Quickest Changepoint Detection in General Multistream Stochastic Models: Recent Results, Applications and Future Challenges

Alexander G Tartakovsky and Valentin Spivak

Abstract Modern information systems generate large volumes of data with anomalies that occur at unknown points in time and have to be detected quickly and reliably with low false alarm rates. The paper develops a general theory of quickest multistream detection in non-i.i.d. stochastic models when a change may occur in a set of multiple data streams. The first part of the paper focuses on the asymptotic quickest detection theory. Nearly optimal pointwise detection strategies that minimize the expected detection delay are proposed and analyzed when the false alarm rate is low. The general theory is illustrated in several examples. In the second part, we discuss challenging applications associated with the rapid detection of new COVID waves and the appearance of near-Earth space objects. Finally, we discuss certain open problems and future challenges.

1 Introduction

The problem of changepoint detection in multiple data streams (sensors, populations, or multichannel systems) arises in numerous applications that include but are not limited to the medical sphere (detection of an epidemic present in only a fraction of hospitals [5, 8, 22, 34]); environmental monitoring (detection of the presence of hazardous materials or intruders [7, 20]); military defense (detection of an unknown number of targets by multichannel sensor systems [1, 30]); near-Earth space informatics (detection of debris and satellites with telescopes [3, 11, 24]); cyber security (detection of attacks in computer networks [23, 26, 32, 33]); detection of malicious activity in social networks [18, 19], to name a few.

Alexander G Tartakovsky
AGT StatConsult, Los Angeles, California, USA, e-mail: alexg.tartakovsky@gmail.com

Valentin Spivak
MIPT, Russia, e-mail: valyas.spivak@gmail.com

In many change detection applications, the pre-change (the baseline or in-control) distribution of observed data is known, but the post-change (out-of-control) distribution is not completely known. As discussed in [29, Ch 3, 6] there are three conventional approaches in this case: (i) to select a representative value of the post-change parameter and apply efficient detection rules tuned to this value such as the Shiryaev rule, the Shiryaev–Roberts rule or CUSUM, (ii) to select a mixing measure over the parameter space and apply mixture-type rules, (iii) to estimate the parameter and apply adaptive schemes. In Chapters 4 and 6 of [29], a single stream case was considered. In this paper, we consider a more general case where the change occurs in multiple data streams and the number and location of affected data streams are unknown.

To be more specific, suppose there are N data streams observed sequentially in time subject to a change at an unknown point in time $\nu \geq 0$, so that the data up to the time ν are generated by one stochastic model and after $\nu + 1$ by another model. The change in distributions may occur at a subset of streams of a size $1 \leq K \leq N$, where K is an assumed maximal number of streams that can be affected, which can be substantially smaller than N. A sequential detection procedure is a stopping time T with respect to an observed sequence. A false alarm is raised when the detection is declared before the change occurs. One wants to detect the change with as small a delay as possible while controlling the false alarm rate.

We consider a fairly general stochastic model assuming that the observations may be dependent and non-identically distributed (non-i.i.d.) before and after the change and that streams may be mutually dependent.

In the case of i.i.d. observations (in pre-change and post-change modes with different distributions), this problem was considered in [6, 15, 25, 31, 32, 35]. Specifically, in the case of a known post-change parameter and $K = 1$ (i.e., when only one stream can be affected but it is unknown which one), Tartakovsky [25] proposed to use a multi-chart CUSUM procedure that raises an alarm when one of the partial CUSUM statistics exceeds a threshold. This procedure is very simple, but it is not optimal and performs poorly when many data streams are affected. To avoid this drawback, Mei [15] suggested a SUM-CUSUM rule based on the sum of CUSUM statistics in streams and evaluated its first-order performance, which shows that this detection scheme is first-order asymptotically minimax minimizing the maximal expected delay to detection when the average run length to false alarm approaches infinity. Fellouris and Sokolov [6] suggested more efficient generalized and mixture-based CUSUM rules that are second-order minimax. Xie and Siegmund [35] considered a particular Gaussian model with an unknown post-change mean. They suggested a rule that combines mixture likelihood ratios that incorporate an assumption about the proportion of affected data streams with the generalized CUSUM statistics in streams and then add up the resulting local statistics. They also performed a detailed asymptotic analysis of the proposed detection procedure in terms of the average run length to a false alarm and the expected delay as well as MC simulations. Chan [4] studied a version of the mixture likelihood ratio rule for detecting a change in the mean of the normal population assuming independence of

data streams and establishing its asymptotic optimality in a minimax setting as well as dependence of operating characteristics on the fraction of affected streams.

In the present paper, we consider a Bayesian problem with a general prior distribution of the change point and multiple data streams with an unknown pattern, i.e., when the size and location of the affected streams are unknown. It is assumed that the observations can be dependent and non-identically distributed in data streams and even across the streams. Furthermore, in contrast to most previous publications where asymptotically stationary models were considered, we consider substantially non-stationary models, even asymptotically. We address two scenarios when the pre- and post-change distributions are completely known and also a parametric uncertainty when the post-change distribution is known up to an unknown parameter. We introduce mixture detection procedures that mix the Shiryaev–Robers-type statistic over the distributions of the unknown pattern and unknown post-change parameter (in the case of prior uncertainty). The resulting statistics are then compared to appropriate thresholds.

The paper is organized as follows. In Section 2, we present a general theory for very general stochastic models, providing sufficient conditions under which the suggested detection procedures are first-order asymptotically optimal. In Section 3, we provide illustrative examples. In Section 4, we evaluate the performance of proposed mixture detection procedures using Monte Carlo simulations for the non-stationary Gaussian model. In Section 5, theoretical results are applied to rapid detection of the COVID-19 outbreak in Australia based on monitoring the percentage of infections in the total population as well as to rapid detection and extraction of faint near-Earth space objects with telescopes. Section 6 concludes the paper with several remarks, including future research challenges.

2 Asymptotic Theory of Multistream Quickest Change Detection for General Non-i.i.d. Models

In this section, we develop the quickest detection theory in the general multistream non-i.i.d. scenario. We design mixture-based change detection procedures which are nearly optimal in the class of change detection procedures with the prespecified average (weighted) probability of false alarm when this probability is small, assuming that the change point is random with a given prior distribution.

2.1 A General Multistream Model and Basic Notations

We begin with a preliminary description of the scenario of interest and general notation.

Consider the multistream scenario where the observations $\mathbf{X} = (X(1), \ldots, X(N))$ are sequentially acquired in N streams, i.e., in the i-th stream one observes a se-

quence $X(i) = \{X_n(i)\}_{n \in \mathbb{N}}$, where $i \in \mathcal{N} := \{1, \ldots, N\}$. The observations are subject to a change at an unknown time $\nu \in \{0, 1, 2, \ldots\}$, so that $X_1(i), \ldots, X_\nu(i)$ are generated by one stochastic model and $X_{\nu+1}(i), X_{\nu+2}(i), \ldots$ by another model when the change occurs in the i-th stream. The change in distributions happens at a subset of streams $\mathsf{B} \subseteq \{1, \ldots, N\}$ with cardinality $1 \leq |\mathsf{B}| \leq K \leq N$, where K is an assumed maximal number of streams that can be affected, which can be and often is substantially smaller than N. A sequential detection rule is a stopping time T with respect to an observed sequence $\{\mathbf{X}_n\}_{n \in \mathbb{N}}$, $\mathbf{X}_n = (X_n(1), \ldots, X_n(N))$, i.e., T is an integer-valued random variable, such that the event $\{T = n\}$ belongs to the sigma-algebra $\mathscr{F}_n = \sigma(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ generated by observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

The observations may have a very general stochastic structure. Specifically, if we let $\mathbf{X}^n(i) = (X_1(i), \ldots, X_n(i))$ denote the sample of size n in the i-th stream and if $\{f_{\theta_i,n}(X_n(i)|\mathbf{X}^{n-1}(i))\}_{n \geq 1}$, $\theta_i \in \Theta_i$ is a parametric family of conditional densities of $X_n(i)$ given $\mathbf{X}^{n-1}(i)$, then when $\nu = \infty$ (there is no change) the parameter θ_i is equal to the known value $\theta_{i,0}$, i.e., $f_{\theta_i,n}(X_n(i)|\mathbf{X}^{n-1}(i)) = f_{\theta_{i,0},n}(X_n(i)|\mathbf{X}^{n-1}(i))$ for all $n \in \mathbb{N}$ and when $\nu = k < \infty$, then $\theta_i = \theta_{i,1} \neq \theta_{i,0}$, i.e., $f_{\theta_i,n}(X_n(i)|\mathbf{X}^{n-1}(i)) = f_{\theta_{i,0},n}(X_n(i)|\mathbf{X}^{n-1}(i))$ for $n \leq k$ and $f_{\theta_i,n}(X_n(i)|\mathbf{X}^{n-1}(i)) = f_{\theta_{i,1},n}(X_n(i)|\mathbf{X}^{n-1}(i))$ for $n > k$. Not only the point of change ν, but also the subset B, its size $|\mathsf{B}|$, and the post-change parameters $\theta_{i,1}$ are unknown. For further details with a certain change of notation see below.

Let P_∞ denote the probability measure corresponding to the sequence of observations $\{\mathbf{X}_n\}_{n \in \mathbb{N}}$ from all N streams when there is never a change ($\nu = \infty$) in any of the components and, for $k = 0, 1, \ldots$ and $\mathsf{B} \subset \mathcal{N}$, let $\mathsf{P}_{k,\mathsf{B}}$ denote the measure corresponding to the sequence $\{\mathbf{X}_n\}_{n \in \mathbb{N}}$ when $\nu = k < \infty$ and the change occurs in a subset B of the set $\mathscr{P}$ (i.e., $X_{\nu+1}(i)$, $i \in \mathsf{B}$ is the first post-change observation). By $\mathsf{H}_\infty : \nu = \infty$ we denote the hypothesis that the change never occurs and by $\mathsf{H}_{k,\mathsf{B}}$ – the hypothesis that the change occurs at time $0 \leq k < \infty$ is the subset of streams $\mathsf{B} \subset \mathscr{P}$. The set $\mathscr{P}$ is a class of subsets of $\mathcal{N}$ that incorporates available prior information regarding the subset B where the change may occur. For example, in applications frequently it is known that at most K streams can be affected, in which case $\mathscr{P} = \mathscr{P}_K = \{\mathsf{B} \subset \mathcal{N} : 1 \leq |\mathsf{B}| \leq K\}$. Hereafter $|\mathsf{B}|$ denotes the size of a subset B (the number of affected streams under $\mathsf{H}_{k,\mathsf{B}}$) and $|\mathscr{P}|$ denotes the size of class $\mathscr{P}$ (the number of possible alternatives in $\mathscr{P}$). Note that $|\mathscr{P}|$ takes maximum value when there is no prior information regarding the subset of affected streams, i.e., when $\mathscr{P} = \mathscr{P}_N$, in which case $|\mathscr{P}| = 2^N - 1$.

The problem is to detect the change as soon as possible after it occurs regardless of the subset B, i.e., we are interested in detecting the event $\cup_{\mathsf{B} \in \mathscr{P}} \mathsf{H}_{k,\mathsf{B}}$ that the change has occurred in some subset but not in identifying the subset of streams where it occurs.

Write $\mathbf{X}^n(i) = (X_1(i), \ldots, X_n(i))$ for the concatenation of the first n observations from the i-th data stream and $\mathbf{X}^n = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ for the concatenation of the first n observations from all N data streams. Let $\{g(\mathbf{X}_n|\mathbf{X}^{n-1})\}_{n \in \mathbb{N}}$ and $\{f_\mathsf{B}(\mathbf{X}_n|\mathbf{X}^{n-1})\}_{n \in \mathbb{N}}$ be sequences of conditional densities of $\mathbf{X}_n$ given $\mathbf{X}^{n-1}$, which may depend on n, i.e., $g = g_n$ and $f_\mathsf{B} = f_{\mathsf{B},n}$. We omit the subscript n for the sake of brevity. For the general non-i.i.d. changepoint model the joint density $p(\mathbf{X}^n|\mathsf{H}_{k,\mathsf{B}})$ under hypothesis

$H_{k,B}$ can be written as follows

$$p(\mathbf{X}^n|H_{k,B}) = \begin{cases} \prod_{t=1}^{n} g(\mathbf{X}_t|\mathbf{X}^{t-1}) & \text{for } \nu = k \geq n, \\ \prod_{t=1}^{k} g(\mathbf{X}_t|\mathbf{X}^{t-1}) \times \prod_{t=k+1}^{n} f_B(\mathbf{X}_t|\mathbf{X}^{t-1}) & \text{for } \nu = k < n, \end{cases} \quad (1)$$

where $B \subset \mathscr{P}$. Therefore, $g(\mathbf{X}_n|\mathbf{X}^{n-1})$ is the pre-change conditional density and $f_B(\mathbf{X}_n|\mathbf{X}^{n-1})$ is the post-change conditional density given that the change occurs in the subset B.

In most practical applications, the post-change distribution is not completely known – it depends on an unknown (generally multidimensional) parameter $\theta \in \Theta$, so that the model (1) may be treated only as a benchmark for a more practical case where the post-change densities $f_B(\mathbf{X}_t|\mathbf{X}^{t-1})$ are replaced by $f_{B,\theta}(\mathbf{X}_t|\mathbf{X}^{t-1})$, i.e.,

$$p(\mathbf{X}^n|H_{k,B},\theta) = \prod_{t=1}^{k} g(\mathbf{X}_t|\mathbf{X}^{t-1}) \times \prod_{t=k+1}^{n} f_{B,\theta}(\mathbf{X}_t|\mathbf{X}^{t-1}) \quad \text{for } \nu = k < n. \quad (2)$$

Notice that the probabilistic models given by (1) and (2) are very general and do not assume that the data streams are mutually independent.

2.2 Optimality Criterion

In the sequel, we assume that the change point ν is a random variable independent of the observations with a prior distribution $\pi_k = P(\nu = k)$, $k = 0,1,2,\dots$ with $\pi_k > 0$ for $k \in \{0,1,2,\dots\} = \mathbb{Z}_+$ and that a change point may take negative values, i.e., $k \in \mathbb{Z}$, but the detailed structure of the distribution $P(\nu = k)$ for $k \in \mathbb{Z}^- = \{-1,-2,\dots,-\infty\}$ is not important. Only the total probability $\pi_- = P(\nu \leq -1)$ of the change being in effect before the observations become available matters.

Let $E_{k,B,\theta}$ and E_∞ denote expectations under $P_{k,B,\theta}$ and P_∞, respectively, where $P_{k,B,\theta}$ corresponds to model (2) with an unknown parameter $\theta \in \Theta$. Define the probability measure on the Borel σ-algebra $\mathscr{B}$ in $\mathbb{R}^\infty \times \mathbb{Z}$ as

$$P_{B,\theta}^\pi(\mathscr{A} \times \mathscr{K}) = \sum_{k \in \mathscr{K}} \pi_k P_{k,B,\theta}(\mathscr{A}), \quad \mathscr{A} \in \mathscr{B}(\mathbb{R}^\infty), \quad \mathscr{K} \in \mathbb{Z}.$$

Under measure $P_{B,\theta}^\pi$ the change point ν has distribution $\pi = \{\pi_k\}$ and the model for the observations is of the form (2), i.e., $\mathbf{X}(t)$ has conditional density $g(\mathbf{X}(t)|\mathbf{X}^{t-1})$ if $\nu \leq k$ and conditional density $f_{B,\theta}(\mathbf{X}(t)|\mathbf{X}^{t-1})$ if $\nu > k$ and the change occurs in the subset B with the parameter θ. Let $E_{B,\theta}^\pi$ denote the expectation under $P_{B,\theta}^\pi$.

For the prior distribution of the change point $\pi = \{\pi_k\}_{k \in \mathbb{Z}}$, introduce the average (weighted) probability of false alarm associated with the change detection procedure T

$$\text{PFA}_\pi(T) = P_{B,\theta}^\pi(T \leq \nu) = \sum_{k=0}^{\infty} \pi_k P_\infty(T \leq k) \quad (3)$$

that corresponds to the risk due to a false alarm. Note that here we took into account that $P_{k,B,\theta}(T \leq k) = P_\infty(T \leq k)$ since the event $\{T \leq k\}$ depends on the observations $\mathbf{X}_1, \ldots, \mathbf{X}_k$ generated by the pre-change probability measure P_∞ (recall that by our convention $\mathbf{X}_k$ is the last pre-change observation if $\nu = k$).

For $\nu = k \in \mathbb{Z}_+$, $B \in \mathscr{P}$, and $\theta \in \Theta$ the risk associated with the detection delay is measured by the conditional expected delay to detection

$$\mathsf{EDD}_{k,B,\theta}(T) = \mathsf{E}_{k,B,\theta}[T - k \,|\, T > k]. \tag{4}$$

Note that if the change occurs before the observations become available, i.e., $k \in \mathbb{Z}^- = \{-1, -2, \ldots\}$, then $\mathsf{EDD}_{k,B,\theta}(T) = \mathsf{E}_{k,B,\theta}[T] \equiv \mathsf{E}_{0,B,\theta}[T]$ since $T \geq 0$ with probability one.

Next, for the prior distribution π and $\alpha \in (0,1)$, define the Bayesian class of changepoint detection procedures with the weighed probability of false alarm $\mathsf{PFA}_\pi(T) = \mathsf{P}^\pi_{B,\theta}(T \leq \nu)$ not greater than a prescribed number α:

$$\mathbb{C}_\pi(\alpha) = \{T \in \mathscr{M} : \mathsf{PFA}_\pi(T) \leq \alpha\} = \left\{ T \in \mathscr{M} : \sum_{k=0}^\infty \pi_k \mathsf{P}_\infty(T \leq k) \leq \alpha \right\}, \tag{5}$$

where $\mathscr{M}$ stands for the totality of Markov times.

