_1 = \cdots = \tau_K$ are the same. This case can be viewed as a single change point affecting all the data streams.

Example 2 (An i.i.d. change-point model). Assume that $\tau_1, \ldots, \tau_K$ are i.i.d. geometrically distributed random variables, with $\Pr(\tau_k = m) = \theta(1-\theta)^m$, for $m = 0, 1, \ldots$ and $\theta \in (0,1)$. In addition, assume that $p_{k,t}(x) = p(x)$ and $q_{k,t}(x) = q(x)$, for all k, t, x. This model is referred to as model $\mathcal{M}_{\mathrm{s}}$ in the rest of the paper.

Note that the geometric distribution assumption is commonly adopted in Bayesian change-point detection (see, e.g., Tartakovsky, Nikiforov and Basseville (2014)). We adopt this assumption for simplicity, because it leads to analytic posterior probabilities, and point out that it can be relaxed to other known distributions. See Section 3.3 for a discussion about the calculation of the posterior probabilities.

2.2. Compound sequential change-point detection

We now introduce a compound sequential change-point detection problem, defined through an index set process, $S_t \subset \{1, \ldots, K\}$, where S_t indicates the set of active streams at time t. Specifically, if $k \in S_t$, then stream k is active at time t; otherwise, it is deactivated. We require the process to satisfy that $S_{t+1} \subset S_t$, for all $t = 1, 2, \ldots$, meaning that a stream is not allowed to be re-activated once turned off. This requirement is consistent with many real-world applications. For example, in standardized educational testing, once an item is found to have been leaked, it will be removed from the item pool permanently. At the beginning of the data collection (i.e., $t = 1$), all the data streams are active, and thus $S_1 = \{1, \ldots, K\}$.

A sequential detection procedure S_t is defined together with an information filtration, where the definition is inductive. We first let $\mathcal{F}_1 = \sigma(X_{k,1}, k = 1, \ldots, K)$. Then, for any $t > 1$, we let $\mathcal{F}_t = \sigma(\mathcal{F}_{t-1}, S_t, X_{k,t}, k \in S_t)$, where $S_t \subset \{1, \ldots, K\}$ is $\mathcal{F}_{t-1}$ measurable. We say $\{\mathcal{F}_t\}_{t=1,2,\ldots}$ is the information filtration, and the index set process $\{S_t\}_{t \geq 1}$ describes a compound sequential change-point detection procedure with respect to this information filtration.

Sometimes, it is more convenient to represent the decision procedure by a random vector $\mathbb{T} = (T_1, \ldots, T_K)$, where $T_k \in \mathbb{Z}_+$ is defined as $T_k = \sup\{t : k \in S_t\}$. It is easy to check that $\{T_k = t\} \in \mathcal{F}_t$ for all t, and thus T_k is a stopping time under the filtration $\{\mathcal{F}_t\}_{t=1,2,\ldots}$. The stopping time T_k indicates the time up to which we collect data from the kth stream. In other words, starting from time $T_k + 1$, the kth stream is deactivated and its data are no longer collected. The index set at time t is given by $S_t = \{k : T_k \geq t\}$.

The sigma field $\sigma(X_{k,s \wedge T_k}, s \leq t, k = 1, \ldots, K)$ is in $\mathcal{F}_t$, meaning that our information filtration at time t contains all the information from the streams when they are active. In addition to the information from the observable data $X_{k,t}$, the filtration $\mathcal{F}_t$ also contains information from the decision history, reflected by the fact that S_s is measurable with respective to $\mathcal{F}_t$, for all $s \leq t$.

In what follows, we introduce two compound performance metrics for this sequential decision problem.

2.3. LFNR

In this sequential decision problem, our primary goal is to control the proportion of post-change streams among the active ones at any time, where a smaller proportion indicates a better overall quality of the active streams. This proportion can be viewed as a false non-discovery proportion (FNP), which is often considered in multiple hypothesis testing (Genovese and Wasserman (2002)), but defined at each time point under the current sequential setting. More precisely, we define the FNP as

$$\text{FNP}_{t+1}(\mathbb{T}) = \frac{\sum_{k \in S_{t+1}} \mathbb{1}(\tau_k < t)}{|S_{t+1}| \vee 1} = \frac{\sum_{k=1}^{K} \mathbb{1}(T_k > t, \tau_k < t)}{\left\{\sum_{k=1}^{K} \mathbb{1}(T_k > t)\right\} \vee 1}, \qquad (2.2)$$

where $t = 1, 2, \ldots$, $a \vee b = \max(a, b)$ and $|S|$ indicates the size of a set S. In this definition, $|S_{t+1}|$ represents the total number of active streams at time $t+1$, and $\sum_{k \in S_{t+1}} \mathbb{1}(\tau_k < t)$ represents the total number of active post-change streams at time $t+1$. By having "$\vee 1$" in the denominator, $\text{FNP}_{t+1}(\mathbb{T})$ is well defined, even when $|S_{t+1}| = 0$. Finally, we let $\text{FNP}_1 = 0$, as $\tau_k \geq 0$ for $k = 1, 2, \ldots, K$.

Ideally, we would like to control the FNP to be below an acceptable threshold at any time point, which is not always possible because the change points are unknown. As an alternative, we control the LFNR which can be viewed as the best estimate of the FNP in the Bayesian sense. The LFNR is defined as

$$\text{LFNR}_{t+1}(\mathbb{T}) = \mathbb{E}(\text{FNP}_{t+1}(\mathbb{T}) \mid \mathcal{F}_t), \ t = 1, 2, \ldots. \qquad (2.3)$$

Because $\text{FNP}_1 = 0$, $\text{LFNR}_1(\mathbb{T})$ is set to zero.

In what follows, we focus on the sequential change-point detection procedures defined in Section 2.2 under the constraint that $\text{LFNR}_t(\mathbb{T}) \leq \alpha$, for all t, for some prespecified level α (e.g., $\alpha = 1\%$). More precisely, for a given $\alpha \in (0, 1]$, we consider the following class of compound sequential change-point detection procedures that control the LFNR to be below or equal to α at any time, $\mathcal{T}_\alpha = \{\mathbb{T} \in \mathcal{T} : \text{LFNR}_t(\mathbb{T}) \leq \alpha \text{ a.s., for all } t = 1, 2, \ldots\}$, where $\mathcal{T}$ denotes the entire set of compound sequential change-point detection procedures.

We provide a few remarks. First, $\text{LFNR}_{t+1}(\mathbb{T})$ is a random variable, measurable with respect to $\mathcal{F}_t$. It depends on both the change-point model and the detection procedure $\mathbb{T}$. Second, it is easy to observe that $\mathbb{E}(\text{FNP}_t(\mathbb{T})) \leq \alpha$ for every t, for any $\mathbb{T} \in \mathcal{T}_\alpha$. That is, the unconditional expectation of FNP is also controlled at the same α level. Finally, by replacing $\tau_k < t$ with $\tau_k \geq t$ and S_{t+1} with $S_t \setminus S_{t+1}$ in the definition of the FNP, we can similarly define the false discovery proportion (FDP) and local false discovery rate (LFDR) as $\text{FDP}_{t+1} = (|S_t \setminus S_{t+1}| \vee 1)^{-1} \sum_{k \in S_t \setminus S_{t+1}} \mathbb{1}(\tau_k \geq t)$ and $\text{LFDR}_{t+1} = \mathbb{E}(\text{FDP}_{t+1}(\mathbb{T})|\mathcal{F}_t)$. The main difference between false non-discovery rate-type and FDR-type risk measures is whether we focus on the remaining streams or the streams to be deactivated. Specifically, the LFNR focuses on the remaining streams, and thus is a preferred measure if the goal is to control the overall quality of the active data streams (e.g., controlling the proportion of leaked items in the item pool of an educational test). On the other hand, the LFDR is calculated based on the streams to be detected and deactivated. It is thus a better metric if the goal is to control the accuracy among the detected streams.

2.4. Stream utilization and optimality criteria

Given a level α, the class $\mathcal{T}_\alpha$ has many elements. We propose comparing them based on their overall use of the data streams. More precisely, we consider the following measure:

$$\text{U}_t(\mathbb{T}) = \sum_{s=1}^{t} |S_s| = \sum_{s=1}^{t} \sum_{k=1}^{K} \mathbb{1}(T_k \geq s),$$

where $\mathbb{T} = (T_1, \ldots, T_K)$ is a sequential change-point detection procedure and $\text{U}_t(\mathbb{T})$ is the total number of data points collected from the beginning to time t.

For two sequential procedures $\mathbb{T}$ and $\mathbb{T}'$ in $\mathcal{T}_\alpha$, we say $\mathbb{T}$ is more efficient than $\mathbb{T}'$ at time t if $\mathbb{E}(\text{U}_t(\mathbb{T})) \geq \mathbb{E}(\text{U}_t(\mathbb{T}'))$. In addition, we say $\mathbb{T}$ is uniformly more efficient than $\mathbb{T}'$ if $\mathbb{E}(\text{U}_t(\mathbb{T})) \geq \mathbb{E}(\text{U}_t(\mathbb{T}'))$, for all $t = 1, 2, \ldots$. Following the

previous discussion, our goal becomes developing an efficient procedure in terms of stream utilization, under the constraint that the LFNR is below a prespecified α level all the time. Specifically, we consider the following two optimality criteria, which will guide our development of the compound detection procedures discussed in Section 3.

Definition 1 (Uniform optimality). We say a sequential change-point detection procedure $\mathbb{T} \in \mathcal{T}_\alpha$ is uniformly optimal in $\mathcal{T}_\alpha$, if $\mathbb{T}$ is uniformly more efficient than $\mathbb{T}'$, for any $\mathbb{T}' \in \mathcal{T}_\alpha$. That is, $\mathbb{E}(\mathrm{U}_t(\mathbb{T})) = \sup_{\mathbb{T}' \in \mathcal{T}_\alpha} \mathbb{E}(\mathrm{U}_t(\mathbb{T}'))$, for all $t = 1, 2, \ldots$.

Ideally, we would like to find this uniformly optimal procedure. However, such a procedure does not necessarily exist, because the most efficient procedure at one time point may be less efficient than another procedure at a different time point. Thus, we also consider a weaker version of optimality, which is referred to as local optimality at a given time point.

Definition 2 (Local optimality). Given $\mathcal{F}_t$ at time t, we say the choice of $S_{t+1} \subset S_t$ is locally optimal at time $t+1$ if S_{t+1} is $\mathcal{F}_t$ measurable, $\mathbb{E}(\sum_{k \in S_{t+1}} \mathbb{1}(\tau_k < t) / (|S_{t+1}| \vee 1) \big| \mathcal{F}_t) \leq \alpha$, and $|S_{t+1}| \geq |S|$ a.s. for any other $S \subset S_t$ that is also $\mathcal{F}_t$ measurable and satisfies $\mathbb{E}(\sum_{k \in S} \mathbb{1}(\tau_k < t) / (|S| \vee 1) \big| \mathcal{F}_t) \leq \alpha$.

Note that the local optimality criterion only considers one step forward. A procedure is locally optimal if it maximizes the stream utilization in the subsequent step. Achieving local optimality in each step does not necessarily lead to uniform optimality, and a uniformly optimal procedure does not necessarily exist; see Example 3 in Section 4.

We provide a discussion on the choice of the performance metric. The expected stream utilization measure is most sensible if the active streams have the same utility at any time point, whether having changed or not. This approximately holds for the application to item pool monitoring in educational testing, when the leaked items are only accessible by a small proportion of test takers, in which case the utilities of the leaked and unleaked items are similar.

A closely related performance measure is the cumulative number of detections $\mathrm{CD}_t = K - |S_t|$ at each time point. This performance metric is sensible when each detection (and thus deactivation) is associated with a fixed cost, in which case, the goal becomes to minimize the total cost up to each time point. This metric may also be sensible for applications to item pool monitoring in educational testing. That is, once an item is deactivated, a new item needs to be developed as a replacement, for which the cost is approximately the same across items.

In some applications, it may be more sensible to consider a performance metric based on the utilization of pre-change streams, defined as $\mathrm{RL}_t(\mathbb{T}) =$

$\sum_{k=1}^{K}(T_k \wedge \tau_k \wedge t)$. The expectation of this metric can be viewed as an online-and-compound version of the average run length to false alarm (Lorden (1971)), a classical performance metric for sequential change detection.