In this section, we are interested in the uniform Bayesian constrained optimization criterion

$$\inf_{\{T : \mathsf{PFA}_\pi(T) \leq \alpha\}} \mathsf{EDD}_{k,B,\theta}(T) \quad \text{for all } k \in \mathbb{Z}_+, \ B \in \mathscr{P} \text{ and } \theta \in \Theta.$$

However, this problem is intractable for arbitrary values of $\alpha \in (0,1)$. For this reason, we will consider the following first-order asymptotic problem assuming that the given PFA α approaches zero: Find a change detection procedure T^* such that it minimizes the expected detection delay $\mathsf{EDD}_{k,B,\theta}(T)$ asymptotically to first order as $\alpha \to 0$ uniformly for all possible values of $k \in \mathbb{Z}_+$, subsets $B \in \mathscr{P}$, and $\theta \in \Theta$. That is, our goal it to design such detection procedure T^* that, as $\alpha \to 0$,

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} \mathsf{EDD}_{k,B,\theta}(T) = \mathsf{EDD}_{k,B,\theta}(T^*)(1 + o(1)) \ \forall\, B \in \mathscr{P}, \ \theta \in \Theta, \ k \in \mathbb{Z}_+, \tag{6}$$

where $\mathbb{C}_\pi(\alpha)$ is the class of detection procedures for which the PFA does not exceed a prescribed number $\alpha \in (0,1)$ defined in (5) and $o(1) \to 0$ as $\alpha \to 0$.

2.3 Multistream Changepoint Detection Procedures

We begin by considering the most general scenario where the observations across streams are dependent.

2.3.1 Parametric Prior Uncertainty

Let $\mathscr{L}_{B,\theta}(n) = f_{B,\theta}(\mathbf{X}_n|\mathbf{X}^{n-1})/g(\mathbf{X}_n|\mathbf{X}^{n-1})$. Note that in the general non-i.i.d. case the statistic $\mathscr{L}_{B,\theta}(n) = \mathscr{L}_{B,\theta}^{(k)}(n)$ depends on the change point $\nu = k$ since the post-change density $f_{B,\theta}(\mathbf{X}_n|\mathbf{X}^{n-1}) = f_{B,\theta}^{(k)}(\mathbf{X}_n|\mathbf{X}^{n-1})$ may depend on k. The likelihood ratio (LR) of the hypothesis $H_{k,B}$ that the change occurs at $\nu = k$ in the subset of streams B against the no-change hypothesis H_∞ based on the sample $\mathbf{X}^n = (\mathbf{X}_1,\ldots,\mathbf{X}_n)$ is given by the product

$$LR_{B,\theta}(k,n) = \prod_{t=k+1}^{n} \mathscr{L}_{B,\theta}(t), \quad n > k$$

and we set $LR_{B,\theta}(k,n) = 1$ for $n \leq k$.

For $B \in \mathscr{P}$ and $\theta \in \Theta$, define the generalized Shiryaev–Roberts (SR) statistic

$$R_{B,\theta}(n) = rLR_{B,\theta}(0,n) + \sum_{k=0}^{n-1} LR_{B,\theta}(k,n), \quad n \geq 1 \tag{7}$$

with the initial condition (non-negative head-start) $R_{B,\theta}(0) = r$, $r \geq 0$.

Now, introduce the probability mass function (mixing measure)

$$\mathbf{p} = \{p_B, B \in \mathscr{P}\}, \quad p_B > 0 \,\forall\, B \in \mathscr{P}, \quad \sum_{B \in \mathscr{P}} p_B = 1, \tag{8}$$

where p_B is the prior probability of the change being in effect on the set of streams B, and also the mixing probability measure

$$\mathbf{W} = \{W(\theta), \theta \in \Theta\}, \quad \int_\Theta dW(\theta) = 1. \tag{9}$$

For a fixed value of θ, introduce the mixture statistic

$$R_{\mathbf{p},\theta}(n) = \sum_{B \in \mathscr{P}} p_B R_{B,\theta}(n)$$

$$= r\Lambda_{\mathbf{p},\theta}(0,n) + \sum_{k=0}^{n-1} \Lambda_{\mathbf{p},\theta}(k,n), \quad n \geq 1, \quad R_{\mathbf{p},\theta}(0) = r, \tag{10}$$

where

$$\Lambda_{\mathbf{p},\theta}(k,n) = \sum_{B \in \mathscr{P}} p_B LR_{B,\theta}(k,n) \tag{11}$$

is the mixture LR.

As discussed in [29, 31], when the parameter θ is unknown there are two main conventional approaches – either to estimate θ (say maximize) or average (mix) over θ. Using the mixing measure (prior distribution) $\mathbf{W} = \{W(\theta), \theta \in \Theta\}$ given in (9), define the double LR-mixture

$$\Lambda_{\mathbf{p},W}(k,n) = \int_{\Theta} \sum_{\mathrm{B} \in \mathscr{P}} p_{\mathrm{B}} LR_{\mathrm{B},\theta}(k,n) \, \mathrm{d}W(\theta) = \int_{\Theta} \Lambda_{\mathbf{p},\theta}(k,n) \, \mathrm{d}W(\theta), \quad k < n \quad (12)$$

and the double-mixture statistic

$$
\begin{aligned}
R_{\mathbf{p},W}(n) &= \int_{\Theta} \sum_{\mathrm{B} \in \mathscr{P}} p_{\mathrm{B}} R_{\mathrm{B},\theta}(n) \, \mathrm{d}W(\theta) \\
&= r\Lambda_{\mathbf{p},W}(0,n) + \sum_{k=0}^{n-1} \Lambda_{\mathbf{p},W}(k,n), \ \ n \ge 1, \ \ R_{\mathbf{p},W}(0) = r
\end{aligned}
\quad (13)
$$

(with a non-negative head-start r).

The corresponding double-mixture LR-based detection procedure is given by the stopping rule which is the first time $n \ge 1$ such that the statistic $R_{\mathbf{p},W}(n)$ exceeds the level $A > 0$:

$$T_A^{\mathbf{p},W} = \inf\{n \ge 1 : R_{\mathbf{p},W}(n) \ge A\}. \quad (14)$$

The main result of our theory is that this changepoint detection procedure is first-order asymptotically optimal under certain very general conditions if threshold $A = A_\alpha$ is adequately selected, i.e., asymptotic equality (6) holds with $T^* = T_A^{\mathbf{p},W}$.

2.3.2 Known Parameters of the Post-Change Distribution

If the value of the post-change parameter θ is known or its putative value is of special interest, representing a nominal change, then it is reasonable to turn the double-mixture procedure $T_A^{\mathbf{p},W}$ in single-mixture procedure $T_A^{\mathbf{p},\theta}$ by taking the degenerate weight function $W(\vartheta)$ concentrated at $\vartheta = \theta$, i.e.,

$$T_A^{\mathbf{p},\theta} = \inf\{n \ge 1 : R_{\mathbf{p},\theta}(n) \ge A\}, \quad (15)$$

and ask whether or not it has first-order asymptotic optimality properties at the point θ, i.e., that under certain conditions asymptotic formula (6) holds for $T^* = T_A^{\mathbf{p},\theta}$.

2.4 Asymptotic Optimality of Mixture-Based Detection Procedures

2.4.1 Basic Conditions

While we consider a general prior and a very general stochastic model for the observations in streams and between streams, to study asymptotic optimality properties we still need to impose certain constraints on the prior distribution $\pi = \{\pi_k\}$ and on the general stochastic model (1) that guarantee asymptotic stability of the detection statistics as the sample size increases.

Regarding the prior distribution, we will assume that the following condition holds:

$$\lim_{n\to\infty} \frac{1}{n} |\log \mathsf{P}(\nu > n)| = \beta \quad \text{for some } \beta \geq 0.$$

However, this condition being quite general does not cover the case where β is positive but may go to zero. Indeed, the distributions with an exponential right tail that satisfy this condition with $\beta > 0$ do not converge as $\beta \to 0$ to heavy-tailed distributions with $\beta = 0$. For this reason, any assertions for heavy-tailed distributions with $\beta = 0$ do not hold if $\beta \to 0$ with an arbitrary rate; the rate must be matched with the PFA probability α, $\alpha \to 0$. Hence, in what follows we consider the scenario with the prior distribution $\pi^\alpha = \{\pi_k^\alpha\}$ that depends on the PFA constraint α and modify the above condition as

CP$_1$. For some $\beta_\alpha \geq 0$ such that $\beta_\alpha \to 0$ as $\alpha \to 0$

$$\lim_{n\to\infty} \frac{1}{n} \left| \log \sum_{k=n+1}^{\infty} \pi_k^\alpha \right| = \beta_\alpha.$$

For $\mathsf{B} \in \mathscr{P}$ and $\theta \in \Theta$, introduce the log-likelihood ratio (LLR) process between the hypotheses $\mathsf{H}_{k,\mathsf{B},\theta}$ $(k = 0, 1, \dots)$ and H_∞:

$$\lambda_{\mathsf{B},\theta}(k,n) = \sum_{t=k+1}^{n} \log \frac{f_{\mathsf{B},\theta}(\mathbf{X}_t | \mathbf{X}^{t-1})}{g(\mathbf{X}_t | \mathbf{X}^{t-1})}, \quad n > k$$

$(\lambda_{\mathsf{B},\theta}(k,n) = 0$ for $n \leq k)$.

Let $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing and continuous function and by Ψ denote the inverse of ψ, which is also increasing and continuous. We also assume that $\lim_{x\to\infty} \psi(x) = \infty$, and thus, $\Psi(x)$ is properly defined on the entire positive real line.

In the rest of the paper, we assume that the strong law of large numbers (SLLN) holds for the LLR with the rate $\psi(n)$, that is, $\lambda_{\mathsf{B},\theta}(k, k+n)/\psi(n)$ converges almost surely (a.s.) under probability $\mathsf{P}_{k,\mathsf{B},\theta}$ to a finite and positive number $I_{\mathsf{B},\theta}$:

$$\frac{1}{\psi(n)} \lambda_{\mathsf{B},\theta}(k,k+n) \xrightarrow[n\to\infty]{\mathsf{P}_{k,\mathsf{B},\theta}-\text{a.s.}} I_{\mathsf{B},\theta} \quad \text{for all } k \in \mathbb{Z}_+, \ \mathsf{B} \in \mathscr{P}, \ \theta \in \Theta. \tag{16}$$

Condition (16) is the first main assumption regarding the general stochastic model for observed data.

Notice that if the data in streams and across the streams are independent (but non-stationary), then

$$I_{\mathsf{B},\theta} = \lim_{n\to\infty} \frac{1}{\psi(n)} \sum_{i\in\mathsf{B}} \sum_{t=k+1}^{k+n} \mathsf{E}_{k,i,\theta_i} \left[\frac{f_{i,\theta_i}(X_t(i))}{g_{i,t}(X_t(i))} \right],$$

where $g_{i,t}(X_t(i))$ and $f_{i,\theta_i}(X_t(i))$ are pre- and post-change densities of the t-th observation in the i-th stream, respectively. In other words, the number $I_{\mathsf{B},\theta}$ can be interpreted as the limiting local Kullback-Leibler divergence.

If the function $\psi(n)$ is non-linear we will say that the model is *non-stationary (even asymptotically)*. However, in many applications the observations are non-

identically distributed but $\psi(n) = n$, in which case we will say that the non-i.i.d. model is *asymptotically stationary*.

In Subsection 2.4.3, we establish the asymptotic lower bound (as $\alpha \to 0$) for the minimal value of the expected detection delay $\mathsf{EDD}_{k,\mathsf{B},\theta}(T)$ in class $\mathbb{C}_\pi(\alpha)$ whenever the almost sure convergence condition (16) holds. To obtain the lower bound it suffices to require the following right-tail condition

$$\lim_{L \to \infty} \mathsf{P}_{k,\mathsf{B},\theta} \left\{ \frac{1}{\psi(L)} \max_{1 \le n \le L} \lambda_{\mathsf{B},\theta}(k, k+n) \ge (1+\varepsilon) I_{\mathsf{B},\theta} \right\} = 0 \;\; \text{for all } \varepsilon > 0.$$

This condition holds whenever the SLLN (16) takes place. However, the SLLN is not sufficient to show that this lower bound is attained for the mixture detection procedures (14) and (15). To this end, the SLLN should be strengthened into a complete convergence version.

Definition 1. We say that the process $\{Y_{k,n}\}_{n \in \mathbb{N}}$, $k \in \mathbb{Z}_+$ converges to a random variable Y uniformly completely under the measure P_k if

$$\sum_{n=1}^{\infty} \sup_{k \in \mathbb{Z}_+} \mathsf{P}_k \left\{ |Y_{k,n} - Y| > \varepsilon \right\} < \infty \;\; \text{for all } \varepsilon > 0.$$

The proof of the upper bound presented in Subsection 2.4.5 shows that if along with the SLLN (16) the following left-tail condition is satisfied

$$\sum_{n=1}^{\infty} \sup_{k \in \mathbb{Z}_+} \mathsf{P}_{k,\mathsf{B},\theta} \left\{ \frac{1}{\psi(n)} \lambda_{\mathsf{B},\theta}(k, k+n) < I_{\mathsf{B},\theta} - \varepsilon \right\} < \infty \quad \text{for all } \varepsilon > 0, \tag{17}$$

where the "information number" $I_{\mathsf{B},\theta}$ is positive and finite, then the detection procedure (15) is asymptotically optimal for the fixed (prespecified) post-change parameter θ.

Obviously, both conditions (16) and (17) are satisfied whenever there exist positive and finite numbers $I_{\mathsf{B},\theta}$ ($\mathsf{B} \in \mathscr{P}$, $\theta \in \Theta$) such that the normalized LLR $\lambda_{\mathsf{B},\theta}(k, k+n)/\psi(n) \to I_{\mathsf{B},\theta}$ uniformly completely under $\mathsf{P}_{k,\mathsf{B},\theta}$-probability. Therefore, this condition turns out to be sufficient for asymptotic optimality of the detection procedure (15) when the post-change parameter θ is known. In the case of the unknown post-change parameter, the left-tail condition is similar to but more sophisticated than condition (17). The details will be given in Subsection 2.4.6.

2.4.2 Heuristic Argument

We begin with a heuristic argument that allows us to obtain approximations for the expected detection delay when the threshold in the mixture detection procedure $T_A^{\mathbf{p},W}$ is large. For the sake of simplicity, the head-start in the detection statistic $R_{\mathbf{p},W}(n)$ is set to zero, $R_{\mathbf{p},W}(n) = r = 0$. This argument also explains the reason why the condition $\mathbf{CP}_1$ on the prior distribution is imposed.

Assume first that the change occurs at $v = k \leq 0$. It is easy to see that the logarithm of the statistic $R_{\mathbf{p},W}(n)$ can be written as

$$\log R_{\mathbf{p},W}(n) = \lambda_{\mathbf{p},W}(0,n) + Y_n,$$

where

$$Y_n = \log \left[\sum_{t=1}^{n-1} \frac{\Lambda_{\mathbf{p},W}(t,n)}{\Lambda_{\mathbf{p},W}(0,n)} \right].$$

Due to the SLLN (16) $\lambda_{\mathbf{p},W}(n)/\psi(n)$ converges almost surely as $n \to \infty$ under $\mathsf{P}_{0,\mathrm{B},\theta}$ to $I_{\mathrm{B},\theta}$, so we can expect that for a large n

$$\lambda_{\mathbf{p},W}(0,n) \approx I_{\mathrm{B},\theta}\, \psi(n) + \xi_n, \tag{18}$$

where $\xi_n/\psi(n)$ converges to 0. Also, Y_n, $n \geq 1$ are "slowly changing" and converge to a random variable Y_∞. Thus, ignoring the overshoot of $\log R_{\mathbf{p},W}(T_A^{\mathbf{p},W})$ over $\log A$, we obtain

$$\log A \approx \log R_{\mathbf{p},W}(T_A^{\mathbf{p},W}) \approx I_{\mathrm{B},\theta}\, \psi\left(T_A^{\mathbf{p},W}\right) + \xi_{T_A^{\mathbf{p},W}} + Y_\infty. \tag{19}$$

If $\psi(n)$ increases sufficiently fast, at least not slower than n, then the last two terms can be ignored, and hence,

$$\psi(T_A^{\mathbf{p},W}) \approx \frac{\log A}{I_{\mathrm{B},\theta}},$$

which implies

$$T_A^{\mathbf{p},W} \approx \Psi\left(\frac{\log A}{I_{\mathrm{B},\theta}} \right).$$

Taking expectation yields

$$\mathsf{E}_{0,\mathrm{B},\theta}[T_A^{\mathbf{p},W}] \approx \Psi\left(\frac{\log A}{I_{\mathrm{B},\theta}} \right).$$

A similar argument leads to the following approximate formula (for a large A) for the expected delay $\mathsf{E}_{k,\mathrm{B},\theta}[(T_A^{\mathbf{p},W} - k)^+]$ when the change occurs at any $v = k \in \mathbb{Z}_+$:

$$\mathsf{E}_{k,\mathrm{B},\theta}[(T_A^{\mathbf{p},W} - k)^+] \approx \Psi\left(\frac{\log A}{I_{\mathrm{B},\theta}} \right), \quad k \in \mathbb{Z}_+. \tag{20}$$

Next, in Lemma 1 (see Subsection 2.4.4) it is established that the probability of a false alarm of the procedure $T_A^{\mathbf{p},W}$ satisfies the inequality

$$\mathrm{PFA}_\pi(T_A^{\mathbf{p},W}) \leq \bar{v}_A/A,$$

where $\bar{v}_A = \sum_{k=1}^\infty k \pi_k^A$ is the mean of the prior distribution π^A. If we assume that $\bar{v}_A/A \to 0$ as $A \to \infty$, then this inequality along with approximate equality (20) yields the following approximation for the expected delay to detection:

$$\mathsf{E}_{k,\mathsf{B},\theta}[T_A^{\mathbf{p},W} - k | T_A^{\mathbf{p},W} > k] = \frac{\mathsf{E}_{k,\theta}[(T_A^W - k)^+]}{1 - \mathrm{PFA}_\pi(T_A^{\mathbf{p},W})} \approx \Psi\left(\frac{\log A}{I_{\mathsf{B},\theta}}\right). \qquad (21)$$

In subsequent sections, this approximation is justified rigorously.

There are two key points we would like to address. The first one is that if we impose condition $\mathbf{CP}_1$ on the prior distribution of the change point with $\beta > 0$ that does not depend on α and does not converge to 0, then the lower bound for the expected delay to detection in class $\mathbb{C}_\pi(\alpha)$ has the form

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} \mathrm{EDD}_{k,\mathsf{B},\theta} \geq \Psi\left(\frac{|\log\alpha|}{I_{\mathsf{B},\theta} + \beta}\right)(1 + o(1)) \quad \text{as } \alpha \to 0. \qquad (22)$$

In this case, this lower bound is not attained by the procedure $T_A^{\mathbf{p},W}$ since if we take $A = A_\alpha \sim |\log\alpha|$, then it follows from (21) that

$$\mathsf{E}_{k,\mathsf{B},\theta}[T_A^{\mathbf{p},W} - k | T_A^{\mathbf{p},W} > k] \sim \Psi\left(\frac{|\log\alpha|}{I_{\mathsf{B},\theta}}\right) > \Psi\left(\frac{|\log\alpha|}{I_{\mathsf{B},\theta} + \beta}\right).$$

Hereafter we use a conventional notation $y_a \sim z_a$ as $a \to a_0$ if $\lim_{a \to a_0}(y_a/z_a) = 1$. This can be expected since the detection statistic $R_{\mathbf{p},W}(n)$ is based on the uniform prior on a positive half line. However, if $\beta = \beta_\alpha \to 0$, as we assumed in condition $\mathbf{CP}_1$, then the lower bound is

$$\Psi\left(\frac{|\log\alpha|}{I_{\mathsf{B},\theta}}\right)(1 + o(1))$$

and it is attained by $T_A^{\mathbf{p},W}$.