As shown in Section 5, similar optimality results hold based on the performance measures $\mathrm{RL}_t(\mathbb{T})$ and $\mathrm{CD}_t(\mathbb{T})$.

3. Proposed Method

3.1. One-step update rule

We first propose a one-step update rule for controlling the LFNR to be below a prespecified level. Let a certain sequential change-point detection procedure be implemented from time 1 to t, and let $\mathcal{F}_t$ be the current information filtration. A one-step update rule decides the index set $S_{t+1} \subset S_t$ based on the up-to-date information $\mathcal{F}_t$, so that the LFNR at time $t+1$ is controlled below the prespecified level α. This update rule tries to maximize the size of S_{t+1} to optimize stream utilization. The details of the proposed one-step update rule are provided in Algorithm 1 below.

Algorithm 1. One-step update rule.

Input: Threshold α, the current index set S_t, and posterior probabilities $(W_{k,t})_{k \in S_t}$, where $W_{k,t} = \Pr(\tau_k < t | \mathcal{F}_t)$.

1: Sort the posterior probabilities in an ascending order. That is, $W_{k_1,t} \leq W_{k_2,t} \leq \cdots \leq W_{k_{|S_t|},t}$, where $S_t = \{k_1, \ldots, k_{|S_t|}\}$. To avoid additional randomness, when there exists a tie ($W_{k_i,t} = W_{k_{i+1},t}$), we require $k_i < k_{i+1}$.

2: For $n = 1, \ldots, |S_t|$, define $R_n = \sum_{i=1}^{n} W_{k_i,t}/n$. and define $R_0 = 0$.

3: Find the largest $n \in \{0, 1, \ldots, |S_t|\}$ such that $R_n \leq \alpha$.

Output: $S_{t+1} = \{k_1, \ldots, k_n\}$ if $n \geq 1$ and $S_{t+1} = \emptyset$ if $n = 0$.

This algorithm contains three steps. In the first step, the stream-specific posterior probabilities are sorted in ascending order. We tend to select streams with small posterior probabilities into S_{t+1}, because they are more likely to be pre-change streams. In the second step, we calculate the cumulative averages of the sorted posterior probabilities. Finally, we find the largest n such that the corresponding cumulative average is no greater than α. The corresponding streams are kept in S_{t+1}, and the rest are deactivated. The cumulative average of the n streams gives the LFNR for S_{t+1}.

The proposed one-step update rule controls the LFNR under the general model in (2.1), as formally described in Proposition 1.

Proposition 1. *Suppose that we obtain the index set S_{t+1} using Algorithm 1, given the index set S_t and the information filtration $\mathcal{F}_t$ at time t. Then, the*

LFNR at time $t+1$ satisfies $\mathbb{E}(\sum_{k \in S_{t+1}} \mathbb{1}(\tau_k < t)/(|S_{t+1}| \vee 1)|\mathcal{F}_t) \le \alpha$.

3.2. Proposed compound sequential change-point detection procedure

The proposed procedure adaptively applies the above one-step update rule. That is, at each time point t, we select the active set S_{t+1} using Algorithm 1, given the information available at time t, including the current active set S_t and the corresponding posterior probabilities $(W_{k,t})_{k \in S_t}$. This method is formally described in Algorithm 2 below. We later refer to this procedure as $\mathbb{T}^*$.

Algorithm 2. Proposed Procedure ($\mathbb{T}^*$).

Input: Threshold α.

1: Let $S_1 = \{1, \ldots, K\}$ and $W_{k,1} = \Pr(\tau_k < 1|\mathcal{F}_1)$ for $k \in S_1$.

2: For $t = 1, 2, 3, \ldots$, input α, S_t and $(W_{k,t})_{k \in S_t}$ to Algorithm 1, and obtain S_{t+1} and $W_{k,t+1} = \Pr(\tau_k < t+1|\mathcal{F}_{t+1})$ for $k \in S_{t+1}$, where $\mathcal{F}_{t+1} = \sigma(\mathcal{F}_t, S_{t+1}, X_{k,t+1}, k \in S_{t+1})$.

Output: $\{S_t\}_{t=1,2,\ldots}$, or equivalently, $\mathbb{T}^* = (T_1, \ldots, T_K)$, where $T_k = \sup\{t : k \in S_t\}$.

Using of Proposition 1, it is easy to show that the proposed procedure controls the LFNR at each step under the general change-point model described in (2.1). This result is summarized in Proposition 2.

Proposition 2. *Let* $\mathbb{T}^*$ *be defined in Algorithm* 2. *Then,* $\mathbb{T}^* \in \mathcal{T}_\alpha$.

3.3. Calculation of posterior probabilities

The proposed update rule relies on the posterior probability $W_{k,t} = \Pr(\tau_k < t|\mathcal{F}_t)$, which is the conditional probability that the change point has occurred to stream k before the current time point t. In general, this posterior probability depends on data from all the streams. Thus its evaluation may be computationally intensive when K is large and $(\tau_1, \ldots, \tau_K)$ has a complex dependence structure. In that case, a Markov chain Monte Carlo method may be needed to evaluate this posterior probability. Under the special case of model $\mathcal{M}_s$ described in Example 2, this posterior probability is easy to evaluate using an iterative update rule, as given in Lemma 1 below.

Lemma 1. *Under model* $\mathcal{M}_s$, *described in Example* 2, $W_{k,0} = 0$ *for* $1 \le k \le K$, *and* $W_{k,t}$ *can be computed using the following update rule for* $1 \le k \le K$:

$$W_{k,t+1} = \begin{cases} \frac{q(X_{k,t+1})/p(X_{k,t+1})}{(1-\theta)(1-W_{k,t})/(\theta+(1-\theta)W_{k,t})+q(X_{k,t+1})/p(X_{k,t+1})} & \text{if } 1 \le t \le T_k - 1, \\ W_{k,T_k} & \text{if } t \ge T_k. \end{cases}$$

(3.1)

Note that the iteration in the above lemma is a slight modification of a classical result for Bayesian sequential change-point detection (Shiryaev (1963)). Indeed, with a single data stream, the statistic $W_{k,t}$ is known to be the test statistic for the Shiryaev procedure, a sequential change-point detection procedure that has been proven the Bayes rule for minimizing the average detection delay, while controlling the probability of a false alarm. A slight difference here is that $W_{k,t}$ stays the same after T_k, owing to the control process that deactivates the data streams.