Yet another key point is that to obtain the lower bound (22) with $\beta = \beta_\alpha \to 0$ but $\beta_\alpha > 0$, as shown in the proof of Theorem 1, we need the function $\psi(x)$ to be either linear or super-linear. Indeed, if now we define the statistic

$$S_{\mathbf{p},W}^\pi(n) = \frac{1}{\mathsf{P}(\nu \geq n)}\left[\pi_- \Lambda_{\mathbf{p},W}(0,n) + \sum_{k=0}^{n-1} \pi_k \Lambda_{\mathbf{p},W}(k,n)\right]$$

and the corresponding detection procedure

$$\widetilde{T}_A^{\mathbf{p},W} = \inf\left\{n \geq 1 : S_{\mathbf{p},W}^\pi(n) \geq A\right\},$$

then instead of (19) we have

$$\log A \approx \log S_{\mathbf{p},W}(\widetilde{T}_A^{\mathbf{p},W}) \approx \beta \widetilde{T}_A^{\mathbf{p},W} + I_{\mathsf{B},\theta}\,\psi\left(\widetilde{T}_A^{\mathbf{p},W}\right) + \xi_{\widetilde{T}_A^{\mathbf{p},W}} + Y_\infty.$$

So if $\psi(x) < x$ is sub-linear, i.e., $\Psi(n) \gg n$ for large n, we expect that the prior distribution gives much more contribution than the observations and

$$\mathsf{E}_{k,\mathsf{B},\theta}[\widetilde{T}_A^{\mathbf{p},W} - k | \widetilde{T}_A^{\mathbf{p},W} > k] \approx \Psi\left(\frac{\log A}{\beta}\right)$$

as long as $\beta > 0$, i.e., the prior distribution has an exponential right tail.

Despite the simplicity of the basic ideas and the approximate calculations, the rigorous argument in proving these results is rather tedious.

2.4.3 Asymptotic Lower Bound for Expected Detection Delay

For establishing asymptotic optimality of changepoint detection procedures, we first obtain, under the a.s. convergence condition (16), the asymptotic lower bound for expected detection delay $\mathsf{EDD}_{k,\mathsf{B},\theta}(T) = \mathsf{E}_{k,\mathsf{B},\theta}[T - k | T > k]$ of any detection rule T from class $\mathbb{C}_\pi(\alpha)$. In the following subsections, we show that under certain additional conditions associated with complete convergence of LLR these bounds are attained for the mixture procedures $T_A^{\mathbf{p},W}$ and $T_A^{\mathbf{p},\theta}$.

For $\varepsilon \in (0,1)$ and $\delta > 0$, define

$$N_\alpha = N_\alpha(\varepsilon,\delta,\mathsf{B},\theta) = \Psi\left(\frac{(1-\varepsilon)|\log\alpha|}{I_{\mathsf{B},\theta} + \beta_\alpha + \delta}\right). \tag{23}$$

The following theorem specifies the asymptotic lower bound for the expected detection delay.

Theorem 1. *Let the prior distribution of the change point satisfy condition* $\mathbf{CP}_1$ *and assume that for some positive and finite numbers* $I_{\mathsf{B},\theta}$ *(*$\mathsf{B} \in \mathscr{P}$*,* $\theta \in \Theta$*) the a.s. convergence condition* (16) *holds. Suppose in addition that the function* $\psi(x)$ *increases not slower than x, i.e.,*

$$\psi(x) \geq x, \quad x > 0. \tag{24}$$

Then for all $0 < \varepsilon < 1$ *and* $\delta > 0$

$$\lim_{\alpha\to0} \sup_{T\in\mathbb{C}_\pi(\alpha)} \mathsf{P}_{k,\mathsf{B},\theta}\{k < T \leq k + N_\alpha(\varepsilon,\delta,\mathsf{B},\theta)\} = 0 \ \text{for all } \mathsf{B} \in \mathscr{P}, \theta \in \Theta, \tag{25}$$

and, as a result, for all $\nu = k \in \mathbb{Z}_+$*,* $\mathsf{B} \in \mathscr{P}$*, and* $\theta \in \Theta$

$$\inf_{T\in\mathbb{C}_\pi(\alpha)} \mathsf{EDD}_{k,\mathsf{B},\theta}(T) \geq \Psi\left(\frac{|\log\alpha|}{I_{\mathsf{B},\theta}}\right)(1 + o(1)) \ \text{as } \alpha \to 0. \tag{26}$$

Proof. In the asymptotically stationary case where $\psi(n) = n$, the methodology of the proof is analogous to that used by Tartakovsky [27, 28] in the proofs of the lower bounds in a single stream change detection problem with slightly different assumptions on the prior distribution. A generalization of the proof in the substantially non-stationary case has several technical details. We present complete proof.

To begin, consider an arbitrary stopping time $T \in \mathbb{C}_\pi(\alpha)$ and note that by Markov's inequality

$$\mathsf{EDD}_{k,\mathsf{B},\theta}(T) \geq \mathsf{E}_{k,\mathsf{B},\theta}\left[(T-k)^+\right] \geq N_\alpha \, \mathsf{P}_{k,\mathsf{B},\theta}\left\{(T-k)^+ > N_\alpha\right\}$$
$$= N_\alpha \, \mathsf{P}_{k,\mathsf{B},\theta}\left(T > k + N_\alpha\right).$$

If assertion (25) holds, then

$$\lim_{\alpha \to 0} \inf_{T \in \mathbb{C}_\pi(\alpha)} \mathsf{P}_{k,\mathsf{B},\theta}\left(T > k + N_\alpha\right) = 1. \tag{27}$$

Indeed,

$$\mathsf{P}_{k,\mathsf{B},\theta}\left(T > k + N_\alpha\right) = \mathsf{P}_\infty\left(T > k\right) - \mathsf{P}_{k,\mathsf{B},\theta}\left(k < T \leq k + N_\alpha\right),$$

where we used the identity $\mathsf{P}_{k,\mathsf{B},\theta}(T > k) = \mathsf{P}_\infty(T > k)$. Next, since for any change detection procedure $T \in \mathbb{C}_\pi(\alpha)$,

$$\alpha \geq \sum_{i=k}^\infty \pi_i^\alpha \mathsf{P}_\infty(T \leq i) \geq \mathsf{P}_\infty(T \leq k) \sum_{i=k}^\infty \pi_i^\alpha,$$

it follows that

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} \mathsf{P}_\infty(T > k) \geq 1 - \alpha/\Pi_{k-1}^\alpha, \quad k \in \mathbb{Z}_+, \tag{28}$$

where $\Pi_\ell^\alpha = \mathsf{P}(\nu > \ell) = \sum_{k=\ell+1}^\infty \pi_k^\alpha$. Hence, we obtain

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} \mathsf{P}_{k,\mathsf{B},\theta}\left(T > k + N_\alpha\right) \geq 1 - \alpha/\Pi_{k-1}^\alpha - \sup_{T \in \mathbb{C}_\pi(\alpha)} \mathsf{P}_{k,\mathsf{B},\theta}(k < T \leq k + N_\alpha).$$

This inequality yields (27), and we obtain the asymptotic inequality

$$\mathsf{EDD}_{k,\mathsf{B},\theta}(T) \geq N_\alpha(1 + o(1)) \text{ as } \alpha \to 0,$$

which holds for arbitrary values of $\varepsilon \in (0,1)$ and $\delta > 0$. By our assumption, the function $N_\alpha = N_\alpha(\varepsilon, \delta, \mathsf{B}, \theta)$ is continuous, so we may take a limit $\delta, \varepsilon \to 0$, which implies inequality (26). Consequently, to show the validity of asymptotic inequality (26), it suffices to prove the equality (25).

Let

$$\lambda_{\mathsf{B},\theta}^*(t) = \log \frac{f_{\mathsf{B},\theta}(\mathbf{X}_t|\mathbf{X}^{t-1})}{g(\mathbf{X}_t|\mathbf{X}^{t-1})}$$

and notice that for $n > k$

$$\frac{d\mathsf{P}_\infty^{(n)}}{d\mathsf{P}_{k,\mathsf{B},\theta}^{(n)}} = \frac{1}{LR_{\mathsf{B},\theta}(k,n)} = \exp\left\{-\sum_{t=k+1}^n \lambda_{\mathsf{B},\theta}^*(t)\right\},$$

where hereafter $\mathsf{P}^{(n)}$ denotes a restriction of the measure P to the sigma-algebra $\mathscr{F}_n$. Therefore, changing the measure $\mathsf{P}_\infty \to \mathsf{P}_{k,\mathsf{B},\theta}$ and using Wald's likelihood ratio

identity, we obtain for any $C > 0$:

$$
\begin{aligned}
\mathsf{P}_\infty \left(k < T \le k + N_\alpha \right) &= \mathsf{E}_{k,\mathsf{B},\theta} \left[\mathbb{1}_{\{0 < T - k \le N_\alpha\}} \frac{d\mathsf{P}_\infty^{(T)}}{d\mathsf{P}_{k,\mathsf{B},\theta}^{(T)}} \right] \\
&= \mathsf{E}_{k,\mathsf{B},\theta} \left[\mathbb{1}_{\{0 < T - k \le N_\alpha\}} e^{-\sum_{t=k+1}^{T} \lambda_{\mathsf{B},\theta}^*(t)} \right] \\
&\ge \mathsf{E}_{k,\mathsf{B},\theta} \left[\mathbb{1}_{\{0 < T - k \le N_\alpha, \sum_{t=k+1}^{T} \lambda_{\mathsf{B},\theta}^*(t) < C\}} e^{-\sum_{t=k+1}^{T} \lambda_{\mathsf{B},\theta}^*(t)} \right] \\
&\ge e^{-C} \mathsf{P}_{k,\mathsf{B},\theta} \left\{ 0 < T - k \le N_\alpha, \max_{0 < n-k \le N_\alpha} \sum_{t=k+1}^{n} \lambda_{\mathsf{B},\theta}^*(t) < C \right\} \\
&\ge e^{-C} \left[\mathsf{P}_{k,\mathsf{B},\theta} \{0 < T - k \le N_\alpha\} - \mathsf{P}_{k,\mathsf{B},\theta} \left\{ \max_{0 \le n < N_\alpha} \sum_{t=k+1}^{k+n+1} \lambda_{\mathsf{B},\theta}^*(t) \ge C \right\} \right],
\end{aligned}
$$

where the last inequality follows from the fact that $\mathsf{P}(\mathscr{A} \cap \mathscr{B}) = \mathsf{P}(\mathscr{A}) - \mathsf{P}(\mathscr{B}^c)$ for any events $\mathscr{A}$ and $\mathscr{B}$, where $\mathscr{B}^c$ is the complement event of $\mathscr{B}$. Setting

$$
C = \psi(N_\alpha) I_{\mathsf{B},\theta}(1 + \varepsilon) = (1 + \varepsilon) K_\alpha, \quad K_\alpha = \frac{(1-\varepsilon)|\log \alpha| I_{\mathsf{B},\theta}}{I_{\mathsf{B},\theta} + \beta_\alpha + \delta}
$$

yields

$$
\mathsf{P}_{k,\mathsf{B},\theta} \left(k < T \le k + N_\alpha \right) \le \kappa_{k,\alpha}(T) + \sup_{k \ge 0} \gamma_{k,\alpha}, \tag{29}
$$

where

$$
\kappa_{k,\alpha}(T) = \kappa_{k,\alpha}(\varepsilon, \delta, \theta, \mathsf{B}, T) = e^{(1+\varepsilon)K_\alpha} \mathsf{P}_\infty \{0 < T - k \le N_\alpha\}
$$

and

$$
\gamma_{k,\alpha} = \gamma_{k,\alpha}(\varepsilon, \mathsf{B}, \theta) = \mathsf{P}_{k,\mathsf{B},\theta} \left\{ \max_{0 \le n < N_\alpha} \sum_{t=k+1}^{k+n+1} \lambda_{\mathsf{B},\theta}^*(t) \ge (1 + \varepsilon) I_{\mathsf{B},\theta} \psi(N_\alpha) \right\}.
$$

By the a.s. convergence condition (16),

$$
\lim_{\alpha \to 0} \sup_{k \ge 0} \gamma_{k,\alpha} = 0. \tag{30}
$$

To evaluate $\kappa_{k,\alpha}(T)$ it suffices to note that, by condition $\mathbf{CP}_1$, for all sufficiently small α, there exists a small δ such that

$$
\frac{|\log \Pi_{k-1+N_\alpha}^\alpha|}{k - 1 + N_\alpha} \le \beta_\alpha + \delta,
$$

and to use inequality (28), which yields that for all sufficiently small α

$$
\kappa_{k,\alpha}(T) \le e^{(1+\varepsilon)K_\alpha} \mathsf{P}_\infty(T \le k + N_\alpha) \le \alpha\, e^{(1+\varepsilon)K_\alpha} / \Pi_{k-1+N_\alpha}^\alpha
$$

$$\leq \exp\left\{(1+\varepsilon)K_\alpha - |\log\alpha| + (k-1+N_\alpha)\frac{|\log\Pi^\alpha_{k-1+N_\alpha}|}{k-1+N_\alpha}\right\}$$

$$\leq \exp\left\{(1+\varepsilon)K_\alpha - |\log\alpha| + (k-1+N_\alpha)(\beta_\alpha+\delta)\right\}.$$

Since by condition (24),

$$N_\alpha \leq K_\alpha/I_{\mathsf{B},\theta} = \frac{(1-\varepsilon)|\log\alpha|}{I_{\mathsf{B},\theta}+\beta_\alpha+\delta}$$

it follows that

$$\kappa_{k,\alpha}(T) \leq \exp\left\{-\frac{I_{\mathsf{B},\theta}\varepsilon^2+(\beta_\alpha+\delta)\varepsilon}{I_{\mathsf{B},\theta}+\beta_\alpha+\delta}|\log\alpha| + (\beta_\alpha+\delta)k\right\}$$

$$\leq \exp\left\{-\varepsilon^2|\log\alpha| + (\beta_\alpha+\delta)k\right\} := \overline{\kappa}_{\alpha,k}(\varepsilon,\delta).$$

for all $\varepsilon \in (0,1)$ and every small δ. Hence, for all $\varepsilon \in (0,1)$ and sufficiently small δ,

$$\kappa_{k,\alpha}(T) \leq \overline{\kappa}_{k,\alpha}(\varepsilon,\delta), \tag{31}$$

where $\overline{\kappa}_{k,\alpha}(\varepsilon,\delta)$ does not depend on the stopping time T and goes to 0 as $\alpha \to 0$ for any fixed $k \in \mathbb{Z}_+$, which implies that

$$\sup_{T\in\mathbb{C}_\pi(\alpha)} \kappa_{k,\alpha}(T) \to 0 \quad \text{as } \alpha \to 0 \text{ for every fixed } k \in \mathbb{Z}_+.$$

The proof of the lower bound (26) is complete. □

Remark 1. Super-linear condition (24) for $\psi(x)$ can be generalized to asymptotic super-linearity, i.e., it suffices to require that the growth rate of $\psi(x)$ is faster than a linear function for sufficiently large x.

2.4.4 Probabilities of False Alarm of Mixture Change Detection Procedures

An important question is how to select threshold A in change detection procedures $T_A^{\mathbf{p},\theta}$ and $T_A^{\mathbf{p},W}$ defined in (15) and (14), respectively, to imbed them into class $\mathbb{C}_\pi(\alpha)$.

The following lemma provides the upper bounds for the PFA of these procedures.

Lemma 1. *For all $A > 0$ and any prior distribution $\pi = \{\pi_k\}$ of the change point ν with finite mean $\bar{\nu} = \sum_{k=1}^\infty k\pi_k$, the weighted probabilities of false alarms of the detection procedures $T_A^{\mathbf{p},\theta}$ and $T_A^{\mathbf{p},W}$ satisfy the inequalities*

$$\mathsf{PFA}_\pi(T_A^{\mathbf{p},\theta}) \leq \frac{rb+\bar{\nu}}{A}, \tag{32}$$

$$\mathsf{PFA}_\pi(T_A^{\mathbf{p},W}) \leq \frac{rb+\bar{\nu}}{A}, \tag{33}$$

where $b = \sum_{k=1}^{\infty} \pi_k$. *Hence, if* $A = A_\alpha = (rb + \bar{\nu})/\alpha$, *then* $\mathsf{PFA}_\pi(T_{A_\alpha}^{\mathbf{p},\theta}) \leq \alpha$ *and* $\mathsf{PFA}_\pi(T_{A_\alpha}^{\mathbf{p},W}) \leq \alpha$, *i.e.,* $T_{A_\alpha}^{\mathbf{p},\theta} \in \mathbb{C}_\pi(\alpha)$ *and* $T_{A_\alpha}^{\mathbf{p},W} \in \mathbb{C}_\pi(\alpha)$.

Proof. We have

$$\mathsf{E}_\infty[R_{\mathbf{p},\theta}(n)|\mathscr{F}_{n-1}] = \sum_{\mathsf{B} \in \mathscr{P}} p_\mathsf{B} + \sum_{\mathsf{B} \in \mathscr{P}} p_\mathsf{B} R_{\mathsf{B},\theta}(n-1) = 1 + R_{\mathbf{p},\theta}(n-1),$$

and hence, the statistic $\{R_{\mathbf{p},\theta}(n)\}_{n \in \mathbb{N}}$ is a non-negative $(\mathsf{P}_\infty, \mathscr{F}_n)$-submartingale with mean

$$\mathsf{E}_\infty[R_{\mathbf{p},\theta}(n)] = r + n, \quad n \in \mathbb{N}.$$

By Doob's maximal submartingale inequality, for any $k \geq 1$,

$$\mathsf{P}_\infty(T_A^{\mathbf{p},\theta} \leq k) = \mathsf{P}_\infty\left\{ \max_{0 \leq n \leq k} R_{\mathbf{p},\theta}(n) \geq A \right\} \leq \frac{\mathsf{E}_\infty[R_{\mathbf{p},\theta}(k)]}{A} = \frac{r+k}{A}, \qquad (34)$$

which implies

$$\sum_{k=1}^{\infty} \pi_k \mathsf{P}_\infty(T_A^{\mathbf{p},\theta} \leq k) \leq \frac{r \sum_{k=1}^{\infty} \pi_k + \sum_{k=1}^{\infty} k \pi_k}{A},$$

and inequality (32) follows.

To prove inequality (33) it suffices to note that the double mixture statistic $\{R_{\mathbf{p},W}(n)\}_{n \in \mathbb{N}}$ is also a $(\mathsf{P}_\infty, \mathscr{F}_n)$-submartingale with mean $\mathsf{E}_\infty[R_{\mathbf{p},W}(n) = r + n$ since

$$\mathsf{E}_\infty[R_{\mathbf{p},W}(n)|\mathscr{F}_{n-1}] = \mathsf{E}_\infty\left[\int_\Theta R_{\mathbf{p},\theta}(n)\mathrm{d}W(\theta)|\mathscr{F}_{n-1} \right] = \int_\Theta \mathsf{E}_\infty\left[R_{\mathbf{p},\theta}(n)|\mathscr{F}_{n-1} \right] \mathrm{d}W(\theta)$$

$$= 1 + \int_\Theta R_{\mathbf{p},\theta}(n-1)\mathrm{d}W(\theta) = 1 + R_{\mathbf{p},W}(n-1).$$

Applying Doob's submartingale inequality yields

$$\mathsf{P}_\infty(T_A^{\mathbf{p},W} \leq k) \leq (r+k)/A, \quad k = 1, 2, \ldots, \qquad (35)$$

which implies (33) and the proof is complete. $\square$

2.4.5 Asymptotic Optimality of Mixture Detection Procedure $T_A^{\mathbf{p},\theta}$ for Known Post-change Parameter

We now proceed with establishing asymptotic optimality properties of the mixture detection procedures $T_A^{\mathbf{p},\theta}$ in class $\mathbb{C}_\pi(\alpha)$ as $\alpha \to 0$ assuming that the post-change parameter θ is prespecified (known or selected).

Throughout this subsection, we assume that there exist positive and finite numbers $I_{\mathsf{B},\theta}$ ($\mathsf{B} \in \mathscr{P}$, $\theta \in \Theta$) such that the normalized LLR $\lambda_{\mathsf{B},\theta}(k, k+n)/\psi(n)$ converges as $n \to \infty$ to $I_{\mathsf{B},\theta}$ uniformly completely under probability measure $\mathsf{P}_{k,\mathsf{B},\theta}$, i.e.,

$$\sum_{n=1}^{\infty} \sup_{k\in\mathbb{Z}_+} \mathsf{P}_{k,\mathsf{B},\theta} \left\{ \left| \frac{\lambda_{\mathsf{B},\theta}(k,k+n)}{\psi(n)} - I_{\mathsf{B},\theta} \right| > \varepsilon \right\} < \infty \quad \text{for all } \varepsilon > 0. \qquad (36)$$

2.4.5(a) Asymptotic Operating Characteristics of Detection Procedure $T_A^{\mathbf{p},\theta}$ for Large Threshold Values

The following theorem provides asymptotic operating characteristics of the mixture detection procedure $T_A^{\mathbf{p},\theta}$ for large values of threshold A. Since no constraints are imposed on the false alarm rate (FAR) – neither on the PFA nor any other FAR measure – this result is universal and can be used in a variety of optimization problems. We need to impose some constraints on the behavior of the head-start r_A, which may depend on A and approach infinity as $A \to \infty$ in such a way that

$$\lim_{A\to\infty} (r_A A^{-\varepsilon}) = 0 \quad \text{for any } \varepsilon > 0. \qquad (37)$$

We also need to assume that the function $\psi(x)$ increases not too slowly, at least faster than the logarithmic function. Specifically, we assume the following condition on the inverse function $\Psi(x)$

$$\lim_{x\to\infty} \frac{\log \Psi(x)}{x} = 0. \qquad (38)$$

Theorem 2. *Suppose that conditions* (37) *and* (38) *hold and there exist positive and finite numbers $I_{\mathsf{B},\theta}$, $\mathsf{B} \in \mathscr{P}$, $\theta \in \Theta$, such that the uniform complete convergence condition* (36) *is satisfied. Then the following asymptotic as $A \to \infty$ approximation holds*

$$\mathsf{EDD}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},\theta}) = \Psi \left(\frac{\log A}{I_{\mathsf{B},\theta}} \right) (1 + o(1)) \quad \text{for all } k \in \mathbb{Z}_+, \ \mathsf{B} \in \mathscr{P}. \qquad (39)$$

Proof. For $\varepsilon \in (0,1)$, let $N_A = N_A(\varepsilon, \mathsf{B}, \theta) = \Psi \big((1-\varepsilon)(\log A)/I_{\mathsf{B},\theta} \big)$. Inequality (34) implies that for all $k \in \mathbb{Z}_+$

$$\mathsf{P}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},\theta} > k) = \mathsf{P}_\infty(T_A^{\mathbf{p},\theta} > k) \geq 1 - \frac{k + r_A}{A}.$$

Hence, using Markov's inequality, we obtain

$$\begin{aligned}
\mathsf{EDD}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},\theta}) &\geq \mathsf{E}_{k,\mathsf{B},\theta} \left[(T_A^{\mathbf{p},\theta} - k)^+ \right] \\
&\geq N_A \, \mathsf{P}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},\theta} - k > N_A) \\
&\geq N_A \left[\mathsf{P}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},\theta} > k) - \mathsf{P}_{k,\mathsf{B},\theta}(k < T_A^{\mathbf{p},\theta} < k + N_A) \right] \\
&\geq N_A \left[1 - \frac{r_A + k}{A} - \mathsf{P}_{k,\mathsf{B},\theta}(k < T_A^{\mathbf{p},\theta} < k + N_A) \right]. \qquad (40)
\end{aligned}$$

Similarly to (29)

$$P_{k,B,\theta}\left(k < T_A^{\mathbf{p},\theta} \le k + N_A\right) \le \kappa_{k,A}(T_A^{\mathbf{p},\theta}) + \gamma_{k,A}, \tag{41}$$

where

$$\kappa_{k,A}(T_A^{\mathbf{p},\theta}) = e^{(1-\varepsilon^2)\log A} P_\infty\left(0 < T_A^{\mathbf{p},\theta} - k \le N_A\right)$$

and

$$\gamma_{k,A} = P_{k,B,\theta}\left\{\max_{0 \le n < N_A} \lambda_{B,\theta}(k, k+n) \ge (1+\varepsilon)I_{B,\theta}\psi(N_A)\right\}.$$

Using inequality (34), we obtain

$$P_\infty\left(0 < T_A^{\mathbf{p},\theta} - k < N_A\right) \le P_\infty\left(T_A^{\mathbf{p},\theta} < k + N_A\right) \le (k + r_A + N_A)/A,$$

and consequently,

$$\kappa_{k,A}(T_A^{\mathbf{p},\theta}) \le \frac{k + r_A + \Psi((1-\varepsilon)I_{B,\theta}^{-1}\log A)}{A^{\varepsilon^2}}. \tag{42}$$

Since, by conditions (37) and (38), for any $\varepsilon > 0$

$$\frac{r_A + \Psi((1-\varepsilon)I_{B,\theta}^{-1}\log A)}{A^{\varepsilon^2}} \to 0 \quad \text{as } A \to \infty.$$

it follows from (42) that $\kappa_{k,A}(T_A^{\mathbf{p},\theta}) \to 0$ as $A \to \infty$ for any $k \in \mathbb{Z}_+$. Also, $\gamma_{k,A} \to 0$ as $A \to \infty$ by condition (36). Therefore,

$$P_{k,B,\theta}\left(0 < T_A^{\mathbf{p},\theta} - k < N_A\right) \to 0 \quad \text{as } A \to \infty$$

for any fixed $k \in \mathbb{Z}_+$. It follows from (40) that as $A \to \infty$

$$\mathsf{EDD}_{k,B,\theta}(T_A^{\mathbf{p},\theta}) \ge \Psi\left(\frac{(1-\varepsilon)\log A}{I_{B,\theta}}\right)(1 + o(1)),$$

where ε can be arbitrarily small. This yields the asymptotic lower bound (for any fixed $k \in \mathbb{Z}_+$, $B \in \mathscr{P}$)

$$\liminf_{A\to\infty} \frac{\mathsf{EDD}_{k,B,\theta}(T_A^{\mathbf{p},\theta})}{\log A} \ge \frac{1}{I_{B,\theta}}. \tag{43}$$

To prove (39) it suffices to show that this bound is attained by $T_A^{\mathbf{p},\theta}$, i.e.,

$$\limsup_{A\to\infty} \frac{\mathsf{EDD}_{k,B,\theta}(T_A^{\mathbf{p},\theta})}{\log A} \le \frac{1}{I_{B,\theta}}. \tag{44}$$

For $0 < \varepsilon < I_{\mathsf{B},\theta}$, define

$$M_A = M_A(\varepsilon, \mathsf{B}, \theta) = 1 + \left\lfloor \Psi\left(\frac{\log A}{I_{\mathsf{B},\theta} - \varepsilon}\right) \right\rfloor. \tag{45}$$

Recall the definitions of the mixture SR statistic $R_{\mathbf{p},\theta}(n)$ and mixture LR $\Lambda_{\mathbf{p},\theta}(n)$ given in (10) and (11), respectively. Obviously, for any $n \in \mathbb{N}$ and $\mathsf{B} \in \mathscr{P}$,

$$\log R_{\mathbf{p},\theta}(k+n) \geq \log \Lambda_{\mathbf{p},\theta}(k,k+n) \geq \lambda_{\mathsf{B},\vartheta}(k,k+n) + \log p_{\mathsf{B}},$$

and we have

$$\mathsf{P}_{k,\mathsf{B},\theta}\left(T_A^{\mathbf{p},\theta} - k > n\right) \leq \mathsf{P}_{k,\mathsf{B},\theta}\left\{\frac{1}{\psi(n)}\log R_{\mathbf{p},\theta}(k+n) < \frac{1}{\psi(n)}\log A\right\}$$

$$\leq \mathsf{P}_{k,\mathsf{B},\theta}\left\{\frac{1}{\psi(n)}\lambda_{\mathsf{B},\theta}(k,k+n) < \frac{1}{\psi(n)}\left(\log A + |\log p_{\mathsf{B}}|\right)\right\},$$

where for any $n \geq M_A$ the last probability does not exceed the probability

$$\mathsf{P}_{k,\mathsf{B},\theta}\left\{\frac{1}{\psi(n)}\lambda_{\mathsf{B},\theta}(k,k+n) < I_{\mathsf{B},\theta} - \varepsilon + |\log p_{\mathsf{B}}|/\psi(M_A)\right\}.$$

Therefore, for $I_{\mathsf{B},\theta} > \varepsilon > 0$, all $n \geq M_A$ and all sufficiently large A such that $|\log p_{\mathsf{B}}|/\psi(M_A) < \varepsilon/2$, we have

$$\mathsf{P}_{k,\mathsf{B},\theta}\left(T_A^{\mathbf{p},\theta} - k > n\right) \leq \mathsf{P}_{k,\mathsf{B},\theta}\left\{\frac{1}{\psi(n)}\lambda_{\mathsf{B},\theta}(k,k+n) < I_{\mathsf{B},\theta} - \varepsilon/2\right\}. \tag{46}$$

Now, by Lemma A1 in Tartakovsky [29, page 239], for any $\ell \geq 1$, $k \in \mathbb{Z}_+$, and $\mathsf{B} \in \mathscr{P}$,

$$\mathsf{E}_{k,\mathsf{B},\theta}\left[(T_A^{\mathbf{p},\theta} - k)^+\right]^{\ell} \leq M_A^{\ell} + \ell 2^{\ell-1} \sum_{n=M_A}^{\infty} n^{\ell-1} \mathsf{P}_{k,\mathsf{B},\theta}\left(T_A^{\mathbf{p},\theta} > n\right), \tag{47}$$

which along with inequality (46) yields

$$\mathsf{E}_{k,\mathsf{B},\theta}\left[\left(T_A^{\mathbf{p},\theta} - k\right)^+\right] \leq 1 + \left\lfloor \Psi\left(\frac{\log A}{I_{\mathsf{B},\theta} - \varepsilon}\right) \right\rfloor$$

$$+ \sum_{n=M_A}^{\infty} \mathsf{P}_{k,\mathsf{B},\theta}\left\{\frac{1}{\psi(n)}\lambda_{\mathsf{B},\theta}(k,k+n) < I_{\mathsf{B},\theta} - \varepsilon/2\right\}. \tag{48}$$

Write

$$\Upsilon_A(\varepsilon, \mathsf{B}, \theta) = \sum_{n=M_A}^{\infty} \sup_{k\in\mathbb{Z}_+} \mathsf{P}_{k,\mathsf{B},\theta}\left\{\frac{1}{\psi(n)}\lambda_{\mathsf{B},\theta}(k,k+n) < I_{\mathsf{B},\theta} - \varepsilon\right\}.$$

Using (48) and the inequality $\mathsf{P}_{\infty}(T_A^{\mathbf{p},\theta} > k) > 1 - (r_A + k)/A$ (see (34)), we obtain

$$\mathsf{EDD}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},\theta}) = \frac{\mathsf{E}_{k,\mathsf{B},\theta}\left[\left(T_A^{\mathbf{p},\theta}-k\right)^+\right]}{\mathsf{P}_\infty(T_A^{\mathbf{p},\theta}>k)}$$

$$\leq \frac{1+\left\lfloor \Psi\left(\frac{\log A}{I_{\mathsf{B},\theta}-\varepsilon}\right)\right\rfloor + \varUpsilon_A(\varepsilon/2,\mathsf{B},\theta)}{1-(r_A+k)/A}. \tag{49}$$

Since due to condition (37) $r_A/A \to 0$ and, by complete convergence condition (36), $\lim_{A\to\infty}\varUpsilon_A(\varepsilon/2,\mathsf{B},\theta)=0$ for all $\varepsilon>0$ and $\mathsf{B}\in\mathscr{P}$, inequality (49) implies the asymptotic inequality

$$\mathsf{EDD}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},\theta}) \leq \left(\frac{\log A}{I_{\mathsf{B},\theta}-\varepsilon}\right)(1+o(1)) \quad \text{as } A\to\infty.$$

Since ε can be arbitrarily small the asymptotic upper bound (44) follows and the proof of the asymptotic approximation (39) is complete. $\square$

2.4.5(b) Asymptotic Optimality

Since we assume that the prior distribution may depend on the prescribed PFA α the mean $\bar{v}_\alpha = \sum_{k=0}^\infty k\,\pi_k^\alpha$ depends on α. We also suppose that the head-start $r=r_\alpha$ may depend on α and may go to infinity as $\alpha \to 0$. We further assume that r_α and $\bar{v}_\alpha$ approach infinity with such a rate that

$$\lim_{\alpha\to 0}\frac{\log(r_\alpha+\bar{v}_\alpha)}{|\log\alpha|}=0. \tag{50}$$

The following theorem establishes the first-order asymptotic (as $\alpha \to 0$) optimality of the mixture detection procedure $T_A^{\mathbf{p},\theta}$ for the fixed value of θ in the general non-i.i.d. case. In this theorem, we assume that $\psi(x)$ satisfies condition (24) in contrast to Theorem 2 where this function can be practically arbitrary.

Theorem 3. *Let the prior distribution of the change point satisfy condition* $\mathbf{CP}_1$ *and let the mean of the prior distribution* v_α *and the head-start* r_α *of the statistic* $R_{\mathbf{p},\theta}(n)$ *go to infinity with such a rate that condition* (50) *hold. Assume that for some* $0 < I_{\mathsf{B},\theta} < \infty$ *(*$\mathsf{B}\in\mathscr{P}$*) the uniform complete convergence condition* (36) *holds. If threshold* $A = A_\alpha$ *is so selected that* $\mathsf{PFA}_\pi(T_{A_\alpha}^{\mathbf{p},\theta}) \leq \alpha$ *and* $\log A_\alpha \sim |\log\alpha|$ *as* $\alpha\to 0$, *in particular as* $A_\alpha = (r_\alpha+\bar{v}_\alpha)/\alpha$, *then for all* $k\in\mathbb{Z}_+$ *and* $\mathsf{B}\in\mathscr{P}$ *as* $\alpha\to 0$

$$\mathsf{EDD}_{k,\mathsf{B},\theta}(T_{A_\alpha}^{\mathbf{p},\theta}) = \Psi\left(\frac{|\log\alpha|}{I_{\mathsf{B},\theta}}\right)(1+o(1)); \tag{51}$$

$$\inf_{T\in\mathbb{C}_\pi(\alpha)}\mathsf{EDD}_{k,\mathsf{B},\theta}(T) = \mathsf{EDD}_{k,\mathsf{B},\theta}(T_{A_\alpha}^{\mathbf{p},\theta})(1+o(1)). \tag{52}$$

That is, the detection procedure $T_{A_\alpha}^{\mathbf{p},\theta}$ is first-order asymptotically optimal as $\alpha \to 0$ in class $\mathbb{C}_\pi(\alpha)$, minimizing average detection delay $\mathsf{EDD}_{k,\mathsf{B},\theta}(T)$ uniformly for all $k \in \mathbb{Z}_+$ and $\mathsf{B} \in \mathscr{P}$.

Proof. If threshold A_α is so selected that $\log A_\alpha \sim |\log \alpha|$ as $\alpha \to 0$, then asymptotic approximation (51) follows from Theorem 2. This asymptotic approximation coincides with the asymptotic lower bound (26) in Theorem 1 if condition (24) on $\psi(x)$ holds. Hence, the lower bound is attained by $T_{A_\alpha}^{\mathbf{p},\theta}$, proving (52) and the assertion of the theorem.