4. Theoretical Results

4.1. Optimality results

In what follows, we establish optimality results for the proposed one-step update rule and the proposed procedure $\mathbb{T}^*$, under the optimality criteria given in Section 2.4. The proposed update rule is locally optimal under the general change-point model (2.1), following Definition 2 for local optimality.

Proposition 3. *Given LFNR level α and information filtration $\mathcal{F}_t$, the index set S_{t+1} given by Algorithm 1 is locally optimal at time $t + 1$.*

In general, having local optimality in each step does not necessarily lead to uniform optimality, and a uniformly optimal procedure may not even exist. However, Theorem 1 shows that a uniformly optimal procedure exists under change-point model $\mathcal{M}_s$ and, furthermore, the proposed procedure is uniformly optimal. In other words, in this case, a myopic decision rule that maximizes the next-step stream utilization under the LFNR constraint is also uniformly optimal throughout time.

Theorem 1. *Under model $\mathcal{M}_s$, the proposed method $\mathbb{T}^*$ is uniformly optimal in $\mathcal{T}_\alpha$.*

Although model $\mathcal{M}_s$ seems relatively simple, the uniform optimality result established in Theorem 1 is highly nontrivial and requires nonstandard technical tools for the proof, such as the monotone coupling on a partially ordered space for comparing stochastic processes of different dimensions. Part of the challenge is from the compound nature of the problem. Below, we intuitively explain why standard techniques for justifying the optimality of single-stream sequential change-point detection methods do not apply to our problem. Heuristically, for a given t, a larger value of $W_{k,t} = \Pr(\tau_k \leq t - 1 | \mathcal{F}_t)$ suggests a higher chance that a change point has already taken place for the kth data stream. This is

why the proposed procedure chooses to detect streams with the largest posterior probabilities $W_{k,t}$. Indeed, this update rule has been proven to be optimal for a single change detection problem under a Bayesian formulation (Shiryaev (1963)), and is locally optimal according to Proposition 3. However, the local optimality does not necessarily imply uniform optimality. To show uniform optimality, one needs to look into the future. More specifically, we need to deal with the situation where a large value of $W_{k,t}$ is due to random noise, and the posterior probability of the stream may become small at a future time point. In other words, supposing that $W_{k_1,t} > W_{k_2,t}$, we need to show that it is more optimal to detect k_1 than k_2 at time t under our optimality criteria, even though $W_{k_1,t+s} < W_{k_2,t+s}$ can happen with high probability, for some $s > 0$. To establish the uniform optimality, we need the $W_{k,t}$ process generated by the proposed procedure to have some stochastically monotone property. A proof sketch for Theorem 1 and a complete proof are given in the Supplementary Material, where some new techniques are established for the monotone coupling of stochastic processes on a partially ordered space.

In Theorem 1, the assumptions required by the model $\mathcal{M}_s$ may be relaxed. By examining the current proof and the fact that the updating rule (3.1) for the posterior probabilities can be extended to nongeometric priors, we believe that the uniform optimality can still be proved if the change points are i.i.d. following some prior distribution with support $\{0, 1, 2, \dots\}$, for example, a negative binomial distribution. Similarly, the optimality results may be extended to the case where $p_{k,t} = p_t$ and $q_{k,t} = q_t$, for some time-dependent functions p_t and q_t. On the other hand, we believe that it is necessary to assume the data streams $\{X_{k,t}\}_{t \geq 1}$ are identically distributed for different k for the proposed method to be uniformly optimal. Indeed, if the processes $\{X_{k,t}\}_{t \geq 1}$ are not identically distributed, then there may not exist a uniformly optimal procedure. One such example is given below.

Example 3 (Non-existence of uniformly optimal procedure)**.** Let $K = 4$ and τ_ks be independent, for $k = 1, 2, 3, 4$. The change-point distributions satisfy $\Pr(\tau_k \geq 4) = 0$, for $k = 1, 2, 3, 4$. For $m = 0, 1, 2, 3$ and $k = 1, 2, 3, 4$, the probabilities $\Pr(\tau_k = m)$ are given below. In addition, let $X_{k,t}|t \leq \tau_k \sim \text{Bernoulli}(0.5)$ and $X_{k,t}|t > \tau_k \sim \text{Bernoulli}(0.51)$ for $k = 1, 2, 3, 4$. Finally, we set $\alpha = 0.34$.

This model is not in $\mathcal{M}_s$, because the change points are not identically distributed. Enumerating all elements in $\mathcal{T}_\alpha$, we have

$$\sup_{\mathbb{T} \in \mathcal{T}_\alpha} \mathbb{E}\left(U_2(\mathbb{T})\right) = 7 \text{ and } \sup_{\mathbb{T} \in \mathcal{T}_\alpha} \mathbb{E}\left(U_4(\mathbb{T})\right) = 10.$$

Table 1.

$\Pr(\tau_k = m)$	$m = 0$	$m = 1$	$m = 2$	$m = 3$
$k = 1$	0.1	0	0	0.9
$k = 2$	0.4	0.6	0	0
$k = 3$	0.43	0.57	0	0
$k = 4$	0.55	0	0	0.45

However, there is no such sequential procedure maximizing stream utilization at both $t = 2$ and $t = 4$. Consequently, there does not exist a uniformly optimal procedure in this example. The calculation for this example is provided in the Supplementary Material.

Remark 2. Note that a similar algorithm can be given for controlling LFDR_t, while achieving a similar local optimality property. However, because the LFDR is calculated based on the stopped data streams, rather than the active ones, the current techniques for proving uniform optimality no longer apply. The theoretical properties of the LFDR-control procedure are left for future investigation.

4.2. Asymptotic theory

In modern multi-stream change-point detection problems, the number of data streams can be large. To enhance our understanding of the proposed method in large-scale applications, we study the asymptotic properties of the proposed method when the number of streams K goes to infinity.