In particular, if threshold is selected as $A_\alpha = (r_\alpha + \bar{\nu}_\alpha)/\alpha$, then by Lemma 1 $\mathsf{PFA}_\pi(T_{A_\alpha}^{\mathbf{p},\theta}) \le \alpha$ and $\log A_\alpha \sim -\log \alpha$ due to condition (50). □

Remark 2. If the prior distribution π of the change point does not depend on α and condition $\mathbf{CP}_1$ is satisfied with $\beta \ge 0$, then the assertion of Theorem 3 holds if, and only if, $\beta = 0$, i.e., for heavy-tailed prior distributions, but not for priors with exponential right tail with $\beta > 0$. This follows from the fact that for $\beta > 0$ the lower bound has the form

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} \mathsf{EDD}_{k,\mathsf{B},\theta}(T) \ge \Psi\left(\frac{|\log \alpha|}{I_{\mathsf{B},\theta} + \beta}\right)(1 + o(1)) \quad \text{as } \alpha \to 0.$$

2.4.6 Asymptotic Optimality of the Double-Mixture Procedure $T_A^{\mathbf{p},W}$ for Unknown Post-change Parameter

Consider now the case where the post-change parameter $\theta \in \Theta$ is unknown. The goal is to show that the double-mixture detection procedure $T_A^{\mathbf{p},W}$ defined in (14) is asymptotically optimal to first order.

Recall that in the case of the known parameter θ, the sufficient condition for asymptotic optimality (imposed on the data model) is the uniform complete version of the SLLN (36). The proofs of Theorem 1 and Theorem 2 show that to establish asymptotic optimality of procedure $T_A^{\mathbf{p},\theta}$ it suffices to require the following two (right-tail and left-tail) conditions: for all $\varepsilon > 0$, $k \in \mathbb{Z}_+$ and $\mathsf{B} \in \mathscr{P}$

$$\lim_{L \to \infty} \mathsf{P}_{k,\mathsf{B},\theta}\left\{\frac{1}{\psi(L)} \max_{0 \le n < L} \lambda_{\mathsf{B},\theta}(k,k+n) \ge (1+\varepsilon)I_{\mathsf{B},\theta}\right\} = 0 \qquad (53)$$

and

$$\sum_{n=1}^\infty \mathsf{P}_{k,\mathsf{B},\theta}\left\{\frac{1}{\psi(n)} \lambda_{\mathsf{B},\theta}(k,k+n) < I_{\mathsf{B},\theta} - \varepsilon\right\} < \infty. \qquad (54)$$

Note that the SLLN (16) guarantees right-tail condition (53) and complete version (36) guarantees both conditions (53) and (54).

If the post-change parameter θ is unknown to obtain the upper bound for the expected detection delay the left-trail condition (54) has to be modified as follows. For $\delta > 0$, define $\Gamma_{\delta,\theta} = \{\vartheta \in \Theta : |\vartheta - \theta| < \delta\}$ and assume that there exist positive

and finite numbers $I_{\mathsf{B},\theta}$ ($\mathsf{B} \in \mathscr{P}$, $\theta \in \Theta$) such that for any $\varepsilon > 0$ and for all $\mathsf{B} \in \mathscr{P}$ and $\theta \in \Theta$

$$\lim_{\delta \to 0} \sum_{n=1}^{\infty} \sup_{k \in \mathbb{Z}_+} \mathsf{P}_{k,\mathsf{B},\theta} \left\{ \frac{1}{\psi(n)} \inf_{\vartheta \in \Gamma_{\delta,\theta}} \lambda_{\mathsf{B},\vartheta}(k,k+n) < I_{\mathsf{B},\theta} - \varepsilon \right\} < \infty. \tag{55}$$

We ignore parameter values with W-measure null, i.e., without special emphasis we will always assume that

$$W(\Gamma_{\delta,\theta}) > 0 \ \text{ for all } \theta \in \Theta \text{ and } \delta > 0.$$

The following theorem generalizes Theorem 2 to the case of the unknown post-change parameter.

Theorem 4. *Suppose that conditions* (37) *and* (38) *hold and there exist positive and finite numbers* $I_{\mathsf{B},\theta}$, $\mathsf{B} \in \mathscr{P}$, $\theta \in \Theta$, *such that right-tail and left-tail conditions* (53) *and* (55) *are satisfied. Then the following asymptotic as* $A \to \infty$ *approximation holds*

$$\mathrm{EDD}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},W}) = \Psi\left(\frac{\log A}{I_{\mathsf{B},\theta}} \right)(1 + o(1)) \quad \text{for all } k \in \mathbb{Z}_+, \ \mathsf{B} \in \mathscr{P}, \ \theta \in \Theta. \tag{56}$$

Proof. The proof of the asymptotic lower bound (for any fixed $k \in \mathbb{Z}_+$, $\mathsf{B} \in \mathscr{P}$ and $\theta \in \Theta$)

$$\mathrm{EDD}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},W}) \geq \Psi\left(\frac{\log A}{I_{\mathsf{B},\theta}} \right)(1 + o(1)) \quad \text{as } A \to \infty \tag{57}$$

is essentially identical to that used to establish the lower bound (43) for the expected delay to detection $\mathrm{EDD}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},\theta})$ in the proof of Theorem 2. So it is omitted.

To prove asymptotic approximation (56) it suffices to show that the lower bound (57) is attained by $T_A^{\mathbf{p},W}$, i.e.,

$$\limsup_{A \to \infty} \frac{\mathrm{EDD}_{k,\mathsf{B},\theta}(T_A^{\mathbf{p},W})}{\log A} \leq \frac{1}{I_{\mathsf{B},\theta}}. \tag{58}$$

Let M_A be as in (45). Since for any $n \geq 1$,

$$\log R_{\mathbf{p},W}(k+n) \geq \log \Lambda_{\mathbf{p},W}(k,k+n) \geq \inf_{\vartheta \in \Gamma_{\delta,\theta}} \lambda_{\mathsf{B},\vartheta}(k,k+n) + \log W(\Gamma_{\delta,\theta}) + \log p_{\mathsf{B}},$$

using essentially the same argument as in the proof of Theorem 2 that have led to inequality (46) we obtain that for all $n \geq M_A$

$$\mathsf{P}_{k,\mathsf{B},\theta}\left(T_A^{\mathbf{p},W} > n \right) \leq \mathsf{P}_{k,\mathsf{B},\theta} \left\{ \frac{1}{\psi(n)} \inf_{\vartheta \in \Gamma_{\delta,\theta}} \lambda_{\mathsf{B},\vartheta}(k,k+n) < I_{\mathsf{B},\theta} - \varepsilon \right.$$

$$\left. - \frac{1}{\psi(M_A)} \left(\log p_{\mathsf{B}} + \log W(\Gamma_{\delta,\theta}) \right) \right\}.$$

Hence, for all $n \geq M_A$ and all sufficiently large A such that $|\log[p_\mathrm{B} W(\Gamma_{\delta,\theta})]|/\psi(M_A) < \varepsilon/2$,

$$\mathsf{P}_{k,\mathrm{B},\theta}\left(T_A^{\mathbf{p},W} - k > n\right) \leq \mathsf{P}_{k,\mathrm{B},\theta}\left(\frac{1}{\psi(n)} \inf_{\vartheta \in \Gamma_{\delta,\theta}} \lambda_{\mathrm{B},\vartheta}(k,k+n) < I_{\mathrm{B},\theta} - \frac{\varepsilon}{2}\right). \quad (59)$$

Lemma A1 in Tartakovsky [29, Page 239] yields the inequality (for any $k \in \mathbb{Z}_+$, $\mathrm{B} \in \mathscr{P}$, and $\theta \in \Theta$)

$$\mathsf{E}_{k,\mathrm{B},\theta}\left[(T_A^{\mathbf{p},W} - k)^+\right] \leq M_A + \sum_{n=M_A}^{\infty} \mathsf{P}_{k,\mathrm{B},\theta}\left(T_A^{\mathbf{p},W} > n\right). \quad (60)$$

Using (60) and (59), we obtain

$$\mathsf{E}_{k,\mathrm{B},\theta}\left[\left(T_A^{\mathbf{p},W} - k\right)^+\right] \leq 1 + \Psi\left(\left\lfloor \frac{\log A}{I_{\mathrm{B},\theta} - \varepsilon} \right\rfloor\right) + \Upsilon_A(\varepsilon/2, \mathrm{B}, \theta), \quad (61)$$

where

$$\Upsilon_A(\varepsilon, \mathrm{B}, \theta) = \sum_{n=M_A}^{\infty} \sup_{k \in \mathbb{Z}_+} \mathsf{P}_{k,\mathrm{B},\theta}\left\{\frac{1}{\psi(n)} \inf_{\vartheta \in \Gamma_{\delta,\theta}} \lambda_{\mathrm{B},\vartheta}(k,k+n) < I_{\mathrm{B},\theta} - \varepsilon\right\}.$$

Next, inequality (61) along with the inequality $\mathsf{P}_\infty(T_A^{\mathbf{p},W} > k) > 1 - (r_A + k)/A$ (see (35)) implies the inequality

$$\mathrm{EDD}_{k,\mathrm{B},\theta}(T_A^{\mathbf{p},W}) \leq \frac{1 + \Psi\left(\left\lfloor \frac{\log A}{I_{\mathrm{B},\theta}-\varepsilon} \right\rfloor\right) + \Upsilon_A(\varepsilon/2, \mathrm{B}, \theta)}{1 - (r_A + k)/A}. \quad (62)$$

Since $r_A/A \to 0$ (see condition (37)) and, by condition (55), $\Upsilon_A(\varepsilon/2, \mathrm{B}, \theta) \to 0$ as $A \to \infty$ for all $\varepsilon > 0$, $\mathrm{B} \in \mathscr{P}$, and $\theta \in \Theta$ inequality (62) implies the asymptotic inequality

$$\mathrm{EDD}_{k,\mathrm{B},\theta}(T_A^{\mathbf{p},W}) \leq \left(\frac{\log A}{I_{\mathrm{B},\theta} - \varepsilon}\right)(1 + o(1)) \quad \text{as } A \to \infty,$$

where ε can be taken arbitrarily small so that the asymptotic upper bound (58) follows and the proof of the asymptotic approximation (56) is complete. $\square$

Using asymptotic approximation (56) in Theorem 4 and the lower bound (26) in Theorem 1, it is easy to prove that the mixture procedure $T_A^{\mathbf{p},W}$ is asymptotically optimal to first order as $\alpha \to 0$ in class $\mathbb{C}_\pi(\alpha)$. We silently assume that $\psi(x)$ is either a linear or super-linear function (see (24)).

Theorem 5. *Let the prior distribution of the change point satisfy condition* $\mathbf{CP}_1$. *Assume also that the mean value $\bar{\nu} = \bar{\nu}_\alpha$ of the prior distribution and the head-start $r = r_\alpha$ of the statistic $R_{\mathbf{p},W}(n)$ approach infinity as $\alpha \to 0$ with such a rate*

that condition (50) *holds. Suppose further that there exist numbers* $0 < I_{\mathsf{B},\theta} < \infty$
($\mathsf{B} \in \mathscr{P}, \theta \in \Theta$) such that conditions (53) *and* (55) *are satisfied. If threshold* A_α *is*
so selected that $\mathsf{PFA}_\pi(T_{A_\alpha}^{\mathbf{p},W}) \le \alpha$ *and* $\log A_\alpha \sim |\log \alpha|$ *as* $\alpha \to 0$, *in particular as*
$A_\alpha = (r_\alpha + \bar{v}_\alpha)/\alpha$, *then for all* $k \in \mathbb{Z}_+$, $\mathsf{B} \in \mathscr{P}$ *and* $\theta \in \Theta$ *as* $\alpha \to 0$

$$\mathsf{EDD}_{k,\mathsf{B},\theta}(T_{A_\alpha}^{\mathbf{p},W}) = \Psi \left(\frac{|\log \alpha|}{I_{\mathsf{B},\theta}} \right) (1 + o(1)); \tag{63}$$

$$\inf_{T \in \mathbb{C}_\pi(\alpha)} \mathsf{EDD}_{k,\mathsf{B},\theta}(T) = \mathsf{EDD}_{k,\mathsf{B},\theta}(T_{A_\alpha}^{\mathbf{p},W})(1 + o(1)), \tag{64}$$

that is, the procedure $T_{A_\alpha}^{\mathbf{p},W}$ *is first-order asymptotically optimal as* $\alpha \to 0$ *in class*
$\mathbb{C}_\pi(\alpha)$.

Proof. If A_α is so selected that $\log A_\alpha \sim |\log \alpha|$ as $\alpha \to 0$, then asymptotic approx-
imation (63) follows from asymptotic approximation (56) in Theorem 4. Since this
approximation is the same as the asymptotic lower bound (26) in Theorem 1, this
shows that the lower bound is attained by the detection rule $T_{A_\alpha}^{\mathbf{p},W}$, so (64) follows
and the proof is complete.

If, in particular, $A_\alpha = (r_\alpha + \bar{v}_\alpha)/\alpha$, then $\log A_\alpha \sim |\log \alpha|$ and by Lemma 1
$T_{A_\alpha}^{\mathbf{p},W} \in \mathbb{C}_\pi(\alpha)$. $\square$

Remark 3. The assertion of Theorem 5 holds when $\bar{v}$ and r do not depend on α and
the condition $\mathbf{CP}_1$ is satisfied with $\beta \equiv 0$, i.e., for priors π with heavy tails. This
could be expected since the detection statistic $R_{\mathbf{p},W}(n)$ uses an improper uniform
prior distribution of the change point on the whole positive line.

2.4.7 The Case of Independent Streams

Notice that so far we considered a very general stochastic model where not only the
observations in streams may be dependent and non-identically distributed, but also
the streams may be mutually dependent. In this very general case, computing statis-
tic $R_{\mathbf{p},W}(n)$ is problematic even when the statistics in data streams can be computed.
Consider now still a very general scenario where the data streams are mutually in-
dependent (but have a general statistical structure), which is of special interest for
many applications. The computational problem becomes more manageable when
the data between data streams are independent.

2.4.7(a) Computational Issues

In the case where the data across streams are independent, the model has the form

$$p(\mathbf{X}^n|H_{k,\mathrm{B}},\boldsymbol{\theta}_{\mathrm{B}}) = \prod_{t=1}^{n}\prod_{i=1}^{N} g_i(X_t(i)|\mathbf{X}^{t-1}(i)) \quad \text{for } \nu = k \geq n,$$

$$p(\mathbf{X}^n|H_{k,\mathrm{B}},\boldsymbol{\theta}_{\mathrm{B}}) = \prod_{t=1}^{k}\prod_{i=1}^{N} g_i(X_t(i)|\mathbf{X}^{t-1}(i)) \times \tag{65}$$

$$\prod_{t=k+1}^{n}\prod_{i\in\mathrm{B}} f_{i,\theta_i}(X_t(i)|\mathbf{X}^{t-1}(i))\prod_{i\notin\mathrm{B}} g_i(X_t(i)|\mathbf{X}^{t-1}(i)) \quad \text{for } \nu = k < n,$$

where $g_i(X_t(i)|\mathbf{X}^{t-1}(i))$ and $f_{i,\theta_i}(X_t(i)|\mathbf{X}^{t-1}(i))$ are conditional pre- and post-change densities in the i-th data stream, respectively, $\theta_i \in \Theta_i$ is the unknown post-change parameter (generally multidimensional) in the i-th stream ($i \in \{1,\dots,N\} = \mathscr{N}$), and $\boldsymbol{\theta}_{\mathrm{B}} = (\theta_i, i \in \mathrm{B})$ is the vector of parameters in the set B. So the LR processes are

$$LR_{\mathrm{B},\boldsymbol{\theta}_{\mathrm{B}}}(k,n) = \prod_{i\in\mathrm{B}} LR_{i,\theta_i}(k,n), \quad LR_{i,\theta_i}(k,n) = \prod_{t=k+1}^{n} \mathscr{L}_{i,\theta_i}(t), \quad n > k, \tag{66}$$

where $\mathscr{L}_{i,\theta_i}(t) = f_{i,\theta_i}(X_t(i)|\mathbf{X}^{t-1}(i))/g_i(X_t(i)|\mathbf{X}^{t-1}(i))$.

Recall that $\mathscr{P}_K = \{\mathrm{B} : 1 \leq |\mathrm{B}| \leq K\}$ is the subclass of $\mathscr{P}$ for which the cardinality $|\mathrm{B}|$ of the sets where the change may occur does not exceed $K \leq N$ streams, and by $\mathscr{P}_K^{\star} = \{\mathrm{B} : |\mathrm{B}| = K\}$ denote the subclass of $\mathscr{P}$ for which the change occurs in exactly K streams.

Assume, in addition, that the mixing measure is such that

$$p_{\mathrm{B}} = C(\mathscr{P}_K)\prod_{i\in\mathrm{B}} p_i, \quad C(\mathscr{P}_K) = \left(\sum_{\mathrm{B}\in\mathscr{P}_K}\prod_{i\in\mathrm{B}} p_i\right)^{-1}.$$

Then the mixture LR is

$$\Lambda_{\mathbf{p},\boldsymbol{\theta}}(k,n) = C(\mathscr{P}_K)\sum_{i=1}^{K}\sum_{\mathrm{B}\in\mathscr{P}_i^{\star}}\prod_{j\in\mathrm{B}} p_j LR_{j,\theta_i}(k,n),$$

and its computational complexity is polynomial in the number of data streams. Moreover, in the special, most difficult case of $K = N$ and $p_j = p$, we obtain

$$\Lambda_{\mathbf{p},\boldsymbol{\theta}}(k,n) = C(\mathscr{P}_N)\left[\prod_{i=1}^{N}\left(1 + p LR_{i,\theta_i}(k,n)\right) - 1\right], \tag{67}$$

so its computational complexity is only $O(N)$. The representation (67) corresponds to the case when each stream is affected independently with probability $p/(1+p)$, the assumption that was made in [35].