We first study the structure of $\mathbb{T}^*$ under model $\mathcal{M}_\mathrm{s}$. We define the following process:

$$V_0 = 0 \text{ and } V_{t+1} = \frac{q(X_{1,t+1})/p(X_{1,t+1})}{(1-\theta)(1-V_t)/(\theta + (1-\theta)V_t) + q(X_{1,t+1})/p(X_{1,t+1})},$$

where the parameter θ and the densities $p(\cdot)$ and $q(\cdot)$ are given by the model $\mathcal{M}_\mathrm{s}$. We further define $\lambda_0 = 1$ and

$$\lambda_t = \sup \left\{ \lambda : \lambda \in [0,1] \text{ and } \mathbb{E}(V_t \mid V_t \leq \lambda, V_s \leq \lambda_s, 0 \leq s \leq t-1) \leq \alpha \right\}, \quad (4.1)$$

for $t = 1, 2, \ldots$ Theorem 2 shows that when K grows to infinity, the proposed procedure $\mathbb{T}^*$ converges to a limiting procedure $\mathbb{T}^\dagger$, for which the choice of index set $S_{t+1}^\dagger$ is given by $S_{t+1}^\dagger = \left\{ k \in S_t^\dagger : W_{k,t} \leq \lambda_t \right\}$. This suggests that when K is large, we can replace the proposed procedure $\mathbb{T}^*$ by the limiting procedure $\mathbb{T}^\dagger$. The latter is computationally faster, because the thresholds λ_t can be computed offline and the updates for the streams can be computed in parallel. We make

the following technical assumption.

A1. For Z_1 following density function $p(\cdot)$ and Z_2 following density function $q(\cdot)$, the likelihood ratios $q(Z_1)/p(Z_1)$ and $q(Z_2)/p(Z_2)$ have continuous and strictly positive density functions over $\mathbb{R}_+$ (with respect to the Lebesgue measure).

The above assumption is easily satisfied by continuous random variables. For example, it is satisfied when $p(\cdot)$ and $q(\cdot)$ are two normal density functions with different means and/or variances.

Theorem 2. *Assume that model $\mathcal{M}_s$ holds and Assumption A1 is satisfied. To emphasize the dependence on K, we denote the proposed procedure by $\mathbb{T}_K^*$, the corresponding information filtration at time t by $\mathcal{F}_{K,t}^*$, and the index set at time t by $S_{K,t}^*$. Then, the following results hold for each $t \geq 1$:*

1. $\lim_{K\to\infty} \widehat{\lambda}_{K,t} = \lambda_t$ *a.s., where $\widehat{\lambda}_{K,t} = \max\{W_{k,t} : k \in S_{K,t+1}^*\}$ is the threshold used by $\mathbb{T}_K^*$.*

2. $\lim_{K\to\infty} \mathrm{LFNR}_{t+1}(\mathbb{T}_K^*) = \mathbb{E}(V_t \mid V_s \leq \lambda_s, 0 \leq s \leq t)$ *a.s. Moreover,*

$$\mathbb{E}(V_t \mid V_s \leq \lambda_s, 0 \leq s \leq t) = \begin{cases} 1 - (1-\theta)^t, & t < \frac{\log(1-\alpha)}{\log(1-\theta)}, \\ \alpha, & t \geq \frac{\log(1-\alpha)}{\log(1-\theta)}. \end{cases} \qquad (4.2)$$

3. $\lim_{K\to\infty} K^{-1}|S_{K,t+1}^*| = \mathrm{Pr}\left(V_1 \leq \lambda_1, \ldots, V_t \leq \lambda_t\right)$ *a.s.*

Note that according to the definition of λ_t and the second statement of Theorem 2, when $t < \log(1-\alpha)/\log(1-\theta)$, $\lim_{K\to\infty} \mathrm{LFNR}_{t+1}(\mathbb{T}_K^*) < \alpha$ a.s. and no deactivation of streams is needed yet. Otherwise, $\lim_{K\to\infty} \mathrm{LFNR}_{t+1}(\mathbb{T}_K^*) = \alpha$ a.s., which is achieved by deactivating suspicious streams.

We also provide asymptotic theory for a special case of Example 1 when the change points are completely dependent; that is, $\tau_1 = \cdots = \tau_K = \tau_0$. We make the following assumption.

A2. For Z_1 following density $p(\cdot)$ and Z_2 following density $q(\cdot)$, the density functions satisfy $\mathbb{E}\left(\log(p(Z_1)/q(Z_1))\right) > 0, \mathbb{E}\left(\log(q(Z_2)/p(Z_2))\right) > 0, \mathbb{E}(\log(p(Z_1)/q(Z_1)))^2 < \infty$, and $\mathbb{E}\left(\log(q(Z_2)/p(Z_2))\right)^2 < \infty$.

Note that $\mathbb{E}\left(\log(p(Z_1)/q(Z_1))\right)$ and $\mathbb{E}\left(\log(q(Z_2)/p(Z_2))\right)$ are the Kullback-Leibler divergence between $p(\cdot)$ and $q(\cdot)$. Requiring them to be positive is the same as requiring $p(\cdot)$ and $q(\cdot)$ to be densities of two different distributions.

Theorem 3. *Suppose that data follow a special case of the model given in Example 1 when $\eta = 1$ and $\tau_0 \sim Geom(\theta)$, and Assumption A2 holds. Let*

$$W_t = \Pr(\tau_0 < t \mid X_{k,s}, 1 \le k \le K, 1 \le s \le t), \quad T = \min\{t : W_t > \alpha\}.$$

Then, $\mathbb{T}_K^* = (T, \ldots, T)$. Moreover, the following asymptotic results hold:

1. $\lim_{K \to \infty} (T - \tau_0) = 1$ *a.s.,*

2. $\lim_{K \to \infty} \mathrm{LFNR}_{t+1}(\mathbb{T}_K^*) = 0$ *a.s.,*

3. $\lim_{K \to \infty} K^{-1} |S_{K,t+1}^*| = \mathbb{1}(\tau_0 \ge t)$ *a.s.*

According to the above theorem, the detection time in the proposed procedure is the same for all the data streams. This detection rule is the same as the classical Shiryaev procedure (Shiryaev (1963)) for a single data stream. It thus shares all the optimality properties of the Shiryaev procedure. We further remark that the last limit in the above theorem is non-degenerate in the sense that it is a Bernoulli random variable, rather than a constant, as in Theorem 2.

5. Additional Theoretical Results

In this section, we give extensions of Theorem 1. We first extend the uniform optimality result in Theorem 1 to two other performance measures, $\mathrm{RL}_t(\mathbb{T}) = \sum_{k=1}^{K} (T_k \wedge \tau_k \wedge t)$ and $\mathrm{CD}_t = K - |S_t|$, as discussed in Section 2.4.

Theorem 4. *Under model $\mathcal{M}_s$, the following equations hold for all t:*

$$\mathbb{E}(RL_t(\mathbb{T}^*)) = \sup_{\mathbb{T} \in \mathcal{T}_\alpha} \mathbb{E}(RL_t(\mathbb{T})) \text{ and } \mathbb{E}(CD_t(\mathbb{T}^*)) = \inf_{\mathbb{T} \in \mathcal{T}_\alpha} \mathbb{E}(CD_t(\mathbb{T})). \tag{5.1}$$

We then extend Theorem 1 by comparing an arbitrary sequential procedure in $\mathcal{T}_\alpha$ and a procedure that switches from this procedure to the proposed procedure after a certain time point. This result provides further insights into the proposed procedure. Specifically, we use $\mathbb{T}^{\mathbf{A}} \in \mathcal{T}_\alpha$ to denote an arbitrary sequential procedure which controls the LFNR. We further consider a procedure $\mathbb{T}^{\mathbf{AP}_{t_0}}$, which takes the same procedure as $\mathbb{T}^{\mathbf{A}}$ for $t = 1, \ldots, t_0$. After time $t_0 + 1$, each step of $\mathbb{T}^{\mathbf{AP}_{t_0}}$ follows the proposed update rule in Algorithm 1. Theorem 5 compares four sequential procedures, $\mathbb{T}^{\mathbf{A}}$, $\mathbb{T}^{\mathbf{AP}_{t_0}}$, $\mathbb{T}^{\mathbf{AP}_{t_0+1}}$, and $\mathbb{T}^*$.

Theorem 5. *Let $\mathbb{T}^{A} \in \mathcal{T}_\alpha$ be an arbitrary sequential procedure. Furthermore, let $\mathbb{T}^{AP_{t_0}}$ and $\mathbb{T}^{AP_{t_0+1}}$ be the switching procedures described above, with switching times t_0 and $t_0 + 1$, respectively, for some $t_0 \ge 0$. Then, for all $t = 1, 2, \ldots,$*

$\mathbb{T}^{\boldsymbol{AP}_{t_0}}, \mathbb{T}^{\boldsymbol{AP}_{t_0+1}} \in \mathcal{T}_\alpha$ *and under model* $\mathcal{M}_s$,

$$\mathbb{E}\left(\mathrm{U}_t(\mathbb{T}^{\boldsymbol{A}})\right) \leq \mathbb{E}\left(\mathrm{U}_t(\mathbb{T}^{\boldsymbol{AP}_{t_0+1}})\right) \leq \mathbb{E}\left(\mathrm{U}_t(\mathbb{T}^{\boldsymbol{AP}_{t_0}})\right) \leq \mathbb{E}\left(\mathrm{U}_t(\mathbb{T}^*)\right).$$

The above theorem implies that, under model $\mathcal{M}_s$, $\mathbb{T}^{\boldsymbol{AP}_{t_0}}$ is uniformly better than $\mathbb{T}^{\boldsymbol{A}}$. It also suggests switching to the proposed procedure as soon as possible, if one cannot use the proposed procedure at the beginning for practical constraints. Theorems 1 and 5 are implied by the next theorem.

Theorem 6. *Suppose that model* $\mathcal{M}_s$ *holds. For any* $t_0, s \geq 0$, *and any sequential detection procedure* $\mathbb{T}^{\boldsymbol{A}} \in \mathcal{T}_\alpha$, *let* $\mathcal{F}_t^{\boldsymbol{A}}$ *be the information filtration and* $S_t^{\boldsymbol{A}}$ *be the set of active streams at time t given by* $\mathbb{T}^{\boldsymbol{A}}$. *Then,*

$$\mathbb{E}\left[|S_{t_0+s}^{\boldsymbol{A}}|\big|\mathcal{F}_{t_0}^{\boldsymbol{A}}\right] \leq \mathbb{E}\left[|S_{t_0+s}^{\boldsymbol{AP}_{t_0}}|\big|\mathcal{F}_{t_0}^{\boldsymbol{A}}\right] \quad a.s. \tag{5.2}$$

6. Numerical Experiment

We evaluate the proposed procedure using a simulation study under the change-point model $\mathcal{M}_s$. Two stream sizes $K = 50$ and 500 are considered, representing problems of different scales. For all the data streams, we let the pre- and post-change distributions be $N(0,1)$ and $N(1,1)$, respectively. We consider two settings for the change-point distribution, with $\theta = 0.01$ and 0.05 in the geometric distribution, respectively. We set the threshold to be $\alpha = 0.05$ to control the LFNR. The combinations of K and θ lead to four different settings. For each setting, we run 5,000 independent replications.

We consider two procedures: (1) the adaptive procedure given in Algorithm 2, and (2) a procedure in which a stream k is deactivated if the posterior probability $W_{k,t}$ is greater than the non-adaptive threshold λ_t (see (4.1)) given by the asymptotic results. The non-adaptive threshold λ_t is approximated using a simulation with 1,000,000 streams.