2.4.7(b) Asymptotic Optimality of Detection Procedures

Since the data are independent across streams, for an assumed value of the change point $v = k$, stream $i \in \mathcal{N}$, and the post-change parameter value in the i-th stream θ_i, the LLR of observations accumulated by time $k + n$ is given by

$$\lambda_{i,\theta_i}(k, k+n) = \sum_{t=k+1}^{k+n} \log \frac{f_{i,\theta_i}(X_t(i)|\mathbf{X}^{t-1}(i))}{g_i(X_t(i))|\mathbf{X}^{t-1}(i))}, \quad k \in \mathbb{Z}_+, \ n \in \mathbb{N}.$$

Define $\Gamma_{\delta,\theta_i} = \{\vartheta \in \Theta_i : |\vartheta - \theta_i| < \delta\}$,

$$\gamma_{k,L}^{(i)}(\varepsilon, \theta_i) = \mathsf{P}_{k,i,\theta_i} \left\{ \frac{1}{\psi(L)} \max_{1 \leq n \leq L} \lambda_{i,\theta_i}(k, k+n) \geq (1+\varepsilon) I_{i,\theta_i} \right\},$$

$$\Upsilon^{(i)}(\varepsilon, \theta_i) = \lim_{\delta \to 0} \sum_{n=1}^{\infty} \sup_{k \in \mathbb{Z}_+} \mathsf{P}_{k,i,\theta_i} \left\{ \frac{1}{\psi(n)} \inf_{\vartheta \in \Gamma_{\delta,\theta_i}} \lambda_{i,\vartheta}(k, k+n) < I_{i,\theta_i} - \varepsilon \right\},$$

and assume that the following right-tail and left-tail conditions are satisfied for local LLR statistics in data streams: There exist positive and finite numbers I_{i,θ_i}, $\theta_i \in \Theta_i$, $i \in \mathcal{N}$, such that for any $\varepsilon > 0$

$$\lim_{L \to \infty} \gamma_{k,L}^{(i)}(\varepsilon, \theta_i) = 0 \quad \text{for all } k \in \mathbb{Z}_+, \ \theta_i \in \Theta_i, \ i \in \mathcal{N}; \tag{68}$$

and

$$\Upsilon^{(i)}(\varepsilon, \theta_i) < \infty \quad \text{for all } \theta_i \in \Theta_i, \ i \in \mathcal{N}. \tag{69}$$

We also assume that $W(\Gamma_{\delta,\theta_i}) > 0$ for all $\delta > 0$ and $i \in \mathcal{N}$.

Let $I_{\mathsf{B},\boldsymbol{\theta}_\mathsf{B}} = \sum_{i \in \mathsf{B}} I_{i,\theta_i}$. Recall that

$$\gamma_{k,L}(\varepsilon, \mathsf{B}, \theta) = \mathsf{P}_{k,\mathsf{B},\theta} \left\{ \frac{1}{\psi(L)} \max_{1 \leq n \leq L} \lambda_{\mathsf{B},\theta}(k, k+n) \geq (1+\varepsilon) I_{\mathsf{B},\theta} \right\}.$$

Since the LLR process $\lambda_{\mathsf{B},\boldsymbol{\theta}_\mathsf{B}}(k, k+n)$ is the sum of independent local LLRs, $\lambda_{\mathsf{B},\boldsymbol{\theta}_\mathsf{B}}(k, k+n) = \sum_{i \in \mathsf{B}} \lambda_{i,\theta_i}(k, k+n)$ (see (66)), it is easy to see that

$$\gamma_{k,L}(\varepsilon, \mathsf{B}, \boldsymbol{\theta}_\mathsf{B}) \leq \sum_{i \in \mathsf{B}} \gamma_{k,L}^{(i)}(\varepsilon, \theta_i),$$

so that local conditions (68) imply global right-tail condition (53). This is true, in particular, if $\lambda_{i,\theta_i}(k, k+n)/\psi(n)$ converge $\mathsf{P}_{k,i,\theta_i}$-a.s. to I_{i,θ_i}, $i = 1, \ldots, N$, in which case the SLLN for the global LLR (16) holds with $I_{\mathsf{B},\boldsymbol{\theta}_\mathsf{B}} = \sum_{i \in \mathsf{B}} I_{i,\theta_i}$. Also,

$$\Upsilon(\varepsilon, \mathsf{B}, \boldsymbol{\theta}_\mathsf{B}) \leq \sum_{i \in \mathsf{B}} \Upsilon^{(i)}(\varepsilon, \mathsf{B}, \theta_i),$$

which shows that local left-tail conditions (69) imply global left-tail condition (55).

Theorem 5 implies the following results on asymptotic properties of the mixture procedure $T_A^{\mathbf{p},W}$.

Corollary 1. *Assume that for some positive and finite numbers* I_{i,θ_i}, $\theta_i \in \Theta_i$, $i = 1\ldots,N$, *right-tail and left-tail conditions* (68) *and* (69) *for local data streams are satisfied. If threshold* A_α *is so selected that* $\mathsf{PFA}_\pi(T_{A_\alpha}^{\mathbf{p},W}) \leq \alpha$ *and* $\log A_\alpha \sim |\log \alpha|$ *as* $\alpha \to 0$, *in particular as* $A_\alpha = (r_\alpha + \bar{v}_\alpha)/\alpha$, *and if conditions* (24) *and* (50) *are satisfied, then asymptotic formulas* (63) *and* (64) *hold with* $I_{\mathsf{B},\theta} = I_{\mathsf{B},\boldsymbol{\theta}_\mathsf{B}} = \sum_{i \in \mathsf{B}} I_{i,\theta_i}$. *Therefore, detection procedure* $T_{A_\alpha}^{\mathbf{p},W}$ *is first-order asymptotically optimal as* $\alpha \to 0$ *in class* $\mathbb{C}_\pi(\alpha)$.

Remark 4. The assertions of Corollary 1 also hold for different distribution functions W_i, $i \in \mathcal{N}$ in streams if we assume that $W_i(\Gamma_{\delta,\theta_i}) > 0$ for all $\delta > 0$ and $i \in \mathcal{N}$. A modification in the proof is trivial.

Remark 5. In the case where the post-change parameters $\theta_1, \ldots, \theta_N$ are known, essentially similar corollary follows from Theorem 3 if we require the uniform complete version of the SLLN for all streams $i \in \mathcal{N}$:

$$\sum_{n=1}^{\infty} \sup_{k \in \mathbb{Z}_+} \mathsf{P}_{k,i,\theta_i} \left\{ \left| \frac{\lambda_{i,\theta_i}(k,k+n)}{\psi(n)} - I_{i,\theta_i} \right| > \varepsilon \right\} < \infty \quad \text{for all } \varepsilon > 0. \qquad (70)$$

3 Examples

3.1 Detection of Changes in the Mean Values of Multistream Autoregressive Non-stationary Processes

This example related to detecting changes in unknown means of multistream non-stationary AR(p) processes has a specific real application in addition to several other applications in mathematical statistics. Specifically, it arises in multi-channel sensor systems (such as radars and electro-optic imaging systems) where it is required to detect an unknown number of randomly appearing signals from objects in clutter and sensor noise (cf., e.g., [1, 30, 31]).

Observations in the i-th channel have the form

$$X_n(i) = \theta_i S_n(i) \mathbb{1}_{\{n > v\}} + \xi_n(i), \quad n \geq 1, \ i = 1, \ldots, N, \qquad (71)$$

where $\theta_i S_n(i)$ are deterministic signals with unknown "amplitudes" $\theta_i > 0$ that appear at an unknown time v in additive noises $\xi_n(i)$ in an N-channel system. Assume that noise processes $\{\xi_n(i)\}_{n \in \mathbb{N}}$ ($i = 1, \ldots, N$) are mutually independent p-th order Gaussian autoregressive processes AR(p) that obey recursions

$$\xi_n(i) = \sum_{j=1}^{p} \rho_{i,j} \xi_{n-j}(i) + w_n(i), \quad n \geq 1, \qquad (72)$$

where $\{w_n(i)\}_{n\geq 1}, i = 1, \ldots, N$, are mutually independent i.i.d. normal $\mathcal{N}(0, \sigma_i^2)$ sequences ($\sigma_i > 0$), so the observations in channels $X_n(1), \ldots, X_N(n)$ are independent of each other. For simplicity, let us set zero initial conditions $\xi_{1-p}(i) = \xi_{2-p}(i) = \cdots = \xi_0(i) = 0$. The coefficients $\rho_{i,1}, \ldots, \rho_{i,p}$ and variances σ_i^2 are known and all roots of the equation $z^p - \rho_{i,1}z^{p-1} - \cdots - \rho_{i,p} = 0$ are in the interior of the unit circle, so that the AR(p) processes are stable.

For $n \in \mathbb{N}$, define the p_n-th order residuals

$$\widetilde{S}_n(i) = S_n(i) - \sum_{j=1}^{p_n} \rho_{i,j} S_{n-j}(i), \quad \widetilde{X}_n(i) = X_n(i) - \sum_{j=1}^{p_n} \rho_{i,j} X_{n-j}(i),$$

where $p_n = p$ if $n > p$ and $p_n = n$ if $n \leq p$. It is easily shown that the conditional pre-change and post-change densities in the i-th channel are

$$g_i(X_n(i)|\mathbf{X}^{n-1}(i)) = f_{0,i}(X_n(i)|\mathbf{X}^{n-1}(i)) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{\widetilde{X}_n(i)^2}{2\sigma_i^2}\right\},$$

$$f_{i,\theta_i}(X_n(i)|\mathbf{X}^{n-1}(i)) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{-\frac{(\widetilde{X}_n(i) - \theta_i\widetilde{S}_n(i))^2}{2\sigma_i^2}\right\}, \quad \theta_i \in (0, \infty),$$

and that for all $k \in \mathbb{Z}_+$ and $n \in \mathbb{N}$ the LLR in the i-th channel has the form

$$\lambda_{i,\theta_i}(k, k+n) = \frac{\theta_i}{\sigma_i^2} \sum_{t=k+1}^{k+n} \widetilde{S}_t(i)\widetilde{X}_t(i) - \frac{\theta_i^2 \sum_{t=k+1}^{k+n} \widetilde{S}_t(i)^2}{2\sigma_i^2}.$$

Since under measure $\mathsf{P}_{k,i,\vartheta_i}$ the random variables $\{\widetilde{X}_n(i)\}_{n\geq k+1}$ are independent Gaussian random variables $\mathcal{N}(\vartheta_i\widetilde{S}_{n(i)}, \sigma_i^2)$, under $\mathsf{P}_{k,i,\vartheta_i}$ the LLR $\{\lambda_{i,\theta_i}(k, k+n)\}_{n\in\mathbb{N}}$ is a Gaussian process (with independent but non-identically distributed increments) with mean and variance

$$\mathsf{E}_{k,i,\vartheta_i}[\lambda_{i,\theta_i}(k, k+n)] = \frac{2\theta_i\vartheta_i - \theta_i^2}{2\sigma_i^2} \sum_{t=k+1}^{k+n} \widetilde{S}_t(i)^2,$$

$$\mathsf{Var}_{k,i,\vartheta_i}[\lambda_{i,\theta_i}(k, k+n)] = \frac{\theta_i^2}{\sigma_i^2} \sum_{t=k+1}^{k+n} \widetilde{S}_t(i)^2. \tag{73}$$

Assume that

$$\lim_{n\to\infty} \frac{1}{\psi(n)} \sup_{k\in\mathbb{Z}_+} \sum_{t=k+1}^{k+n} \widetilde{S}_t(i)^2 = Q_i,$$

where $0 < Q_i < \infty$. In a variety of signal processing applications this condition holds with $\psi(n) = n$, e.g., in radar applications where the signals $\theta_i S_{i,n}$ are the sequences of harmonic pulses. In some applications such as detection, recognition, and tracking of objects on ballistic trajectories that can be approximated by polynomials of order $m = 2 - 3$, the function $\psi(n) = n^m$, $m > 1$. Then for all $k \in \mathbb{Z}_+$ and $\theta_i \in (0, \infty)$

$$\frac{1}{\psi(n)}\lambda_{i,\theta_i}(k,k+n) \xrightarrow[n\to\infty]{\mathsf{P}_{k,i,\theta_i}-\text{a.s.}} \frac{\theta_i^2 Q_i}{2\sigma_i^2} = I_{i,\theta_i},$$

so that the SLLN (16) takes place, and therefore, the right-tail condition (68) holds. Furthermore, since the second moment of the LLR is finite it can be shown that uniform complete convergence conditions (70) as well as the left-tail condition (69) also hold. Thus, by Corollary 1, the mixture detection procedure $T_{A_\alpha}^{\mathbf{p},W}$ minimizes as $\alpha \to 0$ the expected delay to detection and asymptotic formulas (63) and (64) hold with

$$I_{\mathsf{B},\theta} = I_{\mathsf{B},\boldsymbol{\theta}_\mathsf{B}} = \sum_{i\in\mathsf{B}} \frac{\theta_i^2 Q_i}{2\sigma_i^2}.$$

3.2 Change Detection in the Spectrum of the AR(p) Multistream Process

Consider the problem of detecting the change of the correlation coefficient in the p-th order AR process which in the i-th stream satisfies the recursion

$$X_n(i) = \sum_{j=1}^{p} \rho_{i,j}^{(n)} X_{n-j}(i) + w_n(i), \quad n \geq 1, \tag{74}$$

where

$$\rho_{i,\ell}^{(n)} = \theta_{i,\ell}^* \mathbb{1}_{\{n\leq\nu\}} + \theta_{i,\ell} \mathbb{1}_{\{n>\nu\}}$$

and $\{w_n(i)\}_{n\geq 1}$ are i.i.d. (mutually independent) Gaussian random variables with $\mathsf{E}[w_1(i)] = 0$, $\mathsf{E}[w_1^2(i)] = 1$. Additional notation:

$$\theta_i^* = (\theta_{i,1}^*, \ldots, \theta_{i,p}^*)^\top, \quad \theta_i = (\theta_{i,1}, \ldots, \theta_{i,p})^\top, \quad \mathbf{X}_i^{n-1,n-p} = (X_{n-1}(i), \ldots, X_{n-p}(i)).$$

($\top$ denotes transpose).

It is easy to see that the pre-change and post-change conditional densities $g_i(X_{i,n}|\mathbf{X}_i^{n-1}) = g_i(X_n(i)|\mathbf{X}_i^{n-1,n-p})$ and $f_{i,\theta_i}(X_n(i)|\mathbf{X}_i^{n-1}(i)) = f_{i,\theta_i}(X_n(i)|\mathbf{X}_i^{n-1,n-p})$ are given by

$$g_i(X_n(i)|\mathbf{X}_i^{n-1,n-p}) = \frac{1}{(2\pi)^{p/2}} \exp\left\{ -\frac{(\eta_i^*(X_n(i),\mathbf{X}_i^{n-1,n-p})^2}{2} \right\},$$

$$f_{i,\theta_i}(X_n(i)|\mathbf{X}_i^{n-1,n-p}) = \frac{1}{(2\pi)^{p/2}} \exp\left\{ -\frac{(\eta_{\theta_i,i}(X_n(i),\mathbf{X}_i^{n-1,n-p}))^2}{2} \right\}, \tag{75}$$

where $\eta_i^*(y,x) = y - (\theta_i^*)^\top x$ and $\eta_{\theta_i,i}(y,x) = y - (\theta_i)^\top x$ ($x \in \mathbb{R}$, $y \in \mathbb{R}^p$). Therefore, for any $\theta_i \in \mathbb{R}^p$, the LLR is

$$\lambda_i(k, k+n) = \sum_{t=k+1}^{k+n} \lambda_i^*(t)$$

where

$$
\begin{aligned}
\lambda_i^*(t) &= \log \frac{f_{i,\theta_i}(X_t(i)|\mathbf{X}_i^{t-1,t-p})}{g_i(X_t(i)|\mathbf{X}_i^{t-1,t-p})} \\
&= X_t(i)(\theta_i - \theta_i^*)^\top \mathbf{X}_i^{t-1,t-p} + \frac{((\theta_i^*)^\top \mathbf{X}_i^{t-1,t-p})^2 - (\theta_i^\top \mathbf{X}_i^{t-1,t-p})^2}{2}.
\end{aligned}
\tag{76}
$$

The process (74) is not Markov, but the p-dimensional processes

$$\Phi_{i,n} = (X_{i,n}, \dots, X_{i,n-p+1})^\top \in \mathbb{R}^p, \quad i = 1, \dots, N$$

are Markov. Now, for any $\vartheta = (\vartheta_1, \dots, \vartheta_p) \in \mathbb{R}^p$, define the matrix

$$
\Lambda(\vartheta) =
\begin{pmatrix}
\vartheta_1 & \vartheta_2 & \dots & \vartheta_p \\
1 & 0 & \dots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \dots 1 & 0
\end{pmatrix}.
$$

Note that

$$
\Phi_{i,n} =
\begin{cases}
\Lambda(\theta_i^*)\Phi_{i,n-1} + \widetilde{w}_{i,n} & \text{for } n \le \nu, \\
\Lambda(\theta_i)\Phi_{i,n-1} + \widetilde{w}_{i,n} & \text{for } n > \nu,
\end{cases}
\tag{77}
$$

where $\widetilde{w}_{i,n} = (w_{i,n}, 0, \dots, 0)^\top \in \mathbb{R}^p$. Obviously,

$$
\mathsf{E}[\widetilde{w}_n \widetilde{w}_n^\top] = B =
\begin{pmatrix}
1 & \dots & 0 \\
\vdots & \ddots & \vdots \\
0 & \dots & 0
\end{pmatrix}.
$$

Assume that all eigenvalues of the matrices $\Lambda(\theta_i^*)$ in modules are less than 1 and that θ_i belongs to the set

$$
\Theta_i^{st} = \{\theta_i \in \mathbb{R}^p : \max_{1 \le j \le p} |\mathbf{e}_j(\Lambda(\theta_i))| < 1\} \setminus \{\theta_i^*\},
\tag{78}
$$

where $\mathbf{e}_j(\Lambda)$ is the j-th eigenvalue of the matrix Λ. Using (77) it can be shown that in this case the processes $\{\Phi_{i,n}\}_{n>\nu}$ $(i = 1, \dots, N)$ are ergodic with stationary normal distributions $\mathcal{N}(0, \mathsf{F}_i(\theta_i))$, where

$$
\mathsf{F}_i(\theta_i) = \sum_{n=0}^\infty (\Lambda(\theta_i))^n B(\Lambda^\top(\theta_i))^n.
$$

Taking into account that $\max_{1 \le i \le N} \sup_{t \ge 1} \mathsf{E}_\infty |X_t(i)|^m < \infty$ for any $m > 0$ and using techniques developed in [16] it can be shown that the uniform complete con-

vergence conditions (70) as well as the left-tail conditions (69) hold. Therefore, Corollary 1 implies that the mixture detection procedure $T_{A_\alpha}^{\mathbf{p},W}$ minimizes as $\alpha \to 0$ the expected delay to detection for any compact subset of $\Theta = \Theta_1^{st} \times \cdots \times \Theta_N^{st}$ and asymptotic formulas (63) and (64) hold with

$$I_{\mathsf{B},\theta} = I_{\mathsf{B},\boldsymbol{\theta}_\mathsf{B}} = \frac{1}{2} \sum_{i \in \mathsf{B}} (\theta_i - \theta_i^*)^\top \mathrm{F}_i(\theta_i)(\theta_i - \theta_i^*).$$

In particular, in the purely Markov scalar case where $p = 1$ in (74),

$$I_{\mathsf{B},\theta} = I_{\mathsf{B},\boldsymbol{\theta}_\mathsf{B}} = \sum_{i \in \mathsf{B}} \frac{(\theta_i - \theta_i^*)^2}{2(1 - \theta_i^2)}.$$

4 Monte Carlo

In this section, we perform MC simulations for the example considered in Subsection 3.1, assuming for simplicity (and with a very minor loss of generality) that the noise processes $\xi_n(i) = w_n(i)$ in (71) are i.i.d. Gaussian with mean zero and variances σ_i^2, which is equivalent setting $p = 0$ in (72). We also assume that $S_n(i) = n^{1.1}$, $\sigma_i^2 = 4$, $\theta_i = 0$ pre-change and $\theta_i = \theta = 0.1$ in the post-change mode when the change occurs in the i-th stream.

We suppose that the change in each channel occurs independently with probability q, so the probability of the event $\{M = m\}$ that m streams (out of N) are affected is

$$\mathsf{P}(M = m) = \frac{N!}{m!(N-m)!} q^m (1-q)^{N-m}.$$

The change occurs at time ν according to the geometric prior distribution

$$\mathsf{P}(\nu = k) = \rho(1-\rho)^k, \quad k = 0, 1, 2, \ldots$$

with a parameter $\rho \in (0,1)$.

In MC simulations, we set $N = 10$ for the total number of streams, $\rho = 0.1$, $q = 1/N = 0.1$, and consider three cases: (a) the change occurs in a single stream with $\theta = 0.1$, (b) the change occurs in two streams with $\theta = 0.1$ in each, and (c) the change occurs in three streams with $\theta = 0.1$ in each.

For each MC run, using equations (66) and (67), we compute the statistics (10) and (13) and get the stopping times for the cases where the parameter of the post-change distribution is known (15) and when the parameter of the post-change distribution is unknown (14). We take uniform prior $W(\theta)$ on $[0.1, 0.3]$ with the step 0.01. The number of Monte Carlo runs is 10^6, which ensures very high accuracy of the estimated characteristics.
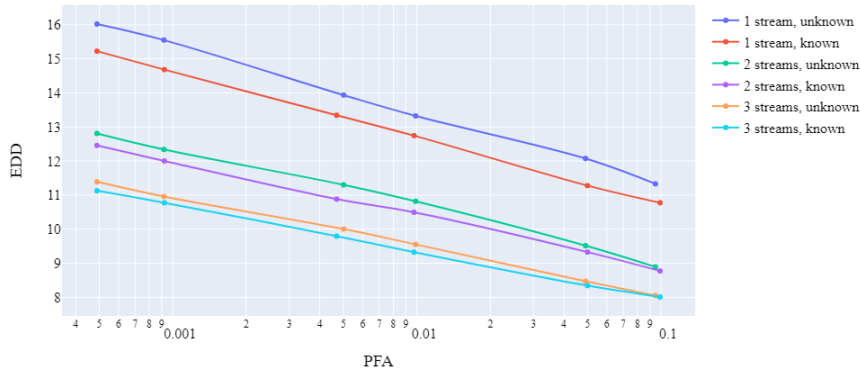
The results are shown in Table 1 and Fig. 1.

In Fig. 1, PFA (x-axis) is presented in a logarithmic scale. Marks on the figure $2, 3, \ldots, 9$ mean how many times the PFA value is greater than the previous decimal

Table 1 Operating characteristics of the mixture detection procedures for the number of streams $N = 10$ with parameters $S_n(i) = n^{1.1}$ and $\theta_1 = \theta_2 = \theta_3 = \theta = 0.1$. The number of MC runs is 10^6.

$\mathrm{PFA}_\pi(T)$	10^{-1}	$5 \cdot 10^{-2}$	10^{-2}	$5 \cdot 10^{-3}$	10^{-3}	$5 \cdot 10^{-4}$
$\mathrm{EDD}_{B_1,\theta}(T_A^{\mathbf{p},\theta})$	10.77	11.28	12.76	13.34	14.68	15.22
$\mathrm{EDD}_{B_1,\theta}(T_A^{\mathbf{p},W})$	11.33	12.07	13.32	13.93	15.55	16.02
$\overline{\mathrm{EDD}}_{B_1,\theta}$	15.99	16.92	18.84	19.50	21.05	21.63
$\mathrm{EDD}_{B_2,\theta}(T_A^{\mathbf{p},\theta})$	8.77	9.33	10.50	10.88	12.00	12.46
$\mathrm{EDD}_{B_2,\theta}(T_A^{\mathbf{p},W})$	8.90	9.52	10.82	11.30	12.34	12.81
$\overline{\mathrm{EDD}}_{B_2,\theta}$	11.78	12.66	14.43	15.02	16.38	16.88
$\mathrm{EDD}_{B_3,\theta}(T_A^{\mathbf{p},\theta})$	8.01	8.34	9.33	9.79	10.77	11.12
$\mathrm{EDD}_{B_3,\theta}(T_A^{\mathbf{p},W})$	8.05	8.47	9.55	9.99	10.96	11.39
$\overline{\mathrm{EDD}}_{B_3,\theta}$	7.90	9.20	11.35	12.00	13.45	13.96

Fig. 1 Operating characteristics of the detection procedures for the total number of streams $N = 10$ and for the number of affected streams $m = 1, 2, 3$: $\mathrm{EDD}_{B_m,\theta}(T_A^{\mathbf{p},\theta})$ vs $\log \mathrm{PFA}_\pi$ and $\mathrm{EDD}_{B_m,\theta}(T_A^{\mathbf{p},W})$ vs $\log \mathrm{PFA}_\pi$. Number of MC runs 10^6. x-axis in logarithmic scale

mark (0.0001, 0.001, or 0.01). It is seen that the detection algorithm has good performance even with such a low signal-to-noise ratio. As expected, in the case when the parameter of the post-change distribution is not known, the algorithm works worse than when it is known, but only slightly – the difference is small. When the change occurs in multiple streams, the expected detection delays $\mathrm{EDD}_{B_2,\theta}(T_A^{\mathbf{p},W})$ and $\mathrm{EDD}_{B_3,\theta}(T_A^{\mathbf{p},W})$ decrease compared to $\mathrm{EDD}_{B_1,\theta}(T_A^{\mathbf{p},W})$ in the case (a), and also $\mathrm{EDD}_{B_3,\theta}(T_A^{\mathbf{p},W})$ becomes smaller than $\mathrm{EDD}_{B_2,\theta}(T_A^{\mathbf{p},W})$, as expected from theoretical results (see, e.g., (56)). Here $\mathrm{EDD}_{B_m,\theta}(T_A^{\mathbf{p},W})$ denotes the expected detection delay when the change occurs in m streams ($m = 1, 2, 3$).

It is also interesting to compare the MC estimates of the expected detection delay with theoretical asymptotic approximations given by (56). In our example, these approximations reduce to

$$\overline{\mathsf{EDD}}_{\mathsf{B}_m,\theta} \approx \left(\frac{\log A_m}{I_{\mathsf{B}_m,\theta}} \right)^{1/3.2}$$

with $I_{\mathsf{B}_m,\theta} = m(\theta^2/6.4\sigma^2)$, $m = 1, 2, 3$. Note that thresholds A_m are different for different m to guarantee the same PFA. The data in Table 1 show that the first-order asymptotic approximation is not too bad but not especially accurate.

5 Applications

In this section, we consider two different applications – rapid detection of new COVID-19 epidemic waves and extraction of tracks of low-observable near-Earth space objects in optical images obtained by telescopes.

5.1 Application to COVID Detection

We consider the problem of detecting the emergence of new COVID-19 epidemic waves in Australia based on real data. Note that in [17] the algorithm of joint disorder detection and identification has been used not only to detect the start of COVID-19 in Italy but also to identify a region where the outbreak occurs. Here we do not consider identification of the region in which the outbreak occurs. We need only to decide on the occurrence of an epidemic in the whole country based on data from various regions. Also, in contrast to [17] where a stationary Markov model has been used, we now propose a substantially different non-stationary model taking into account that a new wave of COVID typically spreads faster than a linear law. This fact has been recently noticed in [13].

We selected data on COVID in Australia, which has 8 regions: Australian Capital Territory, New South Wales, Northern Territory, Queensland, South Australia, Tasmania, Victoria, Western Australia. We propose to use the non-stationary model (71) considered in Subsection 3.1 with super-linear functions $S_n(i) = C_i n^\gamma$, $\gamma > 1$. As an observation, we use the percentage of infections in the total population. We study the COVID outbreak in Australia, which was recorded in the winter of 2021-2022. The data are taken from the World Health Organization and presented in Fig. 2.

A performed statistical analysis shows that the model with parameters $\theta_1 = 0.1$ and $\theta_2 = 0.08$ for New South Wales and Victoria, respectively, describes the beginning of the pandemic outbreak well. We also selected $S_n(i) = n^{1.127}$ and $\xi_n(i) = w_n(i) \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i = 2.4$ in (71). We apply the 8-stream double mixture detection algorithm discussed above when the parameter of the post-change distri-

Distribution of COVID-19 in Australia by Region
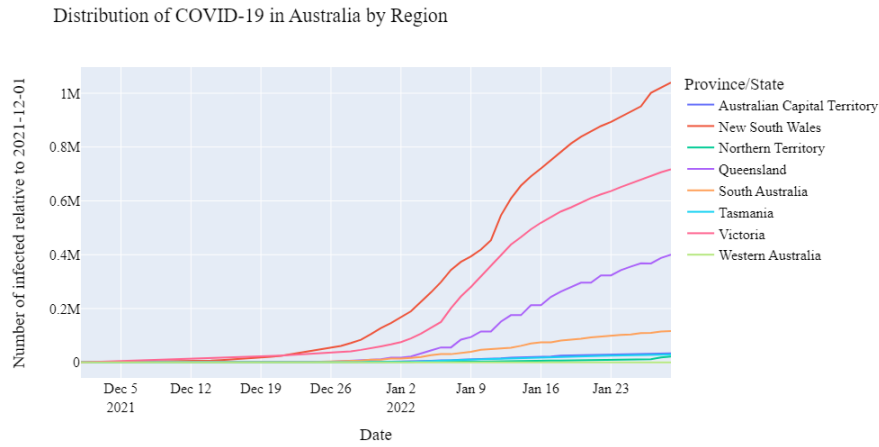
Fig. 2 Distribution of COVID-19 in Australia by region.

bution is not known. We used the discrete uniform prior $W(\theta)$ on $[0.08, \ldots, 0.11]$ with step 0.01. The proposed change detection algorithm with a threshold selected so that the probability of false detection does not exceed 0.01 (average detection delay is about 10) decides on the outbreak of the epidemic on January 9, 2022 (see Fig. 3).

The plots in Fig. 3 illustrate the behavior of the change detection statistic $R_{\mathbf{p},W}(n)$ defined in (13). The COVID wave is detected at the moment when the statistic $R_{\mathbf{p},W}(n)$ crosses the threshold.

Behavior of the change detection statistic over time
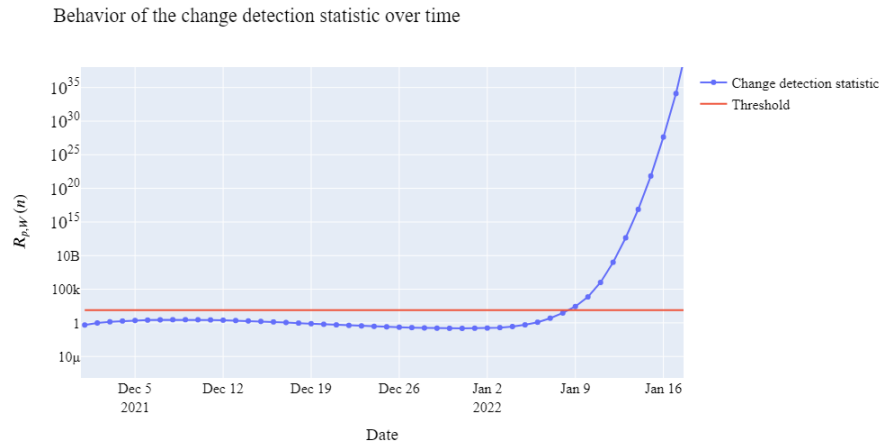
Fig. 3 Behavior of the change detection statistic $R_{\mathbf{p},W}(n)$ over time.

The obtained results complement the existing work on the application of change-point detection algorithms to epidemic detection problems (see, e.g., [2, 13, 17, 36]). The results show that the proposed mixture-based change detection algorithm can be useful for governments when deciding whether to impose a total lockdown across the country without regard to a region in the early stages of a pandemic.

5.2 Application to Detection and Extraction of Faint Space Objects

The problem of rapid detection and extraction of streaks of low-observable space objects with unknown orbits in optical images captured with ground-based telescopes is a challenge for Space Informatics. A typical image (digital frame) with a low-contrast streak with the signal-to-noise ratio (SNR) in pixel approximately 1 is shown in Fig. 4. The object's streak is barely visible, so we drew the rectangle around the streak that marks its position.
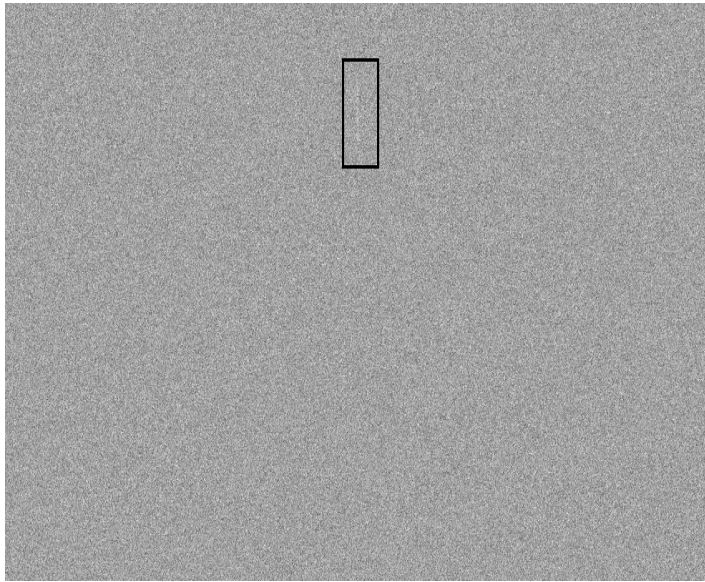
Fig. 4 Digital image with a low-SNR streak. The rectangle marks the streak position.

Since the distribution of observations changes abruptly when the streak starts and ends the problem of object streak extraction can be regarded as the changepoint detection problem in 2-D space (but not in time since we consider a single image).

In realistic situation, the problem is aggravated by the presence of stars and background that produce strong clutter, and special image preprocessing for clutter removal has to be implemented (see, e.g., [30]). We assume that after appropriate pre-

processing, the clutter-removed input frame contains Gaussian noise independent from pixel to pixel. For the sake of simplicity, we also consider a scenario where only one streak may be present in the image. We further assume that the satellite has a linear and uniform motion in the frame. The satellite streak is given by the vector $\mathbf{Y} = (x_0, y_0, x_1, y_1)$, where (x_0, y_0) corresponds to the start point and (x_1, y_1) corresponds to the endpoint. We consider the following model of the observation $X_{i,j}$ in pixel (i, j) of the 2-D frame:

$$X_{i,j} = \theta S_{i,j}(\mathbf{Y}) + \xi_{i,j}, \tag{79}$$

where θ is an unknown signal intensity from the object, $\{S_{i,j}(\mathbf{Y})\}$ are values of the model profile of the streak that are calculated beforehand assuming the point spread function (PSF) is Gaussian with a certain effective width, which is shown in Fig. 5; and $\xi_{i,j} \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise after preprocessing with zero mean and known (estimated empirically) local variance σ^2. Thus, the observation $X_{i,j}$ has normal pre-change distribution $g(X_{i,j}) \sim \mathcal{N}(0, \sigma^2)$ when the streak does not cover pixel (i, j) and normal post-change distribution $f(X_{i,j}) \sim \mathcal{N}(\theta S_{i,j}(\mathbf{Y}), \sigma^2)$ when the streak covers the pixel (i, j).
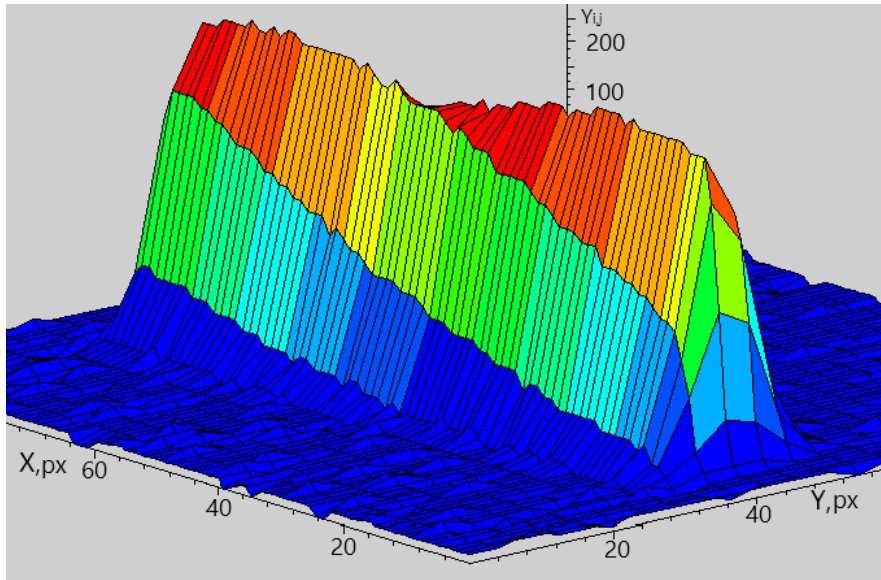

Fig. 5 Model profile of the streak.

The problem is to detect the streak with minimal delay or to make a decision that there is no streak in the frame.