We evaluate these procedures by (1) mean FNP, (2) mean LFNR, (3) mean number of active streams, and (4) mean stream utilization, at each time point. These values are obtained by averaging over the 5,000 independent replications. For example, for each simulation, we calculate the FNP at each time point following equation (2.2). The mean FNP at each time point is then calculated by averaging the corresponding FNP values from the 5,000 independent simulations under each setting. The other metrics are calculated similarly. The results are given in Figures 1 through 4, which correspond to the settings (1) $K = 50$, $\theta = 0.01$, (2) $K = 50$, $\theta = 0.05$, (3) $K = 500$, $\theta = 0.01$, and (4) $K = 500$, $\theta = 0.05$, respectively. We discuss these results below.
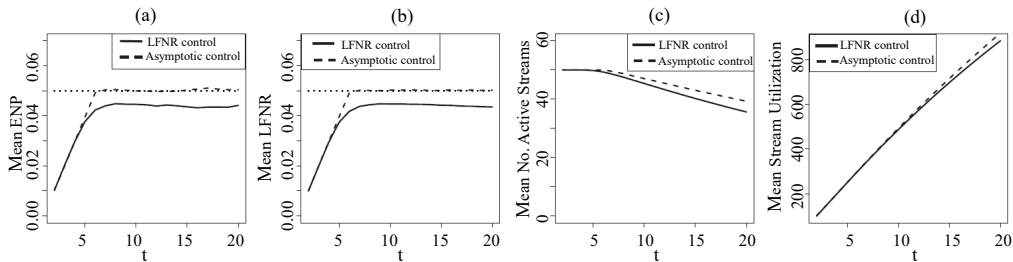
Figure 1. Results under the setting when $K = 50$ and $\theta = 0.01$. Panels (a) through (d) correspond to the four metrics, (1) mean FNP, (2) mean LFNR, (3) mean number of active streams, and (4) mean stream utilization, respectively.

First, for the proposed adaptive procedure, the mean FNP and mean LFNR are always below the 0.05 threshold under all four settings, suggesting that the risk of the active streams is well controlled at the aggregate level; see Panels (a) and (b) of the figures. The control of these quantities is a direct result of the proposed procedure controlling the LFNR at every time point. More specifically, when $\theta = 0.05$, the data streams change relatively more quickly than the case when $\theta = 0.01$. In that case, the proportion of post-change streams quickly exceeds 0.05. The proposed procedure controls this proportion to be slightly below 0.05 by deactivating those with the highest post-change posterior probabilities. As time goes on, the mean FNP decays toward zero as the number of active streams decays to zero; see Panels (c) and (d) of Figures 2 and 4. When $\theta = 0.01$, the data streams change at a much slower rate. Thus, at the beginning, the proportion of post-change streams among the active ones tends to be smaller than 0.05, and the proposed procedure does not tend to detect and deactivate any streams. When the proportion of post-change streams accumulates above 0.05, the proposed procedure starts to deactivate changed streams to control the proportion to be around the targeted level. As the change points occur more slowly, the number of active streams at any given time point tends to be larger than that when $\theta = 0.05$. See Figures 1 and 3 for more details.

Second, as we can see from Panels (a) and (b) of the figures, the non-adaptive procedure based on the asymptotic theory also controls the mean FNP and the mean LFNR to be near or below the targeted level, though the mean FNP may be slightly larger than the targeted level occasionally when $K = 50$. This procedure tends to be slightly more aggressive than the adaptive procedure, because the LFNR can sometimes exceed the targeted threshold α. Overall, the non-adaptive procedure also performs well, in the sense that it tends to control the false non-discovery rate at the targeted level α (i.e., the expected value of the LFNR) at
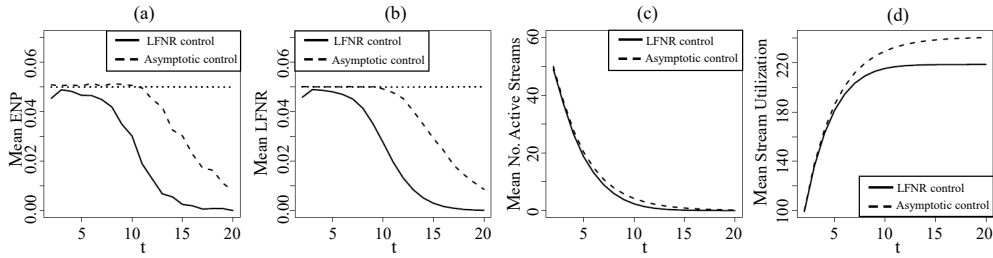
Figure 2. Results under the setting when $K = 50$ and $\theta = 0.05$. The four panels show the same metrics as in Figure 1.
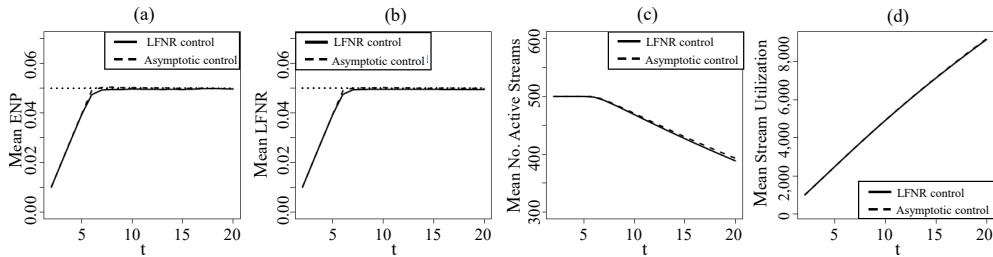

Figure 3. Results under the setting when $K = 500$ and $\theta = 0.01$. The four panels show the same metrics as in Figure 1.
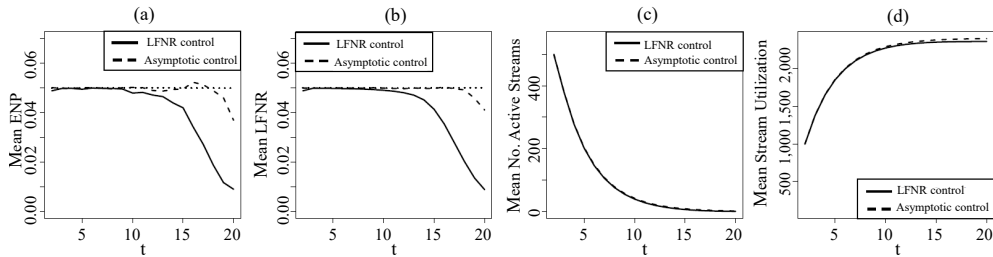

Figure 4. Results under the setting when $K = 500$ and $\theta = 0.05$. The four panels show the same metrics as in Figure 1.

all time points, even though the LFNR itself is not exactly controlled.

Finally, we see that the two procedures tend to perform more similarly when the number of active streams is larger, because the non-adaptive procedure is the limiting case of the adaptive procedure when the number of streams grows to infinity. More specifically, comparing the setting when $K = 500$ (Figures 3 and 4) with that when $K = 50$ (Figures 1 and 2), we see that the two procedures are closer to each other when $K = 500$. For the same value of K, the two procedures tend to be more similar under the setting when $\theta = 0.01$ than that when $\theta = 0.05$, as data streams change more slowly and, thus, there tend to be more active streams at every time point when $\theta = 0.01$. Moreover, for each

setting, the two procedures tend to behave more similarly when t is smaller, as the number of active streams decays with time t.

7. Conclusion

Motivated by real-world applications from various fields, including education, engineering, and finance, we propose a compound decision framework for Bayesian sequential change-point detection in parallel data streams. An easy-to-implement procedure is proposed, for which theoretical properties are established. Specifically, under a class of change-point models, the proposed procedure is shown to be uniformly optimal in a non-asymptotic sense. Numerical experiments show that the proposed procedure can accurately control the aggregated risk of active streams.

The current work can be extended along several directions. First, different optimality criteria may be considered, and the proposed procedure can be extended accordingly. For example, different streams may have different weights, owing to their unequal importance in practice. In that case, more general definitions of the LFNR and stream utilization measure can be given, for which a tailored sequential procedure can be derived.

Second, in some real applications, the change-point distribution and the distributions for the pre- and post-change data may not be known in advance. This problem may be handled by parameterizing the pre- and post-change distributions, and then using a full or empirical Bayes approach that combines the proposed procedure with a sequential estimation of the unknown parameters. Optimality theory may be established when the number of streams K goes to infinity.

Third, optimal sequential procedures remain to be developed under reasonable models for dependent change points. In particular, in many multi-stream change detection problems, the change points may be driven by a low-dimensional latent process, which can be described by a dynamic latent factor model. Several questions remain to be answered under such a change-point model, including the existence of a uniformly optimal procedure, and the construction of such a procedure if it exists.

Finally, a more general setting may be considered that allows new data streams to be added dynamically. For example, in educational testing, once an item is removed from the item pool, a new one needs to be developed to maintain the size of the pool. The inclusion of new data streams changes the information filtration. Under the new information filtration, which contains information from both the original and the new streams, a locally optimal procedure can be

developed under similar optimality criteria. However, it is unclear whether this procedure is still uniformly optimal. This problem is worth future investigation.

Supplementary material

The supplementary material includes proofs of the theoretical results.

Acknowledgments

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300.

Brown, L. D. and Greenshtein, E. (2009). Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *The Annals of Statistics* **37**, 1685–1704.

Cai, T. T., Sun, W. and Wang, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81**, 187–234.

Chan, H. P. (2017). Optimal sequential detection in multi-stream data. *The Annals of Statistics* **45**, 2736–2763.

Chen, H. (2019). Sequential change-point detection based on nearest neighbors. *The Annals of Statistics* **47**, 1381–1407.

Chen, H. and Zhang, N. (2015). Graph-based change-point detection. *The Annals of Statistics* **43**, 139–176.

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

Chen, Y., Wang, T. and Samworth, R. J. (2022). High-dimensional, multiscale online change-point detection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **84**, 234–266.

Choe, E. M., Zhang, J. and Chang, H.-H. (2018). Sequential detection of compromised items using response times in computerized adaptive testing. *Psychometrika* **83**, 650–673.

Cizek, G. J. and Wollack, J. A. (2016). *Handbook of Quantitative Methods for Detecting Cheating on Tests.* Routledge, New York.

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C. and Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. *IEEE Transactions on Neural Networks and Learning Systems* **29**, 3784–3797.

Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statistical Science* **23**, 1–22.

Efron, B. (2019). Bayes, oracle Bayes and empirical Bayes. *Statistical Science* **34**, 177–201.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge University Press, Cambridge.

Fellouris, G. and Sokolov, G. (2016). Second-order asymptotic optimality in multisensor sequential change detection. *IEEE Transactions on Information Theory* **62**, 3662–3675.

Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 499–517.

Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. John Wiley, New York.

Jiang, W. and Zhang, C.-H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics* **37**, 1647–1684.

Lai, T. L. (2001). Sequential analysis: Some classical problems and new challenges. *Statistica Sinica* **11**, 303–351.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics* **42**, 1897–1908.

Mei, Y. (2010). Efficient scalable schemes for monitoring a large number of data streams. *Biometrika* **97**, 419–433.

Page, E. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42**, 523–527.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–115.

Robbins, H. (1956). An empirical Bayes approach to statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (Edited by J. Neyman), 157–163. University of California Press, Berkeley.

Roberts, S. (1966). A comparison of some control chart procedures. *Technometrics* **8**, 411–430.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product*. Van Nostrand, Oxford.

Shiryaev, A. N. (1963). On optimum methods in quickest detection problems. *Theory of Probability & Its Applications* **8**, 22–46.

Song, Y. and Fellouris, G. (2019). Sequential multiple testing with generalized error control: An asymptotic optimality theory. *The Annals of Statistics* **47**, 1776–1803.

Sun, W. and Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association* **102**, 901–912.

Tartakovsky, A., Nikiforov, I. and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. CRC Press, New York.

Thorisson, H. (2000). *Regeneration*. Springer, New York.

van der Linden, W. J. and Lewis, C. (2015). Bayesian checks on cheating on tests. *Psychometrika* **80**, 689–706.

Veerkamp, W. J. and Glas, C. A. (2000). Detection of known items in adaptive testing with a statistical quality control method. *Journal of Educational and Behavioral Statistics* **25**, 373–389.

Xie, Y. and Siegmund, D. (2013). Sequential multi-sensor change-point detection. *The Annals of Statistics* **41**, 670–692.

Zhang, C.-H. (2003). Compound decision theory and empirical Bayes methods. *The Annals of Statistics* **31**, 379–390.

Yunxiao Chen

Department of Statistics, London School of Economics and Political Science, London, WC2A 2AE, UK.

E-mail: y.chen186@lse.ac.uk

Xiaoou Li

School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA.

E-mail: lixx1766@umn.edu