We consider only intra-frame detection of faint streaks of subequatorial satellites with unknown orbits with telescopes mounted at the equator. In this case, a signal

from a satellite (with unknown intensity, start, and end points) is located almost vertically in a small area at the center of the frame, as shown in Fig. 6.
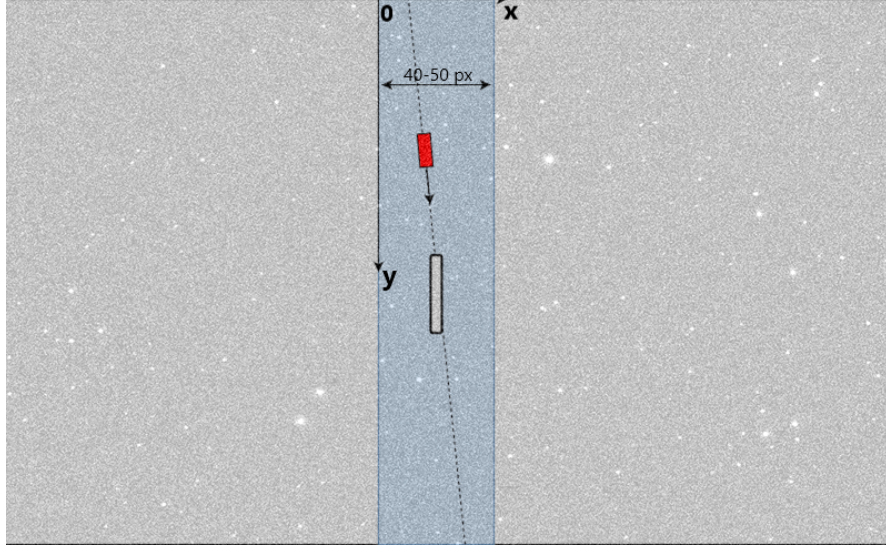
Fig. 6 Search area Ω_S is rectangular. The dotted line shows one of the possible directions. White rectangle – streak, red rectangle – sliding window.

Let Ω_S denote the streak search area. We define N different directions inside Ω_S. Let $d \in \{1,\dots,N\}$ denote a certain direction. Since we assume that there may appear only one satellite's streak in the frame there is no need to mix SR statistics over streak location but rather maximize. However, since the signal intensity θ is not known we still need to mix over the distribution of θ, which is a drawback.

A reasonable way to avoid this averaging is to use the so-called Finite Moving Average (FMA) statistics, as suggested in [24]. Specifically, let $M_d(n), n \geq 1$ stand for a 2-D sliding rectangular window which contains certain pixel numbers (i,j) at each step n in the direction d. Window $M_d(n)$ has a fixed length of L_d pixels and a fixed width of K_d pixels (the choice of the parameters depends on the expected SNR and PSF effective width). For the certain direction d $(d = 1,\dots,N)$, the FMA statistic is defined as

$$V_{M_d(n)}(n) = \sum_{(i,j) \in M_d(n)} S_{i,j}^{(d)} X_{i,j}^{(d)},$$

where $\{S_{i,j}^{(d)}\}$ are values of the Gaussian model profile in the direction d and $X_{i,j}^{(d)}$ are observed data in the direction d. Profile location is given by the vector $\mathbf{Y}_d$ in the direction d. Then the multistream[1] FMA detection procedure is defined as

---

[1] Here "streams" are not streams per se but rather data $X_{i,j}^{(d)}$ in different directions $d \in \{1,\dots,N\}$ in the search area Ω_S.

$$T_h = \inf\left\{ n \geq 1 : \max_{1 \leq d \leq N} V_{M_d(n)}(n) \geq h \right\}.$$

For the Gaussian model, the FMA procedure is invariant to the unknown signal intensity θ, which is a big advantage over the SR-type versions.
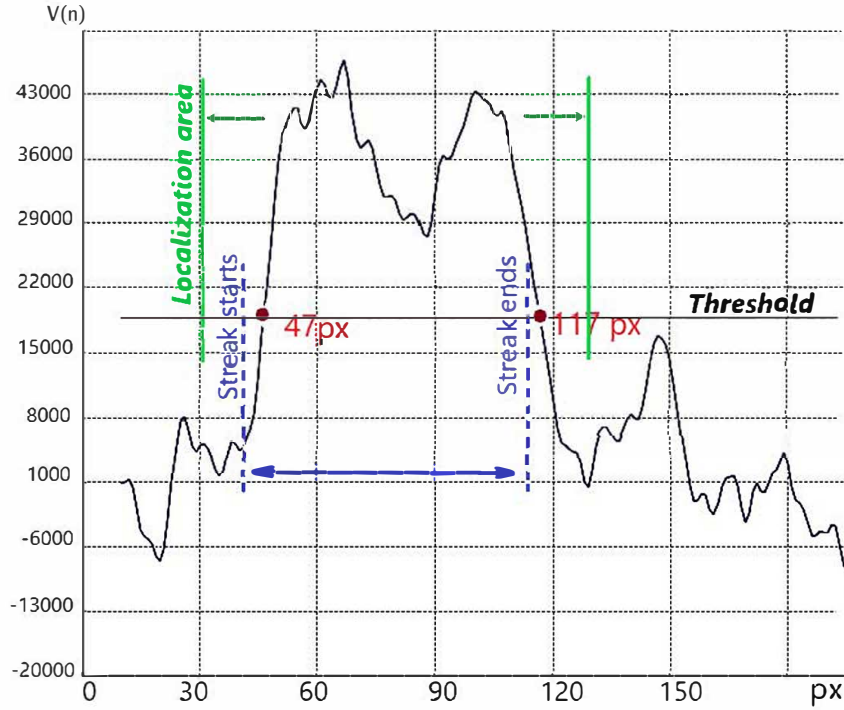
Fig. 7 The behavior of the FMA $V_{M_d(n)}(n)$ statistic along the correct direction.

Fig. 7 shows a typical behavior of the $V_{M_d(n)}(n)$ statistic along the direction containing the streak in the case of a very low SNR $= 0.9$. In this case, the streak is detected with coordinates of start and end at points 47 and 117, respectively, while the true values are 40 and 110 so that the precision is 7 pixels. Experiments show that when sliding the 2-D window in various directions inside Ω_S and then comparing the largest value $\max_d V_{M_d(n)}(n)$ to a threshold we typically determine the approximate position of the streak with an accuracy of 5-10 pixels. Therefore, the proposed FMA version of the change detection algorithm turns out to be efficient – it allows us to rapidly determine a localization area, which with a high probability contains the streak.

6 Concluding Remarks and Future Challenges

6.1 Remarks

1. As we already discussed, for general non-i.i.d. models SR-type statistics cannot be computed recursively even in separate streams, so the computational complexity and memory requirements of the mixture detection procedures $T_A^{\mathbf{p},\theta}$ and $T_A^{\mathbf{p},W}$, especially in the asymptotically non-stationary case, can be quite high. To avoid this drawback it is reasonable to use either one-stage delayed adaptive procedures or window-limited versions of mixture detection procedures where the summation over potential change points $\nu = k$ is restricted to a sliding window of a fixed size ℓ. In the window-limited versions of $T_A^{\mathbf{p},W}$, the statistic $R_{\mathbf{p},W}(n)$ is replaced by the window-limited statistic

$$\widehat{R}_{\mathbf{p},W}(n) = \sum_{k=n-(\ell+1)}^{n-1} \Lambda_{\mathbf{p},W}(k,n) \quad \text{for } n > \ell; \quad \widehat{R}_{\mathbf{p},W}(n) = R_{\mathbf{p},W}(n) \quad \text{for } n \le \ell.$$

Following guidelines of Lai [12] and the techniques developed by Tartakovsky [29, Sec 3.10, pages 116-122] for the single-stream scenario, it can be shown that the window-limited version of the SR mixture also has first-order asymptotic optimality properties as long as the size of the window $\ell(A)$ approaches infinity as $A \to \infty$ with the rate $\ell(A)/\log A \to \infty$. Since thresholds, $A = A_\alpha$, in detection procedures should be selected in such a way that $\log A_\alpha \sim |\log \alpha|$ as $\alpha \to 0$, the value of the window size $\ell(\alpha)$ should satisfy $\lim_{\alpha \to 0}[\ell(\alpha)/|\log \alpha|] = \infty$.

2. Similar asymptotic optimality results can be obtained for the mixture CUSUM procedure based on thresholding of the sum of generalized LR statistic

$$W(n) = \sum_{i=1}^{N} \max_{0 \le k < n} \sup_{\theta_i \in \Theta_i} \lambda_{i,\theta_i}(k,n)$$

and for the corresponding window-limited version.

3. We also conjecture that asymptotic optimality properties hold for the multi-stream Finite Moving Average (FMA) detection procedure given by the stopping time

$$T_h = \inf\left\{ n \ge 1 : \sum_{i=1}^{N} \sup_{\theta_i \in \Theta_i} \sum_{k=\max(1,n-\ell+1)}^{n} \lambda_{i,\theta_i}^*(k) \ge h \right\},$$

where $\lambda_{i,\theta_i}^*(k) = \log[f_{i,\theta_i}(X_k(i)|\mathbf{X}^{k-1}(i))/g_i(X_k(i)|\mathbf{X}^{k-1}(i))]$. In cases where the LLR $\lambda_{i,\theta_i}^*(k)$ is a monotone function of some statistic, the FMA procedure can be appropriately modified in such a way that it is invariant to the unknown parameters θ_i. This is a great advantage over corresponding SR-based and CUSUM-based procedures. See also discussion in Subsection 5.2.

4. The results can be easily generalized to the case where the change points ν_i are different for different streams $i = 1, \ldots, N$.

5. The Bayesian-type results of this paper can be used to establish asymptotic optimality properties of the detection procedures $T_A^{\mathbf{p},\theta}$ and $T_A^{\mathbf{p},W}$ in a non-Bayesian setting. In particular, the optimization problem can be solved in the class of procedures with the upper-bounded maximal local probability of false alarms

$$\sup_{m \geq 0} \mathsf{P}_\infty \left(T \leq m + k | T > m \right) \leq \alpha,$$

both in pointwise and minimax settings, by embedding into the Bayesian class, similar to what was done by Pergamenchtchikov and Tartakovsky [16] for the case of a single stream.

6.2 Future Challenges

1. The results of MC simulations in Section 4 show that first-order approximations to EDD and PFA are typically not especially accurate, so higher-order approximations are in order. However, it is not feasible to obtain such approximations in the general non-i.i.d. case considered in this paper. Higher order approximations to the expected detection delay and the probability of false alarm for the i.i.d. models, assuming that the observations in streams are independent and also independent across streams, can be derived based on the renewal and nonlinear renewal theories. This important problem will be considered in the future.

2. The results of this paper cover the scenario where the number of streams N is fixed so that $\log(N/\alpha) \sim \log(1/\alpha)$ for small α. In Big Data problems, N may be very large and go to infinity. These problems require different approaches and different detection procedures. This challenging problem will be considered in the future.

3. For detecting transient (or intermittent) changes of unknown duration (like object streaks in Subsection 5.2) it is often more reasonable to consider not quickest detection criteria but reliable detection criteria that require minimization of the detection probability in a fixed time (or space) window (see, e.g., [3, 24, 29] and references therein). While certain interesting asymptotic results for single-stream scenarios and i.i.d. data models exist [9, 10, 14, 21, 24, 29], to the best of our knowledge this problem has never been considered in the multistream setting and for non-i.i.d. models.

References

1. Bakut, P.A., Bolshakov, I.A., Gerasimov, B.M., Kuriksha, A.A., Repin, V.G., Tartakovsky, G.P., Shirokov, V.V.: Statistical Radar Theory, vol. 1 (G. P. Tartakovsky, Editor). Sovetskoe Radio, Moscow, USSR (1963). In Russian
2. Baron, M.: Early detection of epidemics as a sequential change-point problem. In: V. Antonov, C. Huber, M. Nikulin, V. Polischook (eds.) Longevity, Aging and Degradation Models in Reliability, Public Health, Medicine and Biology, vol. 2, pp. 31–43. St. Petersburg (2004)
3. Berenkov, N.R., Tartakovsky, A.G., Kolessa, A.E.: Reliable detection of dynamic anomalies with application to extracting faint space object streaks from digital frames. In: 2020 International Conference on Engineering and Telecommunication (EnT-MIPT 2020). Dolgoprudny, Russia (2020)
4. Chan, H.P.: Optimal sequential detection in multi-stream data. Annals of Statistics **45**(6), 2736–2763 (2017)
5. Chang, F.K.: Structural health monitoring: Promises and challenges. In: Proceedings of the 30th Annual Review of Progress in Quantitative NDE (QNDE), Green Bay, WI, USA. American Institute of Physics (2003)
6. Fellouris, G., Sokolov, G.: Second-order asymptotic optimality in multichannel sequential detection. IEEE Transactions on Information Theory **62**(6), 3662–3675 (2016). DOI 10.1109/TIT.2016.2549042
7. Fienberg, S.E., Shmueli, G.: Statistical issues and challenges associated with rapid detection of bio-terrorist attacks. Statistics in Medicine **24**(4), 513–529 (2005)
8. Frisén, M.: Optimal sequential surveillance for finance, public health, and other areas (with discussion). Sequential Analysis **28**(3), 310–393 (2009)
9. Guépié, B.K., Fillatre, L., Nikiforov, I.: Sequential detection of transient changes. Sequential Analysis **31**(4), 528–547 (2012). DOI 10.1080/07474946.2012.719443
10. Guépié, B.K., Fillatre, L., Nikiforov, I.: Detecting a suddenly arriving dynamic profile of finite duration. IEEE Transactions on Information Theory **63**(5), 3039–3052 (2017). DOI 10.1109/TIT.2017.2679057
11. Kolessa, A., Tartakovsky, A., Ivanov, A., Radchenko, V.: Nonlinear estimation and decision-making methods in short track identification and orbit determination problem. IEEE Transactions on Aerospace and Electronic Systems **56**(1), 301–312 (2020). DOI 10.1109/TAES.2019.2911760
12. Lai, T.L.: Information bounds and quick detection of parameter changes in stochastic systems. IEEE Transactions on Information Theory **44**(7), 2917–2929 (1998)
13. Liang, Y., Tartakovsky, A.G., Veeravalli, V.V.: Quickest change detection with non-stationary post-change observations. IEEE Transactions on Information Theory **69**(5), 3400–3414 (2023). DOI 0.1109/TIT.2022.3230583
14. Mana, F.E., Guépié, B.K., Nikiforov, I.: Sequential detection of an arbitrary transient change profile by the FMA test. Sequential Analysis pp. 1–21 (2023). DOI 10.1080/07474946.2023.2171056
15. Mei, Y.: Efficient scalable schemes for monitoring a large number of data streams. Biometrika **97**(2), 419–433 (2010)
16. Pergamenchtikov, S., Tartakovsky, A.G.: Asymptotically optimal pointwise and minimax change-point detection for general stochastic models with a composite post-change hypothesis. Journal of Multivariate Analysis **174**(11), 1–20 (2019)
17. Pergamenchtikov, S.M., Tartakovsky, A.G., Spivak, V.S.: Minimax and pointwise sequential changepoint detection and identification for general stochastic models. Journal of Multivariate Analysis **190**, 1–22 (2022)
18. Raghavan, V., Galstyan, A., Tartakovsky, A.G.: Hidden Markov models for the activity profile of terrorist groups. Annals of Applied Statistics **7**, 2402–24,307 (2013)
19. Raghavan, V., Steeg, G.V., Galstyan, A., Tartakovsky, A.G.: Modeling temporal activity patterns in dynamic social networks. IEEE Transactions on Computational Social Systems **1**, 89–107 (2013)

20. Rolka, H., Burkom, H., Cooper, G.F., Kulldorff, M., Madigan, D., Wong, W.K.: Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: research needs. Statistics in Medicine **26**(8), 1834–1856 (2007)

21. Sokolov, G., Spivak, V.S., Tartakovsky, A.G.: Detecting an intermittent change of unknown duration. Sequential Analysis (2023, to be published)

22. Sonesson, C., Bock, D.: A review and discussion of prospective statistical surveillance in public health. Journal of the Royal Statistical Society A **166**, 5–21 (2003)

23. Szor, P.: The Art of Computer Virus Research and Defense. Addison-Wesley Professional, Upper Saddle River, NJ, USA (2005)

24. Tartakovsky, A., Berenkov, N., Kolessa, A., Nikiforov, I.: Optimal sequential detection of signals with unknown appearance and disappearance points in time. IEEE Transactions on Signal Processing **69**, 2653–2662 (2021). DOI 10.1109/TSP.2021.3071016

25. Tartakovsky, A.G.: Asymptotic performance of a multichart CUSUM test under false alarm probability constraint. In: Proceedings of the 44th IEEE Conference Decision and Control and European Control Conference (CDC-ECC'05), Seville, SP, pp. 320–325. IEEE, Omnipress CD-ROM (2005)

26. Tartakovsky, A.G.: Rapid detection of attacks in computer networks by quickest changepoint detection methods. In: N. Adams, N. Heard (eds.) Data Analysis for Network Cyber-Security, pp. 33–70. Imperial College Press, London, UK (2014)

27. Tartakovsky, A.G.: On asymptotic optimality in sequential changepoint detection: Non-iid case. IEEE Transactions on Information Theory **63**(6), 3433–3450 (2017). DOI 10.1109/TIT.2017.2683496

28. Tartakovsky, A.G.: Asymptotic optimality of mixture rules for detecting changes in general stochastic models. IEEE Transactions on Information Theory **65**(3), 1413–1429 (2019). DOI 10.1109/TIT.2018.2876863

29. Tartakovsky, A.G.: Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules. Monographs on Statistics and Applied Probability 165. Chapman & Hall/CRC Press, Taylor & Francis Group, Boca Raton, London, New York (2020)

30. Tartakovsky, A.G., Brown, J.: Adaptive spatial-temporal filtering methods for clutter removal and target tracking. IEEE Transactions on Aerospace and Electronic Systems **44**(4), 1522–1537 (2008)

31. Tartakovsky, A.G., Nikiforov, I.V., Basseville, M.: Sequential Analysis: Hypothesis Testing and Changepoint Detection. Monographs on Statistics and Applied Probability 136. Chapman & Hall/CRC Press, Taylor & Francis Group, Boca Raton, London, New York (2015)

32. Tartakovsky, A.G., Rozovskii, B.L., Blaźek, R.B., Kim, H.: Detection of intrusions in information systems by sequential change-point methods. Statistical Methodology **3**(3), 252–293 (2006)

33. Tartakovsky, A.G., Rozovskii, B.L., Blaźek, R.B., Kim, H.: A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods. IEEE Transactions on Signal Processing **54**(9), 3372–3382 (2006)

34. Tsui, K., Chiu, W., Gierlich, P., Goldsman, D., Liu, X., Maschek, T.: A review of healthcare, public health, and syndromic surveillance. Quality Engineering **20**(4), 435–450 (2008)

35. Xie, Y., Siegmund, D.: Sequential multi-sensor change-point detection. Annals of Statistics **41**(2), 670–692 (2013)

36. Yu, X., Baron, M., Choudhary, P.K.: Change-point detection in binomial thinning processes, with applications in epidemiology. Sequential Analysis **32**(3), 350–367 (2013